

Das Öl des 21. Jahrhunderts

Daten sind das Öl des 21. Jahrhunderts. Ich kann diesen Satz nicht mehr hören. Kein Zeitungsartikel, kein Forschungsantrag und kein Buch zu KI kommt ohne diesen Satz aus. Natürlich stimmt es, dass die großen Internetunternehmen mit Daten eine ungeheure Wertschöpfung schaffen. Sie raffinieren die Rohdaten zu bisweilen nützlichen Produkten wie Internetsuche, Kaufempfehlungen, Routenplaner, Übersetzungsprogramme oder Sprachassistenten. Und die Hoffnung ist, dass andere Branchen nachziehen werden und ebenso aus der Kombination ihrer traditionellen Produkte mit Daten und KI einen Mehrwert erzeugen. Das autonome Fahren ist wahrscheinlich das bekannteste Beispiel.

Der Vergleich hinkt trotzdem, denn anders als Öl, das eine endliche Ressource ist, führt die Digitalisierung dazu, dass wir zunehmend in Daten ertrinken werden. Eine von einem Festplattenhersteller gesponserte Studie schätzt, dass alleine von 2018 auf 2019 die Menge der Daten auf allen Geräten von 33 auf 40 Zettabyte angewachsen ist.¹ Ein Zettabyte sind 10^{21} Bytes. Das heißt, auf jeden der 8 Milliarden Menschen auf der Erde kam schon 2019 eine 5 Terabyte Festplatte (oder etwa 1000 DVDs). 2025 sind es insgesamt wohl 175 Zettabyte. Kein Mensch kann all diese Daten ohne Computerunterstützung sichten oder gar Schlussfolgerungen aus ihnen ziehen. Computer müssen uns mal wieder helfen, Probleme zu lösen, die wir ohne Computer gar nicht hätten. Da trifft es sich gut, dass maschinelle Lernmethoden, wie künstliche neuronale Netze, erst mit richtig vielen Daten gut funktionieren. Wir brauchen mehr Daten, damit die KI besser wird, und wir brauchen mehr KI, damit wir die Daten besser nutzen können. Das ist der Grund, warum die zwei Buzzwords Big Data und KI so gerne zusammen auftreten.

1 Siehe Reinsel, Gantz & Rydning (2018).

Aber wie wird aus großen Datenmengen durch KI Geld gewonnen? Was rechtfertigt die gewaltigen Investitionen in Big Data, um noch mehr Daten zu sammeln, und in KI, um die Daten zu analysieren? Ein 2013 im Netz viel geteilter Post besagte, dass sich die meisten Firmen bei dem Thema Big Data so verhalten wie Teenager beim Thema Sex: »Jeder spricht darüber, keiner weiß, wie es geht, und jeder denkt, alle anderen tun es, also behaupten sie, es auch zu tun.«²

Wie Streaming die Videotheken verdrängte

Netflix war einer der ersten frühreifen Teenager, die wussten, wie es geht. Heute kennt jeder Netflix als einen Streamingdienst, der mehr oder weniger originelle Serien produziert. Eine rein digitale Internet-firma. Aber in seinen Anfangsjahren war das Unternehmen eigentlich nur eine große Videothek, die DVDs per Post verschickte. An Streaming war technisch noch nicht zu denken, dazu war das Internet viel zu langsam. Früher ist man nach der Schule zu seiner lokalen Videothek geradelt und hat sich aus den Regalen ein Video für den Filmabend mit Freunden ausgesucht. Das Angebot war begrenzt und aktuelle Videos meist schon von jemand anderem ausgeliehen. Dafür gab es einen Videothekar (oft ein studentischer Filmfreak), der einem Klassiker empfehlen konnte, die man unbedingt gesehen haben musste.

Netflix erkannte, dass man dank Post und Internet auf die Filialen und den netten Videothekar verzichten und gleichzeitig einen besseren Service bieten kann. Die Videos lagern in großen Hallen und statt in die Filiale zu kommen, bestellen die Kunden die Filme online (daher der weitsichtige Name der Firma). Dadurch kann man eine größere Menge an Filmen anbieten und gleichzeitig mehr Kunden bedienen. Statt in jeder Videothek genügend Kopien eines aktuellen Films bereitzustellen, kann man zentral die Verteilung planen. Netflix hatte außerdem die geniale Idee, dass man nicht für jeden Film einzeln zahlt, sondern ein Abo abschließt. Der Kunde pflegt online in seinem Profil dafür eine Liste von Filmen, die er irgendwann gerne sehen würde. Sobald er eine

2 Das Zitat stammt von dem Bestseller-Autor Dan Ariely. Im Original lautet es: »Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...« (<https://x.com/danariely/status/287952257926971392>).

DVD fertig angeschaut hat, schickt er sie per Post zurück und bekommt eine neue DVD von seiner Liste zugeschickt. Netflix entscheidet aber welche und kann so die verfügbaren DVDs optimal auf seine Kunden verteilen.

Natürlich setzten aber trotzdem viele Kunden gleichzeitig dieselben aktuellen Filme auf ihre Listen. Damit die Kunden sich nicht ärgern, dass sie vielleicht ein paar Wochen auf einen bestimmten Film warten müssen, und damit all die anderen DVDs nicht einfach im Lager herumliegen, brauchte Netflix nun doch einen Videothekar, der einen guten Film als Lückenfüller empfehlen konnte. Also ließ die Firma seine Kunden die Filme, die sie gesehen hatten, bewerten: von einem Stern für einen nicht so guten Film bis zu fünf Sternen für einen großartigen Film. Und maschinelles Lernen sollte mithilfe all diesen Daten voraussagen, welche Filme die einzelnen Kunden mögen. Aus betriebswirtschaftlicher Sicht ist es bei dem Abo-Modell entscheidend, neue Kunden zu gewinnen und alte zu behalten. Dazu muss man den Geschmack seiner Kunden genau kennen.

2006 erregte Netflix bei KI-Forscherinnen und -Forschern große Aufmerksamkeit, als die Firma einen Preis über eine Million Dollar für denjenigen auslobte, dessen Lernalgorithmus es schafft, die Vorlieben der Nutzer um mindestens zehn Prozent besser vorherzusagen als das Unternehmen selber. Der Datensatz, aus dem diese Vorlieben gelernt werden sollten, bestand aus 100 Millionen Filmurteilen – aus heutiger Sicht nicht besonders viel. Ich war damals Doktorand in Tübingen und für Wochen sprachen wir beim Mittagessen über nichts anderes, als welche Methoden man dafür nutzen könnte. Ich bin mir sicher, dass es an anderen Forschungsinstituten nicht anders war. Zwar gab es stetigen Fortschritt und von Monat zu Monat stellten neue Teams neue Rekorde auf, aber es dauerte fast drei Jahre, bis ein Team das Preisgeld einstrich.

Auch wenn Netflix heute keine DVDs mehr verschickt, so will die Firma doch immer noch genau wissen, was den Kunden gefällt. Bei 150 Millionen Abonnenten im Jahr 2019 kommt da ordentlich was an Daten zusammen. Früher bewerteten die Kunden jeden gesehenen Film, um bessere Empfehlungen zu bekommen. Aber durch das Streaming weiß Netflix heute ohnehin genau, welche Filme und Serien welche Kunden in welchen Ländern wie oft und wie lange anschauen. Aufgrund der Erfahrungen, die das Unternehmen schon sehr früh mit maschinellem Lernen gemacht hat, kann man davon ausgehen, dass es genau weiß,

was man mit diesen Daten macht. Herauszufinden, dass es einen australischen Nutzer gibt, der in einem Jahr 352-mal *Madagaskar 3* anschaut hat, ist dabei nur die Spitze des Dateneisbergs.

Den Kunden die Filme empfehlen, die sie mögen werden, ist sehr wichtig für Netflix, weil die meisten Kunden nicht viel länger als eine Minute mit der Filmauswahl verbringen. Kunden, die nicht schnell etwas finden, was ihnen gefällt, können daher auch schnell wieder verloren gehen. Je häufiger die Kunden den Service nutzen, desto unwahrscheinlicher ist es, dass sie das Abo wieder kündigen (ich scheine die Ausnahme zu sein, die gekündigt hat, weil mich das Streamen zu viel Zeit gekostet hat). Angeblich sind die Verbesserungen an dem Filmempfehlungsalgorithmus von Netflix eine Milliarde Dollar pro Jahr wert.³

Empfehlungen für Neukunden sind dabei besonders schwierig, weil Netflix noch wenig über sie weiß. Daher müssen sie zum Einstieg ein paar Filme bewerten. Als ich mein Abo angefangen habe, habe ich mir als Erstes die neue *Star Trek*-Serie angesehen. Der Algorithmus hat im Anschluss lange gebraucht, bis er gelernt hat, dass ich neben Sci-Fi auch gerne Western gucke. Denn für die Analysten gilt gerade die erste Sendung, die Neukunden anschauen, als besonders wichtig, da die es wahrscheinlich war, die den Kunden angelockt hat. Serien die viele Neukunden gewinnen, sind besonders wertvoll, weil es einfacher ist, einen alten Kunden zu behalten, als einen neuen zu gewinnen. Aufgrund der großen Anzahl an Abonnenten und ihrer Nutzungsdaten kann Netflix präzise Marktanalysen machen, noch bevor eine neue Produktion beschlossen wird. Sie wissen vorher schon, wie groß der Markt für Kunden ist, die Sci-Fi und Western mögen, und deshalb wahrscheinlich für ein Crossover empfänglich sind (ich persönlich ziehe ja die *Serenity* der *Enterprise* vor). Netflix muss daher nicht nur den Mainstream bedienen, sondern kann auch für Nischengeschmack im Voraus abschätzen, welche Produktionen wie viele Zuschauer haben werden und vielleicht sogar neue Kunden anziehen.

Datengetriebene Firmen wie Netflix sammeln aber nicht nur Daten und analysieren sie. Sie nutzen den konstanten Datenstrom, um sich stetig zu verbessern. Dazu machen sie auch gezielt Experimente. Potenzielle Verbesserungen am Empfehlungsalgorithmus, der Benutzeroberfläche oder der ›Filmposter‹ können in einer repräsentativen

3 Siehe Gomez-Urbe & Hunt (2015).

Gruppe mit den alten Varianten verglichen werden. Das nennt man auch »A/B-Test«. Oft kriegen mehrere Gruppen – ohne, dass sie das merken – verschiedene neue Varianten. Nur falls in einer Gruppe signifikant länger Filme geschaut werden und weniger Kunden verloren gehen, werden die Änderungen an alle Kunden ausgerollt. Dieser Prozess kann weitgehend automatisiert werden. Ein KI-Programm kann selbständig mögliche Varianten auf Testgruppen verteilen, deren Effekt beobachten und so selbständig effizient nach der besten Variante suchen. Schon wieder ein Suchproblem! Je mehr Kunden Netflix hat, desto größer ist der Nutzen selbst kleiner Verbesserungen. Und je mehr Kunden man hat und je mehr Daten man deshalb sammeln kann, desto mehr Experimente kann man auch automatisiert durchführen, um selbst kleine Verbesserungen zu finden.

Es würde mich nicht wundern, wenn dieser datengetriebene Ansatz bei Netflix nicht nur zur Weiterentwicklung ihres Filmempfehlungsalgorithmus genutzt wird. Vor einer Weile habe ich an den Bushaltestellen auf dem Weg zur Arbeit Werbetafel für die neue Staffel *Stranger Things* gesehen. Vielleicht wurde in einer anderen Stadt eine andere Serie beworben oder vielleicht wurde in einer weiteren Stadt überhaupt nicht für Netflix geworben. Da das Unternehmen weiß, in welcher Stadt die Kunden gerade ihre Sendungen streamen, kann es den Effekt von Außenwerbung auf die Anzahl ihrer Neukunden und die gesehenen Sendungen sofort in seinen Daten sehen. Netflix würde wahrscheinlich keine Außenwerbung nutzen, wenn die Daten nicht zeigten, dass sie funktioniert.

Das Geschäftsmodell von Netflix beruht darauf, Filme und Serien anzubieten. Ohne Rechte an interessanten Inhalten wäre das Unternehmen nicht so erfolgreich. Aber ohne Digitalisierung und intelligente Datenanalyse eben auch nicht. Netflix hat auf diese Weise das klassische Geschäftsmodell der Videothek revolutioniert. Maschinelle Lernalgorithmen spielten dabei eine wichtige Rolle. Aber hat die Firma deshalb künstliche Intelligenz eingesetzt? Man könnte auch weniger reißerisch von computergestützter Statistik sprechen. Zwar kommen manche Methoden, wie neuronale Netze, aus der KI-Forschung, aber im Grunde geht es nur um die automatische Verarbeitung von statistischen Daten mit der Hilfe von Computern. Eigentlich braucht man nur Kenntnisse in Mathematik, Statistik und Informatik, wie sie jeder Ingenieur und jede Ingenieurin am Anfang des Studiums lernt, um Nutzen aus diesen Methoden zu ziehen. Statt von KI sprechen daher

viele Leute in solchen Fällen lieber von »Data Science«, dem Feld genau an dieser Schnittstelle zwischen Statistik und Informatik. In einem Aspekt geht das, was Netflix macht, aber scheinbar über normale Statistik hinaus. Der Computer wird nicht nur genutzt, um die Daten zu speichern und Statistiken zu berechnen. Der Empfehlungsalgorithmus lernt die Kunden von alleine besser kennen und trifft außerdem selbständig Entscheidungen und handelt entsprechend. Er ersetzt den Videothekar. Diese Personifizierung des Algorithmus lässt die Digitalisierung und Automatisierung von Geschäftsprozessen glamouröser erscheinen als sie in Wirklichkeit ist. Kein Wunder, dass so häufig von KI gesprochen wird.

Der Handel braucht Daten

Mehrere Jahre vor Netflix hat Amazon in ähnlicher Weise den Buchhandel revolutioniert. Natürlich sammelte auch der traditionelle Buchhandel in den Fußgängerzonen schon immer Daten darüber, wie viele und welche Bücher er wann verkauft. Eine penible Buchführung und aussagekräftige Statistiken helfen bei der Planung und Steuerung jedes Geschäftes. Schließlich müssen neue Bücher nachbestellt werden, bevor die Regale leer sind. Am besten die, die die Kunden kaufen wollen. Ein Kunde, der nicht findet, was er sucht, kauft woanders ein. Vielleicht online bequem von zu Hause. Amazon hat durch seine enorme Größe und seine Lagerhallen natürlich den Vorteil, wesentlich mehr Bücher vorhalten zu können als der kleine Buchladen an der Ecke. Da man aber auch im Buchladen die meisten Bücher für den nächsten Tag bestellen kann, kann das nicht Amazons einziger Vorteil sein. Durch seine Größe sieht Amazon auch schneller und besser, was am Markt passiert. Da der stationäre Buchhandel neben ein paar großen Ketten aus vielen kleinen Händlern besteht, weiß erst mal niemand genau, wie viele Exemplare eines bestimmten Romans aktuell gerade verkauft werden. Die *Spiegel*-Bestsellerliste wird daher aufwendig durch Abfragen bei vielen, vielen Buchhändlern bestimmt.⁴ Amazon weiß hingegen sofort, was gerade auf seiner Plattform gut geht – und versorgt den Kunden automatisch mit den entsprechenden Vorschlägen.

4 Siehe <https://www.buchreport.de/spiegel-bestseller/ermittlung-der-bestseller/>.

Nach dem Buchhandel expandierte Amazon in viele andere Branchen. Heute kann man bei Amazon fast alles bestellen. Was Amazon nicht selber im Sortiment hat, bieten andere Händler über Amazons Plattform an. Über diese Plattform beobachtet Amazon auch den Teil des Marktes, den sie selber (noch) nicht direkt bedienen. Dieses Wissen kann ein gewisser Vorteil bei der Planung des eigenen Sortiments sein. Insbesondere, wenn Algorithmen das Sortiment automatisch analysieren und Vorschläge für Verbesserungen machen. Für Kunden, die fast alles bei Amazon kaufen, kann Amazon außerdem detaillierte Profile erstellen: Was hat ein Kunde, der den neuen Harry Potter gekauft hat, noch gekauft? Braucht er vielleicht Rasierseife oder Mascara? Obwohl der Empfehlungsalgorithmus von Amazon manchmal spektakulär daneben liegt, liegt er manchmal auch erschreckend richtig. Und die richtige Empfehlung verführt den Kunden vielleicht zu einem Impulskauf. Sucht er etwas Bestimmtes, ist es auch besser, er findet es schnell, denn sonst sucht er vielleicht woanders weiter. Da Amazon wie Netflix sehr, sehr viele Kunden hat, kann selbst eine kleine Verbesserung des Empfehlungsalgorithmus zu sehr viel mehr Verkäufen führen – auch, wenn der Algorithmus nicht perfekt ist.

Das gleiche gilt selbstverständlich ebenso für große Handelsketten mit echten Geschäften in Fußgängerzonen oder am Stadtrand. Im Vergleich zu Amazon haben sie einen Nachteil: Sie können die Daten von einem Einkauf nicht ganz so leicht dem nächsten Einkauf zuordnen. Gut, dass es Kundenkarten gibt. Mit einem Kundenprofil kann man auch traditionelle Papierwerbung auf die Kunden zuschneiden. Die Offline-Variante von Online-Empfehlungen. Eine Supermarktkette in den USA fand zum Beispiel heraus, dass schwangere Frauen mehr geruchlose Hautcremes, Wattebällchen und große Handtaschen, in die auch Windeln passen, kaufen. Gestresste junge Eltern, die Windeln in einem Laden kaufen, machen den Rest des Einkaufs auch gleich dort. Junge Eltern in die eigenen Läden zu lotsen, ist daher sehr lukrativ. Kann aber auch Ärger machen. In Minnesota platzte ein empörter Vater einer Teenagerin wütend in seine Filiale und regte sich darüber auf, dass seiner minderjährigen Tochter Coupons für Babyklamotten geschickt wurden. Als sich ein Mitarbeiter später telefonisch dafür entschuldigen wollte, entschuldigte sich der Vater für seinen Wutaus-

bruch, denn seine schwangere Tochter konnte die Coupons tatsächlich gut gebrauchen.⁵

Eine bekannte Erfahrung im Internet ist, dass uns Hundefutterwerbung auf dem Computer in der Arbeit angezeigt wird, kurz nachdem wir auf dem Handy einen Artikel über Dackel gelesen haben. Komischer Zufall. Dass man Menschen im Internet über verschiedene Geräte hinweg verfolgen kann, um ein Profil über sie zu erstellen, überrascht wahrscheinlich niemanden mehr.⁶ Da es aber um so private Informationen wie eine Schwangerschaft geht, kann es die eine oder andere vielleicht doch schaudern, dass es der Supermarkt auch schon weiß. Um ein PR-Desaster abzuwenden, würde ich ja die Coupons für Babykleidung ganz unauffällig unter andere Coupons mischen.

Keinen Spaß würden die Kunden wahrscheinlich auch bei bestimmten Arten des ›Dynamic Pricing‹ verstehen. Dass die Preise an Tankstellen und die Preise für Flüge mittlerweile mehrmals täglich ganz automatisch durch Algorithmen an die aktuelle Marktsituation angepasst werden, ist eine verständliche Folge von in Echtzeit verfügbaren Daten und Analysen. Falls nun ein Kunde aufgrund seines Profils wahrscheinlich bereit ist, mehr für ein Produkt zu zahlen, warum den Preis dann nicht extra für ihn hochsetzen? Der Preis ist dann personalisiert und hängt nicht nur von Angebot und Nachfrage ab, sondern vielleicht auch von der Postleitzahl, der Marke des Mobiltelefons, dem Alter, dem vermuteten Einkommen, dem Suchverlauf, seinen früheren Einkäufen oder wer weiß was noch. So eine personalisierte Preisanpassung ist online durch Algorithmen leicht zu bewerkstelligen, und wenn der Kunde es aufgrund der ohnehin schwankenden Preise nicht merkt, dann kann er sich auch nicht besonders empören.

Von der Dating-App Tinder ist bekannt, dass sie die Preise für ihren Premiumdienst automatisch an die Kunden angepasst hat, ohne sie davon zu informieren. Sobald die automatische Datenanalyse es als unwahrscheinlich einschätzte, dass jemand den Premiumdienst zum normalen Preis kauft, wurde der Preis gesenkt. Das macht es für Kunden schwer, Preise von verschiedenen Dating-Apps zu verglei-

5 Siehe Duhigg (2012).

6 Dadurch, dass Sie sich mit Ihrem Facebook- oder Google-Account regelmäßig auf unterschiedlichen Geräten einloggen, wissen die Tech-Unternehmen, dass diese Geräte Ihnen gehören. Ganz ähnlich erlaubt die Paybackkarte den Unternehmen, Ihnen Ihre Einkäufe über verschiedene Geschäfte hinweg zuzuordnen.

chen, weil sie nicht wissen, von welchen Faktoren der Preis abhängt. Dass die Preise auch vom Alter der Kunden abhängen, lässt den Verdacht zu, dass die Preise vielleicht auch davon abhängen könnten, wie oft das Profilbild der Kunden nach rechts gewischt wurde. Weil Verbraucherschutzbehörden festgestellt haben, dass die Praxis von Tinder gegen EU-Verbraucherrecht verstößt, hat Tinder diese eingestellt.⁷ In welchem Ausmaß andere Firmen – insbesondere im Internethandel – Preise für einzelne Kunden personalisieren, ist nicht ganz klar. Aber die automatische Personalisierung von Preisen ist eine verlockende neue Möglichkeit der Preisgestaltung.

Die Maschinenbauer ziehen nach

Es ist ganz normal, dass ein Unternehmen versucht, seinen Gewinn zu maximieren. Dazu gehört, die Kunden und den Markt genau zu kennen und die Preise entsprechend zu gestalten. Ebenso gehört dazu, Daten über interne Prozesse und Kosten zu sammeln. Daten helfen, Geschäftsprozesse sichtbar zu machen und zu optimieren. Das nennt man »Business Intelligence«. Weiß man zum Beispiel, wie viele Ersatzteile noch im Lager liegen und wie häufig man jedes einzelne Teil in der Vergangenheit gebraucht hat, kann man rechtzeitig neue nachbestellen und besser mit dem Lagerplatz planen. Neu daran ist lediglich, dass es durch die Digitalisierung leichter wird, viel mehr und viel detailliertere Daten zu sammeln.

Die Digitalisierung wirkt sich aber nicht nur auf Einkauf, Verkauf und Lagerung aus, sondern auch auf die Produktion. Bei der traditionellen Massenfertigung wird eine Fabrik so gebaut und darauf optimiert, dass sie das gleiche Produkt wieder und wieder herstellt. Schon heute können dank Digitalisierung und Baukastenprinzip verschiedene Varianten nach Bedarf gefertigt werden. Das kennt man vom Autokauf. So wie ein programmierbarer Webstuhl, mit dem sich unterschiedliche Muster weben lassen, soll die smarte Fabrik der Zukunft schnell umprogrammierbar sein, um auch Kleinserien oder gar ganz individuelle Produkte herstellen zu können. Intelligente Roboter und vielseitig einsetzbare 3D-Drucker sollen dabei helfen, die Produktivität so weit zu steigern, dass sich die flexible Fertigung auch in

7 Die Europäische Kommission hat am 7. März 2024 eine entsprechende Pressemitteilung veröffentlicht (Europäische Kommission, 2024).

einem Hochlohnland wie Deutschland rechnet. Nach der Erfindung von Dampfkraft, Elektrizität und Computern soll jetzt die intelligente Automatisierung eine neue industrielle Revolution bringen. In Deutschland verbindet sich diese Hoffnung mit dem Schlagwort Industrie 4.0 (die vierte industrielle Revolution hat noch nicht einmal richtig angefangen und schon wird von Industrie 5.0 gesprochen, die den Menschen stärker in den Mittelpunkt stellen soll und außerdem auch noch nachhaltig ist).

Ein weiteres verwandtes Buzzword in Vorstandsetagen ist das »Internet der Dinge«. Nicht nur Menschen sind in Zukunft ständig online, sondern eben auch Dinge. Jedes Bauteil und jede Maschine in der Fertigung ist dann über ein Funknetz mit dem zentralen Computer einer Fabrik verbunden. Sensoren melden ständig an den Computer, wo sich jedes einzelne Bauteil gerade befindet und ob es die nötigen Qualitätsprüfungen bestanden hat. Auch die Maschinen haben Sensoren und melden, dass sie geölt werden müssen oder eine Störung vorliegt. Diese Daten erlauben es, Produktions- und Geschäftsprozesse weiter zu optimieren.

Nehmen wir als Beispiel einen Maschinenbauer, der Druckmaschinen herstellt – diese altmodischen Dinger, mit denen Zeitungen und Bücher gedruckt werden. Druckmaschinen müssen präzise arbeiten. In komplexen Produktionsprozessen ist bei Problemen mit der Qualität meist nicht leicht zurückzuverfolgen, welche Faktoren dazu geführt haben. Vielleicht war es die Temperatur und ein Bauteil hat sich zu weit ausgedehnt, vielleicht waren die Schrauben minderwertig, vielleicht war die fähigste Mitarbeiterin gerade im Urlaub, vielleicht war es Montag, oder vielleicht war es eine Kombination mehrerer Faktoren. Nur wenn man die entsprechenden Daten gesammelt hat, kann man im Nachhinein nach möglichen Gründen suchen. Mit genügend Daten könnte ein KI-Programm zum Beispiel selbständig lernen, welche Faktoren relevant sind und den Produktionsprozess beobachten und Auffälligkeiten, die einen Einfluss auf die Qualität haben könnten, rechtzeitig melden.

Trotz Qualitätssicherung können später Probleme in der Druckerei auftreten, an die die Druckmaschine verkauft wurde. Ein Ausfall der Maschine kann für eine Druckerei teuer sein. Viele Maschinenbauer bieten deshalb eine »kostenlose« Wartung und einen schnellen Reparaturservice als Verkaufsargument an. Die Wartung der verkauften Maschinen ist daher ein großer Kostenpunkt bei vielen Maschinenbauunternehmen. Eine Druckmaschine, die voller Sensoren ist, kann

Daten über ihren Zustand über das Internet an den Druckmaschinenhersteller melden. Die Analyse dieser Daten wird im Idealfall von KI-Programmen übernommen. Ein neuronales Netz könnte zum Beispiel lernen, aus den Daten vorherzusagen, ob eine Druckmaschine bald Probleme verursachen wird. Der Hersteller (oder die KI) kann dann vorbeugend einen Techniker vorbeisenden. Bei Maschinen, die ohnehin rund laufen, können hingegen Wartungskosten eingespart werden.

In der Hoffnung, dass Statistik und KI aus Daten gewinnbringende Erkenntnisse destillieren, werden immer mehr Daten gesammelt. Mittlerweile telefonieren nicht nur E.T. und Handys nach Hause, sondern auch Autos, Fernseher und Küchengeräte. Aber die Masse der Daten ist nicht alles. Wenn die Luftfeuchtigkeit entscheidend ist, um die Fehlfunktion eines Geräts vorherzusagen, wäre es gut, diese auch zu messen. Dazu muss man aber erst mal an diese Möglichkeit denken und extra einen entsprechenden Sensor einbauen. Wenn Sensoren aber nicht das Richtige messen oder wenn Daten eine schlechte Qualität haben, hilft es auch nicht unbedingt, wenn man viele solcher schlechten Daten hat. Für jede Datenanalyse gilt das GIGO-Prinzip: »garbage in, garbage out.« Gibt man oben Müll hinein, kommt unten Müll raus.

Auch Firmen, die bei der Digitalisierung schon weit fortgeschritten sind und Daten systematisch zur Analyse sammeln (das sind in Deutschland beileibe nicht alle), sammeln oft nur die Daten, die ohnehin anfallen. Hier haben Unternehmen, deren Geschäft sich hauptsächlich im Netz abspielt, einen klaren Vorteil. Während Siemens oder Daimler durch Daten produktiver werden können, sind die Daten jedoch nicht ihr Kerngeschäft und relevante Daten müssen mit zusätzlichem Aufwand gesammelt werden. Das Geschäftsmodell von Netflix und Amazon beruht zwar viel stärker auf Daten, aber sie müssen auch außerhalb des Netzes gute Serien produzieren oder Bücher verschicken. Google und Facebook hingegen verdienen ihr Geld mit Werbung ausschließlich online. Relevante Daten fallen online zuhauf an und können leicht gespeichert werden. Die Kunden von Google und Facebook können dank der Daten, die sie über ihre Nutzer sammeln, zielgerichtet Werbung schalten. In der Metapher des Öls des 21. Jahrhunderts ist unser Online-Leben der Rohstoff, den Google und Facebook raffinieren, um daraus Plastikprodukte zu machen, oder um ihn zu verbrennen. Willkommen im Überwachungskapitalismus.⁸

8 Der Begriff stammt von Zuboff (2019).

