Von der Kunst des Lernens

Einige Bemerkungen zur Intentionalität von In- und Output

Claudius Härpfer, Nadine Diefenbach

1. Einleitung

Der Begriff des maschinellen Lernens suggeriert, dass Maschinen lernen. Wie auch immer dieser Prozess im Detail aussieht (vgl. z. B. Shalev-Shwartz/Ben-David 2014; Russel/Norvig 2016: 693–859), in den meisten Verfahren sind nicht nur Maschinen involviert, sondern auch Menschen, die die Maschinen anleiten, indem sie ihnen gezielt Lerndaten zur Verfügung stellen und gegebenenfalls Optimierungen an einzelnen Parametern wie beispielsweise der Gewichtung von Verbindungen im künstlichen neuronalen Netz vornehmen. Konstruktionsbedingt kann der die Maschine anleitende Mensch die eigentlichen Vorgänge im inneren des künstlichen neuronalen Netzes nicht überblicken (vgl. z. B. Burrell 2016), sondern entwickelt nur eine mehr oder weniger vage Vorstellung davon, was konkret im Inneren abläuft, auf deren Basis er das Lernen der Maschine zu steuern versucht. Der Mensch hat also eine in irgendeiner Form sinnhafte, metaphorisch aufgeladene Vorstellung von dem, was die Maschine tut, ohne wirklich zu wissen, wie effektiv und zielführend seine Lenkungsversuche des Lernprozesses sein werden, während die Maschine in den zur Verfügung gestellten Daten Muster erkennt und diese ordnet, indem sie neue Verknüpfungen bildet oder bestehende löst. Diesem komplexen soziotechnischen Phänomen wollen wir uns widmen, indem wir die relevanten Abläufe als Netzwerk soziotechnischer Relationen begreifen, deren Elemente sich vermittelt durch das jeweilige Design treffen und – in einem Prozess der begonnenen Übersetzung – voneinander lernen, ohne sich jemals verstehen zu können. Neben einschlägigen Arbeiten Herbert Simons (1984; 1994) greifen wir hierbei in Anlehnung an Häußling (2012; 2016; 2020) auf das Vokabular der Netzwerktheorie Harrison Whites (1992; 2008)

zurück und verstehen die Interaktionen als Verstrickungen wechselseitiger Kontrollprojekte in sich in diesem Prozess bildenden Netzwerkdomänen.

2. Das Lernen der Maschinen

Zunächst gilt es einen kurzen Blick auf das Lernen der Maschinen zu werfen. Zwar suggeriert das große alte Narrativ einer künstlichen Intelligenz (vgl. z. B. La Mettrie 2001; Riskin 2007; Domingos 2015), dass Maschinen (irgendwann) in der Lage sind, menschenähnlich zu lernen und dem Menschen ebenbürtig (oder gar überlegen) zu sein. In der Praxis gilt seit Turings Arbeiten in den späten 1940er Jahren (2004) jedoch die Maxime kleiner Schritte und den Menschen, die die Maschinen im Lernprozess unterstützen, ist natürlich bewusst, dass maschinelles Lernen auf probabilistischen Modellen beruht, sodass ein Übermaß an Pädagogik und Entwicklungspsychologie hierzu inkompatibel und deshalb fehl am Platze ist. Wie der Blick in die Anwendungsliteratur zum maschinellen Lernen zeigt, sind die faktischen natürlichen Referenzgrößen nach wie vor weniger Menschen als vielmehr das Lernverhalten von Ratten und Tauben (vgl. z. B. Shalev-Shwartz/Ben-David 2014: 1–3).

Der Blick in die soziologisch-kulturwissenschaftliche Literatur wiederum zeigt die Vielfältigkeit der Anknüpfungspunkte und offenbart ein komplexeres Lernsetting. Ausgehend von der Annahme, dass menschliche Intelligenz aus einem Arrangement der Einbettung des Menschen in seine jeweilige Umwelt, und dem Zusammenspiel all jener Faktoren hervorgeht (vgl. Taffel 2019), scheint gerade die Auseinandersetzung mit dem maschinellen Lernen in seiner soziotechnischen Vernetzung relevant. Dies umfasst nicht allein die Betrachtung des Begriffs Machine Learning, sondern ermöglicht zudem einen Beitrag zur Reflexion des menschlichen Lernens. Lernen im systemischen Sinn meint nach Ziegler (2007) die umfassende Berücksichtigung verschiedener am Lernprozess beteiligter Faktoren. Darunter sind nicht nur Zeit und Raum zu subsumieren, sondern auch die am Lernprozess beteiligten Akteur:innen. Ganz im Sinne von Reigeluth & Castelle (2021) gehört zum Lernen eine Lehrperson, die anleitet und unterstützt. Die beiden Autoren begreifen maschinelles Lernen nicht allein als die Verarbeitung von bereits entschiedenen Informationen als soziotechnischen Prozess, der auf der Basis verschiedener Algorithmen verläuft. Maschinelles Lernen umfasst vielmehr die »Anleitung«, das Trainieren sowie die Interpretation von Ergebnissen und daraus gefolgerte Erkenntnisse des Menschen, der nicht auf den »Daten- und

Mittelgeber« reduziert werden kann. Maschinelles Lernen wird in ihrem Sinn als »cultural activity«, als »soziale[r] Prozess« gefasst (Reigeluth/Castelle 2021: 105).

Angesichts dieses Auseinanderdriftens der Perspektiven, ist es nicht verwunderlich, dass Herbert Simons technisch gehaltene (und damit vermittelnde) Definition maschinellen Lernens bis heute relevant ist (Kaminski/Glass 2019). Simon fasste Lernen als »any change in a system that allows it to perform better the second time on repetition of the same task or on another task drawn from the same population.« Diese Änderung soll zu einem gewissen Grad dahingehend irreversibel sein, dass der Lerneffekt nicht ohne Weiteres von selbst wieder verschwindet. Besser durchgeführt heißt für ihn, dass die Aufgabe oder eine Aufgabe aus der gleichen Grundgesamtheit vom System beim nächsten Mal »more efficiently and more effectively« erledigt wird (Simon 1983: 28). Diese Definition enthält ein Element der Wiederholung, also der Regelmäßigkeit. Daher ist es folgerichtig, wenn Kaminski und Glass (2019: 130) im Anschluss den Aspekt der Regelbildung betonen, den sie mit Blick auf die Mittel, den Zweck und die Lernintention selbst strukturieren. Das Finden von Regeln ist jedoch soziologisch betrachtet nur eine Dimension, sich Wissen anzueignen, und kann je nach Art der Regel hoch heterogen sein (vgl. Schütz 1972). Simons Definition ist damit aber keinesfalls erschöpft, denn auch die Parameter für die Bildung von Regeln werden benannt. Simon verweist (1) auf eine Zustandsveränderung, die (2) innerhalb eines Systems stattgefunden hat, deren Ergebnis (3) als besser eingestuft werden kann, also auf einen vorhandenen – wie auch immer gearteten – Maßstab (inkl. Effektivitäts- und Effizienzkriterien) übertragen wird und (4) durch den gleichen Vorgang oder bei der gleichen Population angewandt wird.

Wenn wir versuchen, uns dem Phänomen soziologisch zu nähern, bietet sich der Rückgriff auf die White'sche Theorie von Identität und Kontrolle an. White baut seinen Blick auf die Welt (1992; 2008) um das Wechselspiel der Begriffe Identität und Kontrolle auf. Identitäten streben nach Kontrolle und die Kontrollversuche der Identitäten dienen dazu, andere Identitäten auf die eigene Prozessualität zu verpflichten. Dahinter stecken ein komplex werdender, mehrstufiger Identitätsbegriff und eine prozessuale relationale Perspektive. Der Identitätsbegriff ist dazu in der Lage, auch nicht-menschliche Einheiten adäquat einzubeziehen, gepaart mit der prozessualen Perspektive erwächst daraus eine techniksoziologisch gehaltvolle Perspektive, die wir im Folgenden am Lernbegriff skizzieren wollen.

Any change ...

Bei seiner Beschreibung von Prozessualität verstrickt sich Simon an anderer Stelle in metaphorisch aufgeladene systemtheoretische Verallgemeinerungen über Evolution und Auslese in Systemen (Simon 1994: 149–162). Soziologisch gehaltvoller scheint es uns hier, den Lernprozess als Wechselspiel von Kontrollprojekten zu verstehen. Liest man Whites Theorie systematisch, kann man darin drei unterschiedliche Formen dieser Kontrollprojekte finden (vgl. Häußling 2012: 269–272). Erstens die sogenannte »social ambage« (White 1992: 106–107), also die direkte oder indirekte Einflussnahme auf das Gegenüber mittels bestehender oder möglicher Relationen. Zweitens die »cultural ambiguity« (White 1992: 103–106), also die Kontextabhängigkeit der Deutung von Tatbeständen, was einerseits Offenheit hinsichtlich möglicher Deutungsmöglichkeiten bedeutet, andererseits aber auch Abhängigkeit von bestehenden Deutungsmöglichkeiten. Drittens das »decoupling« (White 1992: 12–13; 111–112), also das Vermeiden oder Lösen von Relationen. Whites Kontrollbegriff steht sein Identitätsbegriff gegenüber (White 2008: 9–12), mit dem wir in der Lage sind, auch Maschinen als Identitäten zu fassen, ohne sie mit menschlichen Identitäten gleichsetzen zu müssen.

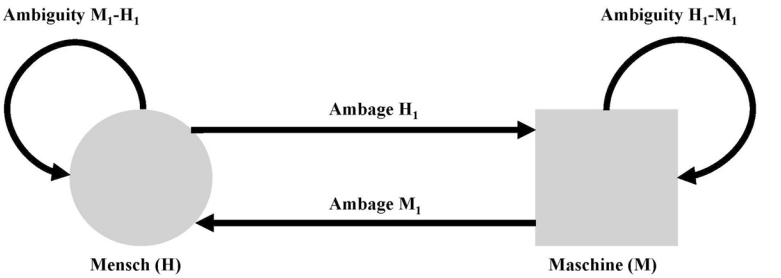
Bezogen auf unser Problem haben wir Formen der Einflussnahme (ambage) durch den Menschen auf die Maschine durch die Eingabe von Daten. Diese Daten interpretiert die Maschine in Abhängigkeit der bisher vorhandenen Daten (ambiguity). Die Maschine wiederrum gibt Informationen über die erkannten Muster aus (ambage), die dann der Mensch auf Basis des ihm bekannten interpretiert (ambiguity) und mit seinen der Maschine gesteckten Lernzielen abgleicht.

Daneben steht die Form der Einflussnahme des Menschen durch Änderung einzelner Parameter, wie der Gewichtung von Verbindungen im neuronalen Netz (ambage). Auf diese Veränderungen »reagiert« die Maschine durch Neuinterpretation der bereits vorhandenen Daten. Ob die Interpretation der Maschine hier nun die Intention des Menschen trifft, ist offen (ambiguity). Auch hier gibt die Maschine wiederrum Informationen über die erkannten Muster aus (ambage), die dann der Mensch auf Basis des ihm bekannten interpretiert (ambiguity).

Das in beiden Fällen Elemente – seien es Datensätze oder Muster – nicht oder nicht mehr berücksichtigt werden (decoupling), liegt in der Natur der Sache. Die Eingabe von Daten seitens des Menschen wollen wir im Folgenden als

Training der Maschine bezeichnen, das Ändern einzelner Parameter hingegen als Lehren bzw. Lernen.

Abbildung 1: Grundschema der Kontrollprojekte



Die Betrachtung der wechselseitigen Bezugnahme der an der Schnittstelle vollzogenen Interpretations- bzw. Lernvorgängen durch die beschriebenen Kontrollversuche hilft dabei, die Kunst des maschinellen Lernens in ihrer soziotechnischen Konstruktion differenziert zu betrachten. Werfen wir also einen genaueren Blick auf den Rest von Simons Definition.

… in a system …

Wenn Simon von einem System spricht, so ist klar, dass er keine soziologische Systemtheorie, sondern eine allgemeine Systemtheorie im weitesten Sinne vor Augen hat (Simon 1994: 144; 1956: 74–79). Er vermeidet eine formale Definition und spricht stattdessen von einem komplexen System als einem Gebilde, »das aus einer großen Zahl von Teilen zusammengesetzt ist, wenn diese Teile nicht bloß in der einfachsten Weise interagieren. In solchen Systemen ist das Ganze mehr als die Summe der Teile – nicht in einem absoluten, metaphysischen Sinn, sondern in dem wichtigen pragmatischen, daß es keine triviale Angelegenheit ist, aus den gegebenen Eigenschaften der Teile und den Gesetzen ihrer Wechselwirkung die Eigenschaften des Ganzen zu erschliessen.« (Simon 1994: 145) Die Einwirkung der Daten, der Interpretation der Ausgaben der Maschine und der weiteren Arbeit an den Daten und Algorithmen sowie der Austausch in der Fachcommunity durch beispielsweise Veröffentlichungen verändert in der Erweiterung des Netzwerkes des lernenden Menschen das (maschinelle) Lernen selbst. Die aus den Lernprozessen resultierende Entwicklung wird folg-

lich erst durch das komplexe Zusammenwirken von Menschen und Maschinen möglich, insofern scheint es plausibel, dass Simon hier eine Einheit mit komplexem Innenleben denkt.

In Whites Terminologie lässt sich eine derartige Einheit als Netdom begreifen. Netdom ist eine Kombination der Begriffe ›Netzwerk‹ und ›Domäne‹ (White 2008: 7) und benennt jene Einheit, die entsteht, wenn durch die Kontrollbemühungen von Identitäten Netzwerke eines bestimmten Inhaltes, und damit einer abgeschlossenen Sinneinheit entstehen. Dass diese abgeschlossene Sinneinheit innerhalb eines Netzwerkkonzeptes steht, womit es sich natürlich nur um eine relative Abgeschlossenheit handeln kann, sollte angesichts von Simons pragmatischem, relativierten Systemverständnis eher eine Akzentverschiebung als ein gravierender Einschnitt sein, zumal Simons Fokus auf dem liegt, was an den »schmalen Schnittstellen« zwischen den Systemen, als den »inneren und äußeren Umgebungen« passiert (Simon 1994: 97).

Als Netdom »Machine Learning« besteht, so gesehen ein Netzwerk als Strukturform von sozialen und technischen Identitäten des maschinellen Lernens, das zudem kulturell unter anderem darüber bestimmt, wie zum Beispiel Wissen über Trainingsmethoden, Learner oder Datensätze geteilt wird. Netdoms setzen Beziehungen unter einem bestimmten Thema und dessen Verständnis voraus, diese sind jedoch, wie angedeutet, nicht absolut, sondern prozessual und transitorisch. Mit einem besonderen Strukturmuster in Verbindung mit spezifischen kulturellen Formen ist ein Netdom ein spezifischer Ausschnitt mit je eigener Einbettung in ein gesamtes Netzwerk. Dabei kann ein Netdom eine dyadische Beziehung abschließen und sehr klein und vergleichsweise flüchtig sein, oder ein komplexes Ganzes mit einer Vielzahl unterschiedlicher Akteur:innen und Beziehungen abbilden.

Das Netdom des »Machine Learnings« lässt sich somit als die dynamisch wechselseitige Prägung von Identitäten mit ihren je verschiedenen Kontrollmaßnahmen fassen. Diese wirken sich sowohl auf die kulturellen als auch die strukturellen, aber auch die Einbettungszusammenhänge aus, die sie mitformen. Das »Machine Learning-Netdom« umfasst dabei nicht nur kommunikativ viele Bereiche von Forschung oder Entwicklung, sondern fasst zudem spezielle Relationen der soziotechnischen Interdependenzen an der schmalen Schnittstelle zwischen Mensch und Maschine entlang von unterschiedlichen

Lern- und Trainingsformen,[1] die in ihrer je spezifischen Einbettung verschiedene Kopplungs- und Entkopplungsprozesse bedeuten können.

... to perform better...

Simon begreift den Lernfortschritt als Effizienz- oder Effektivitätssteigerung, was bei einem klar definierten Problem einfach zu quantifizieren ist, in einem komplexer werdenden Setting natürlich zunehmend schwieriger zu fassen ist. Er selbst spricht das Problem der Verarbeitung natürlicher Sprachen (als »very annoying part of the task«) und die damit verbundene Komplexitätssteigerung an (Simon 1983: 29f.). Im Falle einer derartigen gegebenen Komplexität und der damit verbundenen unendlichen Verbesserbarkeit, lässt sich jede Optimierung zwar theoretisch ins Unendliche iterieren, kommt aber in der Praxis an ihre Grenzen, an deren Ende eher sich stilistisch unterscheidende erwünschte Varianten stehen als klar hierarchisierbare Alternativen, die besser oder schlechter sind (Simon 1994: 112). Weswegen Simon auch eher von »Satisfizierungs-« als von »Optimierungsfragen« spricht (ebd. 26). Bei satisfizierenden Methoden kommt es nicht darauf an, ob das Optimum definiert oder alle Alternativen ausgeschöpft sind, sondern ob ein mit Blick auf Kosten-Nutzen-Erwägungen befriedigendes Ergebnis vorliegt (ebd.: 95–119).

In unserem Anwendungsbeispiel entwickelt der Mensch neben allen Indikatoren, die ihm zur Verfügung stehen, ein Gefühl dafür, ob das neue Ergebnis nun besser ist als das alte und damit gegebenenfalls ein zufriedenstellendes Ergebnis vorliegt oder nicht, und deshalb eine weitere Iteration folgen muss. Dies geschieht in Whites Terminologie über ein Narrativ, über »Stories«, also kontextuierte Verbindungen (White 2008: 20). Die zustande kommenden Netzwerke werden durch sie umgebende Erzählungen gefestigt, sodass ein Maßstab entsteht, der dem Menschen hilft, die Leistung der Maschine einzuschätzen.

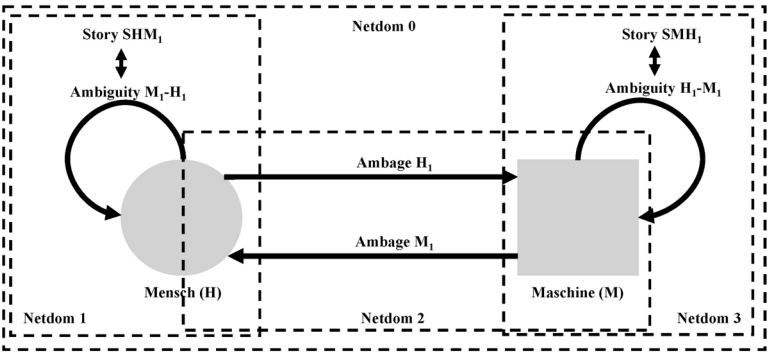
... the second time.

Gleichzeitig ordnen die Stories aber auch die soziale Zeit und grenzen jene Zustände ein, die wiederholt werden. Im Prozess des Vergleichens stellt der Mensch also die Leistung der Maschine zum Zeitpunkt t_1, der Leistung der

1 Wir folgen hier der Unterscheidung von Lernformen und Lernstrategien nach Kaminski/Glass (2019: 130).

Maschine zum Zeitpunkt t_2 gegenüber. Diesen Übergang bezeichnet White als »Switching« (White 2008: 2, 20). Er geht davon aus, dass die Identitäten sich entwickeln, indem sie von einem Netdom zum nächsten wechseln und so eine Spur durch verschiedene Kontexte legen und eine zunehmend komplexere Geschichte schaffen. Bezogen auf den Lernprozess lässt sich das so verstehen, dass unser Modell noch etwas komplexer werden muss.

Abbildung 2: Lernprozess als Switching zwischen Netdoms zum Zeitpunkt t_1



SHM:	Einfluss der Mensch-Maschinen Relation auf die Storyentwicklung
SMH:	Einfluss der Maschinen-Mensch Relation auf die Storyentwicklung
Ambage H:	Kontrollversuch des Menschen auf die Maschine
Ambage M:	Kontrollversuch der Maschine auf den Menschen
Ambiguity H-M:	kontextabhängige Deutung von kontextualisierten Daten durch die Maschine
Ambigutiy M-H:	kontextabhängige Deutung der maschinellen Datenverarbeitung durch den Menschen

Wenn wir also die vorher kurz beschriebenen Veränderungsprozesse nochmals aufgreifen und uns hier auf den ersten der vorhin beschriebenen Fälle beschränken, heißt das holzschnittartig, dass die Identität Mensch zum Zeitpunkt t_1 im Zuge der Eingabe von Daten das Netdom 1 verlässt und durch die Ambage H_1 eine Relation zur Maschine eingeht, also das Netdom 2 aufspannt. Die Identität Maschine wiederum – die durch Ambage H_1 Teil von Netdom 2 wird, verarbeitet diese Daten unter Rückgriff auf die Story SMH_1 in Netdom 3. Woraufhin sie im Rahmen der Ausgabe der erkannten Muster, also Ambage M_1 wieder eine Relation zum Menschen eingeht und Netdom 2 betritt. Der Mensch nimmt diese Ambage auf, wechselt in Netdom 1 und interpretiert die Muster der Maschine unter Rückgriff auf die Story SHM_1, wobei er hier nun vergleicht, ob die erzielten Ergebnisse aus seiner Sicht eine Verbesserung gegenüber seiner Referenzgröße, dem Zustand $t_{0,}$ darstellen und ob sie bereits

zufriedenstellend sind. Wiederholt der Mensch diesen Vorgang, so erhält er in SHM_n immer mehr verknüpfbare Referenzgrößen, ebenso wie die SMH_n der Maschine.

Whites Theorie spricht von Identitäten und hat dabei ein mehrstufiges Konzept im Hintergrund (White 2008: 10–12), was erlaubt, Mensch und Maschine gleichermaßen im Modell zu integrieren, ohne sie jedoch gleich zu machen. Die Identität Mensch ist fraglos entschieden komplexer als die Identität Maschine in diesem Fall, denn die Software ist für diesen Zweck entwickelt, der Mensch hingegen hat ein Leben darüber hinaus. Auch die Stories, mit denen die beiden die Kontrollversuche des jeweils anderen interpretieren, unterscheiden sich signifikant. Die Maschine arbeitet mit einem System statistischer Zusammenhänge, der Mensch interpretiert mit Blick auf aufbereitete statistische Zusammenhänge und geht im Zweifelsfall mit etwas Bauchgefühl darüber hinweg. Für diese Form des Schließens prägte bereits Hermann von Helmholtz Mitte des 19. Jahrhunderts den Begriff der »künstlerische[n] Induction« (Helmholtz 1896: 171), die überall dort Anwendung findet, wo die Gegenstände so komplex sind, dass sie sich nicht mehr klar überschauen lassen. Der Begriff zielt auf herausragende Kunstwerke, die den Prozess in Reinform veranschaulichen. Er wird aber auch überall dort gebraucht, wo die Schließenden ihre Urteile auf Basis ihres Instinktes oder »psychologischem Tactgefühl« treffen (Helmholtz 1986: 172). In der Folge steht die Frage im Mittelpunkt, wie maschinelles Lernen als Produkt der Relation zwischen Mensch und Maschine im Spannungsfeld zwischen dem Training und Lehren/Lernen mit bzw. durch Daten sowie deren Interpretation zu betrachten ist.

3. Formen des Lernens

Wir wollen nun zwei in der Machine Learning-Debatte prominent diskutierte Trainingsmethoden systematisch entlang des skizzierten Lernprozesses betrachten und aufbereiten: das *Supervised Learning* und das *Unsupervised Learning*. In beiden Lern- bzw. Trainingsformen möchten wir die Interdependenz der soziotechnischen Relationen in der Aushandlung von Kontrolle und Identität differenziert nachvollziehen.

Beide Lernformen sind nicht nur eingebettet in das Netdom »Machine Learning«, sondern bilden auf besondere Weise je eigene Netdoms, in deren Prozess weitere Netdoms aufgespannt werden. Verschachtelt wirken sie in

dynamischer Kohärenz von Struktur und Kultur auch in- und aufeinander ein (Abbildung 2). Hinzu kommt, dass die jeweiligen Identitäten und deren Kontrollbemühungen, wie beschrieben, zwischen verschiedenen Netdoms hin- und herspringen. Diese Switchings (White 2008: 7) wiederum beeinflussen die einzelnen Netdoms.

Für die differenzierte Ausleuchtung der Prozesse greifen wir auf neun von Mitarbeiter:innen des Lehrstuhls für Technik- und Organisationssoziologie der RWTH Aachen University 2020 geführte Expert:inneninterviews zum Thema Machine Learning zurück, in denen die Expert:innen auch Auskunft zu einigen Lern- und Trainingsmethoden und ihren Umgang mit diesen geben. Die Ausrichtung der Interviews ist allerdings allgemeiner gehalten, sodass die Aussagen hierzu eher am Rande der aufgespannten Narrative gefallen sind und mehr exemplarischen Charakter haben. Selbstverständlich fokussiert sich der Blick damit auf die Stories der Menschen.

Um nun die einzelnen Kontrollbemühungen in den jeweiligen Lernprozessen konkreter beschreiben zu können, werden diese zunächst als klar voneinander unterschieden markiert und dargestellt, um ihrer entlang von Identitäten engen und komplexen Vernetzung schließlich wieder Rechnung zu tragen. In der idealisierten Trennung werden wir uns auf den Umgang mit Daten als Interface beziehen. Daten werden in diesem Kontext nicht als Substrat objektiven Wissens (Cardoso Llach 2018) betrachtet (vgl. z.B. ebd.; Häußling 2020). Sie bilden vielmehr eine Schnittstelle zwischen den heterogenen Identitäten in und über die einzelnen Netdoms hinaus. Die (Ent-)Kopplungsprozesse werden im Folgendem entlang dieser Netdoms genauer beleuchtet, um die Trainings- und Lernprozesse in ihrer Komplexität als soziotechnische Aushandlung von Identität und Kontrolle auszuleuchten und so dem Einfluss der heterogenen Entitäten auf die Spur zu kommen. Dabei wird zunächst die Lernform Supervised Learning betrachtet, dem die Ausführungen zur Lernform des Unsupervised Learnings folgen.

Das Netdom »Supervised Learning«

Grundsätzlich geht es beim maschinellen Lernen um die Steuerung von Maschinen durch Daten. Die Maschinen sollen in den Daten Muster erkennen und diese wiederum auf neue, noch nicht bekannte Daten anwenden (vgl. z.B. Lenzen 2018; Sudmann 2018). Supervised Learning umfasst dabei eine Form des Lernens, bei der eine nicht bekannte Funktion von einem Algorithmus erlernt werden soll, die als Regel das »Verhältnis von Eingabe- und Ausgabewerten«

(Kaminski/Glass 2019: 130) aufstellt. Bei dieser Form werden dem Algorithmus bekannte, annotierte Eingabe- und Ausgabewerte zur Verfügung gestellt, aus denen die Maschine schlussendlich eine Hypothese ableiten können soll (ebd.). Ein wichtiger Aspekt dabei ist, dass die Hypothese, die sich dem »Verhalten der gesuchten Funktion« (ebd.) annähert, vom Nutzenden aufgestellt werden muss.

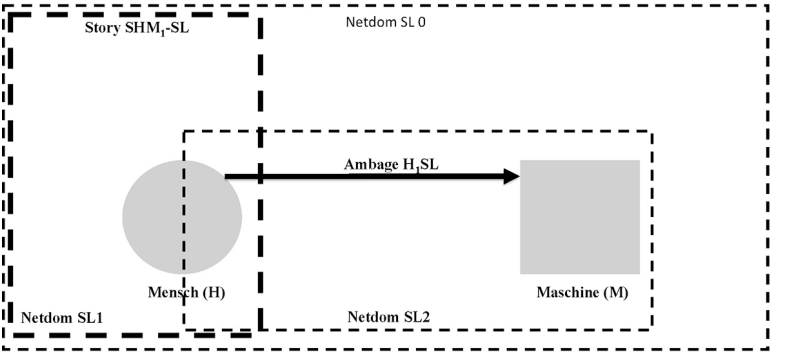
Folglich steht die Frage im Raum, wie groß der Einfluss des Menschen auf den Lernprozess als solchen überhaupt ist und wie stark er das künstlich intelligente Modell durch seine Entscheidungen prägt?

Innerhalb dieses Netdoms (Netdom 0/Supervised Learning (SL)) liegen zunächst Entscheidungen für diese Lernform von Seiten der entwickelnden, aber auch anlernenden menschlichen Akteur:innen (vgl. I04) zugrunde. Die Entscheidung für die Lernform des Supervised Learnings hängt dabei zum Teil von pragmatischen Grundlagen ab, die unter anderem auf der Menge der Daten und deren vermuteter Qualität fußen (vgl. I02, Pos. 48f.).

> »Also ich würde sagen, wenn man Trainingsdaten verfügbar hat, dann ist es wahrscheinlich schon sinnvoll, erstmal Supervised Ansätze zu probieren, weil die einfach relativ, gefühlt zumindest, robustere Ergebnisse dann direkt liefern.« (I07, Pos. 52–53)

Ist diese Entscheidung getroffen, wird eine nächste Einflussquelle im Netdom Supervised Learning (siehe Abbildung 3) offenbar: Der Mensch muss die Daten, die er dem Learner zur Verfügung stellen möchte, zunächst erst einmal annotieren, um das zu lernende Muster für die »Maschine« offensichtlich vorzuprägen. Die Ambage von Seiten des Menschen (H_1) erfordert in diesem Fall also ein hohes Maß an gezielter Vorbereitung des Lehrenden.

Abbildung 3: Netdom Supervised Learning – SL1 (t1)



Netdom SL (0/1/2/...): Netdom Supervised Learning und darin weitere verschachtelte sinnhafte Netzwerk-Domänen
SHM-SL: Einfluss der Mensch-Maschinen Relation auf die Storyentwicklung im Supervised Learning
Ambage H_1SL: Kontrollversuch des Menschen auf die Maschine im Supervised Learning Netdom

Durch diese engere Kontrolle des Menschen kommen im Fall des Supervised Learnings auch stärkere »menschliche Einschätzungen, Vorurteile oder so etwas rein […], andererseits gibt es das für alle Daten, dass die verfälscht sein können und das ist immer eine Frage, die man sich stellen muss, was die Datenqualität ist und was man davon erwarten kann.« (I02, Pos. 49)

Durch Daten erlernte, möglicherweise diskriminierende Muster könnten auch durch neue Datensätze wieder verlernt oder Schiefstellungen überschrieben werden (vgl. I15, Pos. 52–54). Im Supervised Learning könne gerade in der Reflexion dieses Umstandes durch die Lehrenden darauf reagiert werden.

> »Also einfach Trainingsdaten füttern, die das nicht haben oder man kann natürlich auch versuchen, systematisch gegenzusteuern, indem man systematisch trainiert […]. Das ist ja bei einem Menschen auch, wenn man etwas falsch gelernt hat. In der Psychologie, wenn Leute irgendwelche Ängste haben, dann sagt man ja auch: Konfrontation mit der Angst und irgendwie darüber lernen.« (I15, Pos. 52)

Die Vorbereitung der Datensätze für die Ambage H_1SL, verbunden mit der Verantwortung, Schiefstellungen zu identifizieren, wird für die Akteur:innen jedoch dadurch erschwert, das fachfremde Daten nicht leicht zu bewerten sind, gerade wenn diese von Expert:innen anderer Disziplinen annotiert wurden.

> »Man sieht es natürlich, wenn man jetzt irgendwie sagt, er [z. B. ein Biologe] soll jetzt die Struktur ganz genau umranden, oder so, und hat dann einfach irgendwie bisschen schludrig dann irgendwie gezeichnet, dann ist das ein relativ offensichtlicher Fehler, den man auch leicht ausmerzen kann. Teilweise ist es auch so, dass die Bilddaten im Prinzip das nicht hundertprozentig hergeben, dass man es wirklich exakt annotieren kann.« (I10, Pos. 71)

Wie hoch die Genauigkeit der Annotation der Daten ist, kann zum Teil nur durch Experten:innen begutachtet werden. Das wäre der »Idealfall, dass man einfach genau das gleiche Bild, die gleiche Aufgabe einfach verschiedenen Experten gibt und dann so eine Art interrater reliability quasi ausrechnen kann, um dann eben zu sehen, wie stark schwankt im Prinzip die Annotierungsgenauigkeit zwischen den Experten.« (I10, Pos. 73)

Dies ist aber »immer nur begrenzt möglich, weil natürlich auch die Arbeitszeit von diesen Experten teuer« (I10, Pos. 73) ist.

Die Datenannotation ist damit Teil der Storyentwicklung (SHM_1-SL) im Netdom Supervised Learning (SL1), bei der festgelegt wird, mit welchen Daten in welcher Form der Learner trainiert werden soll. Die Verkopplung und Entwicklung einer Story findet dabei nicht nur zwischen den Lehrenden und der »Maschine« statt, sondern wird zum Teil auch in transdisziplinären Teams geschrieben:

> »Also gerade in Kooperation mit Biologen, die dann eben die Daten im Labor erstmal aufnehmen. Dann im Prinzip uns die Rohdaten geben […]. Genau, dann geht es im Prinzip los, dass der Biologe quasi sagt: »Die und die Strukturen interessieren mich«. Und dann basteln wir zum Ersten quasi Tools, mit denen man diese Annotierung überhaupt machen kann […], sowas wie Paint oder Photoshop in die Richtung. Dass man wirklich auf den Bildern quasi zeichnen kann und dann eben so eine Art Maske sich generieren kann von den Objekten, die man haben will. Genau, das ist so der erste Schritt, dass man einfach mal weiß, was will der Biologe überhaupt haben und dann dafür eben dann diese Label erstellt.« (I10, Pos. 12)

Mit der Einflussnahme des Menschen auf die Daten im Prozess der Annotation ist eine »Verantwortung« (I04, Pos. 57) für den gesamten Entwicklungsprozess verbunden. In dieser Phase scheint der Mensch in der Relation zwischen ihm und der »Maschine« die Prozesse zu dominieren.

Der Einfluss der Daten auf die Mustererkennung insgesamt ist im Sinne der Storyentwicklung von den lehrenden Akteur:innen intendiert, aber was an

Mustern vom Learner erkannt wird, also über das Kontrollprojekt der Ambiguity H1-M1-SL, und wie diese auf andere Daten angewendet werden können, ist gerade nicht ersichtlich.

> »Der Algorithmus wird mit Trainingsdaten trainiert, um das zu lernen, was der in den Trainingsdaten sieht. Das ist natürlich nicht unbedingt intendiert, was da rauskommt. Es ist auch nicht intendiert, dass er, sagen wir mal, auch ungewünschte Schiefstellungen in den Trainingsdaten übernimmt. Aber das passiert eben, das ist so. Und das ist auch problematisch.« (I02, Pos. 31)

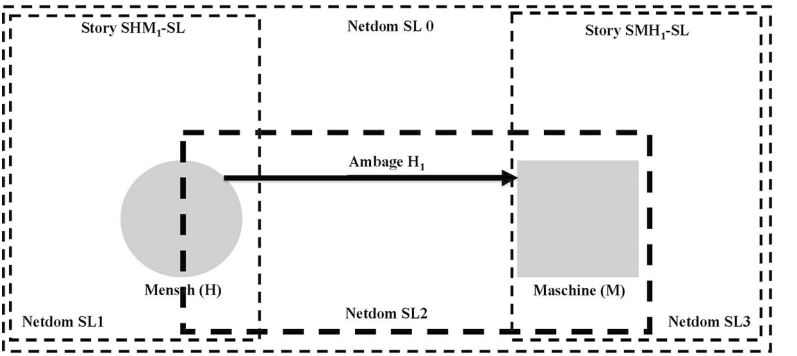
Lehrende Akteur:innen führen damit gerade bei der Datenannotation konsequent auch die Möglichkeit von nicht-intendierten Nebenfolgen mit:

> »Da hat man halt Schwierigkeiten […], dass sich Vorurteile verschieben und verstärken und ähnliches. Das ist schwierig in dem Sinne, weil man das nur sehr schwer kontrollieren kann und das dann auch Dinge sind, die man selber nicht gut versteht. Das ist schwierig, wenn sich diese Sachen tatsächlich verstärken zum einen und dann durch etwas durchgehen, wo man vielleicht eine höhere Objektivität erwartet.« (I02, Pos. 35)

Das Problem des Overfittings durch spezielle Daten(-sätze) und damit die mögliche Verstärkung eines vorhandenen Biases (vgl. I02, Pos. 57) wird bei dieser Lernform erkannt, aber dem soziotechnischen Lernprozess im Supervised Learning förmlich inhärent und als nicht einfach aufzulösen gekennzeichnet. Der Lernprozess im Netdom Supervised Learning kann damit auch als eine Form des Interfacings in Anlehnung an Karafillidis (2018) betrachtet werden, in den sich u.a. die bereits vorab soziotechnisch konstruierten und zu keiner Zeit rohen Daten (vgl. Häußling 2020) als Schnittstellen in die entwickelten Modelle sowie die mit ihnen entworfenen Stories einschreiben. In dieser Daten-Ambiguität und der daran anschließenden tieferliegenden und selbsttätigen Mustererkennung durch den Learner ist der Einfluss des Menschen über die Lernform des Supervised Learnings als Kontrollversuch beschreibbar (Ambage H1 in Netdom SL2). Am Beispiel der Spracherkennung wird beim Supervised Learning der uneinsichtigen Seite des Audiosignals der Spracherkennung durch Vertextung zu begegnen versucht, das heißt, es liegen beim Supervised Learning sowohl die Audiosignale als auch die Transkriptionen dieser vor. Bei der Arbeit mit einem Unsupervised Learner würden eben jene Verschriftlichung fehlen (vgl. I07, Pos. 74).

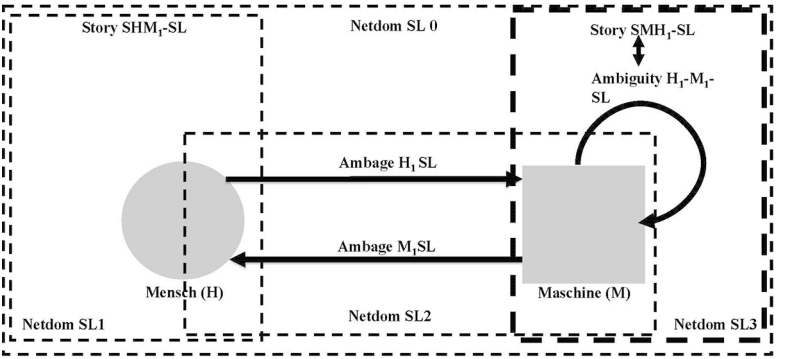
Zurück zu unserer Abbildung. Auf der Basis der entwickelten Story (SHM_1-SL) im Netdom SL1 wird im Idealfall ein »Trainingsdatenset« für die Entwicklung eines Modells zusammengestellt und dem Learner zur Verfügung gestellt. Damit wird im Prozess des Lernens in das Netdom SL2 geswitcht (Ambage H_1SL im Netdom SL2).

Abbildung 4: Netdom Supervised Learning – SL2



Die Ambage H_1 durch den Menschen über die Datensätze im Netdom SL2 geht über in den Prozess der Verarbeitung der Datensätze, der Mustererkennung, durch den Learner. Dieser Prozess bedeutet ein erneutes Switching in das Netdom SL3. Hier wird die Verarbeitung der Daten im Sinne der Ambiguity H_1-M_1-SL verortet. Wie und auf welche Weise Muster durch den Learner erkannt werden, bleibt dabei jedoch intransparent. Die Opazität der Datenverarbeitung wird zwar von Seiten des Menschen reflektiert, kann aber nicht aufgelöst werden. Die Datenverarbeitung und Mustererkennung inklusive der Intransparenzen werden folglich durch ein erneutes Switching in das Netdom SL2 als Ambage M_1SL wirkmächtig. In unserem Abbildung lässt sich der Prozess wie folgt illustrieren:

Abbildung 5: Netdom Supervised Learning – SL3



Der Mensch rechnet gewissermaßen mit der im Netdom SL3 entstehenden Intransparenzen bei der Mustererkennung, welche über die Ambage M_1SL in irgendeiner Art und Weise durch die Datenausgaben wirkmächtig werden.

Die Daten und deren materielle Repräsentation in Datenpaketen, deren Design[2], werden im Netdom SL1 vom Menschen interpretiert. Dieser Vorgang kann als Ambiguity M_1-H_1-SL beschrieben werden. Die soziotechnisch konstruierten Datenpakete sind in diesem Kontext als Interfaces zu verstehen (i.A. Cardoso Llach 2018; Häußling 2020). Es kann damit zudem eine Diskrepanz zwischen den bereits konstruierten Datensätzen in Abhängigkeit von der Story SHM_1-SL und der Datenannotation durch inter-, trans- oder intradisziplinäre Teams, die zur Visualisierung bereitstehen, entstehen.

> »Wir bekommen dann letztendlich die Zeiten ausgegeben […]. Also das sind dann schon, ja ich nenn sie mal nicht Schaltpläne, die wir dann ausgegeben

2 Design soll in diesem Kontext in Anlehnung an Häußling (2012, 2016) als das »Erfinden neuer Sozialpraktiken« (Häußling 2012: 283) betrachtet werden, weshalb hier von einem Arrangement gesprochen wird (vgl. ebd.: 283). Mit den bei Debray entlehnten Begriffen der »›organisierte[n] Materie‹ und ›materialisierte[n] Organisation‹« (Häußling 2016: 33) kann die Mehrdeutigkeit der Materialität in diesem Prozess differenzierter betrachtet werden. Organisierte Materialität umfasst dabei die Manifestation von Symbolen in Form von Materie und materialisierte Organisation hingegen beschreibt die Übertragung von Ideen durch beispielsweise Technik (vgl. ebd.: 34). Der erste Begriff bildet die Seite der Materialität, die die »Übersetzung der menschlichen Welt für die Technik leistet« (ebd.: 38). Im zweiten Fall wird auf die Einsetzung eines gestalteten Objektes für »menschliche und soziale Belange« (ebd.: 39) verwiesen.

> bekommen, aber wir bekommen dann halt ein interpretationsfähiges Bild, mit dem wir dann halt weiterarbeiten können. Und wenn wir das dann wieder auf so einen Schaltplan umrüsten […] [und] man die Schaltzeiten an den und den Positionen ändert, [dass] ist dann, ja, also ein sehr, sehr interpretationsfähiges Bild, auch mit sehr vielen Diagrammen und Graphen, damit man das dann auch, ich sag mal an die Menschen, die da an den Entscheidungshebeln sitzen, herantragen kann.« (I06, Pos. 50)

Zwischen der Datenausgabe und den daraus gezogenen Informationen entsteht in diesem Fall eine Differenz (Koren/Klamma 2018), die ausgeglichen werden muss, damit Daten interpretierbar bleiben. Es liegt hier also nahe, dass zwischen den Learnern in beiden Netdoms über (un-)bestimmte Datenpakete zudem weitere (technische) »Übersetzer« wie zum Beispiel technische Analytic Tools (ebd.) stehen, die die (Weiter-)Verarbeitung von Informationen erst möglich machen.

Das Design der Datenpakete nimmt im soziotechnisch konstruierten Supervised Learning aufgrund der Selbstorganisiertheit der Learner in der Verarbeitung der Datensätze, der Ambiguity H_1-M_1-SL, einen weiteren Einfluss auf die Arbeit am und mit dem Lernmodell ein und offenbart weitere soziotechnische Ver- und Entkopplungsprozesse, die auf den gesamten Lernprozess wirken können. Auch an dieser Stelle offenbart sich ein Übersetzungsproblem zwischen »Maschine« und »Mensch«, dass sich im Kontext künstlich intelligenter Technologie als ein Problem des Berechnens von etwas schwer Berechenbarem zeigt.

> »Also die Idee des ganzen Ansatzes ist gerade die, dass man eben nicht von Anfang an versucht, als Mensch Kontrollstrukturen aufzubauen, weil man sowieso immer etwas vergisst und das einfach nicht hinkriegt.« (I02, Pos. 47)

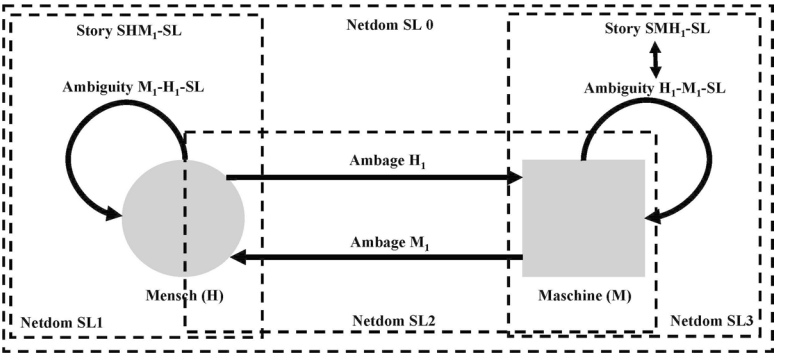
Das Problem des Verstehens, was auf welche Weise berechnet wurde, kann dabei auch im Design des Modells selbst verankert sein (vgl. I02, Pos. 47).

Dieser »zirkuläre Prägungszusammenhang« (Häußling 2012: 265) zwischen den Datensätzen, und den »Ergebnissen« künstlich intelligenter Prozesse ist in Bezug auf künstliche Intelligenz nicht nur, dass durch die Anknüpfung an die interpretierten Daten neue Prozesse gestartet, sondern auch wie diese verstanden und interpretiert werden, um unter anderem KI-Technologie weiterzuentwickeln. Idealerweise können die lehrenden Akteur:innen die Ergebnisse auch durch ein »Validierungsdatenset« prüfen, und

so festzustellen, ob sich das Modell »wirklich auch so halbwegs generalisieren« (I10, Pos. 20) lässt und zufriedenstellende Ergebnisse ermöglicht.

Wieder gibt der Mensch Daten kontrolliert vor, mit deren Hilfe das durch die Maschine entwickelte Modell überprüft werden soll.

Abbildung 6: Netdom Supervised Learning t_2



Hier schließt ein weiterer Zyklus des Lernens an, der über die Switchings in die einzelnen Netdoms im Netdom Supervised Learning (SL 0) zu einem zweiten Zeitpunkt (t_2) führt und so die wechselseitigen Kontrollbemühungen der heterogenen Entitäten umfasst.

Die Lernprozesse werden durch den Menschen im Netdom SL1 zu t_2 über die Ambiguity M_2-H_2-SL erneut interpretiert. Danach schließt sich im besten Fall eine weitere idealtypische »Schleife« der soziotechnischen Konstruktion des Supervised Learnings an. Mithilfe eines unabhängigen »Testsets« im vorgezeichneten Problemhorizont der Story wird eben jenes entwickelte Modell evaluiert und das Ergebnis nach seiner Qualität überprüft (vgl. I10, Pos. 20). Dabei hat der gesamte Lernprozess des Supervised Learnings Einfluss auf die darin fortgeschriebene Story, welche zu jedem Zeitpunkt überprüft und wenn nötig auch angepasst wird.

Die Frage der Kontrolle ist damit eine Frage der Relation und wie bereits angesprochen, keine der Objektivität.

> »Wir haben jetzt diese Modelle und wir haben keine Ahnung, wie die funktionieren. Jetzt können wir aber damit, mit den Modellen, auch wieder arbeiten und uns zum Beispiel vorstellen, dass wir Fragen an diesem Modell stel-

> len, so wie wir Fragen an eine Datenbank stellen. Zum Beispiel so eine Frage wie: Was passiert, wenn ich die Eingabe hier ein bisschen ändere?« (I02, Pos. 47)

Auf die mangelnde direkte Kontrolle von Seiten des Menschen auf die Datenverarbeitung der Learner, der Ambiguity M_3-H_3-SL, werden strategische Versuche gestartet, die im Mindesten Einfluss auf die Ausgaben generieren sollen.

Wenn das Modell nach den Prüf- und Testungen nicht funktionieren sollte, werden Problemanalysen im Netdom SL1 angestellt. Da komme es auf die Expertise der Lehrenden an.

> »Wenn man jetzt irgendwie einen Fehler beim Training oder sowas hat, [muss man] eben einfach so ein Gespür zu haben: an welchen Stellschrauben muss ich jetzt drehen, damit ich den und den Fehler beheben kann. Das ist, glaube ich, einfach Erfahrung oder eben auch viel Literaturrecherche. […] Was haben andere gemacht in der Richtung, was sind vielleicht irgendwie die typischen Probleme, die eigentlich immer bei der speziellen Architektur auftreten.« (I10, Pos. 55)

Bei sehr großen Modellen besteht die Gefahr, dass man »zu viele Stellschrauben hat« (I10, Pos. 22). Das könne zur Folge haben, dass die zur Verfügung gestellten Daten vom Modell schnell auswendig gelernt würden und so nur für das Trainingsdatenset gut funktionieren, »aber nicht auf dem Testset« (ebd.). Eine Möglichkeit, dem zu begegnen, wäre, das Modell anzupassen, es beispielsweise zu verkleinern, damit es »eben gar nicht die Möglichkeit hat, das Trainingsdatenset auswendig zu lernen, sondern gezwungen ist, so ein bisschen allgemeinere Repräsentation der Daten zu finden« (I10, Pos. 22).

Auch die Veränderung der Trainingsdaten kann eine Veränderung des Modells erzeugen:

> »Was auch häufig gemacht wird, ist, dass man sogenannte Datenaugmentierung benutzt. Das heißt, man nutzt die Trainingsdaten, die man schon hat und versucht die irgendwie zu variieren, dass man beispielsweise einfach die Bilder rotiert oder irgendwie noch ein bisschen Rauschen hinzufügt oder die Helligkeit ändert und so weiter, dass man quasi verschiedene Variationen, die so in der Realität auch auftauchen können, versucht, mit reinzunehmen.« (I10, Pos. 22)

Der Prozess der Fehlersuche sei jedoch sehr komplex und gleiche oft eher einer Suche über »Trial-and-Error« (ebd., Pos. 24).

> »Und oft ist es auch so, dass man sich nicht so hundertprozentig erklären kann, woran es eigentlich liegt. Da diese neuronalen Netze eben so komplex sind mittlerweile, dass man eigentlich fast keine Chance hat, das wirklich im Detail zu verstehen, was da genau passiert.« (Ebd.)

Spätestens an dieser Stelle wird die soziotechnische Konstruktion des Lernprozesses der Form des Supervised Learnings besonders deutlich herausgestellt.

Hinzu kommt, dass Entwickler:innen die Daten mithilfe umfassender Dokumentationen zur Unterstützung der Interpretation der Datenverarbeitungsprozesse durch Dritte einsetzen. Sie sprechen hier von einer Art »Veredelung« (vgl. I06, Pos. 42) der Daten, die unter anderem verschiedene Datenquellen und Beschreibung der Datenform ausweisen (vgl. ebd., Pos. 70). Damit konstatieren sie erneute Kontrolle im Sinne der Ambage. Entwickelnde Akteur:innen sprechen bei dieser Lernform unter anderem davon, die »Stellhebel in der Hand« (vgl. ebd., Pos. 54) zu halten, sowohl bei der Interpretation der Daten und der Ableitung der Entscheidungen als auch dabei, die daraus generierten Informationen für Dritte, bspw. für eine spezifische Kundschaft, transparent und verständlich zu machen.

Die Ambiguity, welche im Umgang mit den Datenausgaben vorliegt, wird durch die entwickelnden Akteur:innen über die Einflussnahme in Abhängigkeit von den Stakeholdern oder anderen Zielsetzungen vereindeutigt und so wiederum kontrolliert.

Wir können konstatieren, dass der Einfluss der Akteur:innen im Netdom Supervised Learning das Kontrollprojekt Ambage aus Sicht der Akteur:innen prägend und durch die spezifische Datengrundlage mit eindeutigen Mustern dominant wahrgenommen wird. Die Kontrollbemühungen des Menschen an den vielfältigen Stellen im Lernprozess offenbaren damit eine gewisse Eingriffstiefe in die technologischen Abläufe beim Supervised Learning. Dennoch wird offensichtlich, dass der Lernprozess über die verschiedenen Netdom-Switchings als eine Form des Interfacings die soziotechnische Konstruktion vorprägt und diese wechselseitig miteinander verzahnt ist.

Das Netdom »Unsupervised Learning«

Bei der Lernform des Unsupervised Learnings werden die Learner strenggenommen nicht in der oben beschriebenen Weise durch zuvor definierte Ein- und Ausgaben (vgl. Lenzen 2018) angeleitet, wie und auf welche Weise Muster in den Daten erkannt werden sollen. Vielmehr werden Daten durchsucht und der Versuch unternommen, darin »selbstständig« Muster zu identifizieren (vgl. ebd.).

Die Entscheidung für diese oder eine andere Lernform fällt vordergründig auf der Basis der vorliegenden Daten. Liegen große Datenmengen vor, die »nicht irgendwie schon mal von einem Menschen angefasst« (I02, Pos. 49) wurden, kann der Versuch unternommen werden, mit einem Unsupervised Learning Ansatz aus den Daten »trotzdem etwas« (ebd.) herauszuholen. So können zum Beispiel »sogenannte Auto-Encoder« (vgl. I10, Pos. 51) verwendet werden, um die Netzwerke auf die Lernprozesse inhaltlich einzustellen. Dabei werden zum Beispiel Bilder eingegeben, die es ermöglichen sollen, später »genau das gleiche Bild« (ebd.) wiederzuerkennen. Dieses Training kann dann eine Basis für einen anschließenden Supervised Learning-Pfad bilden.

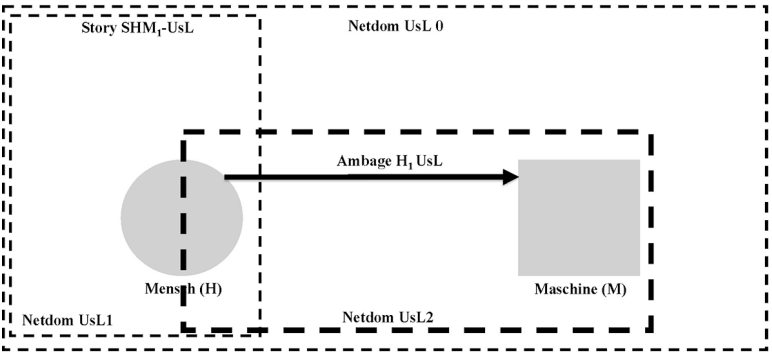
> »Man forciert im Prinzip das Netzwerk, dass es eine komprimierte Repräsentation quasi von diesem Bild lernt, und durch diesen Prozess kann man im Prinzip so eine Art Vortraining von beispielsweise einem neuronalen Netz hinbekommen, das man dann wiederrum verwenden könnte.« (I10, Pos. 51)

Auch beim Unsupervised Learning steht zu Beginn des Lernprozesses die Entwicklung einer Fragestellung, die als Rahmen für die Story SHM_1-UsL von Seiten der Akteur:innen im Netdom UsL1 gestellt wird, diese ist aber viel weniger orientiert als beim Supervised Learning. Doch sinkt damit auch der Einfluss des Menschen auf den Lernprozess im Unsupervised Learning?

Denn im Vergleich zum Supervised Learning, bei dem eine vorherige Annotation der verwendeten Daten notwendig ist, wird beim Unsupervised Learning hingegen der Versuch unternommen, direkt etwas aus »unbearbeiteten« Daten zu lernen, die jedoch, niemals in roher Form vorliegen (s.o. vgl. Häußling 2020).

Wie auch schon im Kapitel Supervised Learning dargestellt, findet auch bei dieser Lernform ein Netdom-Switching in das Netdom UsL2 statt, in dem eine deutliche Ambage (H_1UsL) der Akteur:innen auf die Learner einwirkt.

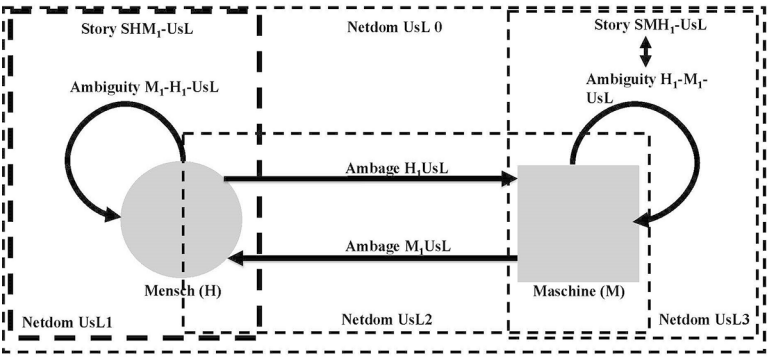
Abbildung 7: Ambage H_1UsL im Netdom UsL2



Netdom UsL (0/1/2/...): Netdom Unsupervised Learning und darin weitere verschachtelte sinnhafte Netzwerk-Domänen
SHM-UsL: Einfluss der Mensch-Maschinen Relation auf die Storyentwicklung in der UsL Netzwerk-Domäne
Ambage H_1UsL: Kontrollversuch des Menschen auf die Maschine im Unsupervised Learning

Eine große, noch nicht klar bearbeitete Datenmenge wird von Seiten der Akteur:innen an die »Maschine« übermittelt, ohne zuvor Muster durch annotierte Daten vorzugeben. Die Learner verarbeiten diese Datensätze und erkennen möglicherweise Muster in den Daten, was als Kontrollprojekt Ambiguity H_1-M_1-UsL im Netdom UsL3 in unserem Lernmodell beschrieben wird.

Abbildung 8: Ambiguity M_1-H_1-UsL



Diese Ver- und Bearbeitung der Daten und insbesondere die Mustererkennung wird wiederum über die Ambage M_1UsL im Netdom UsL2 zurück an die Akteur:innen gespielt, welche vom Menschen im Netdom UsL1 gedeutet wer-

den muss. Die Herausforderung besteht auf Seiten der Menschen gerade darin, dass die durch die Maschine in den Datensätzen erkannten Muster von den anlernenden Akteur:innen selbst nicht mehr erkannt werden können.

> »Die Maschine sagt, das Muster ist da, dann ist das natürlich problematisch. Wenn ich dann aber ausprobieren kann, ob da was ist, weil ich dann ein Experiment machen kann, dann gehe ich diesem Problem aus dem Weg. Wenn ich mich nur auf das verlasse, was die Maschine erkennt, ist das schwieriger.« (I02, Pos. 17)

Die Ambiguity H_1-M_1-UsL im dritten Netdom des Unsupervised Learnings der durch die Learner erkannten Muster, ist folglich eine besondere Herausforderung bei der erneuten Interpretation der Daten durch die anlernenden Akteur:innen, die durch den Prozess der Visualisierung noch einmal verschärft wird, wenn es zum Beispiel um die Eindeutigkeit der Muster geht.

> »Man muss immer ein so ein bisschen tricksen, dass man im Prinzip die Sachen dann auch rausfindet, die man haben will. Also beispielsweise, wenn man jetzt eine Clusterung durchführt, weiß man ja im Prinzip gar nicht, welches Cluster ist was, was bedeuten die Cluster und so weiter. Man hat halt irgendeine Gruppierung im Raum von seinen Daten, muss die aber dann später nochmal irgendwie weiterverarbeiten und eben, ja, rausfinden, was was bedeutet.« (I10, Pos. 53)

Die über die einzelnen Netdoms im Unsupervised Learning hinweg vorhandene wechselseitige Mehrdeutigkeit und die gegenseitig einwirkende Ambage zeigt einen soziotechnischen Lernprozess über verschiedene Netdoms hinweg. Die Kunst des Lernens in der Lernform des Unsupervised Learnings ist damit stark von einer Annäherung durch die Vernetzung von Interpretation gekennzeichnet.

Damit rücken die Auswahl, Interpretation und Verarbeitung der Daten in den Blick, wenn es um die Bedeutung von Intransparenzen bei der Weiterverarbeitung der spezifischen Daten in den verschiedenen Schleifen der wechselseitigen Kontrollprojekte zwischen Mensch und Maschine geht. Es ist nicht nachvollziehbar, was in der Verarbeitung, zum Beispiel der Clusterung von Daten auf der Seite der Maschine passiert, was ein neues Problem der Intransparenz darstellt (vgl. Schmitt/Heckwolf in diesem Band).

> »Also vielleicht in ganz kritischen Bereichen kann man die Systeme nur als Zwischenschritt verwenden, das sind auch Möglichkeiten. Dass man am Ende, das was man dann wirklich in der sicherheitskritischen Anwendung einsetzt vielleicht gar nicht mehr so eine Black-Box ist, sondern die Black-Box verwendet, um das andere zu bauen.« (I02, Pos. 47)

Die Aufnahme der Ambage M_1UsL über das Netdomswitching von N-UsL2 zum Netdom UsL1 kann im Prozess der Interpretation der Datenausgaben einen weiteren Kontrollversuch (Ambiguity H_1-M_1-UsL) vor der Ambage H_2UsL bedeuten. Da die vorständige Kontrolle der Datenverarbeitung der KI durch den Menschen nicht möglich ist, kann der Versuch unternommen werden, interpretationsfähige Daten darüber zu erhalten, was die KI überhaupt identifiziert hat und worauf sie scharf stellt.

Geht man davon aus, dass künstlich intelligente Modelle über ein Unsupervised Learning Muster in den Strukturen erkennen, besteht die Möglichkeit, systematisch Daten zu manipulieren, und zwar so, dass diese Manipulationen allein vom Modell und nicht für den Mensch wahrgenommen werden können, wie zum Beispiel feinste Farbveränderungen in Bildern. Die Frage ist an dieser Stelle wiederum, welche Auswirkungen dies für die Ambage durch die Maschine (M_nUsL) und den Deutungsversuch des Menschen, der Ambiguity M_n-H_n-UsL hat.

Veränderungen der Daten bewirken klare Veränderungen in den Mustererkennungsprozessen der Learner, die wiederum Datenausgaben weitergeben, welche vom Menschen interpretiert werden müssen. Das Rechnen mit der Intransparenz in der Verarbeitung der Daten durch die künstlich intelligenten Modelle, kommt damit in gewisser Weise einem neuen Kontrollprojekt, dem Blackboxing, gleich (vgl. Schmitt/Heckwolf in diesem Band), das im Kontext der Ambiguity sowohl im Netdom 1 des Unsupervised Learnings (UsL1) als auch im Netdom UsL3 wirkmächtig werden kann. Dies ist bei dieser Lernform besonders stark, weil hier keine Datenannotation stattfindet und keine Muster vordefiniert werden, die einen deutlicheren Einfluss durch den Menschen in der Bearbeitung und Orientierung der Daten sowie der Learner bedeutet.

Doch die Reinform des maschinellen Lernens findet in der Forschung und Entwicklung wenig Anwendung. Oft liegen Kopplungen des Unsupervised mit dem Supervised Learning vor, sodass die Kunst des Lernens nicht zuletzt auch eine Kunst des Lehrenden ist, die spezifische Ver- und Entkopplungen der Lernformen miteinander je Kontext anzuwenden.

4. Die Kunst des Lehrens

Die Kunst des Lernens präsentiert sich idealtypisch differenziert, in Abgrenzung der beiden Lernformen zueinander, als einen wechselseitigen Prägungszusammenhang von Kontrollversuchen heterogener Entitäten. Häufig kommt es jedoch zu einer Kopplung der beiden Lernformen. Die beiden Netdoms SLo und UsLo treten dabei in neue Kopplungen, die eine komplexe Verschachtelung der Switchingprozesse in und über die Netdoms hinweg zufolge haben. Um diese komplexen wechselseitigen Ver- und Entkopplungsprozesse der Lernprozesse nachvollziehen zu können, müssen die Entscheidungsprozesse auch im Kontext der Entwicklung der jeweiligen Stories, und damit verbunden für die Beziehung der beiden Lernformen, betrachtet werden. Damit rückt nach der schon beschriebenen Kunst des Lernens die Kunst des Lehrens in den Fokus.

Je nach Forschungsfrage werden im idealen Fall unterschiedliche Lernformen miteinander verbunden, sodass zu Beginn die Storyentwicklung steht, also welche Funktion die künstlich intelligenten Modelle erfüllen sollen. Erst daran orientiert, werden die einzelnen Prozesse und Lernformen je nach Bedarf gekoppelt.

> »Das heißt, das wäre im Prinzip so eine Mischung aus beidem. Das man einmal die Unsupervised Methode benutzt, um die Trainingsdaten zu generieren, also um diesen Simulationsansatz quasi durchzuführen und dann diese Trainingsdaten, die wir da generiert haben, diese künstlich erzeugten, dann im Prinzip für einen Supervised Ansatz benutzt und dabei im Prinzip genau das umgedrehte lernt. Also, dass man vom Rohbild auf die Maske kommt, das dann aber wiederrum auf die realen Daten später anwendet, um dann aus den realen Daten die Masken rauszubekommen. […]. Also im Prinzip so ein bisschen vertrackt, also das ist so eine Mischung aus verschiedenen Ansätzen, die aber […] das eigentliche Ziel eben haben, aus dem Rohbild Daten rauszukriegen.« (I10, Pos. 53)

Je nach Fachgebiet wird in der Forschung der Versuch unternommen, das Verhältnis der beiden Lernformen zu erschließen und im Einsatz zu prüfen. In Abhängigkeit zur Fragestellung werden die Lernformen nicht nur ergänzend eingesetzt, sondern auch in grundsätzlicher Beziehung zueinander erforscht. Im Bereich der Spracherkennung wird das »reine« Unsupervised Learning als »kryptographisches Problem« (I07, Pos. 74) betrachtet. Dieses Problem wird

theoretisch im Bereich der informatischen Spracherkennung untersucht, aber zumeist durch die Kopplung von Unsupervised und Supervised Learning gelöst (ebd.). Die in diesem Fachgebiet angesprochene Verbindung der beiden Lernformen als mögliche Lösung der Herausforderungen deutet zum einen die technischen Schwierigkeiten des reinen Unsupervised Learnings an. Zum anderen werden auch die geführten wissenschaftlichen Diskurse davon beeinflusst und Theoriebildung dadurch angeregt. Interessant ist gerade die Ergänzung von anderen Lernformen im Netdom. Reigeluth und Castelle (2021) haben bereits darauf hingewiesen, dass diese Lernform den Anschein erweckt, ohne »Lehrende« auszukommen. Die Autoren betonen jedoch die kulturelle Einbettung des Lernens und damit auch die Einbettung von Learnern in sozio-kulturelle Arrangements und konstatieren, dass Lernen ohne Anleitung im strengen Sinn unmöglich ist.

Im Prozess des Anlernens gehe es aber gar nicht um das Verstehen, was der Learner genau mit den Datensätzen macht, wie Muster generiert werden, sondern es gehe vielmehr um die Arbeit mit den entstandenen Modellen (vgl. I02, Pos. 47). Damit werden die strategischen Entscheidungen gestärkt, also welche Sets im Supervised und welche Modelländerungen beim Unsupervised Learning vorgenommen werden müssen.

Wir stehen damit vor der Relevanz der Frage nach dem Vertrauen in einerseits die »richtige« Verarbeitung von Daten in den Modellen und der Bedeutung der Intuition und Erfahrung in der Arbeit an den Modellen:

> »Also ich glaube im Moment sind wir im Wesentlichen noch in einem Stadium, wo das erfahrungsgetrieben ist. Also, Leute, die da arbeiten, wissen ungefähr, was sie tun können, und manche wissen es besser und sind am Ende damit erfolgreicher, und das ist zum Teil mysteriös. Und das ist natürlich ein Zustand, der auch nicht so doll ist, ist aber wiederum etwas, worum sich viele Leute Gedanken machen, wie man das vernünftiger machen kann, dass diese Technik letztlich einfacher einsetzbar ist. Da ist man noch ein Stück weit von entfernt.« (I02, Pos. 51)

Andererseits offenbart es auch die Bedeutung der Arbeit mit den Intransparenzen, einem Blackboxing nach Schmitt und Heckwolf (in diesem Band) in der wechselseitigen Aushandlung von Identität und Kontrolle zwischen Mensch und Maschine sowie die Herausforderung der Bedeutung eines grundsätzlichen bestehenden Objektivitätsanspruches von quantitativen Daten, in Anlehnung an Porter (2020) Rechnung zu tragen. Die historische

Quantifizierung in den Wissenschaften als Stärkung einer idealen mechanischen Objektivierung, des nach exakten Methoden entwickelten Wissens, ist verbunden mit dem Wunsch Expert:innenurteile zu ersetzen. Künstlich intelligente Modelle scheinen hier in klarer Linie einer idealen Objektivität zu stehen. Doch wie wir zeigen konnten, ist ganz im Sinne Porters, das »Problem des Vertrauens« (Porter 2020: 214) nicht einfach aufzulösen.

Bereits Helmholtz hatte darauf verwiesen, dass Wissenschaften, die sich der künstlerischen Induktion bedienen, gut beraten sind, zunächst »die Glaubwürdigkeit der Berichterstatter, die ihnen die Thatsachen überliefern« (Helmholtz 1896: 172), also die Entscheidungsgrundlagen liefern, zu prüfen. In gewisser Weise ist die Arbeit an der Wirklichkeit, der Arbeit an und mit künstlich intelligenten Modellen kunstfertig konstruiert (vgl. Porter 2020: 5).

Beispielhaft kann man hier ChatGPT und dessen Entwicklung seit seiner Veröffentlichung im November 2022 heranziehen, bei der sich sowohl Intransparenzen in den letzten Updates der Modelle (ChatGPT 3.5 zu ChatGPT4) zu offenbaren scheinen als auch qualitative Einbußen bemerkt wurden (vgl. Chen et al. 2023: 14f.), welche möglicherweise auch auf den (intentionalen) Eingriff der Entwickler:innen zurückgeführt werden könnten. Diese Veränderungen führten bereits zu einem stärkeren Misstrauen in die Entwicklungsarbeit der Technologie und führten dazu, dass Fachleute zu Skepsis gegenüber den Ergebnissen von ChatGPT aufrufen (vgl. DLF Nova, 24.07.2023). Die Glaub- und Vertrauenswürdigkeit der Entwickler:innen und folglich auch der Ergebnisse des ChatBots selbst scheint dadurch zu leiden und lenkt zudem den Fokus auf die Rolle der Entwickler:innen KI-basierter Technologie.

Die Kunstfertigkeit der Lehre liegt damit nicht nur in der Reflexion der Kunstfertigkeit selbst, sondern im Design eines soziotechnisch konstruierten Lernprozesses, der Lernformen mit- und zueinander, der Daten und Datendokumentation nur in der Kopplung von Mensch und Maschine kennzeichnet. Dies zeigt sich nicht nur darin, dass das Berechnen des nicht Berechenbaren ohne das nicht Berechenbare wirklich ausrechnen zu wollen, von einer soziotechnisch konstruierten künstlichen Intelligenz ausgeht, die sich von der menschlichen unterscheidet, aber nicht ohne diese auskommt. Sondern sie führt gleichzeitig die Bedeutung vor Augen, wie soziotechnisch konstruiert die Eingaben, die Daten, in den Learnern selbst sind und, dass diese einen Effekt haben, sowie der Umgang mit den visualisierten Daten als Ausgaben der Maschine weitere soziotechnische (Ent)Kopplungen prägt. Dabei geht es nicht nur um die Relevanz der Auswahl und Annotierung von Daten, sondern auch um die Verkopplung von Lernformen je nach Kontext und Einbettung.

Die Kunst des Lernens im Machine Learning und auch Deep Learning setzt dabei nicht nur an der soziotechnischen Konstruktion der Daten an, sondern führt mithilfe der Terminologie der White'schen Theorie die artifiziellen Bereiche des An-Lernens vor Augen. Damit gehört zur Kunst des künstlich intelligenten Lernens auch die Kunst des Lehrens von künstlich intelligenten Lernern.

5. Literatur

Burrell, Jenna (2016): »How the machine ›thinks‹: Understanding opacity in machine learning algorithms«, in: Big Data & Society 3, S. 1–12.

Cardoso Llach, Daniel (2018): »Daten als Schnittstelle. Die Poetik des maschinellen Lernens im Design«, in: Christoph Engemann/Andreas Sudmann (Hg.), Machine Learning. Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz, Bielefeld: transcript, S. 195–218.

Chen Lingjiao/Zaharia, Matei/Zou, James (2023): »How Is ChatGPT's Behavior Changing over Time?«, in: arXiv preprint arXiv:2 2307.09009.pdf (arxiv.org) vom 17.08.2023.

DLF Nova, 24.07.2023: »Updates: ChatGPT wird wohl immer dümmer«, in: Dlf Nova (deutschlandfunknova.de).

Domingos, Pedro (2015): The Master Algorithm. How the Quest for the Ultimate Learning Machine Will Remake Our World, New York: Basic Book.

Häußling, Roger (2012): »Design als soziotechnische Relation. Neue Herausforderungen der Gestaltung inter- und transaktiver Technik am Fallbeispiel humanoider Robotik«, in: Stephan Moebius/Sophia Prinz (Hg.), Das Design der Gesellschaft. Zur Kultursoziologie des Designs, Bielefeld: transcript, S. 273–298.

Häußling, Roger (2016): »Zur Rolle von Entwürfen, Zeichnungen und Modellen im Konstruktionsprozess von Ingenieuren. Eine theoretische Skizze«, in: Thomas H. Schmitz/Roger Häußling/Claudia Mareis/Hannah Groninger (Hg.), Manifestationen im Entwurf. Design – Architektur – Ingenieurwesen, Bielefeld: transcript, S. 27–64.

Häußling, Roger (2020): »Daten als Schnittstellen zwischen algorithmischen und sozialen Prozessen. Konzeptuelle Überlegungen zu einer Relationalen Techniksoziologie der Datafizierung in der digitalen Sphäre«, in: Soziale Welt (Sonderband 23), S. 134–150.

Helmholtz, Hermann (1896): »Ueber das Verhältnis der Naturwissenschaften zur Gesamtheit der Wissenschaft. Akademische Festrede gehalten zu Heidelberg beim Antritt des Prorectorats 1862«, in: Vorträge und Reden Bd. 1, Braunschweig: Vieweg, S. 157–186.

Kaminski, Andreas/Glass, Colin W. (2019): »Das Lernen der Maschinen«, in: Kevin Liggeri/Oliver Müller (Hg.), Mensch-Maschine-Interaktion. Handbuch zu Geschichte – Kultur – Ethik, Weimar: J. B. Metzler, S. 128–133.

Karafillidis, Athanasios (2018) »Die Komplexität von Interfaces. Touchscreens, nationale Identitäten und eine Analytik der Grenzziehung«, in: Berliner Debatte Initial 29, 130–146.

Koren, István/Klamma, Ralf (2018): »Enabling visual community learning analytics with Internet of Things devices«, in: Computers in Human Behavior 89, S. 385–394.

La Mettrie (2001): L'Homme machine. Der Mensch eine Maschine, Stuttgart: Reclam.

Lenzen, Manuela (2018): Künstliche Intelligenz. Was sie kann & was uns erwartet, München: C. H. Beck.

Porter, Theodore M. (2020): Trust in Numbers. The Pursuit of Objectivity in Science and Public Life, Princeton: Princeton University Press.

Reigeluth, Tyler/Castelle, Michael (2021): »What Kind of Learning Is Machine Learning?«, in: Jonathan Roberge/Michael Castelle (Hg.), The Cultural Life of Machine Learning. An Incursion into Critical AI Studies, Cham: Palgrave Macmillan, S. 79–116.

Riskin, Jessica (Hg.) (2007): Genesis Redux. Essays in the History and Philosophy of Artificial Life, Chicago/London: University of Chicago Press.

Russel, Stuart J./Norvig, Peter (2016): Artificial Intelligence. A Modern Approach, Boston et al.: Pearson.

Schmitt, Marco/Heckwolf, Christoph (2023): »KI zwischen Blackbox und Transparenz«, in: diesem Band.

Schütz, Alfred, 1972: »Der gut informierte Bürger. Ein Versuch über die soziale Verteilung des Wissens«, in: Alfred Schütz (Hg.), Gesammelte Aufsätze, The Hague: Nijhoff, S. 85–101.

Shalev-Shwartz, Shai/Ben-David, Shai (2014): Understanding Machine Learning. From Theory to Algorithms, Cambridge: Cambridge University Press.

Simon, Herbert (1984): »Why Should Machines Learn?«, in: Ryszard S. Michalski/Jamie G. Carbonell/Tom M. Mitchell (Hg.), Machine Learning. An Artificial Intelligence Approach, Berlin: Springer, S. 25–37.

Simon, Herbert (1994): Die Wissenschaften vom Künstlichen. Zweite Auflage, Wien: Springer.

Sudmann, Andreas (2018): »Zur Einführung. Medien, Infrastrukturen und Technolohgien des maschinellen Lernens«, in: Christoph Engemann/Andreas Sudmann (Hg.), Machine Learning. Medien, Infrastrukturen und Technologien der künstlichen Intelligenz, Bielefeld: transcript, S. 9–23.

Taffel, Sy (2019): »Automating Creativity – Artificial Intelligence and Distributed Cognition«, in: spheres 5, S. 1–9.

Turing, Alan (2004): »Intelligent Machinery (1948)«, in: B. Jack Copeland (Hg.), The Essential Turing. Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life plus The Secrets of Enigma, Oxford: Oxford University Press, S. 410–432.

White, Harrison C. (1992): Identity and control: A structural theory of social action, Princeton/NJ: Princeton University Press.

White, Harrison C. (2008): Identity and control: how social formations emerge, Princeton/NJ: Princeton University Press.

Ziegler, Siegfried (2007): Lernen bei Gregory Bateson. Die Veränderung sozialer Systeme durch organisationales Lernen, Saarbrücken: VDM.