Kapitel 4 – Veränderungen des Innovationswettbewerbs durch selbstlernende Systeme

Der Vorschlag eines Datenzugangsrechts geht auf die konkrete Sorge ein, dass Systeme Künstlicher Intelligenz die Innovationsfähigkeiten einiger Marktteilnehmer und Innovationsanreize aller Marktteilnehmer ernsthaft beschneiden. Das Vorliegen großer, vielfältiger Datenmassen könnte eine Voraussetzung zur Entwicklung selbstlernender Systeme sein und damit die Innovationsfähigkeit potentieller Markteintreter bedingen. Die Besonderheit gegenüber anderen Innovationsvoraussetzungen ist der eingangs schon beschriebene Kreislauf: Diejenigen Machine-Learning-Anwendungen und die dahinterstehenden Unternehmen, die früh gute Dienste anbieten, verbessern sich kontinuierlich mit den ihnen zum Lernen überlassenen Daten[860], ziehen damit mehr Nutzer an und sind damit für neue Nutzer attraktiv. Fraglich ist, inwiefern Feedback-Effekte oder Data Network Effects tatsächlich wirken, ob und wie sie durchbrochen werden können und welche Auswirkungen sie auf den Innovationswettbewerb haben. Möglicherweise gab es in der Vergangenheit vergleichbare Verläufe bei Verbreitungen von Basisinnovationen auf Märkten. Bisher wurde eingehend von Ökonomen und Kartellrechtlern untersucht, ob Datenhoheit in extremen Fällen in Marktmacht münden kann, die wiederum missbraucht worden sein könnte. Vielmehr geht es hier aber darum, ob die enormen Datensammlungen, die die Industrie 4.0 und Anwendungen der Künstlichen Intelligenz hervorbringen, legale, aber unerwünschte Effekte auf den Innovations- und Ideenwettbewerb haben.

Zunächst wird untersucht, welche Rolle Daten für selbstlernende Systeme in der Industrie 4.0 spielen und wie aktuell ihre Zuordnung erfolgt. Die rechtlichen Überlegungen setzen ein Verständnis von der Funktionsweise selbstlernender Systeme voraus. Anschließend sollen exklusive Datensets von offenen Datensets abgegrenzt werden, um den Bedarf ermitteln zu können. Ein Datenzugangsrecht, das sich auf ohnehin verfügbare Daten richtet, kann zu einem Funktionieren der Märkte wenig beitragen. Daher ist es umso wichtiger, herauszustellen, welche Daten als unentbehr-

860 „Überlassene Daten" sind ebenso wie „volunteered data" ein Pleonasmus. Die Bezeichnung als Daten stammt vom lateinischen Verb dare (geben, gewähren, anvertrauen), ein Datum ist (immer) das Gegebene oder die Gabe.

lich zur Entwicklung neuer Produkte oder Dienstleistungen betrachtet werden. Nach einer genaueren Betrachtung von Maschinendaten werden dann die Feedback-Effekte oder Datennetzwerkeffekte in den Blick genommen. Möglicherweise handelt es sich bei ihnen nur um seit Jahren bekanntes „Learning by doing", wie Hal Varian als Ökonom von Google annimmt.[861] Anschließend soll auf die potentiellen Monopolisierungstendenzen, die sich aus Feedback-Effekten ergeben und Anknüpfungspunkt einer marktöffnenden Regulierung sein könnten, eingegangen werden.

A. Einleitung – Die Bedeutung von Daten in der Industrie 4.0

In den letzten Jahren konzentrierten sich die Öffentlichkeit und die rechtswissenschaftliche Diskussion auf die Sammlung personenbezogener Daten. „Big Data" bezog sich meist auf Daten für Unterhaltungsangebote, in Kommunikationsdiensten und Online-Shops. Als Konsequenz wurden vorrangig Themen des Datenschutzes auf die Agenda gesetzt und politisch bearbeitet. Nicht zuletzt wegen der Industrie 4.0 verlagert sich der Fokus aber aktuell auf die wirtschaftliche Dimension von Daten in der produzierenden Industrie. Gerade das Internet der Dinge wird dazu beitragen, dass den „Daten der Dinge", also „nicht-human-zentrischen Daten"[862], mehr Bedeutung zukommt.[863]

Wie schon personenbezogene Daten werden Maschinendaten damit zu einem Anknüpfungspunkt für vielfältigste Spekulationen, ob das „analoge Recht" im digitalen Zeitalter noch genüge. Aufgeworfen werden dabei Fragen des „Dateneigentums", des Datenzugangs und der Datensicherheit. Die zunehmende Erfassung, Speicherung und Auswertung von Daten, die mit den oben genannten Entwicklungen einhergeht, stellt Fragen an verschiedene Rechtsgebiete: das Datenschutzrecht, das Recht des geistigen Eigentums und das Vertrags- und Haftungsrecht. Das Problem der Spekulationen ist, dass sie allein nicht als Grundlage für Gesetzgebung

---

861 *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 15.

862 So *Mitomo*, Data Network Effects: Implications for Data Business, 28th European Regional Conference of the International Telecommunications Society (ITS) 2017, S. 1.

863 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 68; *Surblytė*, WuW 2017, 120 (121); *OECD*, Consumer Policy and the Smart Home, April 2018, S. 19: schon weniger als 10.000 Haushalte, die ein bestimmtes IoT-Produkt nutzen, können täglich 150 Millionen Data Points generieren (übersetzt aus dem Englischen).

und Gesetzesanwendung genügen. Der Gesetzgeber muss die Kosten einer Regulierung mit dem Nutzen aufwiegen können. Hierzu müssen Kosten und Nutzen aber auch immerhin ungefähr kalkulierbar sein. Von der Digitalisierung und Industrie 4.0 versprechen sich Fertigungsindustrie, Dienstleister, Telekommunikation, Politik und Wissenschaft beträchtliche Kostenvorteile, Flexibilität, eine höhere Kundenorientiertheit und Wettbewerbsvorteile. Die Gefahr, durch voreilige, auf vagen Vermutungen beruhende gesetzliche Grenzen diese Vorteile zu versperren, verlangt nach gesicherten Erkenntnissen. Die Bezeichnung der zu erwartenden Entwicklungen der produzierenden Industrie lehnt sich an vergangene industrielle Revolutionen an: Daher liegt es nahe, zu fragen, wie das Recht auf sie reagierte und ob sich diese Reaktion bewährt hat.

I. Begriff der Industrie 4.0

Ein besonderes Augenmerk der deutschen Wirtschaftspolitik liegt auf der produzierenden Industrie. Die Industrie 4.0 beschreibt „die intelligente Vernetzung von Maschinen und Abläufen in der Industrie mit Hilfe von Informations- und Kommunikationstechnologie".[864] Der „Plattform Industrie 4.0" entspricht auf europäischer Ebene die Strategie „Digitising European Industry" der Europäischen Kommission. Eine Besonderheit der Digitalisierung gegenüber vorausgehenden industriellen Revolutionen ist, dass sie global gleichzeitig erfolgt, statt sich mit Zeitverzögerung von einem Punkt ausgehend auszubreiten.[865]

„Industrie 4.0" nimmt Bezug auf eine so bezeichnete vierte industrielle Revolution. Die eigentliche industrielle Revolution wurde zum Ende des 18. Jahrhunderts ausgelöst durch Dampfmaschinen und maschinelle Webstühle. Die zweite industrielle Revolution wird auf das Ende des 19. Jahrhunderts datiert und ist von der Fließbandarbeit geprägt, also einer Kleinschrittigkeit von Produktionsschritten. Die Zunahme der Nutzung von Computern zur Bewältigung von Arbeitsschritten ab 1970 wird heute als die dritte industrielle Revolution bezeichnet. Es kann bezweifelt werden, dass die dritte industrielle Revolution schon alle Branchen und alle Produzenten erfasst hat; trotzdem wird bereits von der vierten industriel-

---

864 *BMWi (Hrsg.)*, Plattform Industrie 4.0, Was ist Industrie 4.0.

865 *Kowitz*, Einführung: Soziale Marktwirtschaft im digitalen Zeitalter, in: Wirtschaftsrat der CDU (Hrsg.), Soziale Marktwirtschaft im digitalen Zeitalter, S. 15–28, 17.

len Revolution gesprochen. Kennzeichnend für diese Stufe ist, dass die Produktion dezentral gesteuert wird und damit ein hohes Maß von Flexibilität aufweist. Mit intelligenten vernetzten Sensoren soll eine dauerhafte und präzise Umwelterfassung möglich werden; Smart Objects sollen interaktiv und eigenständig Produktionsvorgaben erfüllen.[866] Wesensmerkmal der Industrie 4.0 ist die Vernetzung von Maschinen untereinander und mit Menschen. Auffällig ist, dass die zeitlichen Abstände zwischen den „Revolutionen" immer kürzer werden. Darauf, ob es sich tatsächlich um eine Revolution oder bloß eine Evolution der Produktionsbedingungen handelt, soll hier aber nicht weiter eingegangen werden.

„Industrie 4.0" soll optimistisch eine totale Vernetzung aller Produktionsstufen durch den Einsatz von RFID-Systemen und Maschine-zu-Maschine-Kommunikation beschreiben. Dank einer Ausstattung mit Sensoren (Cyber-Physische Systeme, CPS) und selbstlernenden Systemen, „weiß" das Werkstück, wie es wo bearbeitet werden soll. Die technologischen Entwicklungen sollen nicht nur den einzelnen Unternehmen dienen, sondern dank Ressourcenschonung, erhöhter Energieeffizienz und der Stärkung der Wettbewerbsfähigkeit Deutschlands weitergehenden Nutzen haben.

In ihrer Summe werden die Technologien, die die Industrie 4.0 ausmachen, einer Basisinnovation[867] gleichkommen, wie auch der begriffliche Vergleich mit industriellen Revolutionen andeuten soll. IBM vergleicht die Entwicklung von KI-Systemen mit der Entwicklung des Automobils, insofern als dass es jeweils verschiedener Bausteine brauchte, die aufeinander aufbauten, um Massentauglichkeit zu erreichen.[868] Einzelne Bausteine wie der Verbrennungsmotor und die Achse ermöglichten komplementäre Innovationen – entsprechend ermöglichten erst eine breite Datenerfassung und Analytik die Entwicklung von KI. Weitergedacht verfügten auch bei Entwicklung des Automobils zunächst nur wenige über die entsprechenden Bausteine, Markteintreter konnten aber durch Imitationen und daran anknüpfende Entwicklungen aufholen.

---

866 Dazu *Drexl*, Data Access and Control (BEUC), S. 28; *Europäische Kommission*, Weißbuch KI, COM(2020) 65 final, 19. Februar 2020, S. 4.

867 Kapitel 2 B.3. Basisinnovationen und ‚Invention of a New Method of Innovating', S. 59.

868 *R. Thomas/Brehm*, Das A und O für AI ist IA, IBM, 22. März 2018.

II. Besonderheiten der Analyse von Daten in der Industrie 4.0

Die Informationstechnologie zog ab den 1970er Jahren in deutsche Unternehmen ein – zunächst mit PCs und Office-Anwendungen nur in die Büros, mittlerweile auch in die Produktionshallen. Die fertigende Industrie sammelte schon immer Daten über ihre Produktionsprozesse und wertete sie aus, um den wirtschaftlichen Erfolg zu maximieren. Die erfolgten und noch bevorstehenden Entwicklungen im Kontext der Industrie 4.0 machen es besonders einfach und attraktiv, Daten im großen Stil zu speichern und auszuwerten. Die Internetökonomie zeichnet sich dadurch aus, dass sie Nutzer am Innovations- und Entwicklungsprozess der Unternehmen beteiligt.[869] Die Elemente, die die Industrie 4.0 kennzeichnen, sind im Gegensatz zu vorangegangenen industriellen Revolutionen unkörperlich und daher besonders flexibel.[870]

Die prognostizierten und schon eingeleiteten Innovationen ergreifen die Produktwelt von der Erfindung über die Produktion bis zum Vertrieb in all ihren Facetten. Die soeben angesprochene Vernetzung wird von einer Verzahnung von Produktion mit Kommunikations- und Informationstechnik ermöglicht. Die Koordination der Fertigungsprozesse wird dabei maßgeblich auch von intelligenten Maschinen übernommen, wobei Koordination und Kommunikation nicht auf die fabrikinterne Produktion beschränkt sind, sondern über Branchen hinweg ablaufen und sich vom Zulieferer bis zum Konsumenten oder Nutzer erstrecken. Hierbei werden umfassend Daten gesammelt, die ein bedeutender Wirtschaftsfaktor sein können.[871]

Viele Potentiale der Industrie sind abhängig vom Zugang zu Big Data, also enorm großen Datenbeständen.[872] Die Analyse dieser großen (volume), wenig strukturierten Datenmengen wird als Big Data Analytics bezeichnet, sie ist von Echtzeitauswertung (velocity) zur Gewinnung neuer Erkenntnisse aus heterogenen Quellen (variety) geprägt.[873] Es erfolgt eine Verknüpfung von Echtzeitdaten mit historischen Datenbeständen. Service-Roboter, selbstlernende Maschinen und fahrerlose Transportfahrzeuge werden Einzug halten in Produktionshallen. Mithilfe von CPS kann eine durchgängige Informationsverarbeitung ermöglicht werden, anhand

869 *Clement/Schreiber*, Internet-Ökonomie, S. 24ff.
870 Vgl. *Mateos-Garcia*, The Complex Economics of Artificial Intelligence, S. 2, 9.
871 *Ensthaler*, NJW 2016, 3473 (3473).
872 Vgl. *Schweitzer/Peitz,* NJW 2018, 275 (275).
873 Drei V's, *Drexl,* Designing Competitive Markets for Industrial Data, S. 13.

derer Algorithmen der Künstlichen Intelligenz lernen können. Maschinen erfüllen weiter ihre traditionelle Funktion, darüber hinaus dienen sie als Teil des Internet of (Industrial) Things auch der Datenerfassung, -verarbeitung und dem -austausch untereinander. Bisher getrennte Informationsquellen werden durchgängig miteinander verbunden und in einen Kontext gebracht. Das Innovationspotential von großen Datensets liegt in der Offenbarung von Korrelationen.[874] Daten sind daher entscheidend für die Wettbewerbsfähigkeit von Industrie 4.0-Diensten.[875]

III. Einsatz selbstlernender Systeme

Als „Allzweck-Technologie" (General Purpose Technology)[876] wird Künstliche Intelligenz in allen Aspekten der Informationstechnologien zum Einsatz kommen. Für Optimisten ist sie einer der wichtigsten Wachstumstreiber und ein Schlüssel zur Sicherung der Wettbewerbsfähigkeit.

Im Jahr 1950 stellte Alan Turing erstmals die Möglichkeit denkender Computer vor.[877] 47 Jahre später konnte der Supercomputer Deep Blue den Schachweltweister Gary Kasparov schlagen.[878] Wiederum 14 Jahre später wurde Watson von IBM der beste Jeopardy-Spieler und nur weitere vier Jahre später (2015) fuhren Fahrzeuge autonom mit Künstlicher Intelligenz.[879] Spätestens seit diesem Zeitpunkt ist KI einer breiteren Öffentlichkeit bekannt. Die rasante Entwicklung der letzten Jahre wurde – nach allgemeiner Meinung – ermöglicht durch das exponentielle Wachstum der Rechenleistungen, besser geeignete Algorithmen und verstärkte automatisierte Datenerfassung.[880] Heute werden größte Hoffnungen in Künstliche

---

874 *Drexl,* Designing Competitive Markets for Industrial Data, S. 10.

875 *BMWi – Plattform Industrie 4.0 (Hrsg.)*, Marktmacht durch Datenhoheit, in: dass., Industrie 4.0 – Kartellrechtliche Betrachtungen, S. 15–22 (16); *Shelanski*, UPenn Law Review, Vol. 161, S. 1663–1705, 1686 (2013).

876 *McElheran*, Economic Measurement of AI, S. 6; *OECD*, Digital Innovation, S. 24.

877 *Turing,* Computing Machinery and Intelligence, Mind, Vol. LIX, No. 236, Oktober 1950, S. 433–460; dazu *Nilsson*, The Quest for Artificial Intelligence, S. 61.

878 Dazu *Wittpahl*, Vorwort, in: Wittpahl (Hrsg.), Künstliche Intelligenz, S. 7, sowie kritisch *Nilsson*, The Quest for Artificial Intelligence, S. 594.

879 *A. Gal*, It's a Feature, not a Bug: On Learning Algorithms and what they teach us; OECD, DAF/COMP/WD(2017)50, S. 2.

880 Vgl. *Morik*, Daten – wem gehören sie, wer speichert sie, wer darf auf sie zugreifen?, in: Morik/Krämer (Hrsg.), Daten, S. 15–47 (17); *Stoica et al.*, A Berkeley View of Systems Challenges for AI, S. 1; *UK Select Committee on Artificial Intelligence*, AI in the UK, S. 19.

Intelligenz gesetzt: „AI is one of the most important things humanity is working on. It is more profound than electricity or fire."[881], sagte Google-Chef Sundar Pichai im Januar 2018.

Sensorische Datenerfassung und Künstliche Intelligenz werden in der Industrie 4.0 etwa für die Produktentwicklung verstärkt eingesetzt: Bereits seit einigen Jahren werden mit Hilfe selbstlernender Simulationsprogramme innovative Produkte ohne nennenswerten menschlichen Eingriff entwickelt.[882] Beispielsweise hat eine KI-Designsoftware (Generative-Design-Software[883]) einen Wärmetauscher für Klimaanlagen komplett neuartig geplant.[884] Einen ähnlichen Erfolg erzielte Audi: Ein Deep-Learning-System hat eine Metalllegierung vorgeschlagen, die menschliche Entwickler zuvor nicht erwogen haben.[885] Die entwickelte Metalllegierung sei leichter, stabiler und funktioniere in der Umsetzung. Um diese Innovationsempfehlung zu erhalten, wurden die Anforderungen des Endprodukts angegeben und modelliert. Neuronale Netze prognostizieren die mechanischen Eigenschaften gewünschter Produkte und können die optimalen Konstruktionen vorschlagen, was die Bedeutung selbstlernender Systeme steigert.

1. Begriff: Künstliche Intelligenz, Machine Learning, selbstlernende Systeme

Im deutschen Sprachgebrauch wird die „Artificial Intelligence" (AI) als Künstliche Intelligenz (KI) übersetzt. Die englische Bezeichnung wurde wohl im August 1955 erstmals im Paper „A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence" von McCarthy, Minsky, Rochester und Shannon genutzt.[886] Ob „Künstliche Intelligenz" eine

881 Deutsch: Künstliche Intelligenz ist eines der wichtigsten Dinge, an denen Menschen arbeiten. Sie ist bedeutsamer als Elektrizität oder das Feuer; *The Economist*, Playing with Fire, 16. Juni 2018.

882 Allgemein dazu *Abott*, Hal the Inventor, in: Sugimoto et al. (Hrsg.), Big Data Is Not a Monolith, S. 187–198 (187); *Surblytė*, WuW 2017, 120 (124).

883 *McAfee/Brynjolfsson*, Machine, Platform, Crowd, S. 110ff; eine Software, die nicht den Menschen unterstützt, sondern eigenständig auf Grundlage der Anweisungen eine Lösung entwickelt.

884 *McAfee/Brynjolfsson*, Machine, Platform, Crowd, S. 110ff.

885 *Ermert*, Experten: KI-Systeme sollen patentierbar sein, Heise, 1. Juni 2018.

886 *McCarthy et al.*, A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, 31. August 1955; dazu *Mateos-Garcia*, The Complex Eco-

geeignete Übersetzung ist, erscheint zweifelhaft: Intelligence ist nicht gleichbedeutend mit Intelligenz und bedeutet artificial nicht nur künstlich.[887] Ein solcher Begriff beinhaltet immer eine semantische Unschärfe, weshalb sich daneben auch Machine Learning (Maschinelles Lernen) und Unterkategorien etabliert haben. Es mag stimmen, dass die in der deutschen Sprache genutzte Terminologie vage und zirkulär ist[888], der Fokus liegt hier aber ohnehin auf der Lernfähigkeit, weshalb diese Arbeit von „selbstlernenden Systemen" spricht.

Seit mehreren Jahrzehnten beschäftigt sich die Informatik mit der Lernfähigkeit von Systemen, ohne dass sich eine einheitliche Definition für Künstliche Intelligenz herauskristallisiert hat.[889] Ein Definitionsansatz besagt, dass KI vorliegt, wenn Maschinen Aufgaben erfüllen, für die beim Menschen Intelligenz nötig wäre.[890] Ein anderer bezeichnet KI als die Disziplin, die kognitive Systeme zu simulieren versucht. Dies ist in gewissem Maße ein Zirkelschluss: Vielmehr geht es darum, dass Maschinen ohne umfassendes menschliches Zutun auf ihre Umwelt reagieren und komplexe Probleme lösen. Dies kann ihnen ein Mensch beigebracht haben. Der Bezug auf menschliches Denken ist eher eine Analogie: Das menschliche Denken als solches kann – anders als Vorhersagen[891] – einer Maschine nicht beigebracht werden.

In den Anfangsjahren lernte Künstliche Intelligenz Spiele (Schach, Go, Jeopardy)[892] und mathematische Logik. Hieraus haben sich in der Folgezeit das Feld des Machine Learning (ML) und schließlich das Deep Learning entwickelt. Bei Machine Learning bringt sich die Maschine ein intelligentes Verhalten selbst bei: ML ist immer ein Unterfall von KI, aber KI nicht gleichbedeutend mit ML.[893] Die Grundlage für Maschinelles

---

nomics of Artificial Intelligence, S. 6; *Nilsson*, The Quest for Artificial Intelligence, S. 77f.

887 Vgl. *Görz/Nebel*, Künstliche Intelligenz, S. 9.

888 So auch *Herberger*, NJW 2018, 2825 (2826); zur Kritik an der englischen Terminologie: *Nilsson*, The Quest for Artificial Intelligence, S. 79.

889 *Minsky,* Steps towards Artificial Intelligence, Proceedings of the IRE, Vol. 49, No. 1, 1961; *Nilsson*, The Quest for Artificial Intelligence, S. XIII.

890 *Nilsson*, The Quest for Artificial Intelligence, S. XIII; *UK Department for Business, Energy and Industrial Strategy*, Industrial Strategy: Building a Britain fit for the future, S. 37; ähnlich *Europäische Kommission*, Factsheet Artificial Intelligence, S. 1.

891 *Agrawal/Gans/Goldfarb*, Prediction Machines, S. 5.

892 *A. Deng*, An Antitrust Lawyer's Guide to Machine Learning, S. 8.

893 Dazu *Mateos-Garcia*, The Complex Economics of Artificial Intelligence, S. 6.

Lernen wurde nach einer „Innovationsflaute“ (AI Winter)[894] in den 1980er Jahren gelegt: Die Entscheidungsregeln des Programms sollten sich im Sinne einer Rückkopplung flexibel an das Gelernte anpassen können.[895] Anders als klassische Computerprogramme werden sie nicht vollständig von Entwicklern geschrieben, sondern optimieren sich selbst iterativ.[896]

Übergreifend wird Machine Learning so verstanden: Ein Programm sucht mögliche Zusammenhänge zwischen erfassten Daten, dann baut es eine Funktion, die diesen Zusammenhang beschreibt und trifft schließlich eine Prognose. Für das Beispiel der Bilderkennung werden Bilder (z. B. 10.000 Bilder, davon 5.000, die einen Hund zeigen) als Trainingsdaten in ein Programm eingespeist. Die Bilder sind mit der jeweiligen Meta-Information[897] „dies ist ein Hund“/„dies ist kein Hund“ gekennzeichnet. Das Programm versucht, Gesetzmäßigkeiten in den Daten zu erkennen und passt dazu seine Funktionsweise rekursiv an. Idealerweise ist das Programm am Ende des Trainings in der Lage, ein unbekanntes Bild darauf zu prüfen, ob es einen Hund zeigt. Jeder Nutzung dieser Varianten der Künstlichen Intelligenz geht damit ein Zeitraum des Trainings voraus.

2. Überblick: Funktionsweise

Unter den Begriff „Künstliche Intelligenz“ fallen eine große Zahl von Ansätzen und Methoden, um komplexe Probleme zu lösen. Grundsätzlich braucht es dazu hohe Rechenleistungen, große Datensets und intelligente Algorithmen. Die nötige hohe Rechenleistung wurde von dem Fortschritt der Hardware in den letzten Jahren ermöglicht.[898] Die einzelnen Units sind günstiger, kleiner und energieeffizienter geworden, sodass heute Smartphones und kleinste Sensoren damit ausgestattet werden. Weitere Fortschritte scheinen wahrscheinlich und entsprechende Innovationen

894 Zum Begriff: *UK Select Committee on Artificial Intelligence*, AI in the UK, S. 19.
895 So *Kirste/Schürholz*, Einleitung. Entwicklungswege zur KI, in: Wittpahl (Hrsg.), Künstliche Intelligenz, S. 21–35 (24).
896 *OECD*, Algorithms and Collusion, Juni 2017, S. 7; *UK Select Committee on Artificial Intelligence*, AI in the UK, S. 14.
897 Meta-Information: Eine Information über eine Information, hier die Kennzeichnung eines Trainingsdatums.
898 *Mateos-Garcia*, The Complex Economics of Artificial Intelligence, S. 6; *Stoica et al.*, A Berkeley View of Systems Challenges for AI, 16. Oktober 2017, S. 1.

sind nicht zuletzt von der aktuellen Verbreitung und Entwicklung neuer KI-Systeme getrieben.[899]

Algorithmen sind Organisations- oder Rechenregeln, zum Beispiel Rezepte, Spielregeln und Benutzungsanweisungen. Sie sind menschliche Handlungsanweisungen in einem Lösungsweg.[900] Heute wird der Begriff meist im Kontext von Software genutzt. Ein Algorithmus in diesem Sinne ist eine Rechenvorschrift, die so präzise formuliert ist, dass die einzelnen Verarbeitungsschritte eindeutig daraus hervorgehen und so ein mechanisch oder elektronisch arbeitendes Gerät die Regel ausführen kann.[901] Dabei werden Eingaben in Ausgaben umgewandelt.[902] Intelligente Algorithmen sind Dienstprogrammfunktionen, deren Bedingungen zum Zeitpunkt der Codierung noch ungewiss sind. Obwohl die Algorithmen für maschinelles Lernen in den letzten Jahrzehnten raffinierter geworden sind, werden vielmehr die Daten für den KI-Boom verantwortlich gemacht.[903]

Künstlich intelligente Systeme werden mit Daten trainiert. Das „Training" ist dabei die Eingabe von Daten in das Programm, anhand derer der Algorithmus in den Daten ein Muster sucht und daraus die optimale Formel zur Lösung seiner Aufgabe entwickelt.[904] Die Formel wird in Test-Durchläufen überprüft. Aus diesem Grund unterteilt sich das Set der Trainingsdaten üblicherweise in zwei Sets: „Train Data-Set" und „Test Data-Set" (90:10 oder 80:20). Zweiteres wird nicht zum Training, sondern ausschließlich für Tests genutzt. Während des Trainings werden – je nach Lösungsziel – zutreffende und nichtzutreffende Daten in das System eingegeben. Im Sinne eines Trial-and-Error-Prozesses versucht das System, Regelmäßigkeiten aus den Daten abzuleiten. Dabei profitiert das Training von einer großen Datenmenge: Die Treffsicherheit steigt mit der

---

899 *Stoica et al.*, A Berkeley View of Systems Challenges for AI, 16. Oktober 2017, S. 7; z. B. hat Google einen Einplatinenrechner für KI-Projekte entwickelt, die TensorFlow Processing Unit (TPU): *Hansen*, Google bringt eigenen Einplatinenrechner für KI-Projekte, 6. August 2018.

900 *Von Hellfeld*, GRUR 1989, 471 (477); *OECD*, Algorithms and Collusion, Juni 2017, S. 6.

901 *Cormen/Leiserson/Rivest/Stein*, Algorithmen, S. 5; *UK Select Committee on Artificial Intelligence*, AI in the UK, S. 14.

902 Vgl. *Kastl*, GRUR 2015, 136 (137).

903 *A. Gal*, It's a Feature, not a Bug: On Learning Algorithms and what they teach us, OECD, DAF/COMP/WD(2017)50, S. 3: „information explosion".

904 *Sivinski/Okuliar/Kjolbye*, ECJ, Vol. 13, Nos. 2–3, S. 199–227, 203 (2017); *Europäische Kommission*, Mitteilung vom 25. April 2018, COM(2018) 237 final, S. 12; *Europäische Kommission*, Weißbuch KI, COM(2020) 65 final, 19. Februar 2020, S. 9.

Zahl der qualitativ hochwertigen Trainingsdaten, aber mit abnehmendem Grad.[905] Die Qualität der Datenauswertung steigt nicht linear, Trainingsdaten sind nicht bis zur Unendlichkeit nützlich. Insbesondere wird es bei extrem umfangreichen Datenmassen nötig, unnütze Daten herauszufiltern. Die Bereinigung, Validierung und Standardisierung der Daten ist ein zeitaufwendiger Vorgang.[906] Es ist davon auszugehen, dass es für den jeweiligen Zweck des Trainings von Machine-Learning-Systemen ein wirtschaftliches Optimum einer Datenmenge gibt. Das genaue Optimum wird nicht zuletzt auch von dem Preis der Datenspeicherung bestimmt, wozu anzumerken ist, dass heute dank Cloud-Diensten nicht mehr von jedem Datenverarbeiter Datenzentren zu errichten sind. Statt anfänglicher Investitionskosten (sowie nutzungsabhängiger Energie- und Wartungskosten) wird die Datenspeicherung und -verarbeitung nutzungsabhängig gezahlt.

Mögliche Aufgaben, die an selbstlernende Systeme gestellt werden, betreffen Vorhersagen, Korrelationen, Optimierungsmöglichkeiten, die Identifizierung von Anomalien und Bewertungen. Bei einer Prognose gibt das Programm auf Grundlage historischer Daten an, wann und mit welcher Wahrscheinlichkeit ein Ereignis eintreten wird; oft wird dies mit einem Handlungsvorschlag verknüpft. Dabei werden Korrelationen durch Extraktion verborgener Strukturen oder Zusammenhänge aus den Daten ausgegeben. Soll das Ausgabedatum ein Optimierungsvorschlag sein, untersucht das Programm eine bestimmte Zielfunktion, etwa einen ressourcenschonenden und schnellen Transportweg. In einem Ranking werden mehrere Handlungsvorschläge verglichen. Diese Aufgaben werden an selbstlernende System gestellt, weil eine traditionelle Software bei der Lösung zu starr vorgegebenen Funktionen folgt.

Machine Learning wird in drei Kategorien eingeteilt: überwachtes, unüberwachtes und verstärktes Lernen. Überwachtes Lernen (Supervised Machine Learning) arbeitet mit gekennzeichneten Eingabedaten. Es wird zum Beispiel im Bereich der Preisentwicklung, vorausschauenden Instandhaltung und Bilderkennung eingesetzt. Bei dem unüberwachten Lernen (Unsupervised Machine Learning) lernt das Programm aus Daten, ohne dass diese vorher zugeordnet oder gekennzeichnet sind. Es muss selbst in den

905 *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 8f mit Verweis auf http://vision.stanford.edu/aditya86/ImageNetDogs/; z. B. *Bajari et al.*, The Impact of Big Data on Firm Performance, NBER Working Paper No. 24334, S. 39; *Hestness et al.*, Deep Learning Scaling Is Predictable, Empirically, 1. Dezember 2017.

906 *Luong/Chou*, Doing Our Part to Share Open Data Responsibly, Blog Google, 5. März 2019.

Daten Strukturen erkennen und interpretieren. Eine wichtige Variante des Unsupervised Machine Learning ist das Clustering.[907] Beim Clustering werden vom Programm Ähnlichkeiten in den Daten erkannt und die Daten daraufhin in Gruppen geordnet; es kommt für Empfehlungssysteme, etwa in Online-Shops, zum Einsatz. Die dritte bedeutende Gruppe ist das verstärkte Lernen (Reinforcement Learning). Das Computerprogramm lernt aus Erfahrungen und wird für richtige Ergebnisse belohnt.[908] Verstärktem Lernen wird eine wichtige Rolle in der Robotik prophezeit;[909] es eignet sich besonders für dynamische Umgebungen.

Deep Learning ist eine komplexe Variante des maschinellen Lernens und nutzt zum Lernen künstliche neuronale Netze, also Algorithmen, die die Netzstrukturen von Nervenzellen imitieren und dafür Algorithmen in mehreren Ebenen schichten.[910] In der Bilderkennung wären die Eingabewerte für das künstliche neuronale Netz die Farbwerte einzelner Pixel und der Ausgabewert die Antwort auf die Frage, ob das Bild einen bestimmten Gegenstand (z. B. Auto) zeigt. Eingehende Werte werden auf der Eingangsebene verarbeitet, der Ausgabewert zur nächsten Ebene weitergeleitet und dieser Prozess wird wiederholt, bis die finale Ausgabeebene erreicht ist. Dabei beginnen die Algorithmen, Muster zwischen den einzelnen Merkmalen einzelner Daten auf unterschiedlichen Ebenen zu erkennen. Die Justierung der Ebenen und Verknüpfungen ist bei Deep Learning sehr rechenaufwendig: Es wird eine große Menge an Trainingsdaten benötigt, um das Netz richtig einzustellen und die Fehlerhäufigkeit zu senken.[911] Deep Learning lernt besonders schnell und akkurat und ist dem Menschen bei der Analyse von Big Data überlegen. Das System erkennt deutlich mehr Korrelationen, als ein Mensch in eine Software codieren würde; nicht zuletzt, weil sie vom Menschen nicht bewusst wahrgenommen und beim Programmieren bedacht werden. Daher wird es für besonders vielversprechend gehalten. Die Flexibilität und Kreativität des menschlichen Gehirns kann es jedoch nicht abbilden. Außerdem ist es bisher kaum möglich, vom

907 *Kirste/Schürholz,* Einleitung. Entwicklungswege zur KI, in: Wittpahl (Hrsg.), Künstliche Intelligenz, S. 21–35 (27).
908 Mit Beispielen *A. Deng*, An Antitrust Lawyer's Guide to Machine Learning, S. 8.
909 *Kober/Bagnell/Peters*, IJRR, Vol. 32, No. 11, S. 1238–1274 (2013); dazu *Kirste/Schürholz,* Einleitung. Entwicklungswege zur KI, in: Wittpahl (Hrsg.), Künstliche Intelligenz, S. 21–35 (29).
910 Ausführlich: *Del Toro Barba*, ORDO 68, S. 217–248, 219ff (2017).
911 Vgl. *A. Deng*, An Antitrust Lawyer's Guide to Machine Learning, S. 5.

Ergebnis auf den Lernprozess zu schließen.[912] Trotz der bemerkenswerten Entwicklung von Schachcomputern zu künstlichen neuronalen Netzen werden die Optimierung von Hardware, den Algorithmen und der zunehmenden Expertise weiter eine wichtige Rolle spielen.[913]

3. Voraussetzungen zur Entwicklung von selbstlernenden Systemen

Künstliche Intelligenz funktioniert synergistisch. Wie angedeutet, benötigt das Training eines selbstlernenden Systems drei Komponenten, die ohne einander von geringerem Nutzen wären als miteinander: geeignete Hardware, intelligente Algorithmen und Trainingsdaten.[914] Darüber hinaus ist menschliche Expertise zur Programmierung, Justierung und Überwachung der Software unverzichtbar. Soll die Anwendung außerdem extern zur Nutzung angeboten werden, ist eine darauf ausgerichtete Nutzeroberfläche und Infrastruktur unverzichtbar, um den Kunden eine Integration und die Anpassung der Software an ihre Bedürfnisse zu ermöglichen.

Zentral für die Qualität einer Machine-Learning-Anwendung sind Erfahrungen (Feedback) und Daten, die diese Erfahrungen wiedergeben. Ohne Daten kann der Lernprozess nicht eingeleitet werden. Die Treffergenauigkeit und Qualität der Prognosen und Empfehlungen hängt entscheidend vom Umfang und der Qualität der Trainingsdaten ab. Der Algorithmus ist allein wenig wert, sondern wird erst dann effektiv, wenn eine kritische Menge an guten Daten zum Training eingegeben werden. Die Wichtigkeit der Datenqualität ist offensichtlich: Eine KI-Anwendung ist nur so gut, wie die Daten, auf denen sie basiert. Dies illustriert der von Microsoft entwickelte Chat-Bot „Tay“: Kurze Zeit nach seiner Veröffentlichung im März 2016 befürwortete der Bot auf Twitter Völkermord und gab antisemitische und rassistische Nachrichten aus.[915] Bestimmte Nutzergruppen haben dem Bot gezielt antisemitische und rassistische Aussagen zugeführt, die der Chat-Bot wiederum als Lerngrundlage im Training nutzte.[916] Uner-

912 *A. Gal*, It's a Feature, not a Bug: On Learning Algorithms and what they teach us, OECD, DAF/COMP/WD(2017)50, S. 5; *Mateos-Garcia*, The Complex Economics of Artificial Intelligence, S. 8.
913 Vgl. *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 8f.
914 *Del Toro Barba*, ORDO 68, S. 217–248, 222 (2017); *Sivinski/Okuliar/Kjolbye*, ECJ, Vol. 13, Nos. 2–3, S. 199–227, 202 (2017).
915 Vgl. *Steiner*, Was Microsoft durch „Tay“ gelernt hat, FAZ, 26. März 2016.
916 *Beuth*, Twitter-Nutzer machen Chat-Bot zur Rassistin, Die Zeit, 24. März 2016.

wünschte Eingabedaten werden daher heute in der Regel mit Missbrauchsfiltern daran gehindert, zur Lerngrundlage zu werden. Die Künstliche Intelligenz selbst kann sie nicht ohne Weiteres identifizieren und lernt aus „schlechten" ebenso wie aus „guten" Daten.[917]

Qualitativ hochwertige Daten sind solche, die nicht veraltet, irrelevant, dupliziert, falsch gekennzeichnet oder unstrukturiert sind. Die Sicherung einer gewissen Datenqualität setzt mindestens Aktualität, Vollständigkeit und zeitliche Lückenlosigkeit voraus, um den Produktionsablauf abzubilden. Weiterhin ist eine Vielzahl identischer Signale weniger wert als multidimensionale Daten:[918] Der optimale Lerneffekt ergibt sich nicht aus einem großen Set sich wiederholender gleicher Daten, sondern vielmehr einer großer Zahl von Signalen mit unterschiedlichem Informationsgehalt, die das selbstlernendende System auf vielfältige Situationen vorbereitet. Die Qualität ist relativ in Bezug auf den Verarbeitungszweck und nicht statisch.[919]

Wenn historische Daten vorurteilsbeladen sind, kann das Lernergebnis nur vorurteilsbelastet sein: Dies wird als Data Bias bezeichnet.[920] Ein Beispiel hierfür ist ein selbstlernendes System, das Bewerber für eine Stellenanzeige filtert. Nutzt das System Daten, in denen sich bewusste oder unbewusste Vorurteile – bezüglich Alter, Geschlecht oder sozio-ökonomischem Hintergrund – widerspiegeln, werden diese perpetuiert.[921] Data Biases können gesellschaftliche Ziele der Gleichbehandlung unterminieren.[922] Die Auswirkungen wiegen bei Personenbezug besonders schwer; ein Data Bias ist aber auch bei Maschinendaten denkbar, beispielsweise, wenn mit mangelhaften Materialien gearbeitet wurde und der Produktionsablauf gestört wurde, dies aber nicht aus den Daten erkennbar ist.

Neben der Qualität der Daten ist die Größe des Trainingsdatensets von entscheidender Bedeutung. Zu einem gewissen Grad kann das Volumen

---

917 *Mateos-Garcia*, The Complex Economics of Artificial Intelligence, S. 9: „careless".

918 *Nuys*, WuW 2016, 512 (514f); *Sivinski/Okuliar/Kjolbye*, ECJ, Vol. 13, Nos. 2–3, S. 199–227, 221 (2017).

919 Vgl. *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 103; *Datenethikkommission*, Gutachten 2019, S. 83; *A. Deng*, An Antitrust Lawyer's Guide to Machine Learning, S. 5; *Hoeren*, ZD 2016, 459 (461ff) zu den Qualitätsgrundsätzen der OECD 1980 und des Pipeda Act 2000; *OECD*, Data-Driven Innovation, S. 194.

920 *UK Select Committee on Artificial Intelligence*, AI in the UK, S. 41.

921 Beispiel siehe *UK Select Committee on Artificial Intelligence*, AI in the UK, S. 41f.

922 *Himel/Seamans*, Artificial Intelligence, Incentives to Innovate, And Competition Policy, CPI Antitrust Chronicle December 2017, S. 9; *Hoeren*, MMR 2016, 8 (8f).

des Datensets Qualitätsmängel auffangen, insofern beeinflussen sich beide Variablen. Die Art der benötigten Daten wird dabei von dem Verarbeitungsziel bestimmt: Eine Spracherkennungssoftware lernt anhand von menschlichen Stimmen und eine Bilderkennungssoftware lernt anhand von Bilddateien. Für die Abläufe der Industrie 4.0 werden Sensortechnologien die Menge der zu analysierenden Daten exponentiell erhöhen; das Internet of Things trägt dazu bei, dass zusätzliche Objekte und Ereignisse digital erfasst werden.[923] Die Daten müssen erst erfasst, organisiert und in ein Datenlager (data warehouse) eingespeist werden, bevor sie zum Training von Machine-Learning-Algorithmen genutzt werden können.

Das Volumen der benötigten Daten richtet sich wiederum nach der Methode des Machine Learnings: Deep Learning ist die wohl „datenhungrigste" Variante.[924] Für jedes selbstlernende System gibt es eine individuelle mindestoptimale Datenmenge, die ohne exakte Kenntnis der Datenqualität nicht mit Sicherheit kalkuliert werden kann. Grundsätzlich gilt, dass ein Dienst mit geringem Anspruch an die Datenmenge schneller und einfacher umgesetzt werden kann. Dies kann kaum gesetzgeberisch oder behördlich in kurzer Zeit beurteilt werden: Einerseits wird die technische Expertise fehlen, andererseits ist die Kalkulation hoch prognostisch. Bei aktualisierungsbedürftigen Anwendungen ist darüber hinaus zu bedenken, dass das Training nie richtig abgeschlossen ist. Die Eingabe aktueller Daten ist eine Voraussetzung dafür, dass das Programm die gestellten Aufgaben weiterhin bewältigt und sich flexibel an sich verändernde Bedingungen anpasst. Insofern kann die Voraussetzung nicht nur ein initiales Trainingsdatenset sein, sondern ein weiterer Strom aktueller Daten. Dies gilt nicht für die Erkennung von Hunden auf Bildern, aber etwa für selbstlernende Systeme im Aktienhandel.

Verallgemeinernd kann angenommen werden, dass für besonders komplexe Aufgaben auch besonders große Datensets nötig sind. Baidu hat etwa 50.000 Stunden an Audioaufnahmen als Trainingsdaten für die Spracherkennung und es wird angenommen, dass das Training weiterer 100.000 Stunden Audiomaterial mit über zehn Jahren Abspielzeit geplant ist.[925] Für die Gesichtserkennung gab Baidu wohl 200 Millionen Bilder zur Gesichtserkennung in das System ein, während die prominentesten

923 *Drexl*, Designing Competitive Markets for Industrial Data, 2016, S. 10; *Kerber*, GRUR Int. 2016, 989 (990).
924 Vgl. *Bhardwaj/Di/Wei*, Deep Learning Essentials, S. 8.
925 Angaben nach *Bhardwaj/Di/Wei*, Deep Learning Essentials, S. 15.

wissenschaftlichen Studien dazu etwa 15 Millionen Bilder nutzten.[926] Die enormen Datensets, über die nur der Staat oder besonders große Unternehmen verfügen, haben die Entwicklung selbstlernender Systeme mit Beginn der „Big Data Era" entscheidend vorangetrieben. Gerade datenreiche Unternehmen mit publizierenden Forschungsabteilungen haben zur Weiterentwicklung der Disziplin beigetragen.

Banko und Brill haben das Verhältnis von Datenmenge und Treffgenauigkeit für Natural Language Disambiguation untersucht. Die für das Training notwendigen Texte sind im Internet mit Kennzeichnung (Labeled Data) gratis und ohne großen Aufwand zu sammeln.[927] Sie untersuchten die Lerneffekte mit unterschiedlich großen Datensets. Das Ergebnis der Untersuchung war, dass größere Trainingsdatensets zu einem erfolgreicheren Lernergebnis beitragen. Ein größeres Datenset enthält in der Regel mehr und vielfältigere Erkenntnisse, anhand derer Algorithmen richtige und falsche Antworten erkennen und akkurater bewerten können.[928]

Schließlich müssen die Daten mit der Software kompatibel sein, also ein bestimmtes Format haben.[929] Daten für Supervised Machine Learning müssen zudem mit Meta-Informationen versehen sein (Labeled Data[930]). Nicht zuletzt ist die Konsistenz der Variablen nötig, also die Einheitlichkeit der Angaben wie die Verwendung einer einheitlichen Währung für finanzielle Variablen. Dies ist problematisch für sehr alte Datensets, die erst in ein bestimmtes Format gebracht werden müssen. In Echtzeit erfasste Daten können unmittelbar Machine-Learning-freundlich gespeichert werden. Das Datenset muss schließlich der Verarbeitung angepasst werden. An dieser Stelle ist die Expertise von Data Scientists erforderlich[931] – Rohdaten können nicht blind in ein System eingegeben werden.

Neben dem Volumen von Datensets hat auch eine Optimierung der Software die Entwicklungen auf dem Gebiet der selbstlernenden Systeme

---

926 Angaben nach *Bhardwaj/Di/Wei*, Deep Learning Essentials, S. 15.

927 Zum Folgenden *Banko/Brill*, Scaling to Very Very Large Corpora for Natural Language Disambiguation, S. 1, 7.

928 *Petit*, Technology Giants, S. 36; *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 13f, 80.

929 *Luong/Chou*, Doing Our Part to Share Open Data Responsibly, Blog Google, 5. März 2019.

930 Beispiele nach *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 2f: OpenImages, ein Datenset bestehend aus 9,5 Millionen gekennzeichneten Fotos; sowie das Stanford Dog Dataset mit 20.580 Fotos von 120 Hundearten.

931 *Mateos-Garcia*, The Complex Economics of Artificial Intelligence, S. 10.

verursacht: Mit wachsender Expertise und den Lerneffekten der letzten 50 Jahre konnte sich die Disziplin der Künstlichen Intelligenz auf die bevorstehenden Datenmassen vorbereiten. Dies ist für diese Betrachtung relevant, weil der Wert, die Portabilität und die Qualität von Daten entscheidend von der Methode ihrer Verarbeitung bestimmt werden. Durch die Verbesserung der Werkzeuge gewinnen auch die Daten als Rohstoffe an Wert. Je universeller sie nutzbar sind, desto eher werden Rufe nach dem Zugang zu fremden Maschinendaten laut. Dank wissenschaftlicher Veröffentlichungen und der Bereitstellung zahlreicher selbstlernender Systeme in der Open Source ist die technologische Grenze in dieser Disziplin transparent. Mehrere prominente Unternehmen haben sich dazu entschieden, die Algorithmen zugänglich zu machen – entweder in der Open Source oder zur zahlungspflichtigen Nutzung. Nicht für alle Anwendungsabsichten funktionieren diese „vorgefertigten" Lösungen. Deep Learning braucht etwa Algorithmen für neurale Netze[932]; diese müssen je nach Anwendungsziel gewichtet werden.

Derzeit sind Algorithmen nicht ohne Weiteres als geistiges Eigentum schutzfähig.[933] Ein fehlender Immaterialgüterschutz bedeutet keine freie Zugänglichkeit, aber ermöglicht die Imitation der Algorithmen. Nicht zuletzt deswegen werden sich zahlreiche Technologie-Unternehmen dazu entschieden haben, ihre KI-Systeme und Algorithmen in der Open Source bereitzustellen. Geeignete Trainingsdaten werden von jenen Unternehmen nur ausschnittsweise freigegeben, was darauf hindeutet, dass ihre Exklusivität den jeweiligen Technologieunternehmen aktuell bedeutender erscheint.

Diese Software muss wiederum auf geeigneter Hardware laufen. Die Rechenleistung der Computer potenzierte sich in den letzten Jahren: Was 1995 nur Supercomputer konnten, kann heute jedes Smartphone (100 Milliarden Rechenoperationen pro Sekunde). Es wird prognostiziert, dass die Leistungsstärke von Mikrochips bis 2040 noch einmal um den Faktor 1000 zunehmen wird.[934] Dieser Fortschritt ist nicht zuletzt getrieben von den Ansprüchen, die selbstlernende Systeme an Hardware stellen. Gerade Deep Learning benötigt spezielle leistungsstarke Hardware, um die Algo-

---

932 Beispiele nach *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 2f: TensorFlow, Caffe, MXNet, Theano.

933 *Drexl et al.*, GRUR Int. 2016, 914 (916); *Richter/Surblytė/Wiedemann*, Zugang zu Daten in der datengetriebenen Wirtschaft, MPG, Forschungsbericht 2017.

934 Siehe *Eberl*, Feature: Künstliche und natürliche Intelligenz – was werden smarte Maschinen leisten können?, 19. Februar 2018.

rithmen ablaufen zu lassen.[935] Die Leistung von GPUs hat sich ebenfalls in zehn Jahren vertausendfacht.[936] Entlang der Lieferkette werden die Sensoren immer kleiner, leistungsstärker und dank sinkender Materialkosten günstiger. Hardware ist daher – bis auf die Notwendigkeit begrenzt vorkommender Ressourcen – kein limitierender Faktor für die Entwicklung selbstlernender Systeme.

Schließlich beruhen die Qualität der Algorithmen und der Eingabedaten auf menschlicher Expertise, nämlich den Leistungen von Softwareentwicklern, Data Scientists und Forschern der Wirtschaft und der Universitäten.[937] Talent ist eine begrenzte und rivale Ressource. Langfristig können in Deep Learning und anderen Disziplinen des Machine Learning qualifizierte Entwickler ausgebildet werden, kurzfristig stehen aber nur die bereits qualifizierten Arbeitskräfte zur Verfügung. Obwohl bestimmte Software in der Open Source verfügbar ist, kann sie nicht jede ungelernte Person anwenden, sie muss angepasst werden. Ebenso wichtig sind die Entwickler, die die Daten managen und Algorithmen moderieren. Hier herrsche aktuell noch der dringendste Kapazitätsengpass des Prozesses.[938] Für das Management von Daten sind darüber hinaus auch Arbeitskräfte erforderlich, die sich mit der jeweiligen Materie auskennen und mögliche Lücken und Ungenauigkeiten im Datenset aufzeigen können.[939] Wegen der geringen Verbreitung solcher Spezialisten können sie überdurchschnittliche Gehälter und Arbeitsbedingungen verlangen, die nicht jedes Unternehmen bieten kann.[940]

Damit ergeben sich zwei limitierende Faktoren, nämlich Talent und Daten. Diese sind lediglich kurzfristig begrenzt: Es können neue Daten erfasst und neue Experten ausgebildet werden. Die Limitierung besteht jeweils sowohl in quantitativer als auch qualitativer Hinsicht. Insbesondere ist das Vorliegen aller Voraussetzungen zur Entwicklung selbstlernender

---

935 Beispiele nach *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 2f: CPUs (Central Processing Units), GPUs (Graphical Processing Units), TPUs (Tensor Processing Units).

936 *Bhardwaj/Di/Wei*, Deep Learning Essentials, S. 14.

937 *Bourreau/de Streel/Graef*, Big Data and Competition Policy, S. 7; *Manne/Morris/Stout/Auer*, Comments of ICLE, 7. Januar 2019, S. 15; *Rock*, Engineering Value: The Returns to Technological Talent and Investments in AI, S. 1ff.

938 *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 3.

939 Dazu *Davenport/Patil*, Data Scientist: The Sexiest Job of the 21st Century, HBR, Oktober 2012.

940 Siehe *Rock*, Engineering Value: The Returns to Technological Talent and Investments in AI, S. 3ff.

Systeme in einem ausgewogenen Zusammenspiel nötig. Einzeln sind der jeweilige Wert für das entwickelnde Unternehmen sowie die wettbewerbliche Bedeutung der einzelnen Ressource niedriger.

B. Exklusive Daten in Abgrenzung zu offenen Daten (Open Data)

Die Frage nach dem Zugang stellt sich nur dort, wo ein begehrtes Gut unzugänglich ist. Unzugänglichkeit von Informationen kann rechtlicher oder faktischer Art sein. Daten, von deren Nutzung andere ausgeschlossen sind, werden als exklusiv bezeichnet. Die Europäische Kommission nutzte diese Bezeichnung etwa in der Zusammenschlusssache *Facebook/WhatsApp*.[941] Exklusivität ist von Wesentlichkeit zu unterscheiden: Eine Ressource ist nie per se essentiell, sondern nur in Beziehung zu einem bestimmten Ziel – etwa, wenn sie dazu dient, ein bestimmtes Produkt oder eine bestimmte Leistung anzubieten. Beispielsweise sind Echtzeitdaten für Stauvorhersagen wesentlich, aber nutzlos für Versicherungen, die vielmehr historische Daten benötigen.[942] Weiterhin müssen die Exklusivität von Daten (syntaktisch) und die Exklusivität von Wissen (semantisch) differenziert werden: Die Exklusivität hinsichtlich eines Datums oder Datensets beinhaltet nicht die Exklusivität über die spezifische Information, die es offenbart.[943] Bezüglich selbstlernender Systeme stellt sich folglich die Frage, ob ein bestimmter Lerneffekt nur mittels eines ganz bestimmten Datensets zu erreichen wäre. Mit einer prognostizierten Gesamtdatenmenge von 163 Zettabytes im Jahr 2025[944] mangelt es nicht an Daten, sondern es gelingt nicht, die erfassten Daten jedem interessierten Unternehmen zugänglich zu machen.

Datengetriebene Geschäftsmodelle lassen sich grundsätzlich in zwei Gruppen einteilen: Eine Gruppe handelt mit Daten als Produkt (z. B. Datenbroker), die zweite Gruppe verarbeitet Daten als Input. Dort, wo mit Daten gehandelt wird, sind sie nie exklusiv nur einem Datenverarbeiter

941 Europäische Kommission, Entscheidung vom 3. Oktober 2014, COMP/M.7217 Rn. 189 – *Facebook/WhatsApp*: „not within Facebook's exklusive control".

942 Ein PKW-Hersteller, der historische Daten kontrolliert und so selbst Kfz-Versicherungen anbietet, könnte einen Wettbewerbsvorteil haben, so: *Haucap*, Big Data aus wettbewerbs- und ordnungspolitischer Perspektive, in: Morik/Krämer (Hrsg.), Daten, S. 95–142 (98).

943 *Colangelo/Maggiolino*, ECJ, Vol. 13, S. 249–281, 256 (2017).

944 *Reinsel/Gantz/Rydning*, Data Age 2025, IDC White Paper 2017, S. 3.

zugänglich; dort, wo mit ihnen gearbeitet wird, ist Exklusivität eher zu vermuten.

I. Exklusivität – Begriff

„Exklusiv" ist etwas, wovon ausgeschlossen wird.[945] Die Exklusivität setzt dabei die potentielle Ausschließbarkeit und die Realisierung der Ausschließbarkeit voraus. Informationen sind grundsätzlich nicht ausschließbar. Die Ausschließbarkeit des Datenzugangs ergibt sich aus dem grundsätzlich begrenzten Zugriff auf das jeweilige physische Speichermedium. Der Zugriff kann sowohl physisch (z. B. durch Übergabe des Speichermediums) oder durch Eröffnung eines Zugangs ermöglicht werden. Grundsätzlich kann bei bereits erfassten Daten von einer Ausschließbarkeit ausgegangen werden.[946] Die Herrschaft über geschäftliche Informationen, die sich aus der Herrschaft über physische Produktionseinheiten ergibt, begründet zumeist auch die Ausschließbarkeit Dritter von diesen Informationen.[947] Ein Unternehmen könnte exklusiven Zugriff auf digitale Daten haben, weil es das einzige war, das den von den Daten erfassten Vorgang wahrgenommen hat. Ein hierfür gern gewähltes Beispiel ist ein Unternehmen, das als einziges ein digitalisiertes Bild eines Ortes hat, der von einem Naturdesaster zerstört wurde.[948] Auf die Industrie 4.0 übertragen gilt dies für ein produzierendes Unternehmen, das mithilfe von Sensoren seine Produktion digital aufzeichnet. Die sich daraus ergebenden Informationen könnte allerdings jedes andere Unternehmen mit ähnlichem Produktionsablauf erlangen. In der Regel sollten auf semantischer Ebene Substitute zu finden sein, wie die (kurze) Geschichte der Datenwirtschaft gezeigt hat.[949]

Der Zugang hängt von einer Entscheidung des datenerfassenden Unternehmens ab. Möglicherweise sind die Daten replizierbar oder die in ihnen enthaltenen Informationen aus anderen Quellen zu erlangen. Dies ist nicht der Fall, wenn es auf die Erfahrung eines bestimmten Produktes ankommt. Beispielsweise hat der Betreiber einer Smartphone-App, die die Geo-Daten ihrer Nutzer speichert, exklusiven Zugriff auf die Daten

945 Lat. exclusio = Ausschluss.
946 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 19.
947 *Louven,* NZKart 2018, 217 (220); *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 56.
948 *Rubinfeld/M. Gal*, Access Barriers to Big Data, S. 13.
949 *Duch-Brown/Martens/Müller-Langer*, The economics of ownership, access and trade in digital data, S. 20.

seiner App. Insgesamt gibt es aber zahlreiche Apps, die entsprechende Geo-Daten erfassen, sodass keine einzige App als Monopolist für Geo-Daten bezeichnet werden könnte. Solche Daten wären jeweils syntaktisch exklusiv, aber nicht semantisch exklusiv, weil die in ihnen gespeicherte Information aus mehreren Quellen erlangt werden kann. Insofern kann auch eine „Datenquelle" exklusiv sein.[950] Mit der Frage nach dem Zugang zu exklusiven Daten im Hinblick auf Trainingsdaten ist in der Regel nicht ein spezifisches Datenset gemeint, sondern jedes Datenset, das sich für ein spezifisches Lernziel anbietet. Die verschiedenen Bezugspunkte einer Exklusivität, nämlich Daten, Informationen und Lernziele, erschweren ein einheitliches Verständnis. Während es leicht ist, andere von der Nutzung eines spezifischen Datensets auf einem physischen Speichermedium auszuschließen, ist es vergleichsweise kompliziert, andere von der Erfassung ähnlicher Daten abzuhalten.[951] Spezifische Echtzeit-Daten können oft aus mehreren Quellen erlangt werden,[952] schwieriger dürfte es sich im Hinblick auf historische Daten darstellen.

Aus wirtschaftlicher Perspektive führt die technische Ausschließbarkeit der Nutzung durch Dritte zu der Bewertung von Daten als Clubgut.[953] Clubgüter sind solche Güter, die ausschließbar sind, ohne rival zu sein. Nach ihrer Veröffentlichung geht die Ausschließbarkeit verloren und sie werden zu einem öffentlichen Gut. Exklusivität wirkt sich wirtschaftlich positiv aus, indem sie Anreize für die Generierung und Sammlung von Daten setzt, weil der Dateninhaber die Gewinne vereinnahmen kann.[954]

Der Begriff „exklusiv" wurde in Bezug auf Datenzugang ohne nähere Erläuterung in der Gesetzesbegründung der Regierung zu § 18 Abs. 3a Nr. 4 GWB verwendet.[955] Nach dem Zweck des Gesetzes solle der „exklusive" Zugang zu wettbewerbsrelevanten Daten berücksichtigt werden. Die Exklusivität ist als Kriterium im neuen § 18 Abs. 3a Nr. 4 GWB allerdings nicht enthalten. Dies könnte nahelegen, dass sie als Tatbestandsmerkmal

---

950 *Surblytė*, Data as a Digital Resource, MPI for Innovation and Competition Research Paper No. 16–12, S. 5.

951 So auch *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 6: „While databases may be proprietary, the underlying data usually is not".

952 So auch *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 7; *Drexl*, Designing Competitive Markets for Industrial Data, S. 40: „moving target".

953 *Dewenter/Lüth*, Big Data aus wettbewerblicher Sicht, Wirtschaftsdienst 2016, S. 648–654 (649).

954 *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 88.

955 BT-Drucks. 18/10207, S. 49, 51.

im Hinblick auf generelle Datensammlungen ungeeignet schien. Es wird jeweils festzustellen sein, ob ein Zugangspentent von der Nutzung eines konkreten syntaktischen Datums oder von der Erlangung sich aus entsprechenden Daten ergebender semantischer Erkenntnisse ausgeschlossen wird.

II. Vorkehrungen zur Wahrung von Exklusivität

Zur Wahrung der Exklusivität von erfassten Datensets werden meist kumulativ verschiedene Schutzvorkehrungen getroffen. Hierzu zählen die Verschlüsselung der Daten und vertragliche Abreden. Eine Datenteilungspflicht würde diese Schutzmethoden aufbrechen. Damit wäre sie ein Gegenpol zu dem Recht des geistigen Eigentums, das Schutzmethoden für die exklusive Verwertung von immateriellen Ressourcen anbietet. Das Ergebnis der im Hinblick auf Daten unternommenen Isolationsmaßnahmen wird als „Datensilos" bezeichnet. Nicht alle Hürden bei dem Datenzugriff sind bewusst errichtet: Sie können technologisch, rechtlich und verhaltensbezogen sein und an allen Punkten der Wertschöpfungskette – kumulativ – wirken.[956] Die erhöhte Sensibilisierung für die Erfassung personenbezogener Daten ist etwa eine Hürde für die Erfassung entsprechender Daten durch neue Marktteilnehmer.

1. Technologisch

Zunächst steht die Datenhoheit demjenigen zu, der die Daten selbst erhebt.[957] Diese Einheit entscheidet dann, ob sie das Datenset weitergibt, veröffentlicht oder für die eigene Nutzung mit Verschlüsselungsmethoden schützt. Maschinendaten sind häufiger als personenbezogene Daten von exklusiver Natur, weil sie in Bezug zu einer bestimmten Maschine stehen, die diese Daten weniger frei parallel ausgeben kann als Personen. Observed Data, also erfasste Daten, beziehen sich auf ein bestimmtes Ereignis, zum Beispiel die Luftfeuchtigkeit an einem Ort zu einem spezifischen Zeitpunkt. Sie werden sensorisch erfasst und die Möglichkeit ihrer Erfassung endet mit dem Ereignis. Erfasst nur ein Sensor das Ereignis,

---

956 *M. Gal/Rubinfeld*, Access Barriers to Big Data, S. 32.

957 *BMWi – Plattform Industrie 4.0 (Hrsg.)*, Marktmacht durch Datenhoheit, in: dass., Industrie 4.0 – Kartellrechtliche Betrachtungen, April 2018, S. 15–22 (16).

wird das Datum exklusiv kontrolliert von demjenigen, der den Sensor kontrolliert, ohne dass Schutzmechanismen errichtet werden müssten.[958] Die Europäische Kommission bezeichnet dies als „Single-Source-Informationen".[959] Der Hersteller einer Maschine hat einen „Vorsprung", weil er als Entwickler die Schnittstelle kontrolliert und ohne seine technische Hilfe oder vertragliche Zustimmung niemand über die Daten verfügen kann.[960] In der Praxis ließe sich für Hersteller von IoT-Geräten zwar selten durchsetzen, dass Daten nur vom Hersteller der Maschine erfasst werden, aber die initiale faktische Kontrolle verschafft Verhandlungsmacht.[961]

Beabsichtigt ein Unternehmen die exklusive Nutzung dieser Maschinendaten, wird es die Daten zusätzlich mit technologischen Vorrichtungen schützen, um den Anforderungen des Schutzes der Betriebs- und Geschäftsgeheimnisse gerecht zu werden. Der Geheimnisschutz erfordert gemäß § 2 Nr. 1 lit. b GeschGehG den Umständen angemessene Geheimhaltungsmaßnahmen durch ihren rechtmäßigen Inhaber. In dieser Hinsicht stellt der neue Geheimnisschutz strengere Anforderungen an die Manifestation des Geheimhaltungswillens als die bisherige Regelung der §§ 17–19 UWG. Auch wenn damit nicht primär der Schutz von Daten beabsichtigt wird, motiviert dieses rechtliche Erfordernis zu einer Wahrung der exklusiven Kontrolle von Daten.

Beispiele für entsprechende technologische Vorrichtungen sind die Etablierung nicht universell lesbarer Datenformate, Firewalls, Zugangskontrollen mit Security-Token, Zwei-Faktor-Authentifikation und die verschlüsselte Übertragung von Daten. Mit diesen Maßnahmen verfolgen Unternehmen eigene Ziele der Datensicherheit und den Schutz vor dem Zugriff Dritter auf sicherheitsrelevante Daten, zudem sichern sie aber auch faktisch die Exklusivität der von ihnen erfassten Daten.[962] Auf diese Weise können lediglich bereits erfasste Datensets geschützt werden. Technologische Möglichkeiten, um Dritte von der Erfassung ähnlicher Daten auf

---

958 *Colangelo/Maggiolino*, ECJ, Vol. 13, S. 249–281, 256 (2017).

959 *Europäische Kommission*, Prioritätenmitteilung vom 5. Dezember 2008, KOM(2008) 832 endg., S. 28, Fn. 58; vgl. „Sole-Source-Produkte": *Wielsch*, Zugangsregeln, S. 132f.

960 *BMVI*, „Eigentumsordnung" für Mobilitätsdaten?, August 2017, S. 117.

961 *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 24.

962 *Europäische Kommission*, Mitteilung vom 10. Januar 2017, COM(2017) 9 final, S. 11; *Dewenter/Lüth*, Datenhandel und Plattformen, S. 43; *Scheuch*, Eckpunkte der rechtlichen Behandlung von Daten, in: Morik/Krämer (Hrsg.), Daten, S. 49–77 (63); *Zech*, CR 2015, 137 (140); *Drexl et al.*, Data Ownership and Access to Data, MPI for Innovation and Competition Research Paper No. 16–10, Rn. 7.

alternativen Kanälen abzuhalten, bestehen nicht.[963] Das Gegenteil dieser technologischen Vorkehrungen ist das technologische Design für Open Data.[964]

2. Wirtschaftlich

Der Zugang zu relevanten Datensets wird vorrangig über private Datenmärkte gesteuert.[965] Die wirtschaftliche Exklusivität kann bewahrt werden, wenn die Daten zwar grundsätzlich zur Nutzung angeboten werden, die Lizenzgebühren aber so hoch sind, dass der Zugang zu den Datensets in keinem Verhältnis zu ihrem angenommenen wirtschaftlichen Wert steht.[966] Die Exklusivität von Daten hat für ein Unternehmen einen wirtschaftlichen Wert, den es mit Data-Sharing aufgibt. Ein Teil der Gegenleistung muss diese Opportunitätskosten kompensieren.[967] Zu hohe Lizenzgebühren können wiederum potentielle Zugangspetenten abhalten. Ähnlich wirken hohe Transaktionskosten oder hohe Kosten, die sich aus der Errichtung spezifischer Übermittlungssysteme ergeben. Außerdem könnten die vorgeschlagenen Vertragsbedingungen so streng sein, dass sich ein Erwerb des Zugangs nicht lohnt; hier könnte aber die AGB-Kontrolle nach §§ 305 ff BGB korrigierend eingreifen.[968]

Die Bereitstellung hochwertiger, repräsentativer Datensets verursacht allerdings Kosten. Sowohl die Erfassung als auch die Bereinigung und Speicherung erfordern Investitionen, die sich für den Bereitsteller der Daten lohnen müssen. Die Kosten für das Sammeln und Analysieren von Daten sind allerdings in Relation zu anderen Produktionsmitteln so niedrig, dass sich daraus keine generelle wirtschaftliche Exklusivität bezüglich der Datensets für von selbstlernenden Systemen zu erreichende Lernziele ergeben dürfte.[969] Für die an den Daten interessierten Unternehmen ergibt sich das Problem, dass sie möglicherweise eine hohe Zahl einzelner Ver-

963 So auch *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 7.
964 *OECD*, Data-Driven Innovation, S. 189.
965 So *Schweitzer/Peitz*, NJW 2018, 275 (275).
966 Vgl. *M. Gal/Rubinfeld*, Access Barriers to Big Data, S. 26.
967 *Drexl et al.*, Data Ownership and Access to Data, MPI for Innovation and Competition Research Paper No. 16–10, Rn. 29; *Drexl*, Data Access and Control (BEUC), S. 29; *OECD*, Data-Driven Innovation, S. 192.
968 Vgl. zu ungleicher Verhandlungsmacht: *Europäische Kommission*, Mitteilung vom 10. Januar 2017, COM(2017) 9 final, S. 11f.
969 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 5.

träge schließen müssen, um den Datensatz zusammenzustellen, wenn sie nicht bereits in Datenpools oder von Datentreuhändern verwaltet werden. Dies führt zu einer Kumulation von Lizenzgebühren[970], Bürokratie und hohen Rechtsberatungskosten, falls undurchsichtige Vertragsbedingungen gestellt werden. Das Fehlen rechtlicher Vorgaben im Hinblick auf Datenverträge lässt Raum für strategisches Verhalten von datenreichen Unternehmen und Zugangspetenten beiderseits.[971]

Besonders routiniert und etabliert ist der Datenhandel für Marketing-Daten, Risikominderung und Personensuche.[972] Es ist anzunehmen, dass hier anerkannte Industriestandards und Vertragsbedingungen bestehen, die die Transaktionskosten für Datenerwerber senken. Gerade auf Datenmarktplätzen dürften Such- und Bezahlfunktionen und Validationsmechanismen der Plattformen das Finden und Lizenzieren von Datensets vereinfachen. In Eins-zu-Eins-Datentransaktionen dürften demgegenüber die Transaktionskosten am höchsten sein, weil bilateral individuelle Vertragsbedingungen auszuhandeln sind.[973] Gleichzeitig bietet dieses Modell eine hohe Datenqualität und hohes gegenseitiges Vertrauen.

3. Vertraglich

Werden die von einem Unternehmen erfassten Daten geteilt, können die Vertragsbedingungen der Lizenzvereinbarung Vorgaben dazu enthalten, dass die Daten vom Zugangspetenten nicht an Dritte weitergegeben werden. Die Exklusivität eines „De-facto-Dateneigentums“[974] wird aufrechterhalten und der Zugang auf einen kleinen Kreis begrenzt. In der Regel

970 Eine Fragmentierung des Datenhandels erhöht wohl im Ergebnis die Kosten gegenüber dem Erwerb von einem einzelnen Datenhändler mit Monopolstellung. Dazu *Duch-Brown/ Martens/Müller-Langer*, The economics of ownership, access and trade in digital data, S. 39f mwN.

971 *Duch-Brown/Martens/Müller-Langer*, The economics of ownership, access and trade in digital data, S. 23.

972 *Duch-Brown/Martens/Müller-Langer*, The economics of ownership, access and trade in digital data, S. 37 nach einer Studie der FTC aus dem Jahr 2014; *Koutroumpis/Leiponen/ Thomas*, The (Unfulfilled) Potential of Data Marketplaces, S. 3.

973 *Koutroumpis/Leiponen/Thomas*, The (Unfulfilled) Potential of Data Marketplaces, S. 21, 25f mit Figure 1.

974 *Bourreau/de Streel/Graef*, Big Data and Competition Policy, S. 31; *Schweitzer/ Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 66 mwN; mit Beispielen: *Europäische Kommission*, Staff Working Document vom 10. Januar 2017, SWD(2017) 2 final, S. 16.

wird technisch eine Lösung gewählt, bei der die Daten auf den Servern des ursprünglichen Datenerfassers verbleiben und lediglich ein Zugang eröffnet wird. Das Zugangsmanagement[975] für das Speichermedium verbleibt bei dem Data Controller. Eine Schwäche ist dabei, dass die vereinbarten Bedingungen nur zwischen den Vertragsparteien gelten: Erlangt ein Dritter auf widerrechtlichem Weg Zugang zu den vertragsgegenständlichen Daten, entfalten die Bedingungen zur Aufrechterhaltung der Exklusivität ihm gegenüber keine Wirkung.[976] Ein „Free-Riding" Dritter würde bedeuten, dass der Erfasser der Daten auf seinen Investitionen „sitzen bleibt"[977] und das jeweilige Datenset innerhalb und außerhalb des Kreises der Berechtigten an Wert verliert.[978] Somit wird die Vertrauenswürdigkeit des Data Processors von entscheidender Bedeutung sein: Unternehmen, die sich bisher keine Reputation erarbeiten konnten, könnten daher benachteiligt sein. Vertrauensbildend könnte auch eine Kontrollmöglichkeit für die Verwendung der Daten sein.[979] Technische Lösungen zur Verfolgung der Daten können den Handel mit Daten stärken.

In der Regel wird hier das Arrow'sche Information Paradox[980] zum Tragen kommen: Daten können, anders als materielle Vertragsgegenstände, nicht in Augenschein genommen und auf diesem Wege auf Mängel und Lücken untersucht werden. Der Vertragsgegenstand, also das Datenset, ist festzulegen und der Zugangspetent hat ein Interesse daran, seinen Wert für das eigene Geschäftsmodell einzuschätzen. Hierbei stellt sich für den Dateninhaber jedoch die Herausforderung, nicht zu viel von den Daten preiszugeben, weil der Wert von Informationen sinkt, sobald sie bekannt sind. Seine Strategie wird also sein, bei geringstmöglicher Offenbarung während der Vertragshandlungen den größtmöglichen Profit zu erzielen.

---

975 Digital Rights Management, siehe *Europäische Kommission*, Staff Working Document vom 10. Januar 2017, SWD(2017)2 final, S. 16.

976 *Duch-Brown/Martens/Müller-Langer*, The economics of ownership, access and trade in digital data, S. 15; *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 68f.

977 *OECD*, Data-Driven Innovation, S. 192; wiederum zur Weitergeltung von FRAND-Bedingungen gegenüber Patenterwerbern *OLG Düsseldorf*, Urteil vom 22. März 2019 – 2 U 31/16 = GRUR-RS 2019, 6087.

978 *Koutroumpis/Leiponen/Thomas*, The (Unfulfilled) Potential of Data Marketplaces, S. 22.

979 *Kerber*, Rights on Data, S. 11; *Europäische Kommission*, Eine europäische Datenstrategie, COM(2020) 66 final, S. 9ff.

980 *Arrow*, Economic Welfare and the Allocation of Resources of Invention, in: NBER (Hrsg.), The Rate and Direction of Inventive Activity, S. 609–626.

Bei der Erteilung des Datenzugangs ist in der Regel zwischen dem Data Controller und dem Data Processor[981] zu unterscheiden. Der Data Processor ist in der Nutzung der Daten beschränkt und erhält den bedingten Zugang vom Data Controller. Diese Unterscheidung übernahm etwa die Europäische Kommission im Zusammenschlussfall *Microsoft/LinkedIn*:[982] Microsoft sei für die Daten (z. B. E-Mails, Tabellen) der Enterprise-Nutzer lediglich Data Processor.[983]

4. Exklusivität von Daten aus rechtlichen und faktischen Gründen

Neben der technischen, wirtschaftlichen und vertraglichen Wahrung der Exklusivität tragen rechtliche und faktische Aspekte dazu bei, dass die von erfassten Daten ausgehenden Wertschöpfungsmöglichkeiten nicht jedem interessierten Unternehmen offenstehen.

Das Datenschutzrecht setzt der Weitergabe von personenbezogenen Daten enge Grenzen.[984] Die Einholung einer erweiterten Erlaubnis zur Weitergabe von Daten an zunächst unbekannte Zugangspetenten, die nicht als Blankoerlaubnis für künftige Data-Sharing-Vorhaben ausgestaltet sein darf,[985] dürfte einen bürokratischen Aufwand bedeuten, der sich wiederum in höheren Preisen für den Zugang niederschlägt. Ebenso ergeben sich Rechtsunsicherheiten und die Gefahr hoher Geldbußen. Ein strenges Datenschutzrecht, das hohe Anforderungen an Informationspflichten und Opt-In-Einwilligungen stellt, kann Nutzer dazu bewegen, ihre Daten nur mit wenigen Datenverarbeitern zu teilen, um Bürokratie zu umgehen.[986]

Daneben setzt der Know-How-Schutz (Schutz von Betriebs- und Geschäftsgeheimnissen) Anreize zur faktischen Kontrolle.[987] Weitere punktuelle Schutzgesetze, die mittelbar zur Wahrung der Exklusivität beitragen,

---

981 *Sivinski/Okuliar/Kjolbye*, ECJ, Vol. 13, S. 199–227, 206 (2017).
982 Europäische Kommission, Entscheidung vom 6. Dezember 2016, M.8124 Rn. 255 – *Microsoft/LinkedIn*.
983 *Sivinski/Okuliar/Kjolbye*, ECJ, Vol. 13, S. 199–227, 207 (2017).
984 So *Louven*, NZKart 2018, 217 (221).
985 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 38 und Fn. 103 mwN.
986 Vgl. *Campbell/Goldfarb/C. Tucker*, Journal of Economics and Management Strategy, Vol. 24 No. 1, S. 43–73 (2015); relativierend: *Sabatino/Sapi*, Online Privacy and Market Structure: Theory and Evidence, Februar 2019, DICE Discussion Paper No. 308.
987 So *Louven*, NZKart 2018, 217 (221).

sind § 202a StGB (Ausspähen von Daten), § 202b StGB (Abfangen von Daten) sowie § 202d StGB (Datenhehlerei) neben deliktischen und quasinegatorischen zivilrechtlichen Ansprüchen.[988]

Faktisch ist der Zugang zu Datensets oft allein dadurch begrenzt, dass nicht bekannt ist, wer welche Daten erfasst hat. Aus Gründen des Know-How-Schutzes wird nicht preisgegeben, welche Sensoren und Datenanalyseinstrumente genutzt werden. Weiterhin können Außenstehende nicht sicher wissen, welche Daten längerfristig gespeichert werden, weshalb es Entwicklern nicht möglich ist, einen Adressaten für ihr Zugangsverlangen auszumachen. Nur für personenbezogene Daten ist gegenüber den Nutzern eine Aufklärung nötig. Hinzu kommt, dass Daten aus Gründen der Praktikabilität in der Regel in Sets gehandelt werden.[989] Für spezifische Anwendungsideen können standardisierte Sets impraktikabel und unwirtschaftlich sein. Betrachtet man nicht nur die Exklusivität des spezifischen Datensatzes, sondern eines entsprechenden Datensatzes mit ähnlichem Informationsgehalt, kann Exklusivität aufrechterhalten werden, indem andere von der Erfassung entsprechender Daten abgehalten werden. Dies ist durch die faktische Kontrolle des jeweiligen Informationsgegenstandes möglich, zum Beispiel das Verbot der Fotografie für Teilnehmer von Betriebsbesichtigungen. Die faktische Kontrolle von Daten füllt das rechtliche Vakuum, das sich aus dem Fehlen von umfassenden geistigen Eigentumsrechten für Daten und Informationen ergibt. Zuletzt können ethische Gründe datenreiche Unternehmen dazu bewegen, ihre Daten nicht offenzulegen. Ein großer Teil der KI-entwickelnden Unternehmen hat interne Kodexe für die Ziele und Grenzen Künstlicher Intelligenz entwickelt. Das Öffnen der Datensets könnte mittelbar dazu führen, dass Akteure ohne entsprechende ethische Vorgaben die Daten für diskriminierende oder bedrohliche Dienste gebrauchen, die den datenreichen Unternehmen widersprechen.[990]

---

988 Zur Stärkung von faktischer Exklusivität durch das Strafrecht: *Drexl,* Designing Competitive Markets for Industrial Data, 2016, S. 30; *Steinrötter*, MMR 2017, 731 (733).

989 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 22, 57.

990 *Luong/Chou*, Doing Our Part to Share Open Data Responsibly, Blog Google, 5. März 2019; Beispiel der Synthetic Speech Data in der 2019 ASVspoof Challenge.

III. Zwischenergebnis: Exklusive Daten in Abgrenzung zu offenen Daten (Open Data)

Die Zugangsmöglichkeit ist die grundlegende Voraussetzung für die Schaffung von gesellschaftlichem oder wirtschaftlichen Wert mithilfe von Daten.[991] Die dank Nicht-Rivalität grundsätzlich unbegrenzten Wertschöpfungsmöglichkeiten werden mit verschiedenen, grundsätzlich legitimen Instrumenten limitiert. Dies bedeutet nicht, dass generell die Wertschöpfung begrenzt ist: Sowohl die Erfassung neuer Daten als auch der Zugriff auf historische Open Data sind nicht ausschließbar. Einzelne Daten sind teilweise in ihrem Zugang begrenzt. Daher wird sich politisch eine Erweiterung der Zugangsmöglichkeiten gewünscht.[992]

Entsprechende Anreize zur Lockerung der exklusiven Speicherung von Daten können auf verschiedenen Ebenen gesetzt werden. Hierbei ist allerdings auch von Relevanz, dass das Recht für bestimmte Daten indirekt zur Exklusivität zwingt (Datenschutzrecht) oder Anreize für die Wahrung von Exklusivität setzt (Schutz von Betriebs- und Geschäftsgeheimnissen). Der Ausgleich der jeweiligen Interessen wird dabei eine besondere Herausforderung darstellen. Denkbar ist auch, dass das Senken einer Hürde zur Datenerfassung und -analyse Unternehmen dazu motiviert, an einem anderen Punkt der Wertschöpfungskette die Hürden aus strategischen Gründen zu erhöhen.[993] Ein Beispiel wäre eine Erhöhung der Kosten für die Nutzung der Cloud-Dienste als Reaktion auf die Aufweichung der Exklusivität der eigenen Trainingsdaten. Die Förderung der Generierung solcher Daten, die ohnehin zur Veröffentlichung gedacht sind (Open Data) könnte daher vielversprechend sein, weil sich weniger Interessenkonflikte ergeben. Schließlich ist anzumerken, dass es nicht *das eine* Datenset gibt, das zum Training selbstlernender Systeme eingesetzt wird; je nach Anwendungszweck unterscheiden sich die Bedürfnisse. Dank der längeren Geschichte der Veröffentlichung von Texten und Bildern im Internet bestehen viele offen zugängliche Datensets zum Training von Text- und Bilderkennung, während Open Data zur Spracherkennung rar und oft veraltet sind. Den Mangel an Dialogdaten erkennen auch die Entwickler persönlicher Assis-

991 *OECD*, Data-Driven Innovation, S. 187.

992 So *BMWi*, Künstliche Intelligenz, Artikel; *OECD*, Data-Driven Innovation, S. 187; *Drexl et al.*, Data Ownership and Access to Data, MPI for Innovation and Competition Research Paper No. 16–10, Rn. 29f.

993 *M. Gal/Rubinfeld*, Access Barriers to Big Data, S. 33.

tenten an.[994] Die exklusive Nutzung von Aufnahmen natürlicher Sprache ist wohl Teil der Geschäftsstrategie bei der Verwendung von persönlichen Assistenten, deren Funktionsfähigkeit zu einem großen Teil auf der Erkennung der Sprachbefehle beruht.[995] Insofern ergibt sich für unterschiedliche Anwendungsfelder der selbstlernenden Systeme ein unterschiedliches Niveau der Datenexklusivität, weshalb auch mögliche Feedback-Effekte unterschiedlich stark wirken dürften.

C. Maschinendaten und Big Data

Die im Rahmen der Digitalisierung der produzierenden Industrie gesammelten Daten enthalten Informationen über die Maschinen wie auch Sensordaten über die gesamte Geschichte der Produkte, der Stationen des Produktionsprozesses, Fehler und Verbrauchsdaten. „Big Data" bezeichnete ursprünglich Informationsmengen, die zu groß für die Bearbeitung mit (damaligen) Arbeitsspeichern waren und die Entwicklung neuer Analyse-Technologien einforderten. Diese Technologien sind mittlerweile entwickelt worden; heute ist Big Data von vier Dimensionen, den „Vier Vs", gekennzeichnet: Volume, Variety, Velocity und Veracity[996] – also Volumen, Vielfalt (verschiedene Datenformate und -quellen), Geschwindigkeit (der Verarbeitung) und Wahrhaftigkeit (Datenqualität und Genauigkeit). Der besondere Nutzen von Big Data liegt im Hervorbringen bisher verborgener Informationen aus Korrelationen.

Eigentlich sind alle Daten, die wir heute als solche verstehen, Maschinendaten, weil sie durch Sensoren oder Maschinen erfasst, codiert und in vom Menschen bevorzugte Formate übersetzt werden.[997] Die Europäische Kommission versteht unter Maschinendaten alle Daten, die „von Maschi-

---

994 *Byrne et al.*, Taskmaster-1: Toward a Realistic and Diverse Dialog Dataset, S. 1, 3.

995 Allerdings werden auch für das Training von selbstlernenden Systemen für persönliche Assistenten Open Data von privaten Unternehmen bereitgestellt, z. B. Taskmaster-1 von Google AI mit 5.507 gesprochenen und 7.708 schriftlichen Dialogen von Crowdsource-Workern, https://ai.google/tools/datasets/taskmaster-1, und CCPE von Google AI mit 502 Dialogen, https://ai.google/tools/datasets/coached-conversational-preference-elicitation.

996 Teilweise nur drei V's (Volume, Velocity, Variety), vgl. *Boutin/Clemens*, Defining ‚Big Data' in Antitrust, CPI Antitrust Chronicle August 2017, S. 3; *Monopolkommission*, Sondergutachten 68, S. 44, Rn. 67.

997 *Duch-Brown/Martens/Müller-Langer*, The economics of ownership, access and trade in digital data, S. 7.

nen ohne den unmittelbaren Eingriff eines Menschen im Rahmen von Computerprozessen, Anwendungen oder Diensten oder auch durch Sensoren, die Informationen von virtuellen oder realen Geräten oder Maschinen oder von einer Software erhalten", erzeugt werden.[998]

I. Datenquellen

Daten[999] können entweder unmittelbar bei der Datenerzeugung erfasst werden (Primärmarkt[1000]) oder von dem Erfasser der Daten mittelbar erworben werden (Sekundärmarkt). Üblicherweise ist der Erwerb von Rohstoffen von einem Händler einfacher als die vertikale Integration in die Erzeugung.[1001] Dies gilt für Daten nicht entsprechend.

Der schuldrechtliche Rahmen der Datenerfassung kann jeweils mit oder ohne monetäre Gegenleistung ausgestaltet sein. So können Daten auf dem Sekundärmarkt von Data-Brokern, als Open Data oder im Rahmen einer Portierung von Daten entsprechend Art. 20 DSGVO erlangt werden.

Im Falle von Observed Data sind die in Frage stehenden Daten mit einem bestimmten Ereignis, z. B. einer Nutzer- oder Maschinenaktivität, verknüpft, das nach seinem Ende nicht mehr erfasst werden kann. Insofern heben sie sich von Volunteered Data, die vom Datensubjekt ohne großen Aufwand erneut angegeben werden können, und Inferred Data, die erneut im Wege der Datenverarbeitung erlangt werden können, ab. Es wird prognostiziert, dass im Jahr 2025 Observed Data 20 Prozent der Gesamtdatenerfassung ausmachen.[1002] Geht es Unternehmen um die Steigerung der Effizienz in eigenen Abläufen, werden sie im Zweifel die benötigten Daten selbst von seinen Kunden, beziehungsweise Nutzern, erheben können.[1003] Die Selbstbeschaffung von Maschinendaten dominiert als Datenstrategie in diesem Fall.[1004] Soweit dies zu beurteilen ist, stehen in der Regel auch keine Substitute zur Verfügung, die den Daten entsprechen, die unter den Konditionen im eigenen Unternehmen erfasst würden. Insofern wird eine

998 *Europäische Kommission*, Mitteilung vom 10. Januar 2017, COM(2017) 9 final, S. 10.
999 Grundsätzliches zum Datenbegriff: Kapitel 1 A.I.
1000 *Schweitzer/Peitz*, NJW 2018, 275 (275).
1001 *Colangelo/Maggiolino*, ECJ, Vol. 13, S. 249–281, 259 (2017).
1002 *Reinsel/Gantz/Rydning*, Data Age 2025, IDC White Paper 2017, S. 13.
1003 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 22.
1004 *Europäische Kommission*, Staff Working Document vom 10. Januar 2017, SWD(2017)2 final, S. 15; *BKartA*, Big Data und Wettbewerb, S. 4.

Eigenerhebung für unverzichtbar gehalten, auch wenn sie später durch Data Sharing oder Datenhandel angereichert wird. Die Bevorzugung der Eigenerhebung unterscheidet Daten von anderen in der fertigenden Industrie genutzten Rohstoffen.

Daten können das Nebenprodukt oder Zielprodukt einer Operation sein. Ein Beispiel aus dem Bereich der selbstlernenden Systeme ist die Gesichtserkennung bei Facebook. Weil das soziale Netzwerk es seit langer Zeit Nutzern erlaubt, Freunde mit ihren Namen auf Bildern zu markieren, hat Facebook eine umfassende Datenbank von gekennzeichneten Gesichtern. Ob dies der Zweck der Funktion war oder vielmehr der Generierung von Aufmerksamkeit diente, ist nicht klar. Jedenfalls kann Facebook diese Daten nutzen, um seine selbstlernenden Systeme zur Gesichtserkennung zu trainieren. Bei Amazon werden Daten als Nebenprodukt des Verkaufs zur Schulung der Empfehlungsdienste genutzt. Die meisten der heute zum Training von Maschinen genutzten Daten wurden als Nebenprodukte generiert.[1005] Weitere Quellen von Trainingsdaten sind Daten aus Web Scraping, also der Entnahme von Daten aus öffentlich zugänglichen Webseiten bzw. die Nutzung der von Dritten gescrapeten Daten,[1006] und Open Data[1007], die üblicherweise im Standardformat und ohne viel Aufwand herunterzuladen sind. Das Ziel der meisten selbstlernenden Systeme ist, anhand von den Daten, die bei der Nutzung anfallen, weiter zu lernen. Initial kann jedoch das Zusammenstellen eines Trainingsdatensatzes erforderlich sein.

Aus diesem Grund gibt es Daten, die erlangt wurden, weil speziell zu ihrer Erfassung ein Angebot mit greifbarem Mehrwert für Nutzer geschaffen wurde (z. B. Quick, Draw!; Instagram mit Hashtag- und Ortsmarkierungen als Kennzeichnung von Bildern). Programme zur spielerischen Demonstration der Fähigkeiten selbstlernender Systeme verbessern sich bei Nutzung weiter. „Quick, Draw!" fordert den Nutzer etwa dazu auf, an seinem Computer oder Smartphone ein Strichbild zu zeichnen, das die

---

1005 So *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 24; *Duch-Brown/Martens/Müller-Langer*, The economics of ownership, access and trade in digital data, S. 14; *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 30; *M. Gal/Rubinfeld*, Access Barriers to Big Data, S. 41f.

1006 So auch *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 8; *Mueller-Freitag,* 10 Data Acquisition Strategies for Startups, 31. Mai 2016, siehe Nr. 7.

1007 *BMWi,* Wettbewerbsrecht 4.0, S. 46; *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 21f.

Anwendung während des Zeichenprozesses zu erraten versucht.[1008] Wenn ein Internetnutzer beim Anmelden auf einer Website ein „Re-Captcha"[1009] benutzt, generiert er damit Trainingsdaten für Googles Bildererkennungssysteme. So hat Google von Nutzern eine große Menge Trainingsdaten kennzeichnen lassen, die zur Verbesserung der Bilderkennung eingesetzt werden.[1010] Ähnliche Beispiele sind AutoComplete und die Korrektur von Übersetzungsvorschlägen bei Translate (Google)[1011] oder der Sprach-Lernapp Duolingo. Experiments with AI[1012] stellt weitere ML-Experimente vor, die ähnlich spielerisch Lerndaten sammeln.

Darüber hinaus können Personen auch ganz ausdrücklich dazu gebracht werden, Daten zu kreieren oder zu kennzeichnen.[1013] Die Google Crowdsource App belohnt Nutzer dafür, Bilder für lernende Bilderkennungssysteme, Übersetzungen, Bildtranskriptionen und Handschrifterkennungen zu sammeln. Die Bilder werden als Open Data unter CC-BY-4.0-Lizenz bereitgestellt.[1014]

Die Anreize zur Datenkreierung können auch monetärer Art sein. Neben den Beiträgen, die aus Neugier, dem Wunsch zur Unterstützung der KI-Entwicklung und aus „spielerischem Ehrgeiz" geleistet werden, kann das Kreieren und Kennzeichnen von maschinenlesbaren Daten auf Ho-

---

1008 Google, Quick, Draw!: https://quickdraw.withgoogle.com, Zitat: „Help teach it by adding your drawings to the world's largest doodling data set, shared publicly to help with machine learning research."; Quick, Draw! hat mehr als 50 Millionen Strichzeichnungen erfasst und anonymisiert mit Metadaten (Zeichenvorgabe und Ursprungsland) als Datenset zusammengefasst. Dieses Datenset wird als Open Data zur Verfügung gestellt: https://github.com/googlecreativelab/quickdraw-dataset.

1009 Bilder werden angezeigt und der Nutzer markiert die Bilder, die ein Auto oder einen anderen Gegenstand zeigen, um zu beweisen, dass er kein Roboter ist ("I'm not a robot").

1010 Zitat Google Developers, https://developers.google.com/recaptcha/, abgerufen am 9. Mai 2021: „reCAPTCHA makes positive use of this human effort by channeling the time spent solving CAPTCHAs into digitizing text, annotating images, building machine learning datasets. This in turn helps preserve books, improve maps, and solve hard AI problems".

1011 Die Quantität der Daten von Google Translate unterliegt wohl einer höheren Datenqualität sowie der Technologie des deutschen DeepL Translators, so *Smolentceva*, Kölner Startup schlägt Silicon Valley-Giganten, Deutsche Welle, 10. Dezember 2018.

1012 Siehe https://experiments.withgoogle.com.

1013 *Mueller-Freitag,* 10 Data Acquisition Strategies for Startups, Medium, 31. Mai 2016.

1014 Open Images Extended – Crowdsource, https://ai.google/tools/datasets/open-images-extended-crowdsourced/; CC-BY-4.0: kommerzielle Nutzung möglich.

norarbasis bezahlt und die erzeugten Datensätze lizenziert werden. Dies ist das Geschäftsmodell von Clickworker.[1015] Dort werden auftragsbasiert Trainingsdaten angefragt und von Internetnutzern bereitgestellt sowie Chatbots und andere Machine Learning-Systeme trainiert und getestet. So operiert auch das deutsche Startup TwentyBN[1016], das auf die Generierung von Videos zum Training von Machine Learning Algorithmen spezialisiert ist, mit Crowdworkern zum Aufbau von Datenbanken. Zum Ende des Jahres 2018 enthielt die Datenbank von TwentyBN zwei Millionen Videoaufnahmen, für die Clickworker je ca. 3.50 US-Dollar erhielten.[1017] Die Videos zeigen Tätigkeiten wie das Aufheben und Ablegen eines Gegenstandes, das Zeigen eines „Daumen hoch" und das Einschenken von Wasser in ein Glas. Neben der Lizenzierung bestehender Datenbanken können auch Custom-Data-Generation-Dienste gebucht werden.

Bis zu einem gewissen Grad können Trainingsdaten für selbstlernende Systeme auch künstlich von ambitionierten Marktteilnehmern erstellt werden – diese werden als ‚Fake Data'[1018], ‚Dummy Data' oder synthetische Daten bezeichnet. Synthetische Trainingsdaten sind computergenerierte Daten, die empirische Daten nachahmen. Sie können dort, wo empirische Daten nicht vorhanden oder unzugänglich sind, Abhilfe schaffen oder Datensätze anreichern und sind außerdem frei von Datenschutzbeschränkungen. Mithilfe von computergeneriertem Text[1019], virtuellen Bildern[1020] und Videosimulationen können selbstlernende Systeme bis zu einem gewissen Grad trainiert werden. Weil diese Daten maschinengeneriert sind, sei die

---

1015 Siehe https://www.clickworker.de/maschinelles-lernen-ki-kuenstliche-intelligenz/, bei den erstellten Daten handelt es sich um Sprachaufnahmen, Fotos, Videos und Texte.

1016 Siehe https://20bn.com/about; TwentyBN nutzt Amazon Mechanical Turk als Plattform für das Crowdworking, https://www.mturk.com; dazu *Mueller-Freitag*, 10 Data Acquisition Strategies for Startups, 31. Mai 2016.

1017 *Kugoth*, Dieses Startup trainiert seine Künstliche Intelligenz mit der Crowd, WIRED, 11. November 2018.

1018 Z. B. *Simonite*, Some Startups Use Fake Data to Train AI, WIRED, 25. April 2018, Zitat: „If data is the new oil, this is like brewing biodiesel in your backyard".

1019 Vgl. *Jaderberg et al.*, Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition; sowie für Re-Identifikation *Barros Barbosa et al.*, Looking Beyond Appearances; für Fußgängerzählung *Ekbatani/Pujol/Segui*, Synthetic Data Generation for Deep Learning in Counting Pedestrians.

1020 Erkennung von Objekten: *Tremblay et al.*, Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization, NVIDIA.

Kennzeichnung für überwachtes Lernen deutlich einfacher.[1021] Zudem sind die Rahmenbedingungen schon bei der Erzeugung eng definiert und die Daten daher besonders verlässlich. Es wird jedoch angenommen, dass die Lernerfolge mithilfe von Simulationen in ihrer Anwendbarkeit in der Realität limitiert seien.[1022] Diese „Simulation to Reality Gap" habe sich aber so verkleinert, dass es nur noch eines Fine-Tunings mithilfe eines verhältnismäßig kleinen Sets von empirischen Daten bedarf.[1023] Synthetische Daten werden keine Informationen hervorbringen, die echte Daten nicht auch zeigen würden, weil sie ebendiese simulieren. Die Qualität der synthetischen Datensets hänge zudem maßgeblich vom generativen Modell der Datenkreation ab. Zumindest für die vorhersehbare Zukunft sollte gelten, dass empirische Daten verlässlicher und akkurater sind. Für das initiale Training eines ML-Algorithmus werden synthetische Daten aber gerade von neuen Marktteilnehmern und Startups genutzt. In diesem Kontext wird von einer „Demokratisierung" von KI gesprochen.[1024] Auch traditionell datenreiche Akteure wie Waymo (unter dem Dach von Alphabet)[1025], Microsoft[1026], Apple[1027] und Facebook[1028] nutzen synthetische Da-

1021 Siehe *Nisselson*, Deep Learning with Synthetic Data Will Democratize the Tech Industry, TechCrunch.

1022 Siehe Zitat Primack in *Stephens*, Face recognition for galaxies: Artificial intelligence brings new tools to astronomy, University of California Santa Cruz, 23. April 2018.

1023 Allgemein zum Vergleich mit emprischen Daten *Patki/Wedge/Veeramachaneni*, The Synthetic Data Vault; zum Fine-Tuning siehe Zitat Belongie: „the simulation-to-reality gap is rapidly disappearing. At the very minimum, we can pretrain very deep convolutional neural networks on near-photorealistic imagery and fine tune it on carefully selected real imagery.", *Nisselson*, Deep Learning with Synthetic Data Will Democratize the Tech Industry, TechCrunch.

1024 Z. B. *Nisselson*, Deep Learning with Synthetic Data Will Democratize the Tech Industry, TechCrunch.

1025 Zitat „Waymo has collected in excess of one billion miles generated through computer simulation, suggesting that their virtual fleet consists of thousands of simulated vehicles.", *Grzywaczewski*, Training AI for Self-Driving Vehicles: the Challenge of Scale, 9. Oktober 2017, NVIDIA Developer Blog.

1026 *Hassan/Elaraby/Tawfik*, Synthetic Data for Neural Machine Translation of Spoken-Dialects, Microsoft AI.

1027 *Shrivastava et al.*, Learning from Simulated and Unsupervised Images through Adversarial Training, Apple Inc.; das Paper behandelt die Computergenereation realisitischer Bilder von Augen zur Verbesserung von Gaze Detection Software. Einige der beteiligten Forscher arbeiteten an der Gaze-Detection-Software des iPhone X mit; Apple bestätigt nicht, dass die synthetischen Trainingsdaten in das Projekt Eingang fanden.

ten. Daneben bieten sie sich für die universitäre Ausbildung an.[1029] Zu einem gewissen Grad stellen „künstliche Daten für Künstliche Intelligenz" auch die mögliche Exklusivität von Datensätzen in Frage: Wenn bestimmte Datensätze vom Computer simuliert werden können, ist die Exklusivität entsprechender empirischer Datensammlungen ein weniger gewichtiger Wettbewerbsvorteil. Neben synthetische Daten treten „Fake Data", unter denen eher solche Daten zu verstehen sind, die speziell zum Zweck des Trainings selbstlernender Systeme manuell erstellt wurden. Neben den Möglichkeiten der schnellen und günstigen Erstellung großer computergenerierter Datensätze dürften sie aber an Bedeutung verlieren.

Schließlich bleibt die Möglichkeit, auf dem Sekundärmarkt den Zugang zu Daten zu erwerben. Es sind entsprechende Angebote bekannt, aber nach Auffassung der Europäischen Kommission viel zu verhalten.[1030] Insbesondere Maschinendaten würden zu selten angeboten. Die auf dem Sekundärmarkt verbreiteten synthetischen Daten und die gezielt zu KI-Lernzwecken angefertigten Datensets werden auf dem Datenmarkt gehandelt, wenn sie nicht gezielt nur bloßen Inhouse-Verwendung angefertigt wurden. Die Herausbildung dieses Geschäftsmodells dürfte Datenmärkte weiter stärken. Neben der Lizenzierung einzelner Datensets ist schließlich ein Unternehmenskauf mit dem Ziel des Zugangs zu den Daten des erworbenen Unternehmens denkbar. Hierfür muss das begehrte Datenset nach außen bekannt, von höchster Relevanz, aber unzugänglich sein, andernfalls wäre ein Erwerb des gesamten Unternehmens wohl bürokratisch zu aufwendig und unwirtschaftlich.

Neben dem einseitigen Datenerwerb in Form eines „Datenkaufs" ist denkbar, dass Unternehmen sich bilateral Zugang zu Trainingsdatensätzen verschaffen und einen Datenvorteil statt einen geldwerten Vorteil in An-

---

1028 *Borisyuk/Gordo/Sivakumar*, Rosetta: Large Scale System for Text Detection and Recognition in Images, Facebook AI Research; *Edunov et al.*, Understanding Back-Translation at Scale, Facebook AI Research; *Sivakumar/Gordo/Paluri,* Rosetta: Understanding Text in Images and Videos with Machine Learning, Facebook Code, 11. September 2018.

1029 An der Technischen Universität Darmstadt wurde Material aus dem Videospiel Grand Theft Auto genutzt, um den selbstlernenden Algorithmus zu trainieren. Siehe *S. Richter et al.,* Playing for Data: Ground Truth From Computer Games, 8. April 2016.

1030 *Europäische Kommission*, Mitteilung vom 10. Januar 2017, Aufbau einer europäischen Datenwirtschaft, COM(2017) 9 final, S. 11; *Schweitzer/Peitz,* Datenmärkte in der digitalisierten Wirtschaft, S. 22.

spruch nehmen.[1031] SAP, Microsoft und Adobe sind Gründungspartner der Open Data Initiative, die „Datensilos" aufbrechen und zur Wiederverwendung von Nutzerdaten innerhalb der Initiative beitragen soll.[1032] Dies ist auch als Datapooling für Startups denkbar. Jeder einzelne würde die Exklusivität auf die Partner des Datenpools erweitern und könnte den Vorteil von besser trainierten selbstlernenden Systemen genießen. Ein solches Verhalten steht grundsätzlich unter dem Verdacht der Kollusion[1033] und könnte als Verstoß gegen das Kartellverbot gewertet werden, weshalb die Datensets isoliert sein müssten und keine geschäftlichen Informationen preisgeben dürften. Die Zugangsgewährung ist auch derart denkbar, dass ein datenreiches Unternehmen ein Startup in Form einer strategischen Partnerschaft unterstützt und mit Trainingsdaten „investiert". Hierzu müsste das Startup aber bereits so weit sein, dass es das datenreiche Unternehmen mit einer originellen Idee und fundierten Strategie überzeugen kann. Data Sharing wird in der Regel als kostenlos, aber nicht offen und bedingungslos verstanden; der Zugangspetent könnte die Daten nicht weiter veräußern und entgegen den Vertragsbedingungen nutzen. Insbesondere werden solche Daten nicht unter Creative-Commons-Lizenz geteilt.

Offene Datensets, die keine individuelle Vereinbarung mit dem datenreichen Akteur erfordern, werden von öffentlichen (Public Open Data) und privaten Anbietern zur Verfügung gestellt. Besonders prominent ist die nicht-staatliche Bilderdatenbank ImageNet[1034], der ein Anteil an dem Fortschritt von KI der letzten Jahre zugeschrieben wird. ImageNet wurde 2009 veröffentlicht und seitdem auf über 14 Millionen gekennzeichnete Bilddateien ausgebaut. Die Datenbank wird bei dem Training visueller Objekterkennung mit Deep Learning eingesetzt. Ähnlich genutzt wird OpenImages[1035], das über neun Millionen gekennzeichnete Bilder umfasst,

---

1031 *BKartA*, Big Data und Wettbewerb, S. 9; *BMWi*, Wettbewerbsrecht 4.0, S. 59; *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 23; zu Data-Sharing-Möglichkeiten: *Europäische Kommssion*, Staff Working Document vom 10. Januar 2017, SWD(2017) 2 final, S. 16.

1032 Siehe https://www.microsoft.com/en-us/open-data-initiative.

1033 *BKartA*, Big Data und Wettbewerb, S. 9; allgemein zu Datenpools *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 93ff; *Lundqvist*, EuCML 2018, 146.

1034 Siehe http://www.image-net.org und http://image-net.org/about-publication für eine Liste der Publikationen, die ImageNet zitieren; dazu *J. Deng et al.*, What Does Classifying More Than 10,000 Image Categories Tell Us?, 2010.

1035 Siehe https://github.com/openimages/dataset.

und für spezielle Zwecke das Stanford Dogs Dataset[1036] mit 20.580 Bildern von 120 verschiedener Hunderassen. Eine Übersicht über verschiedene Datensets bietet Kaggle.[1037] Google bietet eine Übersicht über die von Alphabet-Tochterunternehmen bereitgestellten Datensets[1038] wie etwa Google Cloud Public Datasets und das Google Patents Public Dataset mit weiteren Ressourcen.

Die Datensets aus verschiedenen Quellen können kombiniert, bereinigt und mit eigenen Daten angereichert werden. Eine Datenstrategie wird sich in der Regel nicht auf einen einzelnen Kanal verlassen. Das Ziel eines selbstlernenden Systems sollte grundsätzlich sein, in der Realität Anwendung zu finden und aus dem Feedback der Nutzer weiter lernen zu können. Auch das Angebot der eigenen ML-Plattform an Dritte gegen Bezahlung kann Feedback an den Entwickler zurückspielen und die Produktevolution vorantreiben.

1. Personenbezogene Daten

Die Unterscheidung von personenbezogenen und nicht-personenbezogenen Daten ist die wichtigste Binnendifferenzierung und wegen der Strahlkraft des Datenschutzes auf datenbasierte Geschäftsmodelle von entscheidender Bedeutung. Der Datenschutz ist in dem Grundrecht auf informationelle Selbstbestimmung verwurzelt. Seit dem 25. Mai 2018 gilt in der Europäischen Union die Datenschutz-Grundverordnung.[1039] Die Eingabe von Daten in selbstlernende Systeme entspricht einer automatisierten Verarbeitung von Daten und unterfällt somit dem sachlichen Anwendungsbereich der Verordnung gemäß Art. 2 Abs. 1 (iVm Art. 4 Nr. 2 DSGVO), soweit es sich bei den Eingabedaten um solche mit Personenbezug handelt (Art. 4 Nr. 1). Die Erhebung von personenbezogenen Daten beschränkte sich wegen der begrenzten Erfassungsmöglichkeiten und des begrenzten Speicherplatzes vor wenigen Jahrzehnten noch auf vergleichsweise kleine

1036 Siehe http://vision.stanford.edu/aditya86/ImageNetDogs/.
1037 Siehe https://www.kaggle.com/datasets.
1038 Siehe https://research.google/tools/datasets/ mit aktuell 94 Ressourcen sowie allgemein die Datenset-Suche https://datasetsearch.research.google.com.
1039 Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG; im Folgenden DSGVO.

Volumina statischer Daten.[1040] Mit der Verbreitung des Internets, sozialer Medien und schließlich der ständigen Datenerfassung durch Smartphones und virtuelle persönliche Assistenten wachsen die pro natürlicher Person erfassten Datenvolumina exponentiell. Personenbezogene Daten werden in den meisten Fällen direkt bei den Individuen, auf die sich die Daten beziehen, erhoben.[1041] Der Handel mit diesen Daten auf Sekundärmärkten ist von geringerer Bedeutung.[1042] Das Datenschutzrecht erlegt ihm enge rechtliche Grenzen auf.[1043] Bei den hier vordergründig betrachteten maschinengenerierten Daten ist nicht auszuschließen, dass trotz des grundsätzlichen Maschinenbezugs eines Datums ein Personenbezug vorliegen kann oder indirekt durch Verknüpfung mit eigentlich nicht personenbezogenen Daten hergestellt werden kann.[1044] Ein Personenbezug liegt insbesondere dann nahe, wenn Aktivitäten der in der Produktion beschäftigten Arbeitnehmer erfasst werden.[1045] Im industriellen Kontext besteht die Gefahr, dass zu Mitarbeitern des Unternehmens Bezüge hergestellt werden oder ihre Aktivitäten anhand der Daten ausgewertet werden können.

Unternehmen, die personenbezogene Daten erheben, verarbeiten oder Dritten übermitteln, benötigen für jeden Akt der Datenverarbeitung eine Erlaubnis.[1046] Dies schließt solche Verarbeitungen ein, bei denen Unternehmen die Daten von einem Erfasser erhalten und selbst weiterverarbeiten.[1047] Die Erlaubnis kann sich aus einer Einwilligung oder einem gesetzlichen Erlaubnistatbestand ergeben. Gemäß Art. 5 Abs. 1 lit. b DSGVO müssen die Zwecke der Datenverarbeitung hinreichend bestimmt sein und bei einer Weitergabe der Personenkreis der potentiellen Empfänger sowie deren Verarbeitungszwecke bekannt sein, weshalb pauschal erteilte Einwilligungen unwirksam sind. Ein Datenhandel scheint unter diesen Vorgaben mit Einwilligungen der Betroffenen nur schwer möglich.

1040 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 12.
1041 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 4.
1042 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 4.
1043 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 88.
1044 Erwägungsgrund 30 der DSGVO; *Grünwald/Nüßing*, MMR 2015, 378 (382); *Schefzig*, K&R 2014, 772 (773f); so auch EuGH, Urteil vom 19. Oktober 2016, Rs. C-582/14 Rn. 49 – *Breyer/Deutschland*.
1045 Art. 88 DSGVO iVm § 26 BDSG berührt nicht die Verarbeitung von Beschäftigtendaten außerhalb der Sphäre des Arbeitsverhältnisses; *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 34.
1046 *Fries/Scheufen*, MMR 2019, 721 (723) mwN; Paal/Pauly/*Frenzel*, DS-GVO, Art. 6 DS-GVO Rn. 11.
1047 *Schweitzer/Peitz*, NJW 2018, 275 (276).

Gemäß Art. 6 Abs. 1 lit. c DSGVO ist die Verarbeitung rechtmäßig, wenn sie zur Erfüllung einer rechtlichen Verpflichtung, der der Verantwortliche unterliegt, erforderlich ist. Absatz 3 schränkt die rechtliche Verpflichtung ein: Sie muss ein im öffentlichen Interesse liegendes Ziel verfolgen und in einem angemessenen Verhältnis zu dem verfolgten legitimen Zweck stehen (Abs. 3 aE). Nicht zuletzt eröffnet die Generalklausel des Art. 6 Abs. 1 lit. f DSGVO einen erheblichen Anwendungsspielraum, indem die erforderliche Weiterverarbeitung zur Wahrung berechtigter Interessen des Verantwortlichen oder eines Dritten als Quasi-Auffangtatbestand[1048] gestattet wird. Wegen der Abwägung der berechtigten Interessen fehlt es an der Rechtssicherheit, die für den routinierten Handel auf Datensekundärmärkten erforderlich wäre.[1049]

Das Kriterium der Identifizierbarkeit erfasst potentiell einen weiten Kreis an Daten und die latente Bedrohung der DSGVO-Anwendbarkeit mit immer raffinierteren Analyseinstrumenten wird zunehmen. Die Weitergabe von personenbezogen Daten unterliegt strengen Schranken und der Datenschutz steht grundsätzlich einer Weitergabe von personenbezogenen Daten zur Förderung der selbstlernenden Systeme Dritter entgegen. Eine potentielle Datenteilungspflicht wird von dem Schutz personenbezogener Daten beschränkt sein müssen, sofern nicht im konkreten Fall überragend wichtige Gemeinschaftsgüter betroffen sind. Zudem steht der Gedanke des Aufbewahrens von Daten, um spätere Datenzugangsansprüche erfüllen zu können, in einem Konflikt mit Grundsätzen der DSGVO wie etwa der Datenminimierung (c) und Speicherbegrenzung (e) gemäß Art. 5 Abs. 1.

Die Verbreitung von Trainingsdatensets mit Personenbezug hängt erheblich von der Bereitschaft von Individuen zur Erteilung ihrer Zustimmung ab. Das Beispiel von TwentyBN legt nahe, dass diese Bereitschaft gegen eine monetäre Gegenleistung zu erreichen ist; insbesondere wenn der höchstpersönliche Bereich und die eigenen Gewohnheiten nicht betroffen sind.

2. Nicht personenbezogene Daten

Nicht personenbezogene Daten sind solche, die sich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen lassen. So wird

1048 Vgl. *Paal/Hennemann*, NJW 2017, 1697 (1700).
1049 Vgl. *Schweitzer/Peitz*, NJW 2018, 275 (276).

die Messung von Luftdruck, Temperatur und Betriebsparametern regelmäßig von der DSGVO nicht erfasst. Der Personenbezug kann nachträglich durch die Verknüpfung nicht-personenbezogener mit personenbezogenen Daten hergestellt werden. Das Datenschutzrecht differenziert nicht zwischen gezieltem und ungezieltem Personenbezug.

Statt des Datenschutzes wird der Wunsch nach Wahrung der Geschäftsgeheimnisse die Weitergabe von nicht-personenbezogenen Daten bremsen. Das Teilen der Daten auf bilateralem Wege beruht auf Vertrauen – nicht selten sind die Daten für eine oder beide Seiten Betriebsgeheimnisse. Sie können Rückschlüsse auf die Performance, den Verbrauch oder die Zuverlässigkeit der Maschine zulassen. In der Regel werden diese Daten außerdem als Nebenprodukt der fertigenden Industrieproduktion anfallen, weshalb auch für aus ihnen erstellte Datensets die gemäß § 87a Abs. 1 Nr. 1 UrhG zur Annahme eines Datenbankschutzes erforderlichen „wesentlichen Investitionen" abzulehnen wären.[1050] Mangels eines immaterialgüterrechtlichen Schutzes ist zu erwarten, dass der Schutz von Betriebs- und Geschäftsgeheimnissen als berechtigtes Interesse[1051] einen hohen Stellenwert einnimmt. Zu beachten ist, dass nur solche Informationen, die kein anderer ebenso einfach erfassen könnte, ein Geheimnis darstellen.[1052] Die Außentemperatur kann jeder erfassen, die Betriebstemperatur einer Maschine auf dem Betriebsgelände ist hingegen in der Regel geheim. Außerdem wird ein einziges Datum wohl mangels eigenen kommerziellen Wertes kein Geschäftsgeheimnis sein. Datensets und die Kombination verschiedener Daten und Informationen erreichen das Niveau eines Geschäftsgeheimnisses.[1053] Der Schutz von Betriebs- und Geschäftsgeheimnissen wirkt nur mittelbar auf den Datenverkehr und bezweckt seinen Schutz nur so, wie er jeden anderen Wirtschaftsbereich vor Eingriffen in die unternehmerische Sphäre schützt.[1054]

Nicht-personenbezogene Daten befinden sich damit in einem Spannungsfeld verschiedener Rechtsgebiete, die diese Daten in ihren Schutzbe-

---

1050 *Drexl et al.*, Data Ownership and Access to Data, MPI for Innovation and Competition Research Paper No. 16–10, Rn. 11; *Ensthaler*, NJW 2016, 3473 (3474).

1051 *Europäische Kommission*, Mitteilung vom 10. Januar 2017, COM(2017) 9 final, S. 14.

1052 Siehe *Drexl et al.*, Data Ownership and Access to Data, MPI for Innovation and Competition Research Paper No. 16–10, Rn. 23.

1053 *Europäische Kommssion*, Staff Working Document vom 10. Januar 2017, SWD(2017) 2 final, S. 20; *Drexl et al.*, GRUR Int. 2016, 914 (916f).

1054 *Drexl et al.*, GRUR Int. 2016, 914 (916).

reich einschließen, aber nicht ihren Schutz oder ihre Verkehrsfähigkeit bezwecken. Infolge der oft unsicheren Abgrenzbarkeit von personenbezogenen Daten ist ein Datenhandel stets von Rechtsunsicherheit belastet. Dies würde auch auf ein mögliches Datenzugangsrecht abfärben und das Vertrauen in verpflichtete Unternehmen senken.

Es ist zu erwarten, dass weiter neue Quellen für die Erfassung von nicht-personenbezogenen Daten erschlossen und genutzt werden.[1055]

3. Anonymisierte Daten

Um selbstlernende Systeme zu trainieren, ist nicht immer erforderlich, dass Bilder, Texte, Aufnahmen oder Informationen bestimmten natürlichen Personen zugeordnet werden können. Das System kann sich bei der Untersuchung von Korrelationen in den meisten Fällen an abstrakten Personen orientieren und individuell erfasste Daten anonym verwenden.[1056] Bestimmte Informationen können aber, sogar, wenn sie von Kontaktinformationen und offensichtlichen Identifikationsmerkmalen bereinigt sind, natürlichen Personen eindeutig zugeordnet werden. Es stellt sich die Frage, inwieweit eine Anonymisierung von Nutzerdaten erfolgen kann, um potentiell die anonymisierten Daten einem weiteren Kreis von Innovatoren zugänglich zu machen.

Die Anonymisierung wird von der DSGVO indirekt in Erwägungsgrund 26 definiert („personenbezogene Daten, die in einer Weise anonymisiert worden sind, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann"). Art. 5 Nr. 4 DSGVO definiert Pseudonymisierung als die Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können. Die Pseudonymisierung entfernt die Personenbezüge nicht, sondern ersetzt sie durch ein Kennzeichen (Pseudonym). Der Datenverarbeiter bewahrt in der Regel ein weiteres Datenset auf, das ihm die Zuordnung der Kennzeichen zu Personen erlaubt. Die Bezüge zu verbundenen Datensätzen, in die das gleiche Kennzeichen eingesetzt wird, bleiben erhalten. Für diesen Datenverarbeiter bleibt der Personenbezug bestehen; die Pseudonymisierung kann allerdings die Datensicherheit erhöhen, vgl.

1055 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 8.

1056 *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 25, 28 und Fn. 24, 25.

Art. 32 Abs. 1 lit. a DSGVO und Erwägungsgrund 28. Die Re-Identifikation von anonymisierten Daten ist unmöglich, während die Re-Identifikation pseudonymisierter Daten möglich bleibt.

Gemäß Erwägungsgrund 26 der DSGVO sollen die Vorgaben der DSGVO auf anonymisierte Daten keine Anwendung finden. Darüber, wann eine Anonymisierung erfolgreich ist und ob ein Erfolg überhaupt technisch möglich ist, wird gestritten. Die Anonymisierung definiert sich über die Unmöglichkeit der Deanonymisierung, hängt damit also sowohl vom Stand der Technik als auch dem Bezugspunkt der „Unmöglichkeit" ab. Es wird insbesondere unterschieden, ob es bei der Wiederherstellung des Personenbezugs auf das Wissen und die Mittel irgendeiner Stelle ankommt (absolute oder objektive Theorie) oder ob auf die Möglichkeiten der verantwortlichen Stelle (relative Theorie) abzustellen ist.[1057] Der EuGH entschied im Jahr 2016 zu der Perspektive auf die Deanonymisierung im Hinblick auf die Zuordnung von IP-Adressen im Vorabentscheidungsverfahren *Beyer/Deutschland*.[1058] Für den Webseitenbetreiber (in diesem Fall die Bundesrepublik Deutschland) allein sei der Nutzer der betreffenden Webseite nicht bestimmbar, obwohl die dynamische IP-Adresse gespeichert würde. Nur ein Dritter, nämlich der Internetzugangsanbieter, sei in der Lage, diese IP-Adresse dem Nutzer zuzuordnen.[1059] Nach der objektiven Theorie wäre damit die IP-Adresse personenbezogen. Die relative Theorie würde für den Internetzugangsanbieter zu einer Identifizierbarkeit kommen, aber diese für den Webseitenbetreiber, der nur die IP-Adresse erfasst, ablehnen. Der EuGH beantwortete die Vorlagefrage so, dass von einer Identifizierbarkeit (damals Bestimmbarkeit) auszugehen sei, wenn der Webseitenbetreiber über rechtliche Mittel verfügt, die es ihm erlauben, die betreffende Person anhand der Zusatzinformationen, über die der Internetzugangsanbieter dieser Person verfügt, bestimmen zu lassen.[1060] Die Entscheidung kann so verstanden werden, dass nur bei Verfügbarkeit der rechtlichen Mittel zur Erlangung der Zusatzinformationen ein personenbezogenes Datum anzunehmen ist; nur dann sei die Identifikation hinreichend wahrscheinlich und der Aufwand nicht unverhältnismäßig hoch. Der EuGH schließt sich mit diesem Urteil eher der relativen Ansicht an. Auch die DSGVO folgt mit Erwägungsgrund 26 dieser Linie: Es sollten zur

1057 Für eine Darstellung des Streitstandes: *Herbst*, NVwZ 2016, 902 (904f); *Jäschke et al.*, Für immer anonym, ABIDA, S. 24; *Kühling/Klar*, NJW 2013, 3611 (3613).
1058 EuGH, Urteil vom 19. Oktober 2016, C-582/14 – *Breyer/Deutschland*.
1059 EuGH, Urteil vom 19. Oktober 2016, C-582/14 Rn. 25 – *Breyer/Deutschland*.
1060 EuGH, Urteil vom 19. Oktober 2016, C-582/14 Rn. 49 – *Breyer/Deutschland*.

Beurteilung der Re-Identifikation „alle Mittel berücksichtigt werden, die von dem Verantwortlichen oder einer anderen Person nach allgemeinem Ermessen wahrscheinlich genutzt werden, um die natürliche Person direkt oder indirekt zu identifizieren, wie beispielsweise das Aussondern."[1061] Auch die Leitlinien zum freien Datenverkehr stimmen dem zu.[1062] Das Abstellen auf die nach allgemeinem Ermessen wahrscheinliche Nutzung zieht zwar alle objektiven Faktoren wie die Kosten, den Zeitaufwand und die technologischen Möglichkeiten heran, lässt aber unverhältnismäßige Maßnahmen außen vor. Maßnahmen der Deanonymisierung, die mit illegalen Mitteln erfolgen, bleiben außer Betracht.[1063] Zu beachten ist zudem, dass die technische Entwicklung voranschreitet und größere Datenmassen eine Deanonymisierung erleichtern. Der aufzubringende Aufwand wird damit geringer und eine Deanonymisierung grundsätzlich wahrscheinlicher. Anonymisierte Daten bleiben dann „personenbeziehbar".[1064] Die Artikel-29-Datenschutzgruppe versteht dieses Kriterium in ihrer Stellungnahme zu den Anonymisierungstechniken[1065] als die Robustheit des Anonymisierungsverfahrens. Es wurde erfolgreich anonymisiert, wenn die Identifizierung vernünftigerweise unmöglich geworden ist.[1066] Wegen der stetigen Weiterentwicklung der Identifikationstechniken müssten Rechtsvorschriften technisch neutral formuliert werden.[1067]

Die Anonymisierungsvorgaben könnten in Ansehung der zur Verfügung stehenden technischen Möglichkeiten der Deanonymisierung nur schwer mit Rechtssicherheit zu realisieren sein.[1068] Bei aggregierten Daten ist das Risiko einer Deanonymisierung höher, je geringer die Menge an identifizierbaren Personen ist.[1069] Dieses Risiko ist mangels einer Prognose über alle aktuell und in Zukunft verfügbaren Daten und Analysefähig-

---

1061 So auch *Jäschke et al.*, Für immer anonym, ABIDA, S. 25.
1062 *Europäische Kommission*, Mitteilung vom 29. Mai 2019, Leitlinien zur Verordnung über einen Rahmen für den freien Verkehr nicht-personenbezogener Daten in der Europäischen Union, COM(2019) 250 final, S. 4ff.
1063 *Kühling/Klar,* NJW 2013, 3611 (3613).
1064 *Drexl*, NZKart 2017, 415 (416); Art. 1 Abs. 1 DSGVO.
1065 *Artikel-29-Datenschutzgruppe*, Stellungnahme 5/2014 zu Anonymisierungstechniken, 0829/14/DE, 10. April 2014.
1066 *Artikel-29-Datenschutzgruppe*, Stellungnahme 5/2014, S. 9.
1067 *Artikel-29-Datenschutzgruppe*, Stellungnahme 5/2014, S. 10; dazu auch *Janeček,* Computer Law and Security Review, Vol. 34, No. 5, S. 1039–1052, 1040 (2018).
1068 *Kühling/Klar,* NJW 2013, 3611 (3613); *Specht*, GRUR Int. 2017, 1040 (1046).
1069 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 32.

keiten wohl hinzunehmen.[1070] Die Anonymisierung von Datenbeständen der produzierenden Industrie, die Beschäftigtendaten beinhalten könnten, sollte regelmäßig nach relativer Ansicht erfolgreich sein: Außerhalb des Unternehmens sollte es keinem Datenverarbeiter gelingen, den Bezug zu einem einzelnen Mitarbeiter wiederherzustellen.[1071]

Die DSGVO lässt zwar ein Restrisiko einer Deanonymisierung zu, aber die verantwortlichen Stellen tragen weiterhin das Risiko und haften bei unerlaubter Weitergabe von deanonymisierbaren Daten. Die Rechtsunsicherheit ist ein negativer Anreiz für das Teilen von Trainingsdaten zur Wiederverwendung. Ihr könnte mit dem Verbot einer Deanonymisierung oder einer unwiderlegbaren gesetzlichen Vermutung für Anonymität entgegengewirkt werden.[1072] Noch fehlen aufsichtsbehördliche Vorgaben.

4. Auswirkungen des Personenbezugs auf Datenhandel und mögliche Datenzugangsrechte

Die das gesamte Datenverkehrsrecht prägende Unterscheidung zwischen personenbezogenen und nicht-personenbezogenen Daten wirkt sich auf den Datenhandel und mögliche gesetzliche Datenzugangsrechte aus. Das Datenschutzrecht kommt zwar als Zuweisungsinstrument in Betracht, die Zuweisung von Daten nach dem Datenschutz ist jedoch nur ideeller Art.[1073]

Grundsätzlich bereitet der Handel mit Daten, die nie einen Personenbezug aufwiesen, die wenigsten Probleme. Anonymisierte Daten können wie nicht-personenbezogene Daten genutzt und weitergegeben werden. Es fehlt allerdings die Rechtssicherheit, ob die Anonymisierung erfolgreich war. Für personenbezogene Daten sieht das Datenschutzrecht mit Art. 20 DSGVO einen Mechanismus vor, der auf Initiative des betroffenen Datensubjekts die Portierung von personenbezogenen Daten zu einem anderen Anbieter vorsieht. Dies kann als ein erster Schritt zur Ökonomisierung

1070 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 28ff.
1071 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 34.
1072 *Specht*, GRUR Int. 2017, 1040 (1047); mit dem Wunsch nach Konturierung von Anonymisierung und Pseudonymisierung *Bundesverband der Deutschen Industrie/Noerr*, Industrie 4.0 – Rechtliche Herausforderungen der Digitalisierung, S. 12; *BDI/Institut der deutschen Wirtschaft*, Datenwirtschaft in Deutschland, S. 41.
1073 *Härting*, CR 2016, 646 (648).

von personenbezogenen Daten verstanden werden.[1074] Fraglich ist, ob Drittanbieter auch ohne Zutun der Datensubjekte Zugang zu erwünschten Datensets erlangen können. Weil Situationen, in denen es zu einer Datenteilungspflicht kommen kann, für den Verantwortlichen schlecht vorauszusehen sind und er die Empfänger der Daten nicht angeben kann, würden nur Blanko- oder Generaleinwilligungen infrage kommen. Dies widerspricht dem Zweck der Einwilligung nach der DSGVO jedoch. Die DSGVO legt die Einwilligung bewusst in die Hände der betroffenen natürlichen Person, die eine informierte Entscheidung in Kenntnis aller Umstände trifft. Die Einholung einer Einwilligung unmittelbar vor der Weitergabe der Daten würde wohl aus bürokratischen Gründen ausscheiden und auf wenig Verständnis bei den Betroffenen stoßen.

Eine verpflichtende Weitergabe personenbezogener Daten infolge einer Datenteilungspflicht wurde bisher nicht erwogen; die Vorschläge scheinen sich lediglich auf nicht-personenbezogene Daten und anonymisierte Daten zu beziehen. Würde ein solches Gesetzesvorhaben personenbezogene Daten einschließen, müsste es sich neben dem Grundgesetz auch an den Vorgaben der DSGVO messen.[1075] Grundsätzlich könnten sich an dieser Stelle Wertungswidersprüche zu den allgemeinen Zielen des Datenschutzes ergeben: Die vom Kunden an Unternehmen A erteilte Einwilligung impliziert nicht, dass er auch mit der Nutzung seiner Daten durch dessen Wettbewerber B einverstanden gewesen wäre.[1076] Weiterhin widerspricht die Speicherung von identischen Datensätzen an mehreren Orten dem Grundsatz der Datenminimierung und Datensparsamkeit. Insofern ist zu erwarten, dass die politische Debatte das Teilen von personenbezogenen Daten weiter außen vor lässt. Es ist davon auszugehen, dass anonymisierte Daten mit der Auflösung des Personenbezugs entscheidend an Wert verlieren und die Nachfrage nach ihnen geringer ist.[1077] Das Bundeskartellamt gibt an, dass Daten von Dritten für Unternehmen ohnehin einen geringeren Wert hätten als Daten, die sie selbst von eigenen Nutzern erhoben haben.[1078]

1074 *Härting*, CR 2016, 646 (648).

1075 Vgl. Paal/Pauly/*Frenzel*, DS-GVO Art. 6 Rn. 16–19; ausführlicher: Kapitel 5 D.II.1. Vereinbarkeit von Datenzugangsrechten mit dem Datenschutzrecht, S. 379.

1076 *Haucap*, Big Data aus wettbewerbs- und ordnungspolitischer Perspektive, in: Morik/Krämer (Hrsg.), Daten, S. 95–142 (99); *Veil*, NVwZ 2018, 686 (695).

1077 *S. Schmidt*, WuW 2018, 549.

1078 *BKartA*, Big Data und Wettbewerb, S. 7.

II. Datenmärkte: Input, Output, Currency

Wie bereits ausgeführt wurde[1079], sind Daten faktisch verkehrsfähig und werden gehandelt, getauscht oder geteilt. Die unterschiedlichen Dimensionen der Datennutzung werden als „Input[1080]“ (Einsatz), „Output (Ausgabe/Produkt)[1081] und „Currency“[1082] (Währung) bezeichnet. Die wettbewerbsbezogene Analyse der Auswirkungen von Datenmacht muss zwischen den unterschiedlichen Nutzungsdimensionen differenzieren.

Daten werden als Eingabedaten genutzt, um die Funktionalität von Datenanalyse-Tools oder anderen Diensten zu verbessern. Ebenfalls sind sie ein Input, wenn sie als Trainingsdaten in selbstlernende Systeme eingegeben werden. Diese Dimension von Daten gewinnt zunehmend an Wichtigkeit, weil datengetriebene Geschäftsmodelle und das IoT auf gut trainierte selbstlernende Systeme zurückgreifen. Die Bewertung von Daten als Essential Facility nach der EFD würde sich nach der Input-Dimension von Daten richten. Dies gilt auch für die Bewertung von Datenzugang als Innovationsvoraussetzung.[1083] Ausgehend von der Input-Dimension der Daten können Marktteilnehmer als Zugangspetenten identifiziert werden. Ebenfalls ermöglicht dies die Identifikation von alternativen, semantisch einander entsprechenden Datensets oder die Feststellung der Einzigartigkeit eines spezifischen Datensets.

Daten sind gleichzeitig das Ausgangsmaterial und das Ergebnis von Datenanalysevorgängen. Idealerweise offenbaren sie als Output neue Informationen oder solche, die den Rohdaten allein nicht zu entnehmen waren. Die Produktdimension von Daten kommt ebenfalls zum Tragen, wenn sie als Ware gehandelt werden und zur Vertragserfüllung der Zugang zum Datenset ermöglicht wird.[1084] Datenhandel kann verschiedenste

---

1079 Siehe Kapitel 4 B.II.2., S. 232.

1080 Z. B. *BKartA*, Big Data und Wettbewerb, S. 4; *BMWi*, Wettbewerbsrecht 4.0, S. 13.

1081 Z. B. *Balto/Lane*, Monopolizing Water in a Tsunami, S. 1.

1082 Z. B. *Burnside*, No Such Thing as a Free Search, CPI Antitrust Chronicle, May 2015 (2), S. 7; *Grossman*, Data Is Currency, Don't Abuse It, Forbes, 9. Juli 2018; *Kuneva*, Roundtable on Online Data Collection, Targeting and Profiling, Keynote-Rede in Brüssel am 31. März 2009: "Personal data is […] the new currency of the digital world."; *Vestager*, Competition in A Big Data World, Rede in München am 17. Januar 2016: "consumers have a new currency […] – our data".

1083 Z. B. Europäische Kommission, Entscheidung vom 14. Mai 2008, COMP/M.4854 Rn. 14, 120, 164f – *TomTom/TeleAtlas*.

1084 *Sivinski/Okuliar/Kjolbye*, ECJ, Vol. 13, S. 199–227, 208 (2017).

Formen annehmen und innerhalb verschiedenster Geschäftsmodelle stattfinden.[1085] Koutroumpis, Leiponen und Thomas unterschieden zwischen vier Varianten des Datenhandels:[1086] Eins-zu-eins-Handel („one-to-one", also bilateral), Einer-an-viele-Handel („one-to-many"), Viele-an-einen-Handel („many-to-one", nur ein einziger Erwerber, „harvesting of data") und Viele-an-viele-Handel („many-to-many", Daten-Marktplätze). Daten-Marktplätze mit jeweils einer Vielzahl von Anbietern und Nachfragern sind dabei mit Transaktionsproblemen rechtlicher und vertraglicher Art konfrontiert.[1087]

In der Vergangenheit wurden Daten in ihrer Dimension als wirtschaftliches Gut von Wettbewerbsbehörden grundsätzlich wie jedes andere wirtschaftliche Gut beurteilt.[1088] Aus Sicht des anbietenden Datenhändlers ist es sein Produkt, während der Erwerber üblicherweise die Daten, wenn er sie nicht selbst weiterveräußert, als Input nutzen wird. Der Wert der Daten erschließt sich oft erst nach der Analyse oder der Eingabe in ein selbstlernendes System. Bei der wettbewerbsrechtlichen Beurteilung von Daten darf daher nur die jeweils relevante Dimension der Daten betrachtet werden, ohne aber von anderen Dimensionen isoliert werden zu müssen. Geht man von einem Datenmarkt aus, würde es schwerfallen, von einem separaten Markt für Trainingsdaten auszugehen.[1089] Potentiell kann anhand aller Daten gelernt werden; das jeweilige Bedürfnis ist aus der Anwendungsabsicht und dem Informationsbedürfnis abzuleiten. Für Informationen im weiteren Sinne bestehen entwickelte Märkte und etablierte Regulierung, beispielsweise für Bücher, Presseerzeugnisse, Musik, Filme sowie Arbeitsmärkte und Beraterdienste. Obwohl es sich semantisch auch um Informationen handelt, sind Märkte für Daten als unorganisierte, maschinenlesbare Informationsmassen weniger entwickelt. Dies könnte daran liegen, dass Rohdaten besonders kontextabhängig sind, schnell veralten und erst verarbeitet zu Informationen ihren Nutzen offenbaren.

1085 *Kerber*, Rights on Data, S. 10.

1086 Zum Folgenden: *Koutroumpis/Leiponen/Thomas*, The (Unfulfilled) Potential of Data Marketplaces, S. 5.

1087 *Koutroumpis/Leiponen/Thomas*, The (Unfulfilled) Potential of Data Marketplaces, S. 19.

1088 So *Sivinski/Okuliar/Kjolbye*, ECJ, Vol. 13, S. 199–227, 209 (2017), z. B. Europäische Kommission, Entscheidung vom 19. Februar 2008, COMP/M.4726 Rn. 62, 100f, 361 – *Thomson/Reuters*.

1089 *Sivinski/Okuliar/Kjolbye*, ECJ, Vol. 13, S. 199–227, 203 (2017).

Von Zeit zu Zeit werden Daten als die Währung (Currency) der „Online-Märkte" bezeichnet.[1090] Dies ist stark vereinfachend und soll darauf hindeuten, dass die Preisgabe von Daten den Zutritt zu und die Nutzung bestimmter Dienste ermöglicht. Tatsächlich geht der Vergleich fehl. Daten sind unerschöpflich und haben keinen beständigen Wert.[1091] Zudem sind sie heterogen im Wert, der je nach Art der Daten und nach Art der intendierten Verwendung schwankt. Anders, als durch Währungspolitik möglich ist, kann die Knappheit von Daten nicht kontrolliert oder gewährleistet werden. Der Vergleich mit Währungen soll vielmehr verdeutlichen, dass die Preisgabe von Daten für Dienste besonders dort beobachtet wird, wo für ihre Inanspruchnahme keine finanzielle Gegenleistung erbracht wird.

1. Wettbewerbliche Charakteristika von Daten

Die Vergleiche von Daten mit „Öl"[1092] und ihre Bezeichnung als Rohstoffe des 21. Jahrhunderts legen nahe, dass sich datengetriebene Branchen kaum von traditionell arbeitenden Branchen unterscheiden. Während diesem Vergleich von der Datenwirtschaft hartnäckig widersprochen wird („nicht Öl, sondern Sonnenlicht"), wird ebenso hartnäckig ein besonderes Regulierungsbedürfnis abgelehnt. Aus wettbewerbspolitischer Sicht variiert die Relevanz von Datensets drastisch.[1093]

Hervorgehoben wird regelmäßig die wettbewerbliche Bedeutung der Non-Rivalität von Daten. Ein Gut ist non-rival, wenn es mehrfach gleichzeitig genutzt werden kann, ohne dass der Nutzen abnimmt. Die Nutzung einer Datensammlung durch A hindert B in keiner Weise an der Nutzung der gleichen Daten und „verbraucht" sie nicht.[1094] Materielle Güter sind

---

1090 Zitat Margrethe Vestager nach *Kanter*, Antitrust Nominee in Europe Promises Scrutiny of Big Tech Companies, 3. Oktober 2014; dazu *Monopolkommission*, Sondergutachten 68, S. 44, Rn. 65; *Stucke*, Georgetown Law Technology Review, S. 275–324, 284 (2018).

1091 *Balto/Lane*, Monopolizing Water in a Tsunami, S. 2; *Drexl*, Data Access and Control (BEUC), S. 64.

1092 Z. B. *The Economist*, The world's most valuable resource is no longer oil, but data, 6. Mai 2017; *Gallagher*, Data Really Is the New Oil, The Wall Street Journal, 9. März 2019.

1093 *Sivinski/Okuliar/Kjolbye*, ECJ, Vol. 13, S. 199–227, 201 (2017).

1094 *M. Gal/Rubinfeld*, Access Barriers to Big Data, S. 37; *Monopolkommission*, Sondergutachten 68, S. 44, Rn. 65; auch *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 25; *Paal/Hennemann*, NJW 2017, 1697 (1698).

grundsätzlich rival und können deshalb nur von einer Person an einem Ort genutzt werden.[1095] Dies trifft auf Speichermedien wie CDs oder USB-Sticks zu – allerdings nicht auf den Zugriff auf eine Cloud. Hieran knüpft die Europäische Kommission in ihrer Mitteilung zum Aufbau einer europäischen Datenwirtschaft an: Wenn Daten nicht rival sind, sollten sie so oft wie möglich genutzt werden.[1096] Sobald sie veröffentlicht wurden, kann niemand an ihrer Nutzung gehindert werden und sie können von einer unbegrenzten Zahl von Akteuren gleichzeitig in neue Produkte oder zu neuen Erkenntnissen verarbeitet werden, ohne an Wert zu verlieren.

Zudem sind Daten ohne großen Aufwand und ohne Qualitätsverlust duplizierbar.[1097] Die Kosten für Transport und Übertragung sind zu vernachlässigen, weshalb sich, anders als bei analog gespeicherten Informationen, hieraus kaum Barrieren zur Nutzung und Weiterverwertung ergeben. Sind Daten so gespeichert, dass der Zugang über ein Zugriffsmanagement gestattet werden kann (z. B. in der Cloud), ist eine Duplizierung oder Übertragung nicht nötig, sondern nur das Einrichten des Zugangs.

Daten sind nur relativ nützlich, also nicht für beliebige Verwendungszwecke von gleichem Wert. Es ist zwischen verschiedenen Datenformaten, semantischen Inhalten und Verwendungszielen zu unterscheiden. Zudem kann die Aktualität ein wertbildender Faktor sein, sodass bestimmte Datensets schnell an Nutzen einbüßen. Die Mehrzahl der von Sensoren und IoT-Geräten erfassten Daten ist von begrenzter zeitlicher Relevanz. Ebenso ist die Aktualität oder Frische der Rohstoffe auf traditionellen, wie auch auf datengetriebenen Märkten oft notwendig für ein nutzbares digitales oder physisches Produkt, beispielsweise Lebensmittel.

Daten sind vor ihrer Veröffentlichung ein ausschließbares Gut und weisen damit die Eigenschaften eines Clubgutes auf.[1098] Veröffentlichte Daten im Sinne syntaktischer Informationen wären entsprechend dieser Definition nicht ausschließbar und damit ein öffentliches Gut.[1099]

Daten weisen keine natürliche Ressourcenknappheit auf. Historische Daten sind aber in der Regel eine begrenzte Ressource, weil sie nicht

---

1095 *Duch-Brown/Martens/Müller-Langer*, The economics of ownership, access and trade in digital data, S. 12.

1096 *Europäische Kommission*, Staff Working Document vom 10. Januar 2017, SWD(2017) 2 final, S. 36; *OECD*, Data-Driven Innovation, S. 187.

1097 Vgl. *Paal/Hennemann*, NJW 2017, 1697 (1698).

1098 *Dewenter/Lüth*, Big Data aus wettbewerblicher Sicht, Wirtschaftsdienst 2016, S. 648–654 (649).

1099 Vgl. *Härting*, CR 2016, 646 (647) mwN.

erneut authentisch erfasst werden können. Digitale Daten verursachen so gut wie keine Grenzkosten in Produktion und Vertrieb.[1100]

Der Wert von Daten steigt durch die Kombination mit anderen Datensets. Informationen sind immer auf Kommunikation und Austausch gerichtet. Dabei ist aber davon auszugehen, dass Daten grundsätzlich einen abnehmenden Grenznutzen haben.[1101] Je komplexer die Anwendung ist, für die sie genutzt werden, desto mehr Daten werden benötigt, bis ihr Nutzen abnimmt. Der Wert von Daten kann kaum eindeutig bestimmt werden. Wertbildende Faktoren sind die Art, das Volumen, Qualität und Aktualität der Daten, ihre Verfügbarkeit, die Möglichkeit zur Auswertung und Monetarisierung der Daten sowie die Möglichkeit, durch Ausschluss von diesen Daten die Wettbewerbsmöglichkeiten anderer Unternehmen einzuschränken.[1102] Nach Kerber ist es die Kombination der wettbewerblichen Eigenarten von Daten, die sie zu einem sehr besonderen Wirtschaftsgut[1103] machen, das die traditionellen rechtlichen und wirtschaftlichen Kategorien herausfordert. Die Balance von Exklusivität und Zugang ist für die Debatte um Datenregulierung zentral und wird von ihren wettbewerblichen Eigenarten bestimmt.

2. Vermögenszuordnung und Rechte an Daten

Die Frage nach einem Recht an oder einer rechtlichen Zuordnung von Daten ergibt sich aus der Feststellung, dass Daten einen wirtschaftlichen Wert[1104] haben, der monopolisiert werden könnte. Das Datenverkehrsrecht ist kein homogenes Regelwerk, sondern adressiert die unterschiedlichen Interessen von Dateninhabern in Zivilrecht, Strafrecht und öffentlichem Recht. Zu diesen Interessen zählen der Integritätsschutz, der Vertraulichkeitsschutz, die klare wirtschaftliche Zuordnung und Nutzungsrechte. Grundsätzlich ist es das Recht, das die verschiedenen Nutzungs- und Zugriffsinteressen in Bezug auf wirtschaftlich nutzbare Güter in Einklang bringt. Im Zuge der Strategie der Europäischen Kommission für

1100 *Shapiro/Varian*, Information Rules, S. 24.
1101 Vgl. *Paal/Hennemann*, NJW 2017, 1697 (1698); auch *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 103 und Fn. 175.
1102 *BMWi – Plattform Industrie 4.0 (Hrsg.)*, Marktmacht durch Datenhoheit, in: dass., Industrie 4.0 – Kartellrechtliche Betrachtungen, April 2018, S. 15–22 (18); sowie *Körber*, NZKart 2016, 303 (305f); *Nuys*, WuW 2016, 512 (516).
1103 „Very special economic good", *Kerber*, Rights on Data, S. 17.
1104 Siehe *Zech*, CR 2015, 137 (137).

einen europäischen Digitalen Binnenmarkt[1105] wurde daher das Konzept eines Dateneigentums umfassend diskutiert.[1106]

a) De lege lata

Mittlerweile ist unumstritten, dass an Daten *per se* keine eigentumsähnlichen Rechte bestehen.[1107] Durch den Datenbankschutz gemäß §§ 87a ff UrhG oder den Schutz von Betriebs- und Geschäftsgeheimnissen können sie jedoch einem speziellen Schutzregime unterfallen. Hierzu treten unter anderem das Telemediengesetz, das Urheberrecht und das Intelligente-Verkehrssysteme-Gesetz (IVSG).[1108] Die Fragmentierung der Regelungswerke wird als problematisch empfunden, bezieht sich aber nicht auf die rechtliche Zuordnung von Daten und führt daher nicht zu ambivalenten Lösungen bezüglich der Datenkontrolle. Unproblematisch können Daten Gegenstand von Verträgen sein (vgl. § 453 BGB). Die darin festgelegten Rechte gelten jedoch nur inter partes. Darüber hinaus – erga omnes – ergeben sich aus ihnen keine Zugangsrechte. Bei reinen Maschinendaten ist der Zugang meist bilateral geregelt.

Die Datenträger als körperliche Gegenstände unterfallen dem Eigentumsbegriff des Zivilrechts, § 903 BGB, und des Strafrechts. Aus dem Eigentum an der Sache folgend wird der Eigentümer des Speichermediums die dem zivilrechtlichen Eigentum zugewiesenen Herrschaftsrechte an den darauf gespeicherten Daten ausüben können. Er kann anderen untersagen, sein Buch zu lesen oder seine SD-Karte auszulesen, aber ihnen nicht verbieten, die einmal hieraus erlangten Informationen weiterzugeben. Die erfassten Daten sind weder Früchte (§ 99 BGB) noch Nutzungen (§ 100 BGB) der Sache.[1109] Der eigentumsrechtliche Schutz schützt den Sachei-

---

1105 Siehe *Europäische Kommission*, Mitteilung vom 10. Januar 2017, COM(2017) 9 final, S. 14 zu möglichen „Rechten des Datenerzeugers".

1106 Siehe Kapitel 4 C.II.2.b) Diskussion um ein Dateneigentum, S. 265.

1107 Z. B. *BMVI*, „Eigentumsordnung" für Mobilitätsdaten?, August 2017, S. 60; *Keßler*, MMR 2017, 589 (590).

1108 Zur Übersicht über das Zusammenspiel verschiedener Regelungswerke am Beispiel von Mobilitätsdaten: *BMVI*, „Eigentumsordnung" für Mobilitätsdaten?, August 2017.

1109 Daten sind keine Früchte der Sache iSd § 99 BGB: *Specht/Rohmer*, PinG 2016, 127 (131); *Zech*, CR 2015, 137 (142); ablehnend zu Daten als Nutzungen iSd § 100 BGB: *BMVI*, „Eigentumsordnung" für Mobilitätsdaten?, August 2017, S. 60.

gentümer zwar vor Zerstörung der Sache und der Beeinträchtigung ihrer Nutzung, aber nicht vor der Vervielfältigung von auf der Sache enthaltenen immateriellen Inhalten.[1110]

Trotzdem könnte ein Recht an Daten als sonstiges Recht nach § 823 Abs. 1 BGB bestehen und einen deliktsrechtlichen Schutz stützen. Die Überschreibung, Vernichtung oder Veränderung von Daten im strukturellen Sinn kann schon eine Verletzung des Eigentums an dem Datenträger darstellen.[1111] Ob die Voraussetzungen für ein eigenständiges, darüber hinausgehendes sonstiges Recht vorliegen, ist umstritten.[1112] Dieser Schutz würde auf Rechtsfolgenseite nicht umfassend sein, sondern eher einem Rahmenrecht entsprechen. Deliktisch gilt zusätzlich § 823 Abs. 2 BGB in Verbindung mit den Strafgesetzen §§ 202a, 303a StGB (Ausspähen von Daten; Datenveränderung). Im Ergebnis zielt auch dieser deliktische Schutz nur auf Vertraulichkeit und Integrität, nicht den Ausschluss Dritter. Schutzlücken hinsichtlich der Integrität der Daten bestehen im Ergebnis nur bei fahrlässigen Beschädigungen von Daten, die sich auf Datenträgern befinden, die nicht im Eigentum des Datenspeichernden stehen.[1113] Herausgabeansprüche ergeben sich in Verbindung mit vertraglichen Abreden über § 242 BGB oder über die Herausgabe ungerechtfertigter Bereicherung gemäß §§ 812 ff. Diese Ansprüche wären aber weder absolut noch insolvenzfest.[1114] Der Schutz von Daten ist daher am ehesten vergleichbar mit dem possessorischen Besitzschutz des BGB.[1115] Das Strafrecht kann schon seinem Wesen nach keine Rechte begründen und Güterzuordnungen vornehmen. Es spricht in § 303a StGB bezüglich Daten vom „Berechtigten" und bezweckt den Schutz der Verfügungsbefugnis über die Integrität von Daten.[1116] Die Ermittlung einer Verfügungsbefugnis setzt eine Zuordnung

---

1110 Vgl. *Härting*, CR 2016, 646 (647).

1111 OLG Karlsruhe, Urteil vom 7. November 1995, 3 U 15/95 = NJW 1996, 200; OLG Oldenburg, Urteil vom 24. November 2011, 2 U 98/11 = MDR 2012, 403.

1112 Dafür *Hoeren*, MMR 2013, 486 (491); *Faustmann*, VuR 2006, 260 (263); *Meier/Wehlau*, NJW 1998, 1585 (1588f); MünchKommBGB/*Wagner*, § 823, Rn. 294f; dagegen: LG Konstanz Urteil vom 10. Mai 1996 – 1 S 292/95, NJW 1996, 2662 (Ls.); *Heymann*, CR 2016, 650 (650ff).

1113 *BMVI*, „Eigentumsordnung" für Mobilitätsdaten?, August 2017, S. 59.

1114 *Deutscher Anwaltverein*, Stellungnahme zur Frage des Eigentums an Daten, 2016, S. 7.

1115 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 66; zu einem „Datenbesitz" statt einem Dateneigentum: *Hoeren*, MMR 2019, 5.

1116 OLG Nürnberg, Beschluss vom 23. Januar 2013, 1 Ws 445/12 = CR 2013, 212; zum strafrechlichen Schutz: *Dewenter/Lüth*, Datenhandel und Plattformen, S. 35ff.

von Daten zu einer Einheit voraus, diese soll wohl „eigentümerähnlich" sein.[1117] Nach herrschender Meinung ist die Person verfügungsbefugt, die den Skripturakt bewirkt hat.[1118]

Immaterialgüterrechtlich könnten sich an Daten, beziehungsweise an den in ihnen verschlüsselten semantischen Informationen, über das Urheberrecht und das Sui-generis-Recht an Datenbanken Verwertungsrechte ergeben. Beide Schutzregime beziehen sich nicht auf Daten als bloße syntaktische Informationen. Sie könnten lediglich den mittelbaren Schutz von Daten als Reflex des Schutzes anderer Rechte bewirken.[1119]

Daten oder Datensammlungen sind ohne Ansehung ihres semantischen Inhalts nicht als Werke im Sinne des § 1 UrhG geschützt.[1120] Werden Einzeldaten zu Datenbankwerken angeordnet, könnte ein Schutz nach § 4 Abs. 2 S. 1 UrhG bestehen. Datenbankwerke können sowohl analog als auch digital und sowohl online als auch offline verfügbar sein.[1121] Die Daten müssen innerhalb des Sammelwerkes systematisch oder methodisch angeordnet sein. Bloße „Datenhaufen" (Data Lakes) genügen dieser Anforderung nicht.[1122] Unstrukturierte Datenmengen für Big-Data-Analysen dürften nicht vom Schutzbereich umfasst sein.[1123] Die systematische Anordnung von Trainingsdatensets erschöpft sich meist in der Kennzeichnung der Daten (labeled training data). Damit bleiben sie hinter der Schöpfungshöhe von Datenbankwerken zurück, für die bereits der Werkcharakter verneint wurde.[1124] Weiterhin können Datenbanken vom Sui-generis-Schutz der §§ 87a ff UrhG erfasst sein, der zur Umsetzung der Datenbankrichtlinie 96/6/EG in das UrhG als Datenbankherstellerrecht aufgenommen wurde. Zu Datenbanken zusammengefasste Daten genießen unter bestimmten Voraussetzungen den speziellen Investitionsschutz nach §§ 87a bis 87e UrhG.[1125] Als zusätzliche Schutzvoraussetzung des Sui-generis-Schutzes muss die Beschaffung, Überprüfung oder Darstellung

1117 BeckOK StGB/*Weidemann*, 40. Edition, § 303a, Rn. 5.
1118 OLG Nürnberg, Beschluss vom 23. Januar 2013, 1 Ws 445/12, CR 2013, 212.
1119 *Härting*, CR 2016, 646 (648).
1120 Vgl. Dreier/Schulze/*Schulze*, § 2 UrhG, Rn. 8; ebenso *Dewenter/Lüth*, Datenhandel und Plattformen, S. 37.
1121 Dreier/Schulze/*Schulze*, § 4 UrhG, Rn. 16.
1122 Dreier/Schulze/*Schulze*, § 4 UrhG, Rn. 17.
1123 Dreier/Schulze/*Schulze*, § 4 UrhG, Rn. 17; *Witte*, Abgrenzung von Datenbank und Datenbankinhalt, in: FS-Schneider, S. 229–233 (233).
1124 Siehe Dreier/Schulze/*Schulze*, § 4 UrhG, Rn. 20; z. B. OLG Düsseldorf, Urteil vom 29. Juni 1999, 20 U 85/98 = MMR 1999, 729, 731.
1125 Zum Normzweck: Erwägungsgründe 7, 9, 10 der RL 96/9/EG.

eine nach Art oder Umfang wesentliche, aus objektiver Sicht nicht unbedeutende[1126], Investition erfordert haben. Diese Investition kann finanzieller Art sein, es genügen aber auch Zeit und Arbeitseinsatz.[1127] Investitionen, die schon für die Erzeugung der Daten getätigt wurde, bleiben bei der Beurteilung jedoch unbeachtet.[1128] Wie auch für § 4 Abs. 2 S. 1 UrhG gilt, dass bloße Sammlungen von Rohdaten als „Datenhaufen" nicht geschützt sind.[1129] Die Evaluation der zugrunde liegenden Richtlinie stellte 2018 fest, dass das Sui-generis-Schutzrecht „im Großen und Ganzen nicht für die Datenwirtschaft gilt (von Maschinen erzeugte Daten, IoT-Geräte, Big Data, KI)".[1130] Auch in der Literatur wird angenommen, dass das Datenbankherstellerrecht nicht die Realität der Industrie 4.0 oder des IoT abbilde.[1131] Es sei zu statisch, um den Anforderungen sich ständig ändernder Datensets mit Echtzeitdaten zu entsprechen.[1132] Einzelne Daten genießen im Ergebnis keinen urheberrechtlichen Werk- oder Leistungsrechtsschutz.[1133] Die Zuordnung eines einzelnen Datums sei nicht Normzweck der §§ 87a ff UrhG.[1134]

Zu erwähnen ist auch der mögliche derivative Schutz für Erzeugnisse patentierter Verfahren gemäß § 9 S. 2 Nr. 3 PatG. Die jeweiligen Daten müssten hierzu das unmittelbar hergestellte Erzeugnis eines patentierten Verfahrens sein. Selbstlernende Systeme sind als Computerprogramme

---

1126 BGH, Urteil vom 1. Dezember 2010, I ZR 196/08 Rn. 23 = GRUR 2011, 724, 725; *Hetmank/Lauber-Rönsberg*, GRUR 2018, 574 (578).

1127 Z. B. BGH, Urteil vom 1. Dezember 2010, I ZR 196/08 = GRUR 2011, 724, 725.

1128 EuGH, Urteil vom 9. November 2004, C-203/02, Slg. 2004, I-10415 – *British Horseracing Board* = GRUR 2005, 244; Dreier/Schulze/*Schulze*, § 87a UrhG, Rn. 13.

1129 Zur Abgrenzung OLG Köln, Urteil vom 15. Dezember 2006, 6 U 229/05 = ZUM 2007, 548, 549f; Dreier/Schulze/*Schulze*, § 87a UrhG, Rn. 7.

1130 *Europäische Kommission*, Arbeitsunterlage vom 25. April 2018, Zusammenfassung der Bewertung der Richtlinie 96/9/EG über den rechtlichen Schutz von Datenbanken, SWD(2018) 146 final, S. 2; dazu *Drexl*, Data Access and Control (BEUC), S. 67f.

1131 *Drexl*, Designing Competitive Markets for Industrial Data, 2016, S. 22; *Hetmank/Lauber-Rönsberg*, GRUR 2018, 574 (579).

1132 Dreier/Schulze/*Schulze*, Vor § 87a UrhG, Rn. 1; *Drexl*, Designing Competitive Markets for Industrial Data, 2016, S. 21; *Witte*, Abgrenzung von Datenbank und Datenbankinhalt, in: FS-Schneider, S. 229–233 (229).

1133 *BMVI*, „Eigentumsordnung" für Mobilitätsdaten?, August 2017, S. 54 mwN; *Ensthaler*, NJW 2016, 3473 (3474).

1134 *Ensthaler*, NJW 2016, 3473 (3475).

grundsätzlich nicht patentierbar.[1135] Sie sind nur als Beitrag zur Lösung eines konkreten technischen Problems mit technischen Mitteln dem Patentschutz zugänglich.[1136] Daten können aber auch unmittelbare Erzeugnisse dieser und anderer patentierbarer Verfahren sein: Der BGH hat festgestellt, dass Datenfolgen – unter strengen Anforderungen – derivative Erzeugnisse im Sinne des § 9 S. 2 Nr. 3 PatG sein können.[1137] Die Datenfolge müsste ihrer Art nach als tauglicher Gegenstand eines Sachpatents in Betracht kommen.[1138] Die in den Daten enthaltene Information ist unbeachtlich; nicht zuletzt, weil der gesetzliche Ausschlusstatbestand des § 1 Abs. 3 Nr. 4 PatG nicht unterlaufen werden darf.[1139] Es bestehen berechtigte Zweifel daran, dass ein derivativer Schutz von Datenfolgen als „Früchte" patentierbarer Verfahren de lege lata von hoher Relevanz sein wird.[1140]

Auch der Schutz von Betriebs- und Geschäftsgeheimnissen nach GeschGehG vermittelt keine Rechte oder Vermögenszuordnungen, weil er lediglich faktische Kontrollinstrumente zur Geheimhaltung verstärkt und das Missachten der Geheimhaltungsmaßnahmen missbilligt.[1141] Er steht dem Deliktsrecht damit näher als dem Recht des geistigen Eigentums. Hinzu kommt, dass sich nach der von Art. 39 des TRIPS-Abkommens übernommen Definition von Geschäftsgeheimnissen[1142] die Frage stellt, ob auch ganze Datensets und nicht nur einzelne Informationen Subjekt von schutzfähigen Geheimhaltungsmaßnahmen sein können.[1143]

Eine Zuordnung von Daten, die bereits die syntaktische Ebene des binären Codes erfasst, gibt es nicht. Als unverarbeitete semantische Informationen sind sie den Schutzzwecken der Immaterialgüterrechte konzeptionell

---

1135 BPatG, Beschluss vom 9. Juni 2015 – 17 W (pat) 37/12 = juris.bundespatentgericht.de.

1136 So *Deutsches Patent- und Markenamt*, Künstliche Intelligenz und Schutzrechte, 4. Februar 2019.

1137 BGH, Urteil vom 27. September 2016, X ZR 124/15 = GRUR 2017, 261, 263 – *Rezeptortyrosinkinase II:* Der Schutz für bloße digitale Codierungen der Testergebnisse wurde vom BGH abgelehnt; *Drexl*, Data Access and Control (BEUC), S. 87f.

1138 BGH, Urteil vom 27. September 2016, X ZR 124/15 = GRUR 2017, 261, 263 – *Rezeptortyrosinkinase II*.

1139 BGH, Urteil vom 27. September 2016, X ZR 124/15 = GRUR 2017, 261, 263 – *Rezeptortyrosinkinase II*.

1140 *Hetmank/Lauber-Rönsberg*, GRUR 2018, 574 (576f).

1141 Vgl. *Kerber*, GRUR Int. 2016, 989 (991).

1142 Art. 2 Abs. 1 der Richtlinie (EU) 2016/943.

1143 Hierzu *Desaunettes/Hilty/Kur/Knaak*, Stellungnahme des MPI vom 17. April 2018, Rn. 12; *Surblytė*, Data as a Digital Resource, MPI for Innovation and Competition Research Paper No. 16–12, S. 9.

fremd.[1144] Dies geht auf die grundlegenden Funktionen des Immaterialgüterrechts zurück, nämlich die Anreiz-, Offenbarungs- und Belohnungsfunktion.[1145] Die potentiell wettbewerbsbeschränkenden Auswirkungen von Immaterialgüterrechten wären in Bezug auf Daten besonders unerwünscht, wie die Debatten für Datenzugang schon für faktische Kontrolle zeigen.[1146] Gerade für datengetriebene Geschäftsmodelle sind große Datensets wichtiger als einzelne, möglicherweise schutzfähige Informationen.[1147] Für die Gewährleistung von Datenströmen könnte der Schutz von Daten durch Ausschließlichkeitsrechte wie etwa das Urheberrecht sogar negativ sein und die Entwicklung von Künstlicher Intelligenz behindern.[1148]

b) De lege ferenda – Diskussion um ein Dateneigentum

Verschiedene Beweggründe haben in den letzten Jahren eine Diskussion um die Einführung eines umfassenden Ausschließungsrechts für Daten motiviert. Dieses wird schlagwortartig als „Dateneigentum" bezeichnet. In den USA wurde bereits im Jahr 1967 ein Informationseigentum vorgeschlagen.[1149] Weil die Diskussion seitens der europäischen und nationalen Gesetzgeber nicht weiter verfolgt wird[1150], ist eine ausführliche Darstellung der Diskussion entbehrlich. Die Diskussion wurde auf einem Spektrum geführt, auf dem an entgegengesetzten Enden ein umfassendes Dateneigentum und am anderen die faktische Kontrolle von Daten standen, während dazwischen ein Leistungsschutzrecht diskutiert wurde.

Aus der bloß faktischen Kontrolle von Daten erwuchs die Besorgnis, dass es für ihre Generierung und Erfassung zu wenig Anreize gebe und die Datenwirtschaft hinter ihren Möglichkeiten zurückbliebe. Zur Absicherung der Geschäftsmodelle stellte sich für einige Akteure der Wirtschaft,

---

1144 *Keßler*, MMR 2017, 589 (590).
1145 *Hetmank/Lauber-Rönsberg*, GRUR 2018, 574 (580).
1146 Vgl. *Hetmank/Lauber-Rönsberg*, GRUR 2018, 574 (580).
1147 *Surblytė*, Data as a Digital Resource, MPI for Innovation and Competition Research Paper No. 16–12, S. 4.
1148 *Surblytė*, WuW 2017, 120 (125).
1149 *Westin*, Privacy and Freedom, S. 324f; zu der in den USA geführten Diskussion um Dateneigentum für Patientendaten und gegen ein solches Recht: *B. Evans*, Harvard Journal of Law and Technology, Vol. 25, No 1, S. 69–130 (2011).
1150 Vgl. *Arbeitsgruppe „Digitaler Neustart" der Konferenz der Justizministerinnen und Justizminister der Länder*, Bericht vom 15. Mai 2017, S. 98: „festzuhalten, dass es keiner gesetzlichen Bestimmung der Rechtsqualität von Daten bedarf".

der Politik und der Wissenschaft die Frage nach der rechtlichen Absicherung der den Geschäftsmodellen zugrundeliegenden Maschinendaten. Es gibt jedoch kein rechtliches Prinzip, das besagt, dass alles, auch Daten, einem bestimmten Subjekt zuzuordnen sein muss.[1151]

Das wichtigste Argument für ein „Dateneigentum" oder Leistungsschutzrechte von Daten ist die Stärkung von Investitionsanreizen durch Einräumung von zeitlich begrenzten Nutzungs- und Verwertungsrechten. Außerdem sollen der Handel mit und die Weiterverwertung von Daten angetrieben werden.[1152] Die Weiterverwendungsmöglichkeiten würden aktuell nicht ausgeschöpft.[1153] Sowohl industrieweit als auch in Multi-Stakeholder-Situationen (Smart Cars, Smart Health, Smart Farming) sei eine breitere Wertschöpfung möglich. Die Feststellung, dass aktuell nicht genügend in die Erfassung von Daten investiert wird, dürfte jedoch schwerfallen.[1154] Die Europäische Kommission scheint auch nicht von zu niedrigen Anreizen bei der Datengenerierung und -erfassung auszugehen.[1155] Die massenhafte Generierung von Daten ist das kennzeichnende Charakteristikum von Big Data Analytics und der Industrie 4.0. Die Kosten für die Erzeugung von Daten dürften zudem in der näheren Zukunft sinken. Die Feststellung, dass zu wenig Daten erfasst würden, setzt eine ökonomische Feststellung einer mindestoptimalen Datenerzeugung voraus, die bisher nicht vorgenommen wurde.[1156]

Den theoretischen Hintergrund zu diesem Argument liefern das Public-Goods-Problem und Arrows Information Paradox[1157]. Das Public-Goods-Problem besagt, dass einem Entwickler Anreize zur Schaffung neuer Innovationen fehlen, wenn diese leicht imitiert oder kopiert werden kön-

---

1151 *Drexl et al.*, Data Ownership and Access to Data, MPI for Innovation and Competition Research Paper No. 16–10, Rn. 6; *Drexl.*, Designing Competitive Markets for Industrial Data, 2016, S. 7; zu der Gefahr der Selbstreferenzialität und Zirkulatrität *Richter/Hilty*, Die Hydra des Dateneigentums, S. 6.

1152 *BMVI*, „Eigentumsordnung" für Mobilitätsdaten?, August 2017, S. 83.

1153 *Kerber*, Rights on Data, S. 14.

1154 *Dewenter/Lüth*, Datenhandel und Plattformen, S. 47; *Drexl,* Designing Competitive Markets for Industrial Data, 2016, S. 30ff; *Duch-Brown/Martens/Müller-Langer*, The economics of ownership, access and trade in digital data, S. 14; *Kerber*, GRUR Int. 2016, 989 (992f); *ders*, Rights on Data, S. 7; *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 68.

1155 *Kerber*, Rights on Data, S. 7.

1156 Dazu *Del Toro Barba*, ORDO 68, S. 217–248, 225f, 228ff (2017) mwN.

1157 Vgl. Fn. 980; erläuternd *Dewenter/Lüth*, Datenhandel und Plattformen, S. 42f; dagegen, dass dies ein Problem für den Datenhandel begründet: *Kerber*, GRUR Int. 2016, 989 (994).

nen.[1158] Ein Public Good, also ein öffentliches Gut, ist ein solches, das gleichzeitig non-rival und nicht ausschließbar ist. Wegen der fehlenden Ausschließbarkeit ist eine Imitation oder Kopie besonders wahrscheinlich. Für Daten ergibt sich die janusköpfige Situation, dass einmal offenbarte Informationen im Gegensatz zu exklusiven Daten nicht ausschließbar sind.[1159] Das Information Paradox und das Public-Goods-Problem werden in der Regel über Immaterialgüterrechte „geheilt", sodass eine Offenbarung des Schutzgegenstands bei gleichzeitiger ausschließlicher Nutzung und Veräußerung möglich bleibt. Insofern überrascht der Vorschlag eines entsprechenden Immaterialgüterrechts nicht. Die Anreiz- und Belohnungsfunktion ist die traditionelle Rechtfertigung für die Etablierung von Ausschlussrechten für immaterielle Güter. Ohne einen generellen Anreizmangel dürfte die Annahme eines generellen strukturellen Marktversagens und damit die Rechtfertigung eines Dateneigentums- oder Datenleistungsschutzrechtes schwerfallen.[1160] Gerade bei Daten, die als Nebenprodukt (By-Product) der Nutzung eines Dienstes anfallen, fällt die Argumentation für ein Datenerzeugerrecht eher schwer.[1161] Zech verortet das Anreizargument eher bei den Anreizen zur Offenbarung von Daten und zur Etablierung von Datenmärkten.[1162]

Hinzu kommt das Argument der Klarstellungsfunktion und Zuordnungsfunktion, das davon ausgeht, dass die aktuelle De-facto-Lösung mit vertraglichen oder faktischen Zugangsmöglichkeiten die „tatsächlich Berechtigten" benachteiligt.[1163] Insbesondere das industrielle Smart Manufacturing wird Modelle hervorbringen, bei denen mehrere Akteure ein Interesse an der Analyse der bei der Produktion erfassten Daten haben. Hierzu zählen die Produzenten, die Hersteller der Produktionsmaschinen sowie die Entwickler der Software für Smart Machines.[1164] Ein entsprechend ausgestaltetes Ausschlussrecht an Daten könnte Korrekturen an der De-facto-Zuordnung vornehmen und sicherstellen, dass nicht grundsätzlich der Codierende der Berechtigte ist, wenn dies zu unliebsamen Ergebnissen

1158 Dazu *Kerber*, GRUR Int. 2016, 989 (992).
1159 So auch *Kerber*, GRUR Int. 2016, 989 (993).
1160 *Kerber*, Rights on Data, in: Lohsse/Schulze/Staudenmayer (Hrsg.), Trading Data, S. 109–133, (117, 120ff); *ders*, GRUR Int. 2016, 989 (993).
1161 *Drexl,* Designing Competitive Markets for Industrial Data, 2016, S. 16. *Kerber*, GRUR Int. 2016, 989 (993) und Fn. 33.
1162 *Zech*, CR 2015, 137 (144); *ders.*, JIPLP 2016, Vol. 11, No. 6, 460 (470); *ders.* Information als Schutzgegenstand, S. 316–323.
1163 *Zech*, CR 2015, 137 (145); *ders.*, JIPLP 2016, Vol. 11, No. 6, 460 (464).
1164 Dazu *Kerber*, GRUR Int. 2016, 989 (995).

führt. Allerdings wird nicht immer klar sein, wem ein solches Recht im konkreten Fall zugewiesen wird. Diese Rechtsunsicherheit könnte zu höheren Transaktionskosten führen, was dem eigentlichen Zweck eines Dateneigentums widersprechen dürfte.[1165] Zukünftige Formen der Beteiligung bei der Generierung von Daten sind noch nicht absehbar und können tatbestandlich nur schwer in einer generellen Vorschrift erfasst werden. Zudem können sich in der Zukunft auch die gesetzgeberische Intention verschieben und ein anderer Beteiligter bei der Datenschöpfung als „besser berechtigt" betrachtet werden.

Schwer fällt bei einem Dateneigentum zudem die Trennung der syntaktischen und der semantischen Ebene der Information.[1166] Es besteht Einigkeit darüber, dass es keine Exklusivrechte an semantischen Informationen geben soll und die gleiche semantische Information erneut generierbar bleibt.[1167] Würde der syntaktische Schutz zu weit gehen, könnte ein Dateneigentum die Umkehr des Grundsatzes des Gemeingebrauchs von Informationen bewirken.[1168]

Ein weiteres Argument ist mit der in dieser Arbeit behandelten Fragestellung verwandt: Ein Dateneigentum soll zur Steigerung der Verkehrsfähigkeit von Daten und damit zur Stärkung des Datenhandels beitragen. Der Gedanke der Senkung von Transaktionskosten wird beispielsweise von der Europäischen Kommission angeführt.[1169] Für traditionelle Immaterialgüter gilt, dass Patente und Urheberrechte die Transaktionskosten senken.[1170] Ein Recht sei leichter zu veräußern als faktische Positionen, könne vertraglich eindeutiger bestimmt und dinglich leichter übertragen werden. Ein gutgläubiger Erwerb von Datenrechten wäre möglich und die Rückabwicklung fehlgeschlagener Verträge erleichtert und mit weniger Unsicherheiten belastet.[1171] Es gibt durchaus Argumente dafür, dass ein Handel mit Daten durch einen sicheren rechtlichen Rahmen vereinfacht

---

1165 Vgl. *Richter/Hilty*, Die Hydra des Dateneigentums, S. 9.

1166 *Kerber*, GRUR Int. 2016, 989 (997).

1167 *Kerber*, GRUR Int. 2016, 989 (997); *Wiebe*, GRUR Int. 2016, 877 (882); *Zech*, CR 2015, 137 (146).

1168 *Wiebe*, CR 2017, 87 (91).

1169 *Europäische Kommission*, Staff Working Document vom 10. Januar 2017, SWD(2017) 2 final, S. 33: Eine Möglichkeit sei die Schaffung eines neuen Datenherstellerrechts "with the objective of enhancing the tradability of non-personal or anonymised machine-generated data as an economic good".

1170 *Kerber*, Rights on Data, S. 9.

1171 *Europäische Kommission*, Mitteilung vom 10. Januar 2017, COM(2017) 9 final, S. 14; dazu *Denga*, NJW 2018, 1371 (1374).

werden kann. Ein Dateneigentum führt aber nicht per se zum Austausch, sondern könnte auch zur Isolation von Daten führen.[1172] Wenn es besonders weit reicht und Dritte von der Datenerfassung ausschließt, würde dies Wissensmonopole befeuern.[1173] Große Unternehmen wären konfrontiert mit territorial begrenzten Rechtssystemen, die den Datenstrom verkomplizieren. Probleme der Patent-Hold-Ups, Trolls, und standardessentieller Patente würden sich entsprechend für Datenmärkte stellen.[1174] Ein Dateneigentum könnte daher Downstream-Märkte in ihrer Entwicklung behindern.[1175] Es würde dann den Datenmarkt nicht liberalisieren, sondern eher die Entstehung marktmächtiger Stellungen („Informationsmonopole") befeuern und Marktzutrittsbarrieren errichten.[1176] Ein Ausschließungsrecht kann ein Machtungleichgewicht einerseits ausgleichen, andererseits aber auch vertiefen.[1177] Die Realität belegt zudem, dass Daten schon heute regelmäßig Gegenstand von Transaktionen und Handelsbeziehungen sind.

Zu den Befürwortern eines Dateneigentums zählen Fezer[1178], Hoeren[1179] und Zech[1180]. Ursprünglich plädierte auch das Bundesministerium für Verkehr und digitale Infrastruktur für ein Dateneigentum.[1181] Mittlerweile überwiegt die Zahl der Gegner eines Ausschlussrechtes an Daten.[1182]

1172 *Härting*, CR 206, 646 (649); *Kerber*, GRUR Int. 2016, 989 (997).

1173 *Kerber*, GRUR Int. 2016, 989 (992);
Schweitzer/Peitz, Datenmärkte in der digitalisierten Wirtschaft, S. 66.

1174 *Kerber*, GRUR Int. 2016, 989 (997).

1175 *Drexl et al.*, Data Ownership and Access to Data, MPI for Innovation and Competition Research Paper No. 16–10, Rn. 6; *Kerber*, GRUR Int. 2016, 989 (997).

1176 *Drexl et al.*, Data Ownership and Access to Data, MPI for Innovation and Competition Research Paper No. 16–10, Rn. 6; *Drexl.*, Designing Competitive Markets for Industrial Data, 2016, S. 7.

1177 *Kerber*, GRUR Int. 2016, 989 (996); *Schweitzer/Peitz*, NJW 2018, 275 (279).

1178 *Fezer*, Repräsentatives Dateneigentum, 2018, S. 6, 32ff; *ders.*, ZD 2017, 99; *ders.*, MMR 2017, 3.

1179 *Hoeren*, MMR 2013, 486 (491): „Die Schaffung eines Dateneigentums in Analogie zum Sacheigentum, wenn auch in begrenzter Weise, ist daher geboten und möglich.".

1180 *Zech*, CR 2015, 137; *ders*., GRUR 2015, 1151; *ders.*, Information als Schutzgegenstand, 2012, S. 423.

1181 *BMVI*, Wir brauchen ein Datengesetz in Deutschland!, Artikel, 2017; außerdem *Becker*, Schutzrechte an Maschinendaten und die Schnittstelle zum Personendatenschutz, in: FS-Fezer, S. 815–833; *Ensthaler*, NJW 2016, 3473 (3478).

1182 *Arbeitsgruppe „Digitaler Neustart" der Konferenz der Justizministerinnen und Justizminister der Länder*, Bericht vom 15. Mai 2017, S. 98; *BDI/Noerr*, Industrie 4.0 – Rechtliche Herausforderungen der Digitalisierung; *Dewenter/Lüth*, Datenhandel und Plattformen, S. 50; *Drexl*, NZKart 2018, 415 (415ff); *ders.*, Data

Die Einführung eines Dateneigentumsrechts würde das sprichwörtliche Pferd von hinten aufzäumen: Ein Ausschließlichkeitsrecht bezwecke nach den meisten Aussagen keine Schaffung eines rechtlichen – über den faktischen hinausreichenden – Schutzes, sondern einen Rahmen für Zugangsrechte.[1183] Richtig ist, dass eine Zuordnungsregelung wohl mit einer Zugangsregelung einhergehen würde, weil sie sonst kaum verfassungsgemäß und wirtschaftlich dienlich wäre im aktuellen Kanon der effizienten Wiederverwendung von Daten. Gerade im Hinblick auf Daten zum Training selbstlernender Systeme würde ein Dateneigentum, das den Status Quo der faktischen Kontrolle und vollkommenen Privatautonomie aufrechterhält, wenig ändern. Inwiefern ein Dateneigentum eine breite Streuung des Nutzens von Trainingsdaten erlaubt hätte, ist unklar. Die Etablierung einer Zugangsregelung bleibt auch ohne den rechtlichen Rahmen, den ein Ausschließlichkeitsrecht bieten würde, möglich.

III. Die wettbewerbliche Bedeutung von Daten und Informationen

Die wettbewerblichen Vorteile, die sich aus der Kontrolle über Daten ergeben, sind nuanciert und komplex.[1184] Sie wiegen je nach Verwendungszweck unterschiedlich schwer und es gibt keine allgemeingültigen Vorteile, die es erlauben, außer Acht zu lassen, wie die jeweiligen Daten erfassen und verarbeitet werden.[1185] Selbstlernende Systeme können neue Prozesse ermöglichen und Unternehmen in bestehenden Märkten wettbewerbsfähiger machen, neue Märkte erschließen und bei der Entwicklung neuer Produkte helfen.[1186] Die aus Daten gewonnenen Informationen werden die

---

Access and Control (BEUC), S. 149; *Drexl et al.*, Data Ownership and Access to Data, MPI for Innovation and Competition Research Paper No. 16–10, Rn. 4; *Drexl et al.*, GRUR Int. 2016, 914 (914); *Kerber*, Rights on Data, in: Lohsse/Schulze/Staudenmayer (Hrsg.), Trading Data, S. 109–133 (129); *ders.*, GRUR Int. 2016, 989 (989ff); *OECD*, Data-Driven Innovation, S. 198; *Richter/Hilty*, Die Hydra des Dateneigentums; *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 58ff; *Steinrötter*, MMR 2017, 731 (736).

1183 *Europäische Kommission*, Mitteilung vom 10. Januar 2017, COM(2017) 9 final, S. 11, 15; *dies.*, Staff Working Document vom 10. Januar 2017, SWD(2017) 2 final, S. 35f; *Zech*, CR 2015, 137.

1184 *M. Gal/Rubinfeld*, Access Barriers to Big Data, S. 37.

1185 Von Datenwettbewerb zu reden, sei daher „lazy": *Balto/Lane*, Monopolizing Water in a Tsunami, S. 1.

1186 *Colangelo/Maggiolino*, ECJ, Vol. 13, S. 249–281, 254 (2017).

Produktentwicklung stark beeinflussen und die Produktqualität erwartbar verbessern, um sie den Nachfragewünschen besser anzupassen.[1187]

Informationen sind Inputs, die Unternehmen ebenso wie Kapital und Arbeitskraft nutzen, um neue Produkte und Prozesse zu entwickeln, indem sie die Bedürfnisse der Nachfrager und die Strategien der Wettbewerber in Echtzeit auswerten.[1188] Die Wertschöpfungskette verläuft von der Datenerfassung über die -speicherung, -aufbereitung und -analyse zur Anwendungsidee.[1189] Machine-Learning-Technologien erlauben es, Korrelationen zwischen scheinbar unzusammenhängenden Variablen festzustellen[1190] und Muster in den Bedürfnissen der Nachfrager zu erkennen, bevor diese vom Nachfrager geäußert werden. Insofern ist korrekt, dass große Datensets bei richtiger Analyse wertvolle Informationsquellen sein können, die sich in Wettbewerbsvorteilen niederschlagen. Wie für jede Narrative gilt, ist auch die soeben erläuterte stark vereinfachend.[1191] Sie vereinfacht die Zusammenhänge insofern, als dass sie angibt, die bloße Generierung großer Datenvolumina genüge, um einen Wettbewerbsvorteil zu erzielen. Tatsächlich sind der Aufbau von Infrastrukturen und qualifizierten Arbeitskräften, die Entwicklung von Technologien und einer Geschäftsstrategie nötig, um den Daten wirtschaftlich wertvolle Informationen zu entnehmen.

Außerdem bedeutet eine große Menge erfasster Daten nicht, dass auch diese Menge vollständig für Geschäftsentscheidungen herangezogen wird: Hal Varian gibt an, dass bei Google wohl in der Regel 0,1 Prozent eines Datensets für die Analyse von Geschäftsdaten genügen.[1192] Zudem können Daten als digitalisierte Informationen falsch oder verzerrend sein. Im Ergebnis ergibt sich die wirtschaftliche Nützlichkeit von Daten weniger aus ihrem Volumen als vielmehr aus den intellektuellen und materiellen Investitionen in die Datenverarbeitung.[1193] Ein Wettbewerbsvorteil ent-

---

1187 *Europäische Kommission*, Eine europäische Datenstrategie, COM(2020) 66 final, S. 2f; *Lundqvist*, EuCML 2018, 146 (148), vgl. schon 2011 zu der Produktivitätssteigerungen durch Datenverarbeitung: *Brynjolfsson/Hitt/Kim*, Strength in Numbers, S. 31.

1188 *Colangelo/Maggiolino*, ECJ, Vol. 13, S. 249–281, 250ff (2017).

1189 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 16.

1190 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 16.

1191 Vgl. *Colangelo/Maggiolino*, ECJ, Vol. 13, S. 249–281, 250 (2017).

1192 *Varian*, Big Data: New Tricks for Econometrics, 2013, S. 3.

1193 *Colangelo/Maggiolino*, ECJ, Vol. 13, S. 249–281, 252 (2017) mwN; dagegen: *Stucke/Grunes*, Debunking the myths over big data and antitrust, CPI Antitrust Chronicle Mai 2015.

steht aus den in Daten enthaltenen semantischen Informationen, weshalb die Verarbeitung ein empfindliches Bindeglied der datengetriebenen Wertschöpfungskette ist. Die Analysefähigkeit würde nach Ansicht der Europäischen Kommission „zu einer wichtigen Voraussetzung für geschäftlichen Erfolg“[1194] und einem Wettbewerbsvorteil für Unternehmen, „denen riesige Datenmengen zur Verfügung stehen und die auch über die technischen Kapazitäten und qualifizierten Mitarbeiter für die Auswertung der Daten verfügen.“[1195]

Ein Datenvorteil muss erst in einen Informationsvorteil übersetzt werden, bevor sich aus ihm ein Wettbewerbsvorteil ergeben kann. Die erfolgreiche Übersetzung ist nicht selbstverständlich und kann nicht vorausgesetzt werden. Ebenso ist nicht jeder Informationsvorteil automatisch ein Wettbewerbsvorteil. Aus demselben Datensatz könnten mehrere Verarbeiter unterschiedliche oder sogar gegenläufige Informationen erlangen, während gleichzeitig die gleiche Information aus unterschiedlichen, nicht kongruenten Datensets erlangt werden kann. Die prominente Stellung, die große Datenmassen in der Debatte einnehmen, überinterpretiert die Bedeutung der Datensammlung im Vergleich zur Datenanalyse.[1196] Als Folgerung hieraus wird teilweise vorgeschlagen, dass das Kartellrecht und insbesondere die Essential-Facilities-Doktrin statt Daten Informationen betrachten sollten.[1197] Der Einfachheit halber und zur Betonung digitaler Phänomene wird eher von „Datenreichtum“ als „Wissensreichtum“ gesprochen. Ultimativ können nur Informationen Produkte verbessern, nicht bereits Datenmassen. Informationen und Daten sind dabei keineswegs kongruent.

Zu bedenken ist auch, dass die Vorteile für die Produktentwicklung, die sich aus einem einzigartigen und exklusiven Datenset ergeben, eine kurze Halbwertszeit haben können.[1198] Rubinfeld und Gal illustrieren dies anhand eines einzigartigen Datensets zu den geologischen Bedingungen für Tiefenbohrungen. Ein Versicherer, der Zugriff zu diesen Daten hat, kann seine Preise so anpassen, dass sie die Risiken der Tiefenbohrungen in einer Region besser abbilden. Diesem Beispiel könnten andere Versicherer folgen, ohne selbst auf das zugrundeliegende Datenset zugegriffen zu

---

1194 *Europäische Kommission*, Mitteilung vom 25. April 2018, COM(2018) 232 final, S. 3.

1195 *Europäische Kommission*, Mitteilung vom 25. April 2018, COM(2018) 232 final, S. 3.

1196 *Colangelo/Maggiolino*, ECJ, Vol. 13, S. 249–281, 278 (2017).

1197 *Colangelo/Maggiolino*, ECJ, Vol. 13, S. 249–281, 277 (2017).

1198 Zum Folgenden: *Rubinfeld/M. Gal*, Access Barriers to Big Data, S. 37.

haben. Preisbeobachtungsalgorithmen ermöglichen eine zügige Reaktion der Wettbewerber. Zumindest dort, wo das Verhalten von Wettbewerbern leicht zu imitieren ist (Preissetzung) oder Reverse Engineering[1199] leicht fällt, sinken der Wert der Daten zur Produktverbesserung und damit auch die Anreize zur Erfassung und Auswertung der Daten.[1200]

Die Partizipation der Nutzer an der Produktentwicklung durch Erfassung ihres Nutzverhaltens verspricht Innovationen.[1201] Dies ist jedoch kein Automatismus, sondern setzt geeignete Analyseinstrumente voraus. Rückschlüsse für die Weiterentwicklung der angebotenen Produkte ergeben sich erst nach Analyse der Informationen. Grundsätzlich zeigt eine hohe Dichte und Vielfalt von Informationen vielversprechende neue Innovationspfade auf.

D. Konzept der Rückkopplungseffekte oder Datennetzwerkeffekte

Hinsichtlich der Möglichkeiten der Produktoptimierung durch Daten und daraus gewonnener Informationen werden im Hinblick auf selbstlernende Systeme, aber auch generell in datengetriebenen Geschäftsmodellen, Rückkopplungseffekte angenommen. Rückkopplungen sind in verschiedensten Systemen technischer, biologischer oder wirtschaftlicher Art zu beobachten.[1202] Grundsätzlich handelt es sich dabei um einen signalverstärkenden Mechanismus: Ausgaben eines Regelkreises werden ihm als Eingaben, gegebenenfalls in modifizierter Form, wieder zugeführt. Hierher stammt auch die englische Bezeichnung Feedback Effect; ein System füttert sich selbst (feed back). Es wird zwischen positiven und negativen Rückkopplungseffekten unterschieden. Positive Rückkopplungseffekte bezeichnen das Phänomen, dass eine Größe oder ein Signal sich selbst verstärkt.[1203] Bei negativen Rückkopplungseffekten schwächt das Ausgangssignal das Eingangssignal ab und wirkt ihm reduzierend entgegen. Daher werden

---

1199 Siehe dazu Kapitel 4 E.III.2. Reverse Engineering.

1200 *Rubinfeld/M. Gal*, Access Barriers to Big Data, S. 38.

1201 *Von Hippel*, Democratizing Innovation, 2005, S. 168f; *Rohracher*, Zukunftsfähige Technikgestaltung als soziale Innovation, in: Sauer/Lang (Hrsg.), Paradoxien der Innovation, S. 175–189 (184).

1202 Beispiele sind das Thermostat, Windmühlen, die Dampfmaschine, der Benjamin-Franklin-Effekt, Arbeitslosigkeit und Reputation.

1203 Vgl. *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 31; *Dietrich*, Wettbewerb in Gegenwart von Netzwerkeffekten, S. 78f; *Parker/Alstyne/Choudary*, Platform Revolution, S. 296.

sie als stabilisierend oder ausgleichend bezeichnet, während positive Rückkopplungseffekte durch exponentielles Wachstum und Verstärkung von Schwankungen eher zu Instabilität führen. Das ökonomische Verständnis von Rückkopplungseffekten ist nicht ganz einheitlich und trennscharf in der Abgrenzung zu Netzwerk- und Skaleneffekten.[1204] Teilweise werden positive Feedback-Effekte ausgehend vom Symptom erklärt, nämlich einer „Aufwärtsspirale" (virtuous cycle) oder einer Winner-takes-all-Situation.[1205] Keinesfalls sind Rückkopplungseffekte ein datenspezifisches Erklärungsmodell. Diejenigen, die regelmäßig an der Lösung eines Problemtyps arbeiten, verbessern ihre Problemlösungsfähigkeiten und werden schließlich zu Spezialisten.

Der kanadische Informatiker Yoshua Bengio bezeichnete Künstliche Intelligenz als eine Technologie, die ihrer Art nach wie alle datenreichen Sektoren zu „Winner takes all"-Märkten neige.[1206] Das Land und das Unternehmen, das die Technologie dominieren, würden mit der Zeit mehr (Markt-)Macht gewinnen. Mehr Daten und ein größerer Nutzerkreis ergäben einen schwer zu überwindenden Vorteil – auch Programmierer wollten zu den stärksten Unternehmen gehen, sodass sich auf lange Sicht (Daten-)Reichtum und Macht konzentrierten. Hiermit spricht Bengio eine Vielzahl potentieller Rückkopplungseffekte an, die jeweils Informationen, Nutzerzahlen, Arbeitskräfte und Kapital betreffen.

Die Debatte um positive Feedback-Effekte und aus ihnen resultierende „Winner takes all"-Märkte ist nicht neu, sondern wurde schon im Vorfeld der 9. GWB-Novelle geführt:[1207] Der Zugang zu großen Datenpools könnte Wettbewerbsvorteile bedeuten und zu Marktmacht beitragen, weil Wettbewerber dadurch im Aufbau eigener Datenpools eingeschränkt seien. Die 9. GWB-Novelle hatte dabei eher Plattformen und klassische internetbasierte Geschäftsmodelle wie soziale Netzwerke und Suchmaschinen vor Augen als selbstlernende Systeme. Für selbstlernende Systeme knüpft der wohl bedeutendste Rückkopplungseffekt bei dem Lernprozess an.

1204 Siehe *Monopolkommission*, Sondergutachten 68, S. 39, Rn. 55; *Dietrich*, Wettbewerb in Gegenwart von Netzwerkeffekten, S. 78: „sich selbst verstärkende Nachfrage nach Netzwerkgütern in Gegenwart starker Netzwerkeffekte"; *J. Weber*, Zugang zu den Softwarekomponenten der Suchmaschine Google, S. 69; *Weck*, NZKart 2015, 290 (294).

1205 *Dietrich*, Wettbewerb in Gegenwart von Netzwerkeffekten, S. 78f; *Parker/Alstyne/Choudary*, Platform Revolution, S. 296.

1206 Zum Folgenden *LeVine*, Artificial Intelligence pioneer calls for the breakup of Big Tech, Axios, 20. September 2017.

1207 *BMWi*, Referentenentwurf 9. GWB-ÄndG, 1. Juli 2016, S. 51.

I. Begriff

Die besonderen Rückkopplungseffekte in Hinblick auf Daten werden als Datennetzwerkeffekte oder Data Network Effects bezeichnet.[1208] Wie dieser Begriff andeutet, verschwimmen hier die positiven Rückkopplungseffekte mit Netzwerkeffekten. Für das gleiche Phänomen wird auch der Begriff der Feedback-Effekte gewählt.[1209]

Der Begriff der Data Network Effects wurde von Matt Turck im Jahr 2016 verbreitet:[1210] Data Network Effects treten auf, wenn das Produkt, üblicherweise angetrieben durch Machine Learning, intelligenter wird, je mehr Daten es von seinen Nutzern erhält.[1211] Je mehr Personen das Produkt nutzen, desto mehr Daten stellten sie bereit; je mehr Daten bereitstehen, desto stärker wird die Produktevolution angetrieben und desto besser erfüllt es die Bedürfnisse der Nutzer, die das Produkt weiter oder verstärkt nutzen und mehr Daten bereitstellen.[1212] Die Grundlage ist, dass Produkte sich verbessern, je mehr Nutzungsdaten ihnen zur Verfügung gestellt werden. Diese Voraussetzung erfüllen selbstlernende Systeme. Abstrahiert auf die Lernfähigkeiten kann angenommen werden, dass jeder Innovation eine Lernphase zu ihrer Anwendung folgt. Die Arbeitskräfte, die jene neue Technologie anwenden, lernen, wie sie effizient und fehlerfrei genutzt wird und wie der größte Nutzen erzielt wird. Diese Lerneffekte sind der Nährboden für neue Verbesserungen.[1213] Der Datennetzwerkeffekt bezieht sich zunächst nur auf den lernenden Dienst oder die lernende Anwen-

---

1208 *Stucke/Grunes*, Debunking the myths over big data and antitrust, CPI Antitrust Chronicle 2015, Vol. 5 No. 6; dazu auch *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 9f; *Mitomo*, Data Network Effects: Implications for Data Business, 28th European Regional Conference of the International Telecommunications Society (ITS) 2017, S. 2; *Sachnow et al.*, Digitale Transformation bei der Kaeser SE, in: Oswald/Krcmar (Hrsg.), Digitale Transformation, S. 99–120 (115); *Schweitzer/Haucap/Kerber/Welker*, Modernisierung der Missbrauchsaufsicht, S. 12.

1209 *Mayer-Schönberger/Ramge*, Das Digital, S. 193; *dies.*, Die Datenklauer, Wirtschaftswoche Nr. 50, 1. Dezember 2017, S. 12.

1210 *Turck*, The Power of Data Network Effects, 4. Januar 2016.

1211 Original: „Data network effects occur when your product, generally powered by machine learning, becomes smarter as it gets more data from your users.", *Turck*, The Power of Data Network Effects, 4. Januar 2016.

1212 *Turck*, The Power of Data Network Effects, 4. Januar 2016; zu Feedback Loops am Beispiel von LinkedIn, *Parker/Alstyne/Choudary*, Platform Revolution, S. 218.

1213 *Bessen*, Learning by Doing, S. 48.

dung: Beispielsweise verbessert die Erfassung der Nutzergewohnheiten auf Netflix den Empfehlungsdienst, aber nicht notwendig das Produkt. Es ist zu bezweifeln, dass die Qualität des Empfehlungsdienstes zentral für den Erfolg des Produktes von Netflix ist. Ähnlich stellt der Empfehlungsdienst von Spotify auf der Grundlage von Daten über den Musikgeschmack der Nutzer Playlists zusammen.[1214] Langfristig sollen Nutzer dank der Empfehlungsdienste bei Netflix und bei Spotify die für sie interessanten Unterhaltungsangebote finden und ihre Abonnements aufrechterhalten.

Es ist kein neues Phänomen, dass Informationen die Produktevolution und Innovationen antreiben. Diese Lerneffekte werden jedoch von selbstlernenden Systemen automatisiert und erheblich beschleunigt, weshalb der Begriff der „Data Network Effects" erst mit der Verbreitung Künstlicher Intelligenz in den letzten Jahren aufkam. Es stellt sich auch die Frage, ob es notwendig ist, für Datennetzwerkeffekte einen eigenen Begriff zu definieren oder ob dieser Effekt nicht für andere Produktionsfaktoren wie etwa Arbeitskraft oder Kapital ebenso gilt. Jegliche Investitionen in ein gutes Produkt verbessern es mit hoher Wahrscheinlichkeit und machen es für mehr Nutzer attraktiv. Zudem handelt es sich bei selbstlernenden Systemen nicht um „Netzwerke" mit direkten Kontakten der Knoten im eigentlichen Sinne. Insofern wären „Datenskaleneffekte" die passendere Beschreibung.[1215]

Zur Beschreibung des gleichen Phänomens wird auch das etablierte Konzept der (positiven) Rückkopplungseffekte herangezogen.[1216] Stark vereinfacht erklären Rückkopplungseffekte die Beziehung von Input und Output. Daten sind – wie bereits erläutert[1217] – gleichzeitig Inputs und Outputs selbstlernender Systeme: Das Feedback auf das Output wird dem System als Input wieder zugeführt. Daher bieten sich Rückkopplungseffekte oder Feedback-Effekte als Erklärungsansatz für Phänomene wie Lerneff-

---

1214 *Swinney*, How Data Is Creating Better Customer Experiences at Spotify, 19. September 2018; auch *Del Toro Barba*, ORDO 68, S. 217–248, 235ff (2017); *Bourreau/de Streel*, Digital Conglomerates and EU Competition Policy, S. 16.

1215 Trotzdem soll es für diese Arbeit bei dem verbreiteten Begriff der Datennetzwerkeffekte bleiben, um an aktuelle Literatur und politische Debatten anzuknüpfen.

1216 *Bundesregierung*, Strategie Künstliche Intelligenz, S. 42; *Bajari et al.*, The Impact of Big Data on Firm Performance, NBER Working Paper No. 24334, S. 3; *Dewenter/Lüth*, Datenhandel und Plattformen, S. 15, 17; *Mayer-Schönberger/Ramge,* Die Datenklauer, Wirtschaftswoche Nr. 50, 1. Dezember 2017, S. 12; ähnlich: „Selbstverstärkungseffekte", so BKartA, Beschluss vom 6. Februar 2019, B6–22/16 Rn. 496 – *Facebook*.

1217 Siehe Kapitel 4 C.II. Datenmärkte: Input, Output, Currency, S. 255.

fekte an. Beispielsweise kann ein persönlicher Assistent wie Alexa oder Bixby daraus lernen, dass der Nutzer nach einem Missverständnis seine Eingabe wiederholt und spezifiziert. Die Rückkopplung führt zu einem Anstieg der Qualität oder Akkuratheit, also einem besseren Produkt, das dann zusätzliche Nutzer anzieht, die mehr Daten als Eingabematerial liefern. Dies setzt einen selbstverstärkenden Kreislauf in Gang. Dieser wird als „Feedback Loop" bezeichnet.[1218] Insofern deckt sich die Erklärung mit der des Datennetzwerkeffekts. Im Jahr 1999 beschrieben Shapiro und Varian den Positive Feedback Effect so: „Die Starken werden stärker und die Schwachen werden schwächer."[1219] Im Extremfall führe dies zu „Winner Take All"-Märkten.[1220] In Märkten der Informationsökonomie gewinnen die Unternehmen mit Technologien, die von positiven Rückkopplungseffekten angetrieben würden.[1221] Dabei sei ein einheitliches Muster, nämlich eine S-förmige Kurve, zu beobachten: Nach Einführung verbreite sich die Technologie kaum, dann steil während des „Takeoffs", wenn die positiven Feedbackeffekte hinzutreten, bis schließlich der Anstieg abnimmt und Sättigung eintritt.[1222] Shapiro und Varian hatten ein anderes Verständnis von Feedback, als hier für datenbezogene Feedbackeffekte angenommen wird: Vor dem zeitlichen Hintergrund der Veröffentlichung überrascht dies nicht; die als Beispiele aufgeführten, im Jahr 1999 innovativen Technologien wie das lineare Farbfernsehen, die CD oder das „Nintendo Entertainment System" übermittelten den Herstellern kein Feedback über die Nutzung. Feedback bezog sich lediglich auf die Rückkopplungseffekte, die positive direkte Netzwerkeffekte antreiben. Feedback in datenbezogenen Märkten bezieht sich auf das tatsächliche Feedback der Nutzer, die durch ihr Handeln dem Dienst zu erkennen geben, ob seine Ausgabe richtig oder falsch, beziehungsweise nützlich oder nutzlos, war. Der Dienst lernt an diesem impliziten Feedback, ob er eine Aufgabe bereits zufriedenstellend erfüllt hat oder ob weitere Lernschritte nötig sind. Viktor Mayer-Schönberger und Thomas Ramge importieren den Begriff der Feedbackeffekte in

1218 Siehe z. B. *BMWi*, Wettbewerbsrecht 4.0, S. 13; *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 31; *De Streel et al.*, CERRE White Paper 2019–2024, Digital, S. 8; *Farboodi et al.*, Big Data and Firm Dynamics, 14. Januar 2019, S. 2; *Furman et al.*, Unlocking Digital Competition, Rn. 1.73.

1219 *Shapiro/Varian*, Information Rules, 1999, S. 175.

1220 *Shapiro/Varian*, Information Rules, 1999, S. 177.

1221 *Shapiro/Varian*, Information Rules, 1999, S. 177 mit Darstellung 7.1.

1222 *Shapiro/Varian*, Information Rules, 1999, S. 178 mit Darstellung 7.2; dieses Muster wurde in der Informationstechnologie bei der CD, dem Farbfernsehen, Videospielplattformen, der E-Mail und dem Internet beobachtet.

die deutsche Sprache und stellen ihn neben Skalen- und Netzwerkeffekten vor.[1223] Feedbackeffekte würden nur eintreten, wenn Systeme Feedbackdaten zum Lernen nutzten.[1224] Sie berufen sich auf Prüfer und Schottmüller, die dieses Phänomen als datengetriebene indirekte Netzwerkeffekte (data-driven indirect network effects) bezeichnen.[1225] Nach Prüfer und Schottmüller führe die erhöhte Nachfrage nach einem Dienst zu niedrigeren Innovationsgrenzkosten für ebendiesen Dienst.

Die Abgrenzung zu (indirekten) Netzwerkeffekten kann unscharf sein. Grundsätzlich verknüpfen Netzwerkeffekte die Nutzerzahl mit der Nützlichkeit, während Datennetzwerkeffekte die Interaktionen mit der Qualität verknüpfen. Vereinfacht bedeutet ein Netzwerkeffekt: Je mehr Menschen einen Dienst oder eine Infrastruktur nutzen, desto nützlicher ist er für sie, weil sie in dem Netzwerk mit mehr Personen interagieren können. Datennetzwerke bedeuten, dass ein Dienst nützlicher wird, je mehr Personen ihn nutzen. Ein Telefonsystem oder ein soziales Netzwerk wird mit einer höheren Nutzerzahl nicht notwendig qualitativ besser, sondern nur nützlicher. Ein Empfehlungsdienst oder persönlicher Assistent auf Grundlage von selbstlernenden Systemen verbessert sich mit einer höheren Zahl von Dateneingaben und Nutzerfeedback, indem er passendere Vorschläge oder Lösungen anbietet.

Der Begriff der Datennetzwerkeffekte oder Data Network Effects sollte, um ihn klar von klassischen Netzwerkeffekten oder allgemeinen Rückkopplungseffekten abgrenzen zu können, eng verstanden werden. So sollte er nur automatisch lernende Systeme erfassen und sich nur auf die qualitative Produktevolution beziehen. Solche Produkte, die eine separate menschliche Auswertung und Sortierung der Daten erfordern, bevor eine Produktverbesserung eintritt, wären nicht erfasst. Ebenfalls ist die gestiegene Nützlichkeit in dieser Definition nur relevant, wenn sie sich aus der Qualitätssteigerung ergibt. Ein weiteres Verständnis würde nur bekannten Effekten einen neuen Namen und neue – vermutlich unverdiente – Aufmerksamkeit verschaffen.

1223 *Mayer-Schönberger/Ramge*, Das Digital, S. 189.

1224 *Mayer-Schönberger/Ramge*, Das Digital, S. 189.

1225 *Prüfer/Schottmüller*, Competing with Big Data, Tilburg Law School Legal Studies Research Paper Series No. 06/2017, S. 2, 6; dazu *Graef/Prüfer*, ESB 2018, Vol. 103, S. 298–301, 298; ähnlich auch *Stucke/Grunes*, Data-opolies, The University of Tennessee College of Law Research Paper #316, S. 6.

II. Hypothesen

Datennetzwerkeffekte würden somit folgendermaßen funktionieren: Die in ein selbstlernendes System eingespeisten Daten verbessern seine Funktionsweise und Treffgenauigkeit, was zu einer verstärkten Nutzung des Systems führt, die wiederum zu einem größeren Volumen von Input-Daten führt, die zur weiteren Systemevolution beitragen.

Inwiefern das Zusammenspiel von Eingabedaten, selbstlernenden Systemen und Nutzungsintensität tatsächlich so funktioniert, ist nicht verlässlich bewiesen. Umfassende, produktübergreifende Studien fehlen. Es dürfte insbesondere schwerfallen, weitere Einflüsse wie Reputation, Marketing, Netzwerkeffekte, die Gestaltung der Benutzeroberflächen und die jeweils zu erlernenden Informationen und Komplexität der Aufgaben von Datennetzwerkeffekten zu isolieren.[1226] Es wurden entweder nur einzelne Schritte bewiesen oder ein (vermeintlicher) Datennetzwerkeffekt an der Gesamtentwicklung eines Produktes – bevorzugt das eines GAFAM-Unternehmens – aufgezeigt.[1227] Schaefer und Sapi nehmen in einem vorläufigen Paper etwa auf der Grundlage von Yahoo-Suchdaten einen Datennetzwerkeffekt an.[1228] Die Verbesserung der Suchqualität aufgrund von Lerneffekten ist aber nicht alleiniger Bestandteil der hier angenommenen Definition von Datennetzwerkeffekten. Diese Verbesserung führt nicht dazu, dass abseits der Yahoo-Suche keine Erhebung von Suchdaten mehr möglich war und der Dienst alleiniger Marktführer wurde. Die Verfügbarkeit von Trainingsdaten ist nach dieser Untersuchung ein Wettbewerbsvorteil und kann eine Quelle von Marktmacht sein, aber sie leitet nicht zwangsläufig Datennetzwerkeffekte ein und garantiert eine marktmächtige Stellung. Das Begriffsverständnis variiert erheblich.

Der Datennetzwerkeffekt, wie er heute bezeichnet wird, umschreibt die Angst vor einer Monopolisierung der Innovationskraft auf solche Unternehmen, die die größte und qualitativ beste Masse an Fertigungs- und Feedbackdaten kontrollieren.[1229] Es wird befürchtet, dass diejenigen, die über Sensor-, Wartungs- oder Nutzungsdaten verfügen, ihre Produkte und

---

1226 Ein Versuch der Isolierung des Einflusses von Datenvolumen auf die Anwendungsqualität: *Schaefer/Sapi/Lorincz*, The Effect of Big Data on Recommendation Quality, S. 2.

1227 Vgl. *Mayer-Schönberger/Ramge*, Das Digital, S. 190; *SPD*, Digitaler Fortschritt durch Daten-für-alle-Gesetz, Diskussionspapier, S. 3, 6.

1228 *Schaefer/Sapi*, Data Network Effects: The Example of Internet Search.

1229 *Mayer-Schönberger/Ramge*, Are the Most Innovative Companies Just the Ones With the Most Data?, HBR, 7. Februar 2018.

Prozesse optimieren können, sodass sie immer mehr genutzt werden, immer mehr Daten sammeln und schließlich sektorspezifische Innovationsmonopole entstehen.

Neben dieser Sichtweise, die eher der Regulierung zuzuordnen ist, werden Datennetzwerkeffekte auch als vielversprechender Baustein einer Geschäftsstrategie von Startups betrachtet.[1230] Die Demokratisierung von Big-Data-Werkzeugen biete Markteintretern die Möglichkeit, mithilfe des Datennetzwerkeffekts in kurzer Zeit Dienste mit akkuraten Ergebnissen anzubieten und zu skalieren. Dies gelinge Startups besser als etablierten Marktteilnehmern.[1231] Das Ziel wäre, ein natürliches Monopol aufzubauen, das auf der uneinholbaren qualitativen Überlegenheit der eigenen Dienste basiert. So wären höhere Gewinnmargen möglich und das Geschäftsmodell nachhaltig wertvoll.[1232] Ein einzigartiges Datenset, das automatisch mit zunehmender Nutzung wachse, sei einer der am besten zu verteidigenden „Moats" (Burggraben, quasi Wettbewerbsvorteil).[1233]

Die den Datennetzwerkeffekten oder Feedback Effects für beide Sichtweisen zugrundeliegende Hypothese ist die Lernfähigkeit einer Anwendung. Dass selbstlernende Systeme auf der Grundlage größerer und qualitativ hochwertigerer Eingabedaten akkuratere Antworten geben, wurde hinreichend dargelegt.[1234] In den meisten Fällen wird dabei ein abnehmender Grenznutzen festgestellt: Die Lernerfolge steigen nicht linear

---

1230 So *Turck*, The Power of Data Network Effects, 4. Januar 2016: „competitive moat"; sowie *Staun-Olsen*, The Value Isn't Your Algorithm, It's Your Data, Creandum, 3. November 2016.

1231 *Turck*, The Power of Data Network Effects, 4. Januar 2016; dazu *Beim*, Learning Effects, Like Network Effects, Can Create Runaway Leaders, 22. September 2017.

1232 So *Beim*, Learning Effects, Like Network Effects, Can Create Runaway Leaders; *Staun-Olsen*, The Value Isn't Your Algorithm, It's Your Data, Creandum, 3. November 2016.

1233 *Staun-Olsen*, Should you build machine learning in-house? Probably not, Creandum, 26. Oktober 2017.

1234 *Banko/Brill*, Scaling to Very Very Large Corpora for Natural Language Disambiguation, S. 7; *Brill,* Processing Natural Language without Natural Language Processing, CICLing 2003, S. 360–369; *Dewenter/Lüth*, Big Data, Datenschutz und Wettbewerb, S. 13; *Junqué de Fortuny/Martens/Provost*, Predictive Modeling with Big Data: Is Bigger Really Better?, Big Data 2013, S. 215–226, 223; *Khosla/Jayadevaprakash/Yao/Fei-Fei*, Stanford Dogs Dataset, http://vision.stanford.edu/aditya86/ImageNetDogs; *Sun/Shrivastava/Singh/Gupta*, Revisiting Unreasonable Effectiveness of Data in Deep Learning Era, S. 8.

mit den zugeführten Datenvolumina.[1235] Folglich erzielen Big-Data-Anwendungen ab einer bestimmten Menge an Daten keine weiteren Qualitätszuwächse mehr.[1236] Insofern gilt es schon an diesem Punkt zu differenzieren, ob überhaupt ein unbegrenzter selbstverstärkender Kreislauf in Gang gesetzt werden kann, wenn sich die Lernfähigkeit ohnehin einer natürlichen Grenze nähert. Grundsätzlich ist jedoch anzunehmen, dass die Verfügbarkeit von Trainingsdaten eine Grundvoraussetzung ist, um einen selbstverstärkenden Kreislauf in Gang zu setzen. Ein größeres und vielfältigeres Datenset ermöglicht ein besseres Training, das bei Vorliegen aller anderen Voraussetzungen das überlegene Produkt hervorbringt. Ob der Einfluss von Datenherrschaft stärker wirkt als der einer Herrschaft über andere jeweils nötige Innovationsvoraussetzungen, ist damit nicht geklärt. Vordergründig ist hier zu untersuchen, ob Datennetzwerkeffekte den Innovationswettbewerb beeinflussen. Ob sie sich in ihrer Wirkung von traditionellen körperlichen Rohstoffen unterscheiden, betrifft den wettbewerbspolitischen Umgang mit einem möglichen Datennetzwerkeffekt.

Ein gut trainiertes selbstlernendes System, das genauere Vorhersagen und Lösungsmöglichkeiten anbietet, dürfte dem durchschnittlichen Nutzerinteresse entsprechen. An diesem Punkt des Kreislaufes spielen für den Nutzer allerdings weitere Kriterien eine Rolle, um ein System zu einem attraktiven Produkt machen. Hierzu zählen die Bedienbarkeit, die Benutzeroberfläche, Datenschutz, Datensicherheit und Verlässlichkeit. Ohne eine Veränderung dieser Kriterien ist jedoch zu erwarten, dass ein „intelligenteres" selbstlernendes System gegenüber dem weniger trainierten vom Nutzer bevorzugt wird. Auf dynamischen Märkten wird der Wettbewerb eher über Produktinnovationen als über den Preis geführt.[1237]

Dem Feedback- oder Datennetzwerkeffekt werden zwei Auswirkungen zugeschrieben: Einerseits verstärke er die Marktmacht derjenigen Unternehmen, die ihren Nutzern bereits selbstlernende Systeme zur Nutzung anbieten und schwäche den Wettbewerb um sie herum. Andererseits fixiere er Daten als Rohstoff künftiger Innovationen in einem Strudel, auf den neue Marktteilnehmer keinen Zugriff haben. Wenn der Erfolg eines

---

1235 *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 8f mwN.

1236 *Dewenter/Lüth*, Big Data aus wettbewerblicher Sicht, Wirtschaftsdienst 2016, S. 648–654, 653; dagegen: *Graef/Prüfer*, ESB 2018, Vol. 103, S. 298–301, 299.

1237 *Monopolkommission*, Sondergutachten 68, S. 31f, Rn. 32; *J. Weber*, Zugang zu den Softwarekomponenten der Suchmaschine Google, S. 100; siehe Kapitel 3 A.II.1. Dynamischer Wettbewerb, S. 113.

Produktes auf seiner Lernfähigkeit beruht und nur die größten Firmen genug Daten zur Gewährleistung der Lernfähigkeit haben, verlieren Innovationen ihre Funktion zur Belebung des Marktes. Weil datengetriebene Märkte zu einem Tipping neigten und zwangsläufig zu einer hohen Anbieterkonzentration führten, seien sie auf dem „inverted-U"-Spektrum von Aghion weit rechts zu verorten und die Innovationsanreize auf diesen Märkten als[1238] gering zu bewerten.[1239] Dies wirke sich auf den Innovationswettbewerb im Markt ebenso aus wie auf den Wettbewerb um den Markt.[1240] Mayer-Schönberger und Ramge formulieren die erste Auswirkung wie folgt: Angesichts der Datenmengen, mit denen „Superstars" ihre Künstliche Intelligenz trainieren können, bräuchten Newcomer kaum zu hoffen, „den Platzhirschen ernsthafte Konkurrenz zu machen – ihre Produkte lernen zu langsam dazu".[1241] Die Vormachtstellung könnte sich dank der Kostenvorteile aus Datennetzwerkeffekten leicht auf weitere Märkte übertragen lassen.[1242] Dem stimmt ein SPD-Positionspapier zu einem Daten-für-alle-Gesetz zu.[1243] Dank der Selbstverstärkungseffekte entstehe ein wesentlicher Vorteil gegenüber potentiellen Wettbewerbern.[1244]

Ein Untersuchungsausschuss des House of Lords im Vereinigten Königreich benennt ebenfalls „Data Network Effects" und das Ergebnis einer „data dominance" (Datenmacht).[1245] Im Extremfall könnte dies je nach Definition des betrachteten Marktes für eine Anwendung zu Monopolstellungen führen.[1246]

Der „KI-Multiplikatoreneffekt" funktioniere insbesondere in einem regulatorischen Umfeld, das die Erfassung von Daten grundsätzlich befür-

---

1238 Siehe Kapitel 2 C.I.3. Aghion und das umgekehrte U-Modell, S. 68.
1239 *Graef/Prüfer*, ESB 2018, Vol. 103, S. 298–301, 298.
1240 *Graef/Prüfer*, ESB 2018, Vol. 103, S. 298–301, 298.
1241 *Mayer-Schönberger/Ramge*, Das Digital, S. 193f.
1242 *Graef/Prüfer*, ESB 2018, Vol. 103, S. 298–301, 299.
1243 *SPD*, Digitaler Fortschritt durch Daten-für-alle-Gesetz, Diskussionspapier, S. 3, 6.
1244 Vgl. *Graef/Prüfer*, ESB 2018, Vol. 103, S. 298–301, 298; *J. Weber*, Zugang zu den Softwarekomponenten der Suchmaschine Google, S. 69f; ähnlich: *Monopolkommission*, Sondergutachten 68, S. 73, Rn. 161, 163; *OECD*, Data-Driven Innovation, S. 184.
1245 *UK Select Committee on Artificial Intelligence*, AI in the UK, S. 29, 44f, Fn. 30; mit Verweis auf *Hall/Pesenti*, Growing the artificial intelligence industry in the UK, Review, S. 45; *M. Lynch*, Written Evidence (AIC0005), Nr. 7.
1246 *Graef/Prüfer*, ESB 2018, Vol. 103, S. 298–301, 298; *Shelanski*, UPenn Law Review Vol. 161, S. 1663–1705, 1681, 1684 (2013).

wortet. Als Beispiele werden die USA und China genannt.[1247] Dort ansässige Unternehmen hätten einen Vorsprung und bessere Chancen, den Selbstverstärkungseffekt zunächst in diesem Umfeld in Gang zu setzen und später auf andere Jurisdiktionen auszudehnen. In einem extremen Szenario bedeute dies, dass wenige, sehr große Unternehmen aus besonders liberalen Regulierungsumfeldern als uneinholbare Gewinner hervorgehen werden.

Dies würde – neben der Monopolisierung des Rohstoffes Daten – die innovationshemmenden Auswirkungen des Feedback-Effekts auslösen. Datenarme Innovatoren würden weniger Anreize zur Investition in Forschung und Entwicklung sehen und die Innovationsvielfalt würde zurückgehen. Mayer-Schönberger und Ramge formulieren diese Hypothese wie folgt: „Dienstleistungen, die auf mit Feedbackdaten gefütterten KI-Systemen basieren, kaufen Innovationen zu Kosten, die in dem Maß sinken, wie die Menge der Daten wächst".[1248] Unternehmen ohne Zugang zu diesen Daten fehle ein Baustein der Produktentwicklung und damit die Voraussetzung für einen Eintritt in den Produktwettbewerb.

Auch Cockburn, Henderson und Stern nehmen an, dass es in begrenzten Anwendungsbereichen möglich ist, dass ein Unternehmen einen wesentlichen und dauerhaften Innovationsvorteil aus seinem Datenzugang zieht. Dieser könne unabhängig von traditionellen nachfrageseitigen Netzwerkeffekten sein.[1249] Die Aussicht auf diesen Vorteil motiviere zunächst zu besonders intensivem Wettbewerb in dem jeweiligen Sektor, bevor dann dauerhafte, wettbewerbspolitisch bedeutende Marktzutrittshürden errichtet würden. Dieses Verhalten könnte im Extremfall zu einer „Balkanisierung"[1250] von Daten innerhalb der jeweiligen Sektoren führen. In der Konsequenz wäre die Innovationstätigkeit des Sektors gemindert, aber auch Spillover-Effekte auf die abstrakte Entwicklung selbstlernender Systeme (z. B. General Deep Learning) und Anwendungen in anderen Sektoren würden unterbunden. Die Autoren erkennen an, dass diese Annahmen höchst spekulativ sind.

In Sektoren, die von datengetriebenen Geschäftsmodellen bestimmt werden, findet ein von hoher Geschwindigkeit geprägter Wettlauf um

---

1247 *Knop*, Ein Weckruf für die Zukunft Europas, FAZ, 24. April 2018.

1248 *Mayer-Schönberger/Ramge*, Das Digital, S. 193.

1249 Zum Folgenden: *Cockburn/Henderson/Stern*, The Impact of Artificial Intelligence on Innovation, NBER Working Paper No. 24449, März 2018, S. 15.

1250 Geopolitischer Begriff für den Prozess der Fragmentierung oder Aufteilung eines Staates in kleinere Staaten, die feindlich oder nicht kooperativ miteinander sind.

Innovationen statt.[1251] Die hohe Geschwindigkeit der Forschung und Entwicklung wirke sich auf die Geschwindigkeit der Selbstverstärkungseffekte aus, weshalb eine Monopolisierung besonders schnell erfolgen könne und Behörden weniger Möglichkeiten und Anknüpfungspunkte zum Eingreifen hätten. Die Konkurrenz würde den Anschluss verlieren, bevor sie durch Aufstockung ihrer Arbeitskräfte, Optimierung der Algorithmen oder den Ausbau der Hardware Schritte zum Aufholen unternehmen könnte.[1252]

Ein Gutachten des ICLE steht der Annahme von Datennetzwerkeffekten ablehnend gegenüber: Nach Manne et al. teilen Daten gerade nicht die selbstverstärkenden Eigenschaften der Netzwerkeffekte. Allerdings sei anzunehmen, dass die Erfassung eines Mindestdatenvolumens eine Voraussetzung für die Teilnahme am Wettbewerb in datengetriebenen Märkten ist.[1253] Dieser „Learning by Doing"-Vorsprung stagniere aber schnell („diminishing returns"[1254]). Vielmehr haben datenreiche Unternehmen dank überlegener Software und Geschäftsmodelle viele Daten sammeln können, als dass sie ihre Überlegenheit nur den Datensammlungen verdanken. Obwohl die Idee eines Datennetzwerkeffekts zunächst einleuchte, stützten bisher keine ökonomischen Modelle diese These.[1255] Insbesondere sei nicht gelungen, darzulegen, dass „Learning by Doing" eine andere Rolle spiele als in anderen Wirtschaftssektoren. Die Betrachtung von Datennetzwerkeffekten ignoriere zudem den Aspekt der Produktqualität.[1256] Das Datenvolumen allein entscheide nach dieser These über den wettbewerblichen Erfolg. Es gäbe nach Manne et al. eine unbegrenzte Zahl von Möglichkeiten, ein überlegenes Produkt anzubieten, ohne auf das größere Datenvolumen zuzugreifen.[1257] Die Annahme von Datennetzwerkeffekten basiere damit auf einem schwachen theoretischen Modell;[1258] eine lineare oder gar super-lineare Beziehung zwischen Datenreichtum und finanziel-

---

1251 Vgl. *Dreher*, ZWeR 2009, S. 149–175, 151f.

1252 *Surblytė*, Data as a Digital Resource, MPI for Innovation and Competition Research Paper No. 16–12, S. 14.

1253 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 9.

1254 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 9.

1255 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 9; *Bajari et al.*, The Impact of Big Data on Firm Performance, NBER Working Paper No. 24334, S. 5; dagegen: *Graef/Prüfer*, ESB 2018, Vol. 103, S. 298–301.

1256 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 11.

1257 Z. B. Preis, Datenschutz, Menge und Aufdringlichkeit der platzierten Werbung, *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 12.

1258 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 12.

len Gewinnen (z. B. Werbeeinnahmen) konnte bisher nicht nachgewiesen werden. Vielmehr handele es sich in den meisten Fällen, in denen von Data Network Effects gesprochen wird, um die Kombination von Learning By Doing und erworbenen unternehmensweiten Fähigkeiten in datenintensiven Märkten.[1259]

Eine Studie anhand der Daten von Amazons Bedarfsvorhersagesystem konnte keine ‚Data Feedback Loops' feststellen, sondern stattdessen die statistischen Theorien zu Vorhersagefehlern bestätigen.[1260] In anderen Arbeiten wird zwar angenommen, dass „Feedback Loops" bestehen, aber ihre Auswirkungen bei niedrigen Kosten für die Datenerfassung sehr gering seien.[1261]

III. Datengetriebene Lerneffekte

Losgelöst von selbstlernenden Systemen wurden Lerneffekte generell bei datengetriebenen Anwendungen beobachtet. Argenton und Prüfer machen eine von ihnen beobachtete Lernkurve des Fortschritts von Suchmaschinen zur Grundlage eines Regulierungsvorschlags.[1262] Das Prinzip der Lerneffekte wurde von Arrow vorgestellt.[1263] Ein solcher Lerneffekt sei ein dynamischer Größenvorteil, weil die in einem Wirtschaftszweig tätigen Unternehmen Erfahrungswerte in den Produktionsprozess und die Produktentwicklung einfließen lassen können.[1264] Ein Informationsvorsprung wird in einen Wissensvorsprung übersetzt. Dieser Wissensvorsprung muss in einen Innovationsvorsprung umgesetzt werden, der einen direkten Nutzervorteil bedeutet und damit den Nutzervorsprung gegenüber anderen

1259 „Firmwide capabilities", *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 26.

1260 *Bajari et al.*, The Impact of Big Data on Firm Performance, NBER Working Paper No. 24334, S. 5: "We note that our results are inconsistent with a naive "data feedback loop," where the addition of new products always results in better models".

1261 *Bourreau/de Streel/Graef*, Big Data and Competition Policy, S. 36.

1262 *Argenton/Prüfer*, Journal of Competition Law and Economics, Vol. 8, No. 1, S. 73–105, 80ff (2012); zustimmend: *Stucke*, Georgetown Law Technology Review, Vol. 2.2, S. 275–324, 283 (2018).

1263 Später aufgegriffen in: *Boston Consulting Group,* Perspectives on Experience, 1972.

1264 *BMWi*, Wettbewerbspolitik für den Cyberspace, Rn. 45.

Unternehmen ausbaut. Insoweit deckt sich die Beschreibung mit der des Datennetzwerkeffekts.[1265]

Hal Varian, Googles „Chefökonom", stimmt zwar zu, dass Unternehmen mit mehr Nutzern mehr Daten sammeln können, mit denen sie ihre Produkte noch weiter verbessern können.[1266] Aus Daten zu lernen, sei kein Nebeneffekt der Skalierung, sondern erfordere konstante Investitionen und der Wert der Daten sei gefährdet durch eigene und fremde technologische Innovationen. Allerdings wäre dies keine potentiell antikompetitive Ausbeutung von Rückkopplungseffekten[1267] und auch kein (Daten-)Netzwerkeffekt, sondern vielmehr „Learning by Doing" nach Arrow[1268]. Der Unterschied zwischen „Learning by Doing" und Datennetzwerkeffekten liegt für Varian im „Doing": Wenn ein Unternehmen große Datenmassen hat, aber sie nicht anwendet, wird kein Wert geschaffen.[1269] Viel problematischer als ein Fehlen von Daten sei ein Fehlen von Expertise, weil dann das Erkennen von Personalbedürfnissen schwierig sei.[1270]

Es stimmt, dass Lerneffekte nichts Neues sind.[1271] Mit der Digitalisierung wurden sie jedoch mit neuer Energie aufgeladen und mit der Verbreitung selbstlernender Systeme werden sie aus dem menschlichen Intellekt ausgelagert. Offline werden Lerneffekte durch intelligente Menschen bewegt: Ein Produkt wird angepasst, wenn ein Weg gefunden wird, es effizienter zu nutzen. Nach Arrow sind Lernprozesse, die im Zusammenhang mit der Entwicklung verbesserter Fertigkeiten durch Produktionserfahrungen stehen, in der Regel weniger ausgeprägt als Netzwerkeffekte.[1272] Zudem ist – anders als bei Netzwerkeffekten – regelmäßig ein abnehmender

---

1265 Z. B. *Stucke/Grunes*, Big Data and Competition Policy, S. 170; *Rubinfeld/M. Gal*, Access Barriers to Big Data, S. 18: „a positive feedback loop can be created".

1266 *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 15.

1267 *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 15; *Arrow*, The Economic Implications of Learning by Doing, The Review of Economic Studies 1962, Vol. 29, No. 3, S. 155–173.

1268 *Arrow*, The Economic Implications of Learning by Doing, The Review of Economic Studies 1962, Vol. 29, No. 3, S. 155–173.

1269 *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 15 mwN.

1270 *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 15.

1271 Siehe Kapitel 2 D.III. Informationen, S. 89; *Graef/Prüfer*, ESB 2018, Vol. 103, S. 298–301, 299.

1272 So *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 13.

Grenznutzen zu beobachten.[1273] Learning by Doing sei damit für sichtbare Lerneffekte in den frühen Stadien der Prozessoptimierung verantwortlich, dieser Verlauf breche aber nach einiger Zeit ein und Knowledge Spillovers erlauben anderen Marktteilnehmern ein Anknüpfen an den Lernprozess.[1274] Ein „Winner Take All"-Markt sei daher keine Konsequenz von „Learning by Doing".[1275] Ebenfalls solle Learning by Doing keine oder geringere Eintrittsbarrieren konstituieren als Netzwerkeffekte, da sich für etablierte und neue Marktteilnehmer die gleichen Kosten bei dem Beschreiten der Lernkurve ergeben.[1276] Arrows Definition ist sechs Jahrzehnte alt. Die Besonderheit datengetriebener Geschäftsmodelle sind interne Feedbackeffekte, die extern durch Nutzererfahrung ermöglicht werden, was nicht der industriellen Realität im Jahr 1962 entspricht. Dazu kommt, dass ein Unternehmen zwar die erfahrenen Entwickler (z. B. Softwareentwickler oder Ingenieure) von Wettbewerbern abwerben kann, diesen aber das digitalisierte Know-how in Form von Nutzerdaten exklusiv bleibt.[1277] Datennetzwerkeffekte können somit nicht Learning by Doing sein, wenn die Definition nicht nachträglich anpasst wird; vielmehr sind sie Learning by Using oder Learning by Interacting. Manne et al. stimmen Varian insofern zu, als dass sie eher ein Learning by Doing als Datennetzwerkeffekte annehmen.[1278]

Mitomo erkennt die Wirkung von Datennetzwerkeffekten in Abgrenzung zu bloßen Lerneffekten an und unterscheidet zwischen Nutzerdaten und Up-to-Date-Daten, wobei Nutzerdaten am ehesten nach dem Schema klassischer Netzwerkeffekte zu untersuchen seien und problemlos festgestellt werden könnten.[1279] Bei Up-to-Date-Daten, die von Sensoren oder unspezifizierten Nutzern gesammelt würden, hänge der Wettbewerbsvor-

1273 *Graef/Prüfer*, ESB 2018, Vol. 103, S. 298–301, 299; *Duch-Brown/Martens/Müller-Langer*, The economics of ownership, access and trade in digital data, S. 10; *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 80; vgl. zum abnehmenden Grenznutzen bei Eingaben in Suchmaschinen: Europäische Kommission, Entscheidung vom 27. Juni 2017, AT.39740 Rn. 289 – *Google Search (Shopping)*.

1274 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 13.

1275 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 13 mwN.

1276 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 13f.

1277 *Graef/Prüfer*, ESB 2018, Vol. 103, S. 298–301, 299.

1278 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 13, 15.

1279 *Mitomo*, Data Network Effects: Implications for Data Business, 28th European Regional Conference of the International Telecommunications Society (ITS) 2017, S. 7.

teil eher von den Abhängigkeiten in den mehrseitigen Märkten ab.[1280] Eine Unterscheidung erscheint zwar sinnvoll, aber leidet in dieser Form wohl daran, dass es bei Datennetzwerkeffekten weniger auf die Zahl der registrierten Nutzer, als auf die Nutzung, also die tatsächliche Interaktion, ankommt: Für ein Unternehmen der Industrie 4.0 wird eine Anwendung nur einmal implementiert, aber es werden während der gesamten Produktionszeit Daten geliefert. Es mag zwar weniger Nutzer geben, aber dafür möglicherweise trotzdem mehr Interaktionen mit der Anwendung und weniger Anreize, falsche oder doppelte Daten anzugeben.

Als dynamische Größenvorteile sind Lerneffekte oder Learning by Doing legitime Erklärungsansätze für das gleiche Phänomen, das Feedback-Effekte oder Datennetzwerkeffekte erklären sollen. In datengetriebenen Geschäftsmodellen mit mehrseitigen Märkten, kostenfrei angebotenen Diensten, Multi-Homing und nicht immer messbaren Nutzerzahlen kann es schwerfallen, die Lerneffekte isoliert nachzuweisen. Insofern erscheint es sinnvoll, einen alternativen Begriff, nämlich Datennetzwerkeffekte, zu nutzen, der diesen Anspruch nicht erhebt und sich auf die qualitative Verbesserung von selbstlernenden Systemen konzentriert.

IV. Verhältnis zu Netzwerk- und Skaleneffekten

Wie bereits angesprochen wurde, wird der Nachweis von Datennetzwerkeffekten von ihrer fehlenden Isolierbarkeit erschwert. Sie treten in der Regel gemeinsam mit Netzwerk- und Skaleneffekten auf. Systemen Künstlicher Intelligenz sind etwa ganz allgemein Economies of Scale, also Größenvorteile, inhärent, weil die parallele Betrachtung mehrerer Datensets schneller, günstiger und aufschlussreicher ist als die sukzessive Analyse.[1281]

1280 *Mitomo*, Data Network Effects: Implications for Data Business, 28th European Regional Conference of the International Telecommunications Society (ITS) 2017, S. 7.

1281 *Duch-Brown/Martens/Müller-Langer*, The economics of ownership, access and trade in digital data, S. 9; ähnlich *Bourreau/de Streel*, Digital Conglomerates and EU Competition Policy, S. 10.

1. Netzwerkeffekte

Als vermeintliches „ökonomisches perpetuum mobile“[1282] sind Netzwerkeffekte seit den achtziger Jahren ein ständiger Bestandteil der Untersuchung digital geprägter Märkte.[1283] Katz und Shapiro bezeichneten 1985 das bei Informations- und Kommunikationstechnologien beobachtete Phänomen, ohne es präzise zu definieren. Positive Netzwerkeffekte entstehen, wenn der Nutzen, den ein Konsument durch den Gebrauch eines Gutes zieht, mit der Zahl weiterer Konsumenten steigt.[1284] Die aktuelle Attraktivität eines Gutes hängt von seiner Verkaufshistorie ab und Nachfrager machen ihre Entscheidung von dem zukünftigen Erfolg des Produktes abhängig.[1285] Es wird zwischen direkten und indirekten Netzwerkeffekten unterschieden. Bei direkten Netzwerkeffekten hängt der Nutzen des Dienstes von der Zahl der anderen Nutzer ab. Das klassische Beispiel ist ein Telefonnetz.[1286]

Indirekte Netzwerkeffekte sind zu beobachten, wenn mit der Zahl der Nutzer eines Gutes oder Netzes auch das Angebot an komplementären Gütern steigt. Je größer die Zahl der Nutzer ist, desto eher lohnt sich das Angebot komplementärer Dienste oder Anwendungen. Je mehr Dienste angeboten werden, desto größer ist wiederum der Nutzen aus dem Anschluss an ein solches Netz.[1287] Der zusätzliche Nutzen, der durch den Hinzutritt eines weiteren Konsumenten entsteht, ist indirekter Art. Positive indirekte Netzwerkeffekte sind das zentrale Element von Plattformen wie App Stores oder Anzeigenblättern: Eine Plattform mit vielen Nutzern zieht Entwickler an, die Dienste oder Angebote für die Plattform bereit-

---

1282 *Pohlmeier*, Netzwerkeffekte und Kartellrecht, S. 19.

1283 Ausgelöst von U. S. v. International Business Machines Corp., 687 F- 2d 591 (2nd Circ 1982), so *Pohlmeier*, Netzwerkeffekte und Kartellrecht, S. 19; zuvor angedeutet in (chronologisch) *Squire*, The Bell Journal of Economics and Management Science, Vol. 4 No. 2, S. 515–525 (1973); *Littlechild*, The Bell Journal of Economics, Vol. 6, No. 2, S. 661–670 (1975); *Rohlfs*, The Bell Journal of Economics and Management Science, Vol. 5 No. 1, S. 16–37 (1974).

1284 *Katz/Shapiro*, The American Economic Review, Vol. 75, No. 3, S. 424–440 (1985), “network externalities” statt Netzwerkeffekten.

1285 *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 23; *Katz/Shapiro*, The American Economic Review, Vol. 75, No. 3, S. 424–440 (1985); *Pohlmeier*, Netzwerkeffekte und Kartellrecht, S. 32.

1286 Ausführlich illustriert: *Pohlmeier*, Netzwerkeffekte und Kartellrecht, S. 29.

1287 *Katz/Shapiro*, The American Economic Review, Vol. 75, No. 3, S. 424–440 (1985); *Pohlmeier*, Netzwerkeffekte und Kartellrecht, S. 30f; *C. Tucker*, Antitrust, Vol. 32, No. 2, S. 72–79, 72 (2018).

stellen. Nutzer fühlen sich wiederum von Plattformen mit einer hohen Zahl von Diensten angezogen. Bei indirekten Netzwerkeffekten kommt es zu positiven Rückkopplungsprozessen[1288], da sich die Nutzerzahl und die Zahl der komplementären Angeboten gegenseitig „hochschaukeln", wodurch endogen technologische Monopolisierungstendenzen entstehen. Netzwerkeffekte führen zu Größenvorteilen auf der Nachfrageseite (demand-side economies of scale), weil der Wert des Produktes mit steigender Zahl der Konsumenten zunimmt.[1289] In der Konsequenz können diese wiederum angebotsseitige Größenvorteile bewirken.[1290] Gerade in komplexeren Abhängigkeitsverhältnissen verstärken Mitläufereffekte als Rückkopplungsschleifen Netzwerkeffekte.[1291]

Netzwerkeffekte spielen insbesondere für personenbezogene Daten, die in der Regel auf Primärmärkten im Rahmen mehrseitiger Systeme erworben werden, eine Rolle und könnten mächtigen Marktteilnehmern Wettbewerbsvorteile verschaffen.[1292] Insofern ist die Überlegung zulässig, dass der Datenzugang betreffend personenbezogener Daten mangels eines starken Sekundärmarktes für Konzentrationen anfälliger ist.[1293]

2. Skaleneffekte

Allgemeine Größenvorteile werden als Skaleneffekte (Economies of Scale) bezeichnet. Der Begriff beschreibt die Abhängigkeit der Produktionsmenge von der Menge der einzusetzenden Produktionsfaktoren. Kostenseitige Größenvorteile liegen vor, wenn aufgrund hoher Fixkosten die Durchschnittskosten mit steigender Produktionsmenge sinken.[1294]

Die industriellen Revolutionen zeigten jeweils eindrucksvoll die Auswirkungen von Skaleneffekten. Software scheint das Paradebeispiel für Skaleneffekte zu sein: Es entstehen anfangs bei der Entwicklung der Software besonders hohe Kosten in Form von Arbeitsaufwand und Beschaffung der

---

1288 *Shapiro/Varian*, Information Rules, 1999, S. 175.
1289 *Shapiro/Varian*, Information Rules, 1999, S. 179.
1290 *Shapiro/Varian*, Information Rules, 1999, S. 182: „double whammy".
1291 *Dietrich,* Wettbewerb in Gegenwart von Netzwerkeffekten, S. 78; *Parker/Alstyne/Choudary*, Platform Revolution, S. 296.
1292 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 5.
1293 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 5.
1294 *Monopolkommission*, Sondergutachten 68, S. 84.

nötigen Hardware. Darauf folgen deutlich geringere Vertriebskosten.[1295] Die Grenzkosten digitaler Angebote gehen gegen Null. Diese Betrachtung lässt jedoch die ständige Notwendigkeit der Entwicklung von Updates zur Schließung von Sicherheitslücken oder besseren Befriedigung der Nutzerbedürfnisse (z. B. Übersetzungen) außer Acht. Die Größenvorteile können auch bei Software niedriger sein, als zunächst anzunehmen ist. Informationstechnologien sind generell in der Einrichtung kostenintensiv, für Systeme der Industrie 4.0 kommen zum Beispiel die Fixkosten für die Installation von Sensoren hinzu.[1296] Sobald das System voll funktionsfähig ist, könne bei selbstlernenden Systemen entsprechend der Größenvorteile zu niedrigen Kosten der Algorithmus weiter verbessert werden.[1297] Es erscheint plausibel, dass große Datenmengen gegenüber kleineren Mengen vorteilhaft sind, also positive interne Skalenerträge vorliegen. Dieser Effekt dürfte aber mit steigendem Volumen der eingesetzten Daten abnehmen und im Einzelfall negativ werden.[1298] Gegenläufig zu den Skaleneffekten können größere Datenmassen in Einzelfällen steigende Kosten bedeuten, weil sie aufwendiger gekennzeichnet und bereinigt werden müssen.[1299] Für unterschiedliche selbstlernende Systeme ergeben sich unterschiedliche mindestoptimale Datenmengen, mit deren Erreichen das System effizient arbeitet. Je größer das mindestoptimale Datenvolumen ist, desto schwerer ist ein System zu realisieren. Für viele Anwendungsfälle aus der Internetwirtschaft gilt, dass „die Vorhersagekraft von Algorithmen auch bei sehr großen Datenmengen noch von einer Zunahme der Datenmenge profi-

---

1295 *Furman et al.*, Unlocking Digital Competition, Rn. 1.66; *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 14, 18; so auch *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 20 und Fn. 8 mit Verweis auf Facebook, wo auf einen Mitarbeiter 65.000 monatliche Nutzer entfallen. Dieser Vergleich hinkt, weil er nur eine Marktseite berücksichtigt und die Nutzungsintensität außer Acht lässt.

1296 *Duch-Brown/Martens/Müller-Langer*, The economics of ownership, access and trade in digital data, S. 32; *OECD*, Big Data, Background Note, DAF/COMP(2016)14, S. 11, Rn. 27: aufgeführte Beispiele sind Datenzentren, Server, Datenanalyse-Software, hohe Kosten für spezialisierte und begehrte Arbeitskräfte.

1297 *OECD*, Big Data, Background Note, DAF/COMP(2016)14, S. 11, Rn. 27: „Once the system is fully operational, the incremental data can 'train' and improve the algorithms at a low cost (thereby also the product or service quality)".

1298 *Dewenter/Lüth*, Big Data aus wettbewerblicher Sicht, Wirtschaftsdienst 2016, S. 648–654 (652).

1299 *Duch-Brown/Martens/Müller-Langer*, The economics of ownership, access and trade in digital data, S. 32.

tiert".[1300] Der bloße Zugang zu Daten genügt jedoch nicht, um interne Skaleneffekte auszulösen; vielmehr sind die Technologie, das technische Know-how und eine geeignete Geschäftsstrategie nötig.[1301]

3. Zusammenspiel mit Rückkopplungseffekten

Netz-, Skalen-, Lern- und Lock-In-Effekte lösen nicht nur eine Verbundwirkung aus, sondern stehen in einer Wechselwirkung zueinander und können sich in Form eines positiven Selbstverstärkungseffekts gegenseitig verstärken.[1302] Insbesondere bei Vorliegen starker Netzwerkeffekte für ein Produkt können sich die bestehenden Effekte potenzieren und eine bestehende marktbeherrschende Stellung verfestigen.[1303] Laut der Monopolkommission sind bei großen Unternehmen meist Skalen- und Verbundvorteile zu beobachten, weshalb ihnen Informationsvorteile durch die Digitalisierung tendenziell stärker zugutekommen.[1304] Datennetzwerkeffekte könnten auch erst nachträglich, nach der Etablierung selbstlernender Systeme in ein Angebot, hinzutreten und auf schon wirkende Netzwerkeffekte und eine bestehende breite Nutzerschaft aufbauen. Durch den Zugang zu dieser Nutzerschaft und damit zu einer größeren Menge an Daten können zielgerichtete oder bessere Angebote an die Nutzer gemacht werden, die diese wiederum an das Unternehmen binden und damit weiterhin Daten generieren.[1305] Netzwerke zeichnen sich generell durch eine höhere Lernfähigkeit aus, weil ihnen andere Wege der kollektiven Informations-

---

1300 *Dewenter/Lüth*, Big Data aus wettbewerblicher Sicht, Wirtschaftsdienst 2016, S. 648–654 (652); mit Beispiel: *Banko/Brill*, Scaling to Very Very Large Corpora for Natural Language Disambiguation, S. 1.

1301 *Dewenter/Lüth*, Big Data aus wettbewerblicher Sicht, Wirtschaftsdienst 2016, S. 648–654 (653) mwN.

1302 *Bourreau/de Streel/Graef*, Big Data and Competition Policy, S. 29; *BKartA*, Big Data und Wettbewerb, S. 8; *Dreher*, ZWeR 2009, 149 (154); *D. Evans/Schmalensee*, Innovation Policy and the Economy 2002, S. 1–49 (11); *Shapiro/Varian*, Information Rules, S. 173ff; *J. Weber*, Zugang zu Softwarekomponenten der Suchmaschine Google, S. 122; *Zimmerlich*, Marktmacht in dynamischen Märkten, S. 97.

1303 *Monopolkommission*, Sondergutachten 68, S. 44, Rn. 65; *BMWi*, Referentenentwurf 9. GWB-ÄndG, 1. Juli 2016, S. 51; *J. Weber,* Zugang zu Softwarekomponenten der Suchmaschine Google, S. 102.

1304 *Monopolkommission*, Sondergutachten 68, S. 187, Rn. 572; *Nuys*, WuW 2016, 512 (513).

1305 *BKartA*, Big Data und Wettbewerb, S. 7f: „Schneeball-Effekt".

generierung offenstehen als anderen Organisationseinheiten.[1306] So bieten sie auch einen besseren Rahmen für das bereits angesprochene Learning by Doing.[1307]

Märkte, die von starken Skalen-, Netzwerk- und Datennetzwerkeffekten bestimmt werden, dürften von einer hohen Anbieterkonzentration geprägt sein.[1308] Für ein etabliertes Unternehmen sind die zum Markteintritt getätigten Investitionen später nicht mehr entscheidungserheblich. Imitatoren ohne ein überlegenes Produkt könnten durch eine drohende Absenkung der Preise auf Grenzkostenniveau durch das etablierte Unternehmen vom Markteintritt abgehalten werden.[1309] Das dominierende Netzwerk muss eher fürchten, von einem besseren Netzwerk als von einem günstigeren Netzwerk abgelöst zu werden.[1310]

Netzwerkeffekte beschleunigen den Aufstieg eines Unternehmens am Markt, aber auch seinen Abstieg.[1311] Insofern wirken sie destabilisierend. Dies bewirkt wegen der in Aussicht stehenden Ablöse eines Netzwerkes besondere Innovationsanreize. Die Bestreitbarkeit eines marktmächtigen Netzwerkes setzt jedoch die grundsätzliche Möglichkeit der Entwicklung von Innovationen voraus. Ein Datennetzwerkeffekt, der zusätzlich bewirkt, dass Daten, die als Grundlage für die Entwicklung neuer selbstlernender Systeme benötigt werden, unerreichbar werden, reduziert folglich die Bestreitbarkeit von Netzwerkmärkten. Auch ohne ein wettbewerbswidriges Verhalten können Netzwerkeffekte Wettbewerber davon abhalten, sich trotz technologischer Überlegenheit am Markt durchzusetzen.[1312]

Die Auswirkungen von Skalen-, Netzwerk- und Datennetzwerkeffekten werden in der Literatur unterschiedlich schwerwiegend bewertet. Dies liegt nicht zuletzt daran, dass Netzwerkeffekte auf verschiedenen, sich berührenden Ebenen funktionieren und einander voraussetzen, aber nicht genau gemessen werden können und kaum mit Sicherheit von anderen

---

1306 *Eifert*, Innovationen in und durch Netzwerkorganisationen, in: Hoffmann-Riem/Eifert (Hrsg.), Innovation und rechtliche Regulierung, S. 88–133 (96).

1307 *Eifert*, Innovationen in und durch Netzwerkorganisationen, in: Hoffmann-Riem/Eifert (Hrsg.), Innovation und rechtliche Regulierung, S. 88–133 (97), Fn. 27 mwN.

1308 Siehe *Dietrich*, Wettbewerb in Gegenwart von Netzwerkeffekten, S. 66.

1309 Siehe *Dietrich*, Wettbewerb in Gegenwart von Netzwerkeffekten, S. 91.

1310 *Dietrich*, Wettbewerb in Gegenwart von Netzwerkeffekten, S. 92; *Körber*, WuW 2015, 120 (123f); *Lerner*, The Role of „Big Data" in Online Platform Competition, 26. August 2014, S. 46.

1311 *C. Tucker*, Antitrust 2018, Vol. 32, No. 2, S. 72–79, 73ff; *dies.*, Digital Data, Platforms and the Usual [Antitrust] Suspects, 31. Januar 2019, S. 4.

1312 *BMWi*, Wettbewerbspolitik im Cyberspace, Rn. 35.

Variablen isoliert werden können. Für Datennetzwerkeffekte gilt dies ebenfalls: Es wird wohl nicht festzustellen sein, ob von eigenen Nutzern generierte Daten oder zugekaufte Daten weitere bestehende Skalen- und Netzwerkeffekte in gleichem Ausmaß verstärken oder abschwächen. Zuletzt wird es selten gelingen, zu klären, ob diese Effekte oder die Produktqualität der Grund für den hohen Marktanteil eines Unternehmens sind.[1313] Hohe Marktanteile können auf einigen Märkten der Beleg eines intakten Marktes sein.[1314] Auf Märkten, die von globalen digitalen Geschäftsmodellen geprägt sind, findet oft kein starker Wettbewerb auf dem Markt, sondern stattdessen ein intensiver Wettbewerb um den Markt statt. Digitale Industrien sind zudem von geringen Fixkosten geprägt.[1315] So werden Cloud-Dienste nach Bedarf in Anspruch genommen anstelle der Errichtung teurer Datencenter und Entwicklung entsprechender Software. Im Vergleich zu traditionellen produzierenden Industrien dürften die Skaleneffekte daher sogar abnehmen und ein Markteintritt erleichtert sein. Dieser Effekt der fortschreitenden Digitalisierung erhöht die Bestreitbarkeit von Märkten und destabilisiert folglich positive Netzwerkeffekte.

V. Zwischenergebnis: Rückkopplungseffekte und Datennetzwerkeffekte

Die Vermutungen zu Datennetzwerkeffekten oder Feedback-Effekten ergeben insofern nichts Neues, als dass sie weiter von einem Vorteil für große Unternehmen und First Mover Advantages ausgehen. Das erste Unternehmen mit einem nutzerfreundlichen Produkt und guten selbstlernenden Systemen kann sich selbst den Weg für weiteres Wachstum und Qualitätsverbesserungen ebnen, während neue Marktteilnehmer einen anderen Weg finden müssen, um aufzuholen. Die Voraussetzung ist jedoch, dass es auch für sie Wege gibt, um ebenfalls Datennetzwerkeffekte in Gang zu setzen. Hierzu müssen sie in die Entwicklung selbstlernender Systeme investieren und können bei der Entwicklung und dem Vertrieb ihrer Systeme aus unterschiedlichsten Gründen erfolglos sein, ohne dass die Schuld dafür bei marktmächtigen Unternehmen zu suchen ist. Möglicherweise gelingt es ihnen nicht, ein bedeutendes selbstlernendes System zu entwickeln; möglicherweise können die Lerneffekte nicht in ein wertvolles Produkt

1313 *Brynjolfsson/McAfee*, The Second Machine Age (deutsch), S. 186.
1314 *Körber*, WuW 2015, 120 (123).
1315 Vgl. *OECD*, Digital Innovation, S. 36; *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 5.

übersetzt werden und schließlich kann die Geschäftsstrategie zum Vertrieb des Produkts fehlschlagen. Schon in diesem vereinfachten Schema ist nur eine von drei Hürden möglicherweise auf das Fehlen von Daten zurückzuführen. Sonstige Netzwerk- und Skaleneffekte könnten bewirken, dass auch ein datenreicher Marktteilnehmer[1316] mit seiner Geschäftsstrategie nicht gegen etablierte Marktteilnehmer ankommen kann und in Folge dessen negative Datennetzwerkeffekte erfährt.

Wegen des Zusammenhangs zwischen Datenreichtum und -qualität und der Akkuratheit bei selbstlernenden Systemen ergibt sich für den Gesetzgeber – wie auch für das entwickelnde Unternehmen – ein „Henne-und-Ei"-Problem: War eine überzeugende Produktqualität und Geschäftsstrategie zuerst da oder eine mindestoptimale Datenmenge? Gespiegelt hierzu muss auch das Unternehmen strategisch entscheiden, ob es den Schwerpunkt initialer Investitionen bei der Programmierung des selbstlernenden Systems oder bei der Lizenzierung von Trainingsdaten setzt. Die Akquise von Nutzern dürfte auch in Gegenwart von (Daten-)Netzwerkeffekten möglich bleiben, was nicht zuletzt die Veränderungen der Marktstruktur der sozialen Netzwerke und Plattformdienste belegen.[1317] Qualität ist kein negativer Lock-In-Effekt: Die Entscheidung der Nutzer für ein Produkt darf dem Produzenten nicht zum Nachteil gereichen. Die Nutzerzahl ist für viele Unternehmen der Internetökonomie die wesentliche Wertquelle und zentraler Baustein des Geschäftsmodells.[1318]

Nicht zu vergessen ist auch ein makroökonomischer Aspekt: Wenn es derartige Rückkopplungseffekte tatsächlich gibt, wirken sie nicht nur für einzelne Unternehmen, sondern auch für die Gesamtwirtschaft, die ein Interesse darin sieht, dass selbstlernende Systeme bestmöglich genutzt werden. Die deutsche Wirtschaftspolitik wird einen Rückkopplungseffekt dann befürworten, wenn er den deutschen oder europäischen entwickelnden Unternehmen zugutekommt. Der Bestand einiger besonders gut trainierter selbstlernender Systeme ist einer Fragmentierung von Datenbeständen und Lernfähigkeiten vorzuziehen. Zentral für eine gesamtwirtschaftliche Innovationsfähigkeit ist dabei, dass sowohl Innovationsanreize als auch alle Innovationsvoraussetzungen vorliegen. Sollte es wirklich in eini-

---

1316 Gerade Unternehmen der Finanz- und Energiewirtschaft verfügen über bedeutende Datenmassen, ohne Unternehmen der Digitalwirtschaft zu sein und in entsprechende Diskussionen einbezogen zu werden.

1317 Siehe Fn. 1333.

1318 *J. Weber*, Zugang zu Softwarekomponenten der Suchmaschine Google, S. 71; *Zimmerlich*, Marktmacht in dynamischen Märkten, S. 82.

gen Sektoren dazu kommen, dass Netzwerk- und Datennetzwerkeffekte sich gegenseitig in einer Form potenzieren, die dazu führt, dass es zu Monopolbildungen kommt, ist zu erwarten, dass die Innovationsanreize abnehmen. Hinzu kommt, dass in dieser Situation die zur Entwicklung nötigen Daten, wenn ein Datennetzwerkeffekt vorliegen sollte, im Stil einer Zentripetalkraft bei einem Unternehmen sammeln. Extern würde dies als innovationsfeindlicher Verdrängungseffekt wirken. Wenn nur wenige Unternehmen große Datenmengen kontrollieren, besteht das Risiko, dass die Vorteile von KI nur wenigen Unternehmen zugutekommen. Kommt in dieser Situation hinzu, dass die Datenströme unzugänglich sind, keine Open Data bereitgestellt werden und die datenreichen Unternehmen nicht zu Kooperationen bereit sind, ist der Zugang zu einer entscheidenden Ressource der Entwicklung selbstlernender Systeme abgeschnitten.

Daten sind ein dynamischer Innovationsrohstoff, sie ändern mit der Zeit ihren Wert und verlieren in der Regel an Verwertbarkeit und Relevanz. Mit der fortschreitenden Verbreitung von Systemen der Künstlichen Intelligenz in industriellen Prozessen werden sie zu einem Produktionsfaktor. Denkbar ist, dass sich die Geschwindigkeit, mit der Daten veralten, sich auf den Verlauf eines möglichen Datennetzwerkeffektes überträgt und einen Anstieg (positiv) oder Abstieg (negativ) von Datenreichtum beschleunigt. Dies könnte wiederum die Geschwindigkeit einer sich selbst verstärkenden Marktmachtverfestigung indirekt beeinflussen.[1319]

Solange die Marktmacht der datenreichen Unternehmen von ihnen selbst als bestreitbar betrachtet wird, dürften sie Anreize zur Entwicklung von KI-Anwendungen verspüren und Innovationen hervorbringen. Wenn sie selbst aber an Datennetzwerkeffekte glauben, minimiert dies die Anreize zur Investition in Forschung und Entwicklung.[1320]

Der isolierte Nachweis von Datennetzwerkeffekten ist bisher nicht gelungen. Die bloße Möglichkeit einer Beeinträchtigung der Innovationstätigkeit genügt als Anlass zur Stärkung der staatlichen Innovationsförderung, aber nicht als Rechtfertigung von grundrechtsrelevanten Eingriffen[1321] in die wirtschaftliche Tätigkeit privater Unternehmen. Zumindest gegenwärtig ist ein hohes Niveau der Innovationstätigkeiten im Bereich der selbstlernenden Systeme zu beobachten, das zudem von einer hohen Zahl von Startups geprägt ist.

1319 *J. Weber*, Zugang zu Softwarekomponenten der Suchmaschine Google, S. 241.

1320 So *Prüfer/Schottmüller*, Competing with Big Data, Tilburg Law School Legal Studies Research Paper Series No. 06/2017, S. 2.

1321 Dazu Kapitel 5 E.I. Verfassungsmäßigkeit und Interessenabwägung.

E. Monopolisierungstendenzen durch Innovationen

Die Aussicht auf Monopolrenten ist ein wichtiger Innovationsanreiz. Insofern dürfte gelten, dass dort, wo die Erlangung einer Monopolstellung möglich ist, die Anreize zur Investition in Forschung und Entwicklung erhöht sind. Wenn für privatwirtschaftliche Unternehmen der größte Gewinn in einer Monopolstellung zu erzielen ist, werden sie diese anstreben und dafür, soweit möglich, Netzwerkeffekte und Datennetzwerkeffekte in Kraft setzen. Je nach Definition eines Marktes kann der erste Anbieter eines neuartigen Produktes eine Monopolstellung erlangen. Das Patentrecht ermöglicht es zudem, zeitlich begrenzte Monopole auf bestimmte Erfindungen zu errichten. Die Rechtsordnung setzt so gezielt Anreize zur Entwicklung neuer Technologien.

Digitale Produkte zeichnen sich durch eine schnelle und günstige Vervielfältigung aus: Eine Innovation kann innerhalb kürzester Zeit der gesamten Marktnachfrage bereitgestellt werden. Dies macht es Wettbewerbern schwer, aufzuholen, weil es schon nach kurzer Zeit keine unbefriedigte Nachfrage mehr geben könnte. Hinzu kommt, dass an digitalen Gütern kein Verschleiß auftritt und sie deshalb nicht routiniert ersetzt werden. Beide Faktoren können einer hohen Marktkonzentration zuträglich sein und – wenn der überwiegenden Literatur zu dem Verhältnis von Wettbewerb und Innovationstätigkeiten gefolgt wird – in neuen Märkten weitere Innovationstätigkeiten anregen.

I. Historische Verläufe im Innovationswettbewerb

Die letzten 150 Jahre der wirtschaftlichen Entwicklung waren von der Automatisierung geprägt. Zunächst wurden routinemäßige Prozesse automatisiert. Mit selbstlernenden Systemen dürfte die Automatisierung spezialisierter, kognitiver Aufgaben gelingen. Selbstlernenden Systemen wird daher von vielen Seiten die gleiche Transformationskraft zugeschrieben wie der Dampfmaschine oder der Elektrizität.[1322] Als Basisinnovation könnte sie zahlreiche Folgeinnovationen, Geschäftsmodelle und nach den Kond-

1322 *Europäische Kommission*, Mitteilung vom 25. April 2018, COM(2018) 237 final, S. 1; *Aghion/Jones/Jones*, Artificial Intelligence and Economic Growth, S. 4; *S. Lynch*, Andrew Ng, Stanford Business, 11. März 2017: „AI is the new electricity".

ratjew-Zyklen[1323] auch erhebliches wirtschaftliches Wachstum initiieren. Während der zweiten industriellen Revolution wurden Erfindungen der Wissenschaft verstärkt in die industrielle Fertigung eingebracht. Mit neuen Erkenntnissen der Chemie und Physik bauten Unternehmen wie BASF, General Electric und Ford aufeinander auf und ermöglichten gegenseitig ihr Wachstum.[1324] Das Patentrecht gestand entwickelnden Unternehmen zur weiteren Setzung von Innovationsanreizen zeitlich begrenzte Verwertungsmonopole zu. Traditionelle Industrien zeichnen sich durch produktionsseitige Größenvorteile (Skaleneffekte) und kostenintensive Markteintritte aus. Diese Größenvorteile waren jedoch erschöpflich, weil sie sich auf physische Anlagen und Produktionskapazitäten bezogen und bei vollständiger Auslastung neue Investitionen nötig wurden. Märkte waren zu den Zeiten der ersten und zweiten industriellen Revolution relativ stabil mit mäßigen Innovationsraten.[1325] In der darauf folgenden Internetökonomie wurden Monopole wiederum verstärkt durch nachfrageseitige Größenvorteile, die von Innovationen der Informationstechnologie ermöglicht wurden. Eine Studie kommt etwa zu dem Ergebnis, dass bis in die 1930er Jahre der technische Fortschritt größere Betriebseinheiten einforderte, aber sich dieser Zusammenhang in den letzten 50 Jahren nicht mehr zeigt.[1326] Die Informationstechnologie ermöglichte weniger kostenintensive Markteintritte[1327], was die von ihr beeinflussten Märkte destabilisierte und Innovationsraten erhöhte. Eine andere Untersuchung stellt die These auf, dass in der digitalen Ökonomie nicht der First Mover am besten positioniert ist, sondern ein „fast second".[1328] Ohnehin ist nicht die Herangehensweise der Datenwirtschaft neu, sondern die Verbreitung: Die Zusammenstellung kollektiver Logbücher und Zugangsgewährung gegen Herausgabe einzelner, historischer Logbücher war schon eine Art spezialisiertes sozia-

1323 Siehe Kapitel 2 B.I.3. Basisinnovationen und ‚Invention of a New Method of Innovating', S. 59.

1324 *Parker/Alstyne/Choudary*, Platform Revolution, S. 19.

1325 So *Wieddekind*, Innovationsforschung, Wettbewerbstheorie und Kartellrecht, in: Eifert/Hoffman-Riem (Hrsg.), Innovation und rechtliche Regulierung, S. 134–170 (165).

1326 *Blair*, Economic Concentration: Structure, Behaviour and Public Policy, S. 87ff.

1327 Dies gilt nicht für alle Geschäftsmodelle der Internetökonomie; die Inbetriebnahme einer Suchmaschine erfordert beispielsweise hohe Investitionen in die Erstellung eines Web-Indexes, vgl. *Petit,* Technology Giants, 20. Oktober 2016, S. 56.

1328 *Geroski/Markides*, Fast Second. Als Beispiele werden IBM, Microsoft, Amazon und JVC genannt.

les Netzwerk von Hydrografen in der Mitte des 19. Jahrhunderts.[1329] Es entsprach einer analogen Form eines Datennetzwerkeffektes, weil die Qualität des kollektiven Logbuches mit der Anzahl der historischen Logbucheinträge, die eingeflossen sind, stieg. Ähnlich funktioniert die Ablösung von Straßenkarten durch Navigationsprogramme wie Waze.[1330] Nutzer teilen in Echtzeit Verkehrs-, Navigations- und Straßeninformationen, die dazu beitragen, Verkehrsmuster der Zukunft besser vorherzusagen. Technologische Rückkopplungseffekte wurden schon 1998 von Carl Shapiro und Hal Varian beobachtet und am Beispiel von Microsoft und Apple illustriert: Positives Feedback habe den Systemen von Microsoft und Intel Rückenwind gegeben.[1331] Mit sinkenden Marktanteilen für Apple befürchteten Nutzer nun, dass Software-Entwickler Apple den Rücken kehren und das System mangels komplementärer Angebote zusammenbrechen würde. Möglicherweise handelte es sich hierbei tatsächlich um einen Feedback Loop, Apple hat aber heute dank disruptiven Innovationen nicht mit sinkenden Marktanteilen und zugrunde gehenden Systemen zu kämpfen. Rückkopplungseffekte im Hinblick auf einem Unternehmen zur Verfügung gestellte Informationen und Innovationen sind kein neues Thema.[1332]

Historische Beispiele für Datennetzwerkeffekte gibt es wegen des jungen Alters dieses Phänomens noch nicht, während es zahlreiche Beispiele für die Wirkung von Netzwerkeffekten gibt. Diese belegen sowohl die positiven Effekte als auch die negativen Effekte, also etwa den zügigen Auf- und Abstieg von sozialen Netzwerken wie Friendster, MySpace, StudiVZ, Facebook und Snapchat.[1333] Die jeweiligen Machtvorsprünge erodierten nach einiger Zeit und der Erosionsprozess erfolgte verglichen mit der in traditionell fertigenden Branchen zu beobachtenden Marktmachterosion besonders schnell. Das unternehmerische Selbstbewusstsein und die Gewissheit einer Beständigkeit im Markt, die von Skalen- und Netzwerkeffekten getragen wird, dürften mittlerweile gesunken sein.[1334] Die Selbst-

1329 *The Economist*, Clicking for Gold, Data, Data Everywhere, 25. Februar 2010.
1330 Siehe https://www.waze.com/de/; dazu *Weiss*, Network Effects Are Becoming Even More Important On Emerging Platforms, Forbes, 18. März 2018.
1331 *Shapiro/Varian*, Information Rules, S. 174.
1332 *Robracher*, Zukunftsfähige Technikgestaltung als soziale Innovation, in: Sauer/Lang (Hrsg.), Paradoxien der Innovation, S. 175–189 (177).
1333 Dazu BKartA, Beschluss vom 6. Februar 2019, B6–22/16 Rn. 433ff – *Facebook*; *D. Evans/Schmalensee*, Regulation Winter 2017/18, 36 (38f); *Körber*, ZUM 2017, 93 (95); *Tamke,* NZKart 2018, 503 (507).
1334 *Petit,* Technology Giants, 20. Oktober 2016, S. 32, siehe Fn. 146.

Disruption aus Angst vor vorbeiziehenden Wettbewerbsteilnehmern sei ein kennzeichnendes Phänomen.[1335] Diese Angst nährt sich daraus, dass eine sinkende Nutzerzahl aufgrund negativer (Daten-)Netzwerkeffekte den Nutzen und die Qualität eines angebotenen Dienstes sinken lässt. Die Qualität eines physischen Produktes sinkt demgegenüber nicht dadurch, dass Kunden sich für ein anderes Produkt der Kategorie entscheiden.[1336] Offen ist, wie sich Künstliche Intelligenz auf die gesamtwirtschaftliche Innovationstätigkeit auswirken wird.[1337]

Gemein ist allen technologischen Umwälzungen der auf sie folgende Ruf nach Regulierung. Die Betrachtung der Geschichte von Innovationsaktivitäten zeigt, dass sie dann erfolgreich waren, wenn die entwickelnden Unternehmen ihr Wissen offen teilten und ihre Innovationen aufeinander aufbauten. Dies dürfte auch Unternehmen der Digitalwirtschaft bewusst sein und sie zu einem ähnlichen Verhalten motivieren.

II. Notwendigkeit von Daten zur Entwicklung innovativer Produkte

Auch wenn dies kein Automatismus ist, können Datennetzwerkeffekte Möglichkeiten zur Blockade von Innovationspfaden schaffen. Dieses Problem wurde bisher vorrangig im Hinblick auf die Software-Industrie diskutiert.[1338] Ausgangspunkt der Diskussion ist, dass Software-Innovationen von kumulativem Charakter sind, da sie auf vielfältige vorausgehende Innovationen aufbauen.[1339] So wurde schon 2009 angenommen, dass ein Softwareunternehmen (Microsoft) nicht „Tüftlern freiwillig das Tor zum Markt öffnet", wenn dies die Gefahr des Verlustes der eigenen Vormachtstellung erhöhe.[1340] Der Kontrolle über große maschinenlesbare Datensets kann vor dem Hintergrund des Lernbedürfnisses selbstlernender Systeme

---

1335 *Petit,* Technology Giants, 20. Oktober 2016, S. 34, 38.

1336 *Stucke*, Georgetown Law Technology Review, S. 275–324, 283 (2018).

1337 *Dreher,* ZweR 2009, 149 (152); *J. Weber*, Zugang zu Softwarekomponenten der Suchmaschine Google, S. 98; *Zimmerlich*, Marktmacht in dynamischen Märkten, S. 94.

1338 EuG, Urteil vom 17. September 2007, T-201/04 – *Microsoft/Kommission*.

1339 *Baake et al.*, Die Rolle staatlicher Akteure bei der Weiterentwicklung von Technologien in deregulierten TK-Märkten, DIW-Politikberatung kompakt 28, 2007, S. 71.

1340 *Heitzer*, Innovation und Wettbewerb aus kartellrechtlicher Sicht, Rede FIW-Symposium 2009, S. 9.

eine wettbewerbsrechtliche Bedeutung zukommen.[1341] Diese ergibt sich unter anderem aus der hohen Innovationsrelevanz. Somit stellt sich die Frage, ab welchem Punkt die Erfassung von Daten einem Unternehmen einen so großen Vorsprung im Innovationswettbewerb verschafft, dass es diesen allein mithilfe seines Datenreichtums verteidigen kann.[1342] Tatsächlich gilt es hier, einerseits zwischen substituierbaren Datenströmen und andererseits zwischen variierenden Innovationspfaden mit gleichen Zielen zu unterscheiden. Vieles spricht dafür, dass unterschiedliche Innovationsarten unterschiedlich von Lern- und potentiellen Datennetzwerkeffekten betroffen sind.

1. Datenverständnis – Disruptive Innovationen

Eine der gravierendsten Unterscheidungen in der Innovationsforschung ist die zwischen disruptiven und inkrementellen Innovationen.[1343] Je nach dem Innovationsziel werden die zugrunde gelegten Informationen aus unterschiedlichen Perspektiven betrachtet. Das Lernen aus historischen Nutzerdaten ermöglicht selbstlernenden Systemen die Verbesserung und Personalisierung ihrer Dienste. Über einen längeren Zeitraum können inkrementelle Innovationen quasi automatisiert aus der Produktevolution hervorgehen.[1344] In der Regel beziehen sich inkrementelle Innovationen auf die Wünsche der aktuellen Nachfrager oder Nutzergruppe. Das Innovationsziel beschränkt sich auf das bessere oder günstigere Erfüllen der Bedürfnisse dieser Zielgruppe.

Disruptive Innovationen ändern demgegenüber den „Job-to-be-done", also die Herangehensweise an die Problemlösung und schaffen eine Nachfrage, statt bestehende Wünsche zu erfüllen. Daten, die Nutzungserfahrungen mit bisherigen Prozessen und Technologien aufzeigen, können schwerlich Verbesserungsmöglichkeiten für neue Prozesse aufzeigen. Statt der Analyse von Nachfragemustern werden Nutzerdaten dann auf Disruptionspotentiale untersucht. Somit ist ein anderes Datenverständnis nötig. Aus historischen Feedback-Daten ist für disruptive Innovationen

1341 So *Schweitzer/Peitz*, NJW 2018, 275 (276).
1342 Ähnlich *Haucap*, Big Data aus wettbewerbs- und ordnungspolitischer Perspektive, in: Morik/Krämer (Hrsg.), Daten, S. 95–142 (96).
1343 Siehe Kapitel 2 B.I. Abgrenzung nach Umfang, S. 56.
1344 *Cockburn/Henderson/Stern*, The Impact of Artificial Intelligence on Innovation, NBER Working Paper No. 24449, März 2018, S. 7.

weniger Nutzen zu schöpfen, weil sie nur die Reaktionen auf bereits bestehende Dienste oder Produkte abbilden.[1345] Folglich dürfte der Feedback-Datenreichtum eines Wettbewerbers einen disruptiven Markteintritt nicht verhindern, weil es gerade auf diese Rohdaten nicht entscheidend ankommt. In einem Zugangsbegehren nach der Essential-Facilities-Doktrin kann kaum die Notwendigkeit eines Datensets zur Entwicklung einer disruptiven Technologie dargelegt werden.[1346] Ebenfalls dürfte es bei der Beurteilung von Zusammenschlüssen schwerfallen, darzulegen, dass die kombinierten Datensets der Zusammenschlussbeteiligten ein Disruptionspotential offenbaren: Die Vorhersehbarkeit des disruptiven Charakters einer erfolgreichen Technologie würde ihr den disruptiven Charakter nehmen. Entwicklungstätigkeit mit dem Ziel der Disruption ist weniger ressourcenbasiert als vielmehr an menschlicher Kreativität und „Unternehmergeist" (entrepreneurship) ausgerichtet.[1347] Wirkliche Kreativität kann Künstliche Intelligenz bisher nicht erlernen. Je einfacher und automatischer inkrementelle Innovationen hervorzubringen sind, desto eher verlieren sie den Innovationscharakter: Kleinschrittige Entwicklungen werden mit fortschreitender Entwicklung selbstlernender Systeme vom Nutzer möglicherweise als so selbstverständlich vorausgesetzt wie regelmäßige Updates.

Die Identifikation von Disruptionsgefahren und -potentialen ist für Unternehmen mit datenbasierten Geschäftsmodellen ein integraler Bestandteil ihrer Strategie. Dabei ist häufig das Ziel, „die Nachfrage der Kunden in kreativer Hinsicht neu zu definieren, weg von der Zentrierung auf etablierte Kategorien von Produkten und Dienstleistungen, hin zu breiter definierten Grundbedürfnissen."[1348] Google, Facebook und Microsoft ebenso wie global agierende Unternehmen außerhalb der Internetökonomie beschreiben die von ihnen wahrgenommene Bedrohung durch disruptive Technologien als besonders hoch.[1349] Die meisten dieser Unternehmen haben selbst bestehende Marktstrukturen aufgebrochen[1350] und nehmen daher ihr Umfeld als instabil wahr. Zu beachten ist, dass die Unterschei-

---

1345 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 10.
1346 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 10.
1347 *Petit*, Technology Giants, 20. Oktober 2016, S. 67; ähnlich *Bethell/Baird/Waksman*, Journal of Antitrust Enforcement 2020, Vol. 8, S. 30–55 (33).
1348 *Schweitzer/Haucap/Kerber/Welker*, Modernisierung der Missbrauchsaufsicht, S. 15.
1349 *Petit*, Technology Giants, S. 18f mit Verweis auf Google 2015 10-K Form, Facebook 2015 10-K Form, Microsoft 2015 10-K Form.
1350 *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 35.

dung zwischen inkrementellen und disruptiven Neuerungen von der Perspektive des Betrachters und dem zeitlichen Horizont abhängig ist.

Der Innovationsbezug bedeutet eine besondere Wettbewerbssensitivität der Daten: Daten, die mit einem bestimmten Informationsziel erfasst wurden und nicht quasi-nebenbei während der Nutzung des Dienstes, können dieses Innovationsziel möglicherweise offenlegen. Ein Datenzugangsrecht, das die Offenlegung solcher Daten erfordert, ermöglicht den Einblick in Geschäftsgeheimnisse und umgekehrt erlaubt die Einrede des Geschäftsgeheimnisses, alle explizit als innovationsrelevant eingestuften Daten zurückzuhalten.

Daten, die an einen bestehenden Dienst oder ein bestehendes Produkt anknüpfen, können erforderlich sein, um diese weiterzuentwickeln oder einen komplementären Dienst auf einem noch nicht bestehenden nachgelagerten Markt (Aftermarket) zu entwickeln. Für die Entwicklung neuer disruptiver Technologien dürften ebendiese Daten von geringerer Relevanz sein. Folglich dürfte ein Versperren des Zuganges zu diesen Daten von unterschiedlicher Innovationsrelevanz sein. Disruptive Technologien wären von einem Datennetzwerkeffekt weniger bedroht, sondern würden vielmehr einen eigenen Kreislauf in Gang setzen.

2. Unterscheidung: Must-Have-Daten und Nice-to-Have-Daten

Auch wenn teilweise angenommen wird, dass in der Informationstechnologie der Zugriff auf Daten bei der Entwicklung innovativer Produkte grundsätzlich unabdingbar ist, gilt dies nicht für alle Datensets in gleichem Maße. Je nach Innovationsziel kann zwischen Must-Have- und „Nice-to-Have"-Daten[1351] unterschieden werden. Ähnlich der Unterscheidung von Innovationsressourcen geht es hier darum, ob die jeweiligen Datensets notwendig (Must Have) oder lediglich nützlich (Nice to Have) sind. Letztere helfen bei der Entwicklung eines Dienstes, sind aber durch andere Datensets, synthetische Daten, eine andere Herangehensweise bei der Programmierung des Dienstes oder ähnliche Umwege annähernd gleichwertig zu ersetzen.[1352] Möglicherweise erfasst nur ein Anbieter aktuell nachgefragte Daten; dies schließt aber nicht aus, dass ein zweiter ohne größere Hindernisse mit der Erfassung ebenso nützlicher Daten beginnen

1351 Formulierung übernommen von *Sivinski/Okuliar/Kjolbye,* ECJ, Vol. 13, S. 199–227, 202 (2017).

1352 *Sivinski/Okuliar/Kjolbye,* ECJ, Vol. 13, S. 199–227, 215 (2017).

könnte. Dabei könnte es schwerfallen, nachzuweisen, dass ein Ersatz nicht ebenso gut als Trainingsdatenset für selbstlernende Systeme genutzt werden kann. Die Unterscheidung zwischen notwendigen und nützlichen Datensets wurde von der Europäischen Kommission etwa in *Microsoft/LinkedIn* vorgenommen.[1353] Insbesondere wurden schon ohne Zugriff auf die in Frage stehenden Daten der Zusammenschlussparteien CRM-Software-Angebote entwickelt.[1354] Somit schien es auch für das Training von Machine-Learning-Anwendungen Alternativen zu deren Datensets zu geben.[1355] Je nach Anwendungsabsicht seien eigene Inhouse-Daten für entsprechende Dienste eine nützliche Grundlage; ebenfalls könnten nach Angaben der Wettbewerber andere Datensets relevanter sein als die der Parteien des Zusammenschlusses.[1356] Es spielten bei der Beurteilung der Notwendigkeit sowohl die Qualität, als auch die Varianz und die Quantität der Daten eine Rolle.

Wenn die Datensets nicht bereits in die Angebote der nachfragenden Entwickler eingebunden sind (z. B. auf Aftermarkets für Reparatur und Wartung), wird es sich bei ihnen selten um kritische Inputs handeln.[1357] Weil in der Regel unbekannt ist, welche Daten innerhalb von industriell fertigenden Unternehmen erfasst und dauerhaft gespeichert werden, ist es höchst riskant, ein Geschäftsmodell zu entwickeln, das nur mit dem Zugang zu diesen Daten funktionieren kann. Die Tatsache, dass ein Unternehmen ein beliebtes Produkt auf Grundlage bestimmter Datenströme entwickelt, bedeutet nicht, dass ein Konkurrenzprodukt auch nur auf Grundlage dieses einen Datenstroms entwickelt werden kann.[1358]

Nur Must-Have-Daten sind essentiell und können den Zugang zu nachgelagerten Märkten versperren, weshalb sich diese Unterscheidung mit dem Kriterium der Essential Facility deckt.[1359] Die Einrichtung eines Zu-

---

1353 Europäische Kommission, Entscheidung vom 6. Dezember 2016, M.8124 Rn. 256ff – *Microsoft/LinkedIn*.

1354 Europäische Kommission, Entscheidung vom 6. Dezember 2016, M.8124 Rn. 275f – *Microsoft/LinkedIn*.

1355 Europäische Kommission, Entscheidung vom 6. Dezember 2016, M.8124 Rn. 253–277 – *Microsoft/LinkedIn*.

1356 Europäische Kommission, Entscheidung vom 6. Dezember 2016, M.8124 Rn. 260f – *Microsoft/LinkedIn*; dazu *Bourreau/de Streel*, Digital Conglomerates and EU Competition Policy, S. 28.

1357 So auch *Sivinski/Okuliar/Kjolbye,* ECJ, Vol. 13, S. 199–227, 214 (2017).

1358 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 10.

1359 Europäische Kommission, Entscheidung vom 6. Dezember 2016, M.8124 Rn. 186 – *Microsoft/LinkedIn*.

gangs zu Nice-to-Have-Daten wäre wettbewerbsfördernd, während der Zugang zu Must-Have-Daten wettbewerbsermöglichend wäre.

3. Open Data und Public Interest Data

Eine übliche Vorgehensweise bei dem Training neu entwickelter selbstlernender Systeme ist die Nutzung öffentlich verfügbarer Daten als Trainingsdatensets.[1360] Dies können Daten aus der Open Source oder Ergebnisse von Web Scraping[1361] sein. Ebenfalls werden Daten von öffentlichen Stellen bereitgestellt. Hierbei kann es sich um Statistiken zur Zusammensetzung der Bevölkerung, Investitionen oder Geodaten handeln. Die Ausweitung der kostenlosen Bereitstellung öffentlicher Daten, wie sie bereits in Form von „Open Data", GovData und der „Datenlizenz Deutschland" praktiziert wird, ist eine Maßnahme, um den Anteil der exklusiven Daten an der Gesamtdatenmasse zu senken, ohne private Unternehmen zur Aufgabe ihrer exklusiven Verfügungsmöglichkeiten zu zwingen. Dieses Vorgehen schlagen der Bericht der Kommission Wettbewerbsrecht 4.0 und der Furman-Report vor.[1362]

Das Verständnis der Offenheit von Open Data ist ähnlich dem von Open Source und Open Access: „Wissen ist offen, wenn jeder darauf frei zugreifen, es nutzen, verändern und teilen kann – eingeschränkt höchstens durch Maßnahmen, die Ursprung und Offenheit des Wissens bewahren."[1363] Die Offenheit betrifft die gemeinfreie Verfügbarkeit, die freie Weiterverwendung und die Offenheit der Lizenz. Zusätzlich soll das Format maschinenlesbar und offen, also ohne monetäre oder sonstige Einschränkungen lesbar, sein. Der Zugang muss diskriminierungsfrei sein. Eine monetäre Gegenleistung dürfe die Grenzkosten der Verbreitung nicht übersteigen.[1364] Sowohl private als auch staatliche Stellen können

1360 *Himel/Seamans,* Artificial Intelligence, Incentives to Innovate, And Competition Policy, CPI Antitrust Chronicle December 2017, S. 9.

1361 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 8; auch *Angwin/ Stecklow*, Scrapers' Dig Deep for Data on Web, The Wall Street Journal, 12. Oktober 2010.

1362 *BMWi*, Wettbewerbsrecht 4.0, S. 6; *Furman et al.*, Unlocking Digital Competition, S. 5f.

1363 Open Knowledge International, Open Definition, http://opendefinition.org/od/2.1/de/, zuletzt abgerufen am 19. Januar 2021.

1364 *OECD*, Data-Driven Innovation, S. 38.

Daten als Open Data bereitstellen.[1365] In der Regel sind Daten der öffentlichen Hand gemeint, also Open Government Data.[1366]

Aus verschiedenen Gründen streben staatliche Stellen an, ihr Open-Data-Engagement zu erhöhen. Die Bundesregierung setzte auf Punkt acht ihrer KI-Strategie das Verfügbarmachen von nutzbaren, qualitativ hochwertigen Daten.[1367] Neben der Schaffung von Anreizmechanismen und Rahmenbedingungen für das freiwillige, datenschutzkonforme Teilen von Daten sowie dem Aufbau einer vertrauenswürdigen Dateninfrastruktur ist das Verfügbarmachen von Daten aus öffentlich finanzierten Forschungsprojekten beabsichtigt.[1368] Außerdem soll die gezielte Förderung offener Trainingsdatensätze geprüft werden. Ein Ansatz der KI-Strategie der Europäischen Kommission ist es, mehr maschinenlesbare Daten des öffentlichen Sektors für gewerbliche Zwecke bereitzustellen.[1369] Die Verfügbarkeit von Daten des öffentlichen Sektors für datengesteuerte Innovationen würde bessere Produkte und Dienste ermöglichen.[1370] Dem stimmt die deutsche Bundesregierung zu: Daten könnten Impulse für neue Geschäftsmodelle und Innovationen setzen.[1371] Der volkswirtschaftliche Mehrwert aus der intelligenten Nutzung staatlicher Daten in Deutschland wird auf jährlich 43 Milliarden Euro geschätzt.[1372]

Auf Ebene der Europäischen Union macht die PSI-Richtlinie[1373] seit dem 31. Dezember 2003 Vorgaben für die Verfügbarkeit von Daten des öffentlichen Sektors. Auf ihrer Grundlage hat sich die Verarbeitung von Daten aus öffentlichen Quellen weiterentwickelt.[1374] Die PSI-Richtlinie wurde in Deutschland mit dem Informationsweiterverwendungsgesetz

---

1365 Siehe z. B. Fn. 1038 für von privatwirtschaftlicher Seite bereitgestellte Open Data.

1366 Vgl. BT-Drucks. 18/11614, Entwurf eines Ersten Gesetzes zur Änderung des E-Government-Gesetzes, S. 11; Hoeren/Sieber/Holznagel/*Hackenberg*, Multimedia-Recht, 47. EL Oktober 2018, Teil 16.7, Rn. 33.

1367 *Bundesregierung*, Strategie Künstliche Intelligenz, S. 33f.

1368 *Bundesregierung*, Strategie Künstliche Intelligenz, S. 35.

1369 *Europäische Kommission*, Mitteilung vom 25. April 2018, COM(2018) 237 final, S. 10; *dies.*, Daten, Pressemitteilung vom 25. April 2018.

1370 M. Gabriel in: *Europäische Kommission*, Daten, Pressemitteilung vom 25. April 2018.

1371 BT-Drucks. 18/11614, S. 1, 11.

1372 Vgl. *Dapp et al.*, Open Data. The Benefits, S. 10, 56.

1373 Richtlinie (EU) 2019/2014 über die Weiterverwendung von Informationen des öffentlichen Sektors.

1374 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 57.

(IWG)[1375] umgesetzt. Die Neufassung der PSI-Richtlinie hat zum Ziel, den Markt für digitale Daten zu fördern sowie Wettbewerbsverzerrungen auf dem Binnenmarkt zu verhindern.[1376] Nachdem der Fokus zunächst auf dem Informationszugang der Bürger und der Transparenz der Verwaltung lag, sollen in der Neufassung die wirtschaftlichen Aspekte der Wiederverwendung von öffentlichen Daten Berücksichtigung finden. Zudem findet der Begriff „Open Data" Eingang in den Titel der Richtlinie. In der Vergangenheit haben auch Entscheidungen zur Essential Facilities-Doktrin wie *Magill*, *IMS Health* und *Microsoft* die Formulierung der PSI-Richtlinie beeinflusst.[1377] Insofern würde es nicht überraschen, wenn die Richtlinie künftig auch wettbewerbspolitische Aspekte des Datenzugangs für Trainingsdaten aufnimmt.

Mit der Bereitstellung von Rohdaten statt Informationen geben öffentliche Stellen ihre Interpretationshoheit über die Daten zu einem gewissen Grad auf. Ein Vorteil an der Nutzung öffentlicher Daten zum Training selbstlernender Systeme ist, dass der Staat per se ein Interesse daran und gemäß Art. 3 Abs. 3 GG eine Verantwortung dafür hat, einen Data Bias auszuräumen. Entsprechendes gilt für die Wahrung des Datenschutzes. Zudem gibt es Daten, die nur der Staat erheben kann, z. B. Steuerdaten.[1378] Andere Daten, beispielsweise Wetterdaten, können Private ebenso gut erheben. Die Öffnung dieser staatlichen Datenströme könnte den Wettbewerb unter privaten Datenerzeugern daher befeuern.[1379] Hinzu kommen Projekte wie INSPIRE[1380] (Geodateninfrastruktur) und COPERNICUS[1381] (Erdbeobachtungsprogramm), die von Steuergeldern co-finan-

---

1375 Gesetz über die Weiterverwendung von Informationen öffentlicher Stellen, 19. Dezember 2006, BGBl. I, S. 2913.

1376 *Europäische Kommission*, Vorschlag für eine Richtlinie über die Weiterverwendung von Informationen des öffentlichen Sektors (Neufassung), COM(2018) 234 final, 2018/0111(COD), 25. April 2018, S. 6; *Europäische Kommission*, Mitteilung vom 12. Dezember 2011, COM(2011) 882 final; *dies.*, Public Sector Information: A Key Resource for Europe, Green Paper on Public Sector Information in the Information Society, COM(1998) 585 final; *Schweitzer*, GRUR 2019, 569 (572).

1377 *Lundqvist*, Big Data, Open Data, Privacy Regulations, Intellectual Propery and Competition Law in an Internet of Things World – The Issue of Access, S. 15.

1378 Vgl. *Podszun*, GRUR Int. 2015, 327 (330).

1379 Vgl. *Podszun*, GRUR Int. 2015, 327 (329).

1380 Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates vom 14. März 2007 zur Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft (INSPIRE), ABl. L 108 vom 25. April 2007, S. 1–14.

1381 Auf Grundlage der Verordnung (EU) Nr. 377/2014 des Europäischen Parlaments und des Rates vom 3. April 2014 zur Einrichtung des Programms

ziert werden. Hieraus ergibt sich das Interesse an der breitestmöglichen Nutzung dieser Daten zur Entwicklung neuer Produkte und Dienste. Ähnliches gilt für die Veröffentlichung von Forschungsergebnissen als Open Research Data im Rahmen der Innovationsunion.

Im internationalen Vergleich hinkt Deutschland trotz aufgenommener Bemühungen nach Ansicht der Open Knowledge Foundation sowohl quantitativ als auch qualitativ hinterher.[1382] Frankreich wird als positives Beispiel genannt: Nach dem Gesetz über die digitale Republik (2016) und vorangegangenen Initiativen sind ein offener Standard und offene Datenformate zu wählen und es sind Metadaten zu generieren. Zudem ist festgelegt, dass Daten regelmäßig zu aktualisieren sind.[1383] Auch das Vereinigte Königreich engagiert sich für offene Daten. Die Bereitstellung der Transport-for-London-Daten seit 2009 habe jährliche Kostenvorteile von 130 Millionen Pfund generiert.[1384] Der Vorteil eines ambitionierten supranationalen Open-Data-Ansatzes, der sich an bereits etablierten und getesteten Modellen orientiert[1385], ist, dass eine Harmonisierung der verwendeten Formate, Standards und Metadaten sich an der technologischen Realität orientieren kann.

Zudem stellen sich Kostenfragen: Die fiskalischen Aspekte von Open Data sollten die Bereitstellung nicht dominieren; gleichzeitig könnte ein staatliches kostenloses Leistungsangebot den (jeweiligen) Markt für Datenerzeugung stören.[1386] Die Datenbereitstellung gegen eine Gebühr in Höhe der Grenzkosten könnte diese Erwägungen ausbalancieren.

Copernicus, ABl. L 122 vom 24. April 2014, S. 44–66; auf Grundlage von Art. 3ff der Delegierten Verordnung (EU) Nr. 1159/2013 der Kommission vom 12. Juli 2013 zur Ergänzung der Verordnung (EU) Nr. 911/2010 des Europäischen Parlaments und des Rates über das Europäische Erdbeobachtungsprogramm (GMES), ABl. L 309 vom 19. November 2013, S. 1–6, werden die Daten offen bereitgestellt: https://code-de.org/de/marketplace/search?filter[0]=type:dataset.

1382 *BMWi*, Wettbewerbsrecht 4.0, S. 45: „unzureichend".

1383 Zirkular (Circulaire) vom 26. Mai 2011 über die Einrichtung des einheitlichen Portals öffentlicher staatlicher Informationen „data.gouv.fr"; verfügbar auf Französisch unter https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000024072788, zuletzt abgerufen am 9. Mai 2021; Datenabruf: www.data.gouv.fr.

1384 *Furman et al.*, Unlocking Digital Competition, Rn. 2.85; zur Bewertung der Open-Data-Anstrengungen des Vereinigten Königreichs: *World Wide Web Foundation*, Open Data Barometer 2017.

1385 Konkret haben z. B. Serbien und Luxembourg den Open-Source-Code von data.gouv.fr verwendet, um eigene Plattformen aufzubauen.

1386 *Podszun*, GRUR Int. 2015, 327 (333).

Von staatlichen Stellen können vielfältige Datensätze zum Training selbstlernender Systeme bereitgestellt werden, allerdings keine personenbezogenen Daten und keine Daten, die Geheimnisse über die wirtschaftliche Tätigkeit von Privaten offenbaren. Dies ist Ausdruck einer Wertentscheidung des Gesetzgebers. Möglicherweise kann zahlreichen Datenzugangsbegehren schon durch Open-Data-Initiativen abgeholfen werden. Staatliche Stellen verfügen zwar nicht über Feedbackdaten, diese dürften aber ohnehin meist entweder Geschäftsgeheimnisse sein oder zu spezifisch, um für andere Systeme einen Trainingswert zu bieten.

III. Begrenzende Effekte

Datennetzwerkeffekte wirken, soweit ihre Wirkung belegt werden kann, nicht unbegrenzt und nicht in jedem System in gleichem Maße. Möglicherweise werden sie von bestehenden tatsächlichen und rechtlichen Gegebenheiten begrenzt. Eine denkbare Begrenzung des Feedbackeffektes ist die besprochene Verringerung des Anteils exklusiver Daten an der Gesamtdatenmasse durch Bereitstellung von Open Data. Datennetzwerkeffekte werden dadurch begrenzt, dass einzelne Schritte der Datenerlangung oder -analyse übersprungen und beschleunigt werden können und so der Eintritt in den selbstverstärkenden Kreislauf ermöglicht wird. Die Erlangung hochwertiger Daten ist kosten- und zeitintensiv. Neue Marktteilnehmer verfügen in der Regel weder über finanzielle Mittel noch über Zeit. Abkürzungen wie etwa das Reverse Engineering könnten den Zeit- und Geldvorsprung etablierter Entwickler selbstlernender Systeme relativieren. Ultimativ verfolgt auch ein diskutiertes Datenzugangsrecht den Zweck, Feedbackeffekte zu begrenzen. Bestehen bereits ausreichend begrenzende Effekte, ist dies ein Argument gegen Datenzugangsrechte, weil die Einführung eines solchen zu einem Ungleichgewicht zuungunsten der Entwicklungsanreize führen würde.

1. Negative Skaleneffekte – Bereinigung der Daten

Eine wachsende Datenmasse erfordert eine verstärkte Zuwendung von Data Scientists, um die Daten zu bereinigen. Der Grenznutzen der Daten nimmt ab, weil hinzukommende Daten seltener neue Erkenntnisse hervorbringen. Ein Großteil der hinzukommenden Informationen wird sich

mit bereits vorhandenen Daten decken.[1387] Das Generieren von tatsächlich lehrreichen Daten ist umso aufwendiger und teurer, je mehr Daten vorhanden sind. Dies unterscheidet Datennetzwerkeffekte von klassischen Netzwerkeffekten, bei denen das Hinzugewinnen neuer Nutzer grundsätzlich die Kosten pro Nutzer senkt. Gleichzeitig ist die Bereinigung, Aufbereitung und Kennzeichnung von Daten im Sinne negativer Skalenerträge umso teurer, je mehr vorhanden sind.[1388] Damit sinkt die Agilität und der wettbewerbliche Vorteil, den die Daten bringen, kann sich in einen Nachteil umkehren.

Hinzu kommt, dass Daten schnell veralten. Veraltete Straßenschilder, veraltete Karten und überholte Formulierungen müssen etwa aus verkehrsbezogenen Datensets entfernt werden, um ein Rauschen zu verhindern, während gleichzeitig aktuelle Daten erfasst werden müssen. Entsprechend könnten sie als historische Daten für andere Anwendungsfälle einen bestehenden Wert haben, daher ist eine Löschung unwahrscheinlich. Historische Daten mit Finanzbezug können etwa eine Inflationsbereinigung erfordern, um für neue selbstlernende Systeme einsetzbar zu sein. Gerade kritische Anwendungen wie autonomes Fahren und Präzisionsmedizin erfordern eine ständige Bereinigung und Kontrolle der Datensets zur Gewährleistung eines hohen Maßes an Akkuratheit. Bereits bei der Entwicklung des selbstlernenden Systems müssen Qualität und Quantität, also Tiefe und Weite, gegeneinander abgewogen werden: Eine besonders große, schwach überwachte Datenmasse kann für viele Fälle akzeptable, aber nicht exakte Prognosen liefern. Zu wenige, gut sortierte Datensets führen dazu, dass Spezialfälle exakt beurteilt werden können, aber ein großer Teil der Nutzeranfragen unbefriedigend beantwortet wird.[1389] Obwohl die Generierung und Kuratierung eines Trainingsdatensets zu Beginn Skaleneffekte auslöst, deutet einiges darauf hin, dass die Skalenerträge mit wachsendem Umfang des Datensets negativ sind. Dass auch datenreiche Unternehmen die ihnen zur Verfügung stehenden Daten nicht in vollem Umfang zum Training der selbstlernenden Systeme nutzen, stützt diese These. Negative Skalenerträge, die Datennetzwerkeffekten entgegenwirken, dürfen bei der Untersuchung nicht außer Acht gelassen werden.

1387 *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 8f mwN.

1388 Vgl. *Casado/Lauten*, The Empty Promise of Data Moats, Andreessen Horowitz; *Nuys*, WuW 2016, 512 (514).

1389 *Casado/Lauten*, The Empty Promise of Data Moats, Andreessen Horowitz.

2. Reverse Engineering

Reverse Engineering bezeichnet den Rückbau und die Analyse eines Objektes mit dem Ziel, das nicht offenkundige Know-How zum Vorschein zu bringen.[1390] Mit der Know-How-Richtlinie[1391] wurde der grundsätzliche Schutz vor einem Reverse Engineering aufgehoben, vgl. Art. 3 Abs. 1 S. 1 lit. b.[1392] Das Umsetzungsgesetz GeschGehG erlaubt in § 3 Abs. 1 Nr. 2 den Rückbau.[1393] Zu erwarten ist somit, dass die Praktik des Reverse Engineering künftig stärker genutzt wird. Vorrangig werden von der Liberalisierung physische Objekte wie Werkzeugmaschinen betroffen sein. Das rechtmäßige Reverse Engineering setzt zukünftig voraus, dass das jeweilige Produkt öffentlich verfügbar gemacht wurde und sich rechtmäßig im Besitz des Rückbauenden befindet, ohne dass er einer Pflicht zur Beschränkung unterliegt.[1394] Bei selbstlernenden Systemen dürfte es sich, sobald sie als Software einem breiteren Publikum angeboten werden, um öffentlich verfügbare Objekte handeln. Die grundsätzliche Legalität der Rückentwicklung eines Computerprogrammes kann von geistigen Eigentumsrechten beschränkt sein. Vertragsbedingungen, die ein Reverse Engineering verbieten, also eine „Pflicht zur Beschränkung der Erlangung des Geschäftsgeheimnisses" darstellen, könnten als Allgemeine Geschäftsbedingungen missbräuchlich sein. Nach Leister wird ein Verbot des Reverse Engineering regelmäßig dem Grundgedanken des Gesetzes widersprechen und damit in AGB unzulässig sein.[1395] Denkbar bleibt es in den Rahmenvorgaben von Entwicklungskooperationen.

Für selbstlernende Systeme wurde lange angenommen, dass ohne Kenntnis von den Algorithmen oder Trainingsdaten hinter dem Machine Learning Model die Technologie nicht zu entschlüsseln ist. Selbstlernende Systeme verbergen sich oft in einer „Black Box" hinter Webseiten, Apps und APIs, ohne dass der Arbeitsprozess des Systems bei der Nutzung offen-

---

1390 *Leister*, GRUR-Prax 2019, 175 (175).
1391 Richtlinie (EU) 2016/943, siehe Fn. 768.
1392 Vgl. RGZ 149, 329 – *Stiefeleisenpresse*; BayObLG, Urteil vom 28. August 1990, RReg. 4 St 250/89 = GRUR 1991, 694 – *Geldspielautomat*; in der Folgezeit gelockert, vgl. OLG Hamburg, Urteil vom 19. Oktober 2000, 3 U 191/98 = GRUR-RR 2001, 137 – *Nachbau einer technischen Vorrichtung nach Ablauf des Patentschutzes*.
1393 Regierungsentwurf BT-Drucks. 19/4724; angenommen in der vom Rechtsausschuss geänderten Fassung, BT-Drucks. 19/8300.
1394 *Leister*, GRUR-Prax 2019, 175 (176).
1395 *Leister*, GRUR-Prax 2019, 175 (176).

bart wird. 2016 wurden in einem Paper sogenannte „Model Extraction Attacks" gegen BigML und Amazon Machine Learning demonstriert.[1396] Durch Beobachtung der Ausgaben auf eine Sequenz von Anfragen an das API eines selbstlernenden Systems konnten interne Parameter akkurat geschätzt und die Funktionsweise nachgeahmt werden. Komplexere Algorithmen sind jeweils schwieriger einzuschätzen. Sobald der Algorithmus „gestohlen" wurde, ist es möglich, die verwendeten Trainingsdaten zu identifizieren. Innerhalb des neuronalen Netzes wird ist allerdings keine „Kopie" der Trainingsdaten zu finden, sondern die gewichteten Merkmale der Daten.[1397] Es ist zu erwarten, dass die Rückanalyse-Möglichkeiten weiterentwickelt werden und spiegelbildlich seitens etablierter Entwickler erschwert werden. Reverse Engineering kann, soweit es keine geistigen Eigentumsrechte verletzt, dazu beitragen, ein Level Playing Field zu schaffen und den Markteintritt zu erleichtern.[1398] Zudem besteht Grund zur Annahme, dass Künstliche Intelligenz selbst das Reverse Engineering von Technologien sektorübergreifend vereinfacht und damit den Imitationswettbewerb stärken dürfte.[1399]

3. Geographische Begrenzungen

Viele Datensätze können sprachlich, kulturell oder kontextuell geographisch begrenzt sein und für globale Anwendungen nicht zur Verfügung stehen. Alle Datensätze, die in irgendeiner Form Sprache beinhalten, etwa Sprachaufnahmen, Bilder von Schrift, Textdateien, setzen voraus, dass der Algorithmus mit dieser Sprache umgehen kann und auf ihr Verständnis trainiert wurde. Beispielsweise würde ein selbstlernendes System, das von links nach rechts und von oben nach unten liest, an Textdateien mit arabischer, chinesischer, hebräischer und japanischer Sprache scheitern. Nicht jedes selbstlernende System wird von Anfang an auf das Verständnis aller Sprachen trainiert sein. Hier bestehen für lokale Unternehmen Möglichkeiten, ohne starke Konkurrenz ihr System für sprachlich begrenzte Verwendungszwecke und kleinere Nutzergruppen zu trainieren und mit

1396 *Tramèr et al.*, Stealing Machine Learning Models via Prediction APIs, 3. Oktober 2016.
1397 Dazu *Winter/Battis/Halvani*, ZD 2019, 489 (492).
1398 Ähnlich schon *Wielsch*, ECLR, Vol. 25, S. 95–106, 103ff (2004).
1399 Vgl. *Aghion/Jones/Jones*, Artifical Intelligence and Economic Growth, S. 32.

deren Feedback zu verbessern, bevor ein Vordringen in weitere geographische Märkte erfolgt.

Zu den geographischen Hürden zählen auch unterschiedliche regulatorische Umfelder. Ein selbstlernendes System muss gegebenenfalls auf regionale Besonderheiten angepasst werden, was Entwickler davon abhalten kann, in dieses Umfeld vorzudringen, und anderen die Möglichkeit eröffnet, diese Lücke auszufüllen und auf diesem Gebiet intensiv zu trainieren und die Erkenntnisse später zu übertragen. Einige Sektoren sind besonders intensiv reguliert, etwa Finanzmärkte und Gesundheit. Zu erwarten ist, dass das Dickicht an bestehender Regulierung und Überwachung einen Datennetzwerkeffekt einerseits ausbremst, andererseits zusätzliche Markteintrittshürden konstituiert, die den Kreislauf vor neuen Marktteilnehmern abschirmen und Kaltstart-Effekte verstärken könnten. Ein geändertes regulatorisches Umfeld kann eine Chance für einen Markteintritt sein für denjenigen, der sich als erster auf diese neue Situation einstellt und mit einem optimalen Angebot einen potentiellen Datennetzwerkeffekt auslöst. Anpassbarkeit und Konzentration auf spezielle Problemstellungen sind Vorteile, die Startups sich zu eigen machen können, um qualitativ hochwertige Datensets aufzubauen, mithilfe derer sie die Systeme etablierter Unternehmen ausstechen können.

4. Hardware, Algorithmen und Humanressourcen

Wenn – wie oben angenommen – Daten für den potentiell überragenden Erfolg bestimmter selbstlernender Systeme verantwortlich sind, aber andere Bestandteile Voraussetzungen dafür sind, ist denkbar, dass ebendiese den Datennetzwerkeffekt begrenzen könnten. Zwar ist Hardware ohne bedeutende Hürden zugänglich oder sogar risikofrei zu mieten. Die Funktionsfähigkeit und Verlässlichkeit können jedoch nicht garantiert werden. Die gestiegene Rechenleistung ebnete der Entwicklung selbstlernender Systeme den Weg. Mit der wachsenden Raffiniertheit der Systeme steigen nun wiederum die Anforderungen an die Hardware. Für komplexe Systeme sind maßgeschneiderte Hardware-Lösungen erforderlich, weil sich etwa die Anforderungen von Bild- und Spracherkennungssoftware grundlegend unterscheiden. Dies setzt wiederum eine gewisse Hardware-Expertise bei entwickelnden Unternehmen voraus: Möglicherweise kann mit einer besonders gut abgestimmten Hardwarelösung ein selbstlernendes System im Vorteil sein. Aus diesem Grund ist die Dynamik des Marktes für Hardware für die Relativierung von Datennetzwerkeffekten relevant. Jedenfalls

liegt der Mehrwert von Daten nicht in bloßer Verfügungsmacht, sondern in der Verarbeitung, sodass die führenden Datenverarbeiter eine starke Verhandlungsposition gegenüber Hardwareherstellern haben dürften. Da diese auch ein großes Interesse an der Kommerzialisierung von Daten haben, ist ein Marktversagen a priori nicht erkennbar. Hardware ist in den letzten Jahren günstiger und flexibler geworden und stellt wohl in der Zukunft kein Bottleneck dar.[1400] Sie bleibt ein Faktor bei der Entwicklung selbstlernender Systeme, sodass ein Hardware-Vorsprung einen Vorsprung gegenüber einem aktivierten Datennetzwerkeffekt bedeuten könnte.

Eine weitere Voraussetzung für die Entwicklung selbstlernender Systeme sind Algorithmen und die dahinter stehende menschliche Entwicklungskraft. Datennetzwerkeffekte wirken nicht automatisch, sondern setzen eine Überwachung und Kommentierung der gesammelten Daten durch menschliche Arbeitskräfte voraus.[1401]

Zu möglichen Datennetzwerkeffekten kommt das Problem der Arbeitnehmeranziehungskraft: Ein datenbezogener Rückkopplungseffekt könnte von einem „Talent Attraction Loop" verstärkt werden. Ein datenreiches Unternehmen ist als Arbeitsplatz für Entwickler attraktiver, wenn sie erwarten, dort bessere Systeme entwickelt zu können. Die Knappheit von spezialisierten Entwicklern führt zu einer „Competition for Brains". Gleichzeitig sind die wesentlichen Variablen für entwicklerischen Erfolg „mitnehmbar". Anders als Produktionsbänder und Fabrikhallen in früheren industriellen Revolutionen können Erfahrungen und Fähigkeiten von KI-Entwicklern zu neuen Arbeitgebern mitgenommen werden und dort in neue selbstlernende Systeme eingebaut werden. Ein möglicher Talent Attraction Loop ist deutlich instabiler als ein Datennetzwerkeffekt, weil es keine langfristigen Vorkehrungen zur Wahrung von Exklusivität für Arbeitskräfte gibt.[1402] Hohe Löhne bei finanzstarken Unternehmen können in vielen Fällen durch mehr entwicklerische Freiheit bei Startups ausgestochen werden.[1403] Gleichzeitig sind in den letzten Jahren KI-Master-Stu-

---

1400 *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 29.

1401 *Varian/Dolmans/Baird/Senges*, Digitale Herausforderungen für die Wettbewerbspolitik, in: Wirtschaftsrat der CDU (Hrsg.), Soziale Marktwirtschaft im digitalen Zeitalter, S. 75–92 (80).

1402 Dazu *Fromer*, NYU Law Review, Vol. 94, No. 4, S. 706–736, 715 (2019).

1403 Vgl. *Varian/Dolmans/Baird/Senges*, Digitale Herausforderungen für die Wettbewerbspolitik, in: Wirtschaftsrat der CDU (Hrsg.), Soziale Marktwirtschaft im digitalen Zeitalter, S. 75–92 (76).

dienprogramme und Online-Kurse etabliert worden, die die Knappheit qualifizierter Entwickler abschwächen dürften.[1404]

Das Problem zu kleiner Trainingsdatensets ist entwickelnden Unternehmen bekannt und dieser Variable wurde sich von öffentlicher und privater Forschung angenommen. Die Entwicklung von Systemen, die mit deutlich kleineren Datensets so gut trainiert werden, dass sie verlässliche Ergebnisse liefern, ist eine logische Konsequenz der datenintensiven selbstlernenden Systeme der aktuellen Generation. Auch sie kann helfen, einen möglichen Datennetzwerkeffekt zu relativieren und den Fokus auf Nutzeroberflächen und Geschäftsideen zu verlagern. Varianten von KI, die mit kleinen Datensets akkurate Ergebnisse liefern, sind Transfer Learning[1405] und Federated Learning[1406]. Grundsätzlich funktionieren Modelle mit wenigen Parametern besser für kleinere Datensets.[1407] Dank der Open-Source-Verfügbarkeit von vielfältigen KI-Programmen operieren viele Wettbewerber auf nahezu identischem technologischem Niveau, was den Raum für Differenzierungen schmälert. Gleichzeitig führt diese Verfügbarkeit dazu, dass Software in der näheren Zukunft kein Bottleneck mehr sein dürfte.[1408] Dies kann die Hürde zum Einstieg in den Kreislauf eines potentiellen Datennetzwerkeffektes senken.

5. Zwischenergebnis – Begrenzende Effekte

Ein potentieller Datennetzwerkeffekt kann durch verschiedene Parameter relativiert werden. Es erscheint möglich, dass mithilfe von Reverse

---

1404 So *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 29.

1405 Transfer Learning ermöglicht es, Erkenntnisse aus der Analyse eines Datensets auf ein anderes Problem und dazugehörige Datensets zu übertragen: *Donahue et al.*, DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition, 6. Oktober 2013.

1406 Federated Learning trainiert ein zentralisiertes Machine Learning-Modell mit dezentralisierten Trainingsdaten, die auf den individuellen Geräten verbleiben und nicht in die Cloud geladen werden. Stattdessen wird nur die Erkenntnis mit dem Modell in der Cloud geteilt. Vgl. *Konečný et al.,* Federated Learning: Strategies for Improving Communication Efficiency, 30. Oktober 2017; *McMahan/Ramage*, Federated Learning: Collaborative Machine Learning without Centralized Training Data, 6. April 2017; *Winter/Battis/Halvani*, ZD 2019, 489 (492).

1407 Vgl. *Forman/Cohen*, Learning from Little: Comparison of Classifiers Given Little Training, in: Boulicaut et al. (Hrsg.), PKKD 2004, S. 161–172.

1408 *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 29.

Engineering erlangte Erkenntnisse auf neue Systeme übertragen werden können. Ebenfalls können weniger datenhungrige selbstlernende Systeme genutzt werden und Startups können von der Expertise erfahrener Software-Entwickler profitieren. Die Verfügbarkeit von großen, hochwertigen Datensets als Open Data schwächt mögliche Monopolisierungstendenzen und die Konzentration von Datenmassen auf wenige Akteure ab. Technologisch ist nicht ausgeschlossen, dass KI irgendwann die Grenze ihrer Fähigkeiten erreicht[1409] und durch überragende Algorithmen und Datensets keine wettbewerblichen Vorsprünge erlangt werden können, sondern vielmehr innovative Geschäftsmodelle und Gestaltungen der Nutzeroberfläche im Wettbewerb vorteilhaft sind. Nach verbreiteter Ansicht bleibt der Zugang zu Daten trotzdem die entscheidende Hürde zur Aufnahme von KI-Entwicklungsaktivitäten.[1410] Je komplexer und vielschichtiger das Entwicklungsziel ist, desto mehr Daten werden grundsätzlich benötigt, um einen Feedback-Effekt zu starten. Ob aufstrebende Unternehmen tatsächlich von Anfang an generelle, hochkomplexe Systeme für breite Nutzergruppen entwickeln werden, ist zu bezweifeln.

IV. Daten als Marktzutrittsschranken (Barriers to Entry)

Bevor ein Unternehmen an dem Wettbewerb auf einem Markt teilnehmen kann, muss es ihm betreten können. Jeder Markteintritt setzt den Erwerb oder Aufbau der in diesem Markt notwendigen Infrastruktur und der dazugehörigen Fähigkeiten voraus.[1411] Die Position in Politik und Literatur, die ein Bestehen starker Datennetzwerkeffekte annimmt, bewertet diese als strukturelle Marktzutrittsschranke. Ein durch den „Datenhunger" selbstlernender Systeme verursachtes Hindernis beruhe auf technologischen Charakteristika des Marktes und sei damit struktureller Natur.[1412] Die Beurteilung von Marktzutrittshürden wird relevant für die Beurteilung von Zusammenschlüssen und marktmächtigen oder marktstarken Stellungen in Missbrauchsfällen (vgl. § 18 Abs. 3 Nr. 5 GWB). Im weitesten

---

1409 Vgl. *Aghion/Jones/Jones*, Artifical Intelligence and Economic Growth, S. 27.

1410 *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 29: „the most important candidate for a competition bottleneck in AI is anonymous access to individual-level data".

1411 Vgl. *Colangelo/Maggiolino*, ECJ, Vol. 13, S. 249–281, 259 (2017) und Fn. 30.

1412 Zur Definition: *BKartA*, Leitfaden zur Marktbeherrschung in der Fusionskontrolle, 29. März 2012, Rn. 64; LMRKM-*Kahlenberg*, Kartellrecht, § 36 GWB, Rn. 77.

Sinne rechtfertigt das Bestehen signifikanter Markzutrittsschranken eine besondere Aufmerksamkeit für einen Markt oder Sektor. Wenn Marktzutrittsschranken als notwendige Konsequenz von Datennetzwerkeffekten nachgewiesen werden, dürfte es für behördliche Entscheidungen genügen, Datennetzwerkeffekte nachzuweisen. Umgekehrt kann dieser Schluss aber auch nur mit dem Nachweis eines eindeutigen Zusammenhangs gelten.

Eine Marktzutrittsschranke ist jeder (finanzielle) Aufwand, den ein Markteintreter hat, der für etablierte Marktteilnehmer nicht anfällt. Diese Umschreibung ist allerdings nur eine Annäherung an den Begriff, nachdem in den letzten Jahren die theoretischen Streitigkeiten um eine Definition aufgegeben wurden.[1413] Das Bundeskartellamt hat zwei Bedingungen formuliert, bei deren kumuliertem Eintreten Daten als Marktzutrittsbarrieren gelten:[1414] Erstens muss der Zugang zu den bestimmten Daten wichtig sein, um auf einem Markt erfolgreich tätig zu werden. Zweitens dürfen andere Marktteilnehmer nicht in der Lage sein, die entsprechenden Daten wie die etablierten Unternehmen selbst zu erheben oder sich über Dritte Zugang zu ihnen zu verschaffen.[1415] Die „Wichtigkeit" als Kriterium der ersten Voraussetzung scheint weniger strenge Anforderungen an die Erforderlichkeit des Zugangsgegenstands zu stellen als „Unentbehrlichkeit" als Kriterium der Essential-Facilities-Doktrin.[1416] Um unerlässlich zu sein, muss die Ressource ein zwingend erforderlicher Bestandteil des auf dem nachgelagerten Markt angebotenen Endproduktes ein.[1417]

Problematisch erscheint die Abgrenzung von „entsprechenden" Daten in diesem Schema. Wegen der Heterogenität und Austauschbarkeit von Daten muss den potentiellen Wettbewerbern zugemutet werden, ähnliche Datensets zu erwerben oder zusammenzustellen. Der dafür zu unternehmende Aufwand darf den des etablierten Unternehmens übersteigen, der meist ohnehin nicht finanziell oder zeitlich zu beziffern ist. Richtigerweise ist eine Betrachtung der Gesamtumstände nötig.[1418] Bei zu enger Betrach-

1413 So *OECD*, Competition and Barriers to Entry, Policy Brief Januar 2007, S. 2.

1414 *BKartA*, Big Data und Wettbewerb, S. 7; *G7 Competition Authorities*, Common Understanding on "Competition and the Digital Economy" 5. Juli 2019, S. 4.

1415 *BKartA*, Big Data und Wettbewerb, S. 7.

1416 *J. Weber*, Zugang zu den Softwarekomponenten der Suchmaschine Google, S. 228.

1417 Vgl. Immenga/Mestmäcker-*Fuchs/Möschel*, Art. 102 AEUV, Rn. 336 mwN.

1418 Vgl. Begründung der Bundesregierung zum Entwurf eines Neunten Gesetzes zur Änderung des Gesetzes gegen Wettbewerbsbeschränkungen, BT-Drucks. 18/10207, S. 51; *Argenton/Prüfer*, Journal of Competition Law & Economics, Vol. 8, No. 1, S. 73–105 (2012).

tung könnte eine Datensammlung schon dann eine Marktzutrittsschranke darstellen, wenn ihre Zusammenstellung für das etablierte Unternehmen günstiger war als für Konkurrenten.[1419] Auf jedem Markt müssen potentielle Markteintreter die nötige Infrastruktur und Management- und Ingenieurfähigkeiten erlangen.

Wernerfelt stellte 1984 einen „Resource-Based View" vor:[1420] Die Kontrolle einer unverzichtbaren und nicht imitierbaren Ressource sei eine besonders effektive Marktzutrittshürde. Ein Unternehmen mit einer solchen Ressource sei sicher vor neuen Markteintretern, die diese Ressource nicht erlangen können.[1421] Zu diesen zählen für Wernerfelt auch Produktionserfahrungen, also Lerneffekte.[1422] Diese Ansicht wurde in den Folgejahren angezweifelt: In Zeiten der Digitalisierung und internetbasierter Geschäftsmodelle seien nachhaltige, ressourcenbegründete Wettbewerbsvorteile eine Illusion.[1423]

Es steht fest, dass individuelle Datensets eine kritische Inputressource für die Entwicklung selbstlernender Systeme sein können.[1424] Sind diese nicht zugänglich oder nicht substituierbar[1425], kann für potentielle Wettbewerber der Weg zur Entwicklung spezieller selbstlernender Systeme abgeschnitten sein.[1426] Der Zugriff auf eine überlegene Datensammlung ist dann problematisch, wenn sich der abgeleitete Wettbewerbsvorteil mit einem potentiellen Datennetzwerkeffekt rückverstärkt.[1427]

1419 *Auer/Manne/Portuese/Schrepel*, Why sound law and economics should guide competi-tion policy in the digital economy, ICLE, September 2018, S. 9; *Mahnke*, Big Data as a Barrier to Entry, CPI Antitrust Chronicle May 2015; *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 12; *Rubinfeld/M. Gal*, Access Barriers to Big Data.

1420 *Wernerfelt*, Strategic Management Journal, Vol. 5, No. 2, S. 171–180 (1984).

1421 *Wernerfelt*, Strategic Management Journal, Vol. 5, No. 2, S. 171–180, 171ff (1984); dazu *Parker/Alstyne/Choudary*, Platform Revolution, S. 209.

1422 *Wernerfelt*, Strategic Management Journal, Vol. 5, No. 2, S. 171–180, 174 (1984).

1423 *Parker/Alstyne/Choudary*, Platform Revolution, S. 209 mwN.

1424 Siehe Kapitel 4 A.III.3. Voraussetzungen zur Entwicklung von selbstlernenden Systemen, S. 221.

1425 Formulierung nach *BKartA*, Leitfaden zur Marktbeherrschung in der Fusionskontrolle, 29. März 2012, Rn. 66 und Fn. 97 und *Schweitzer/Haucap/Kerber/Welker*, Modernisierung der Missbrauchsaufsicht, S. 79.

1426 Vgl. *Himel/Seamans,* Artificial Intelligence, Incentives to Innovate, And Competition Policy, CPI Antitrust Chronicle December 2017, S. 1.

1427 *Krämer,* Herausforderungen bei der Bestimmung von Marktmacht in digitalen Märkten, Wirtschaftsdienst 2016, Heft 4: Wettbewerbspolitik in der digitalen Wirtschaft, S. 231–235 (235).

Der Zusammenhang wird von einer Zahl von Variablen infrage gestellt, beispielsweise dem zügigen Verfall der Daten, ihrem abnehmenden Grenznutzen bei der Auswertung und fehlender Transfermöglichkeit[1428] auf neue Kontexte, sowie von Faktoren wie der Ubiquität und Nicht-Rivalität von Daten relativiert. Zudem können auch winzige Datensätze anhand ihres konkreten Inhalts in wettbewerbspolitisch negativer Weise eingesetzt werden.

Denkbar ist, dass für datenbasierte Dienste jeweils eine „mindestoptimale Datenmenge" zu einem bestimmten Zeitpunkt besteht, die eine Voraussetzung für die Entwicklung eines effizienten Angebotes ist.[1429] Die mindestoptimale Datenmenge muss markt- und anwendungsspezifisch ermittelt werden. Dewenter illustriert dies am Beispiel einer Navigationsapp, die zur effizienten Stauvorhersage in einer großen Stadt eine höhere Nutzerzahl benötigt als in einer kleinen Stadt. Im Vergleich dazu wäre für intelligente Spracherkennungsdienste wegen des dynamischen Charakters von Sprache auch bei sehr großen Datenmassen ein Nutzenzuwachs spürbar. Werden für die Erstellung eines Dienstes nur einige tausend Datensätze benötigt, ist er meist problemlos realisierbar. Für die Bereitstellung eines komplexeren Dienstes, der mehrere hunderttausend Datensätze benötigt, sind die Hürden höher.[1430]

Zu anwendungsspezifischen Schwankungen von mindestoptimalen Datenmengen kommt, dass Daten nicht von homogener Qualität sind und die Verarbeitung von in großen Datenmassen enthaltenen „schlechten" Daten die Verarbeitungskosten entgegen von Skaleneffekten erhöht.[1431] Diese Zusammenhänge können in ihrer Komplexität und Dynamik kaum gesetzgeberisch oder behördlich in kurzer Zeit beurteilt werden – einerseits wird die technische Expertise fehlen, andererseits ist die Kalkulation hoch prognostisch.[1432] Lambrecht und Tucker haben ein Modell von

---

1428 Vgl. *Mateos-Garcia*, The Complex Economics of Artificial Intelligence, S. 8.

1429 Zum Folgenden: *Dewenter*, Digitale Ökonomie – Herausforderungen für die Wettbewerbspolitik, Wirtschaftsdienst 2016, Heft 4: Wettbewerbspolitik in der digitalen Wirtschaft, S. 236–239 (238).

1430 *Dewenter/Lüth*, Big Data aus wettbewerblicher Sicht, Wirtschaftsdienst 2016, S. 648–654 (652).

1431 *Rubinfeld/M. Gal*, Access Barriers to Big Data, S. 16.

1432 Für eine Einzelfallbetrachtung: *Dewenter*, Digitale Ökonomie – Herausforderungen für die Wettbewerbspolitik, Wirtschaftsdienst 2016, Heft 4: Wettbewerbspolitik in der digitalen Wirtschaft, S. 236–239 (238); *OECD*, Competition and Barriers to Entry, Policy Brief Januar 2007, S. 3.

Barney[1433] untersucht, das anhand von vier Bedingungen Wettbewerbsvorteile untersucht und möglicherweise von Behörden zur Beurteilung des Vorliegens von Marktzutrittsschranken genutzt werden könnte: Eine Ressource müsste danach ökonomisch wertvoll, selten, schwer zu imitieren und nicht oder nur schwer substituierbar sein (Valuable, Rare, Inimitable, Non-Substitutable[1434]).[1435] Lambrecht und Tucker kamen zu dem Schluss, dass Big Data nicht unverzichtbar und damit kein nachhaltiger Wettbewerbsvorteil oder „Barrier to Entry" seien.[1436] Zum Aufbau eines Wettbewerbsvorteils genügten nicht Datenmassen, sondern es seien wertschöpfende Nutzungsmöglichkeiten und spezialisierte Arbeitskräfte zu finden.[1437] Kritisiert wurde ihre Bewertung dahingehend, als dass die einzelnen Kriterien eher für spezielle Datensets und spezifische Anwendungszwecke untersucht werden sollten.[1438] Nicht außer Acht zu lassen ist auch der Umstand, dass das Ansammeln von Daten durch bestehende Marktteilnehmer den Markteintritt für potentielle Wettbewerber auf Haupt- oder nachgelagerten Märkten erst ermöglichen oder erleichtern kann.[1439] Manne et al. äußern ausdrückliche Kritik an der generellen Bezeichnung großer Datensets als „Barriers to entry".[1440] Die Annahme, dass Daten Marktzutrittsbarrieren seien, würde, wenn Daten zu Verbesserung von Produkten und Diensten genutzt werden, ultimativ bedeuten, dass die Produktoptimierung ein Hindernis darstelle.[1441] Im Ergebnis würde sich das Wettbewerbsrecht gegen das Ergebnis des Produktwettbewerbs richten.[1442] Der Punkt, ab dem das Volumen eines Datensets als Markt-

---

1433 *Barney*, Journal of Management, Vol. 17, No. 1, S. 99–120 (1991).

1434 Das Akronym dieser Kriterien ergibt die übliche Bezeichnung als VRIN-Kriterien.

1435 *Barney*, Journal of Management, Vol. 17, No. 1, S. 99–120, 99 (1991).

1436 *Lambrecht/C. Tucker*, Can Big Data Protect a Firm from Competition?, CPI Antitrust Chronicle Januar 2017, S. 8; in drei Anwendungsfällen untersucht: *Del Toro Barba*, ORDO 68, S. 217–248 (2017).

1437 *Lambrecht/C. Tucker*, Can Big Data Protect a Firm from Competition?, CPI Antitrust Chronicle Januar 2017, S. 8; so *D. Tucker/Wellford*, Big Mistakes Regarding Big Data, Antitrust Source, Dezember 2014, S. 9.

1438 *Bourreau/de Streel*, Digital Conglomerates and EU Competition Policy, S. 28.

1439 *Auer/Manne/Portuese/Schrepel*, Why sound law and economics should guide competition policy in the digital economy, ICLE, September 2018, S. 9.

1440 *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 26.

1441 Vgl. *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 27.

1442 „Je besser das etablierte Unternehmen ist, desto schwerer ist der erfolgreiche Markteintritt.", *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 27.

eintrittsbarriere beurteilt wird, wäre zudem willkürlich.[1443] Die Überlegenheit von Firmen, die große Mengen von Daten generieren, dürfe nicht undifferenziert als Marktzutrittsschranke bezeichnet werden; außerdem wäre eine darauf reagierende Datenteilungspflicht eine Strafe für solche erfolgreichen Geschäftsmodelle.

Die Kritik von Tucker und Wellford an der Bezeichnung von Daten als „Barriers to Entry" trägt ähnliche Argumente vor: Sie könnten wegen ihrer Non-Rivalität und der fehlenden semantischen Exklusivität keinen Marktzutritt erschweren.[1444] Die bloße Tatsache, dass Wettbewerber auf große Datensammlungen zugreifen könnten, bedeute nicht, dass diese Daten wirklich für einen erfolgreichen Markteintritt nötig seien.[1445] Außerdem spricht gegen die Klassifizierung von Daten als Marktzutrittshürden, dass kleineren Unternehmen andere Formen des mittelbaren und mittelbaren Datenzugriffs offen stünden.[1446] Schließlich könnte auch ein Datenzwischenhändler Zutrittshürden überwinden, indem er die Nachfrage nach Daten bündelt und zentral Daten sammelt, um sie Markteintretern weiterzugeben.[1447]

Die vermittelnde Ansicht dürfte besagen, dass den meisten datenbasierten Märkten Konzentration inhärent sei und hinsichtlich der Marktzutrittsschranken jeweils Einzelfallprüfungen notwendig seien.[1448] Tatsächlich sind viele Markteintritte von Startups mit datenbasierten Geschäftsmodellen zu beobachten, was für einige ein Beweis für niedrige, leicht überwindbare Marktzutrittshürden ist.[1449] Dieser Zusammenhang ist allerdings vereinfachend. Die Einsatzmöglichkeiten von Datenanalyse-Anwendungen und selbstlernenden Systemen sind vielfältig. Zudem entsprechen bloße Marktzutritte noch keinem langfristigen intensiven (Innovations-)Wettbewerb.

Auf die Frage, ob Daten Marktzutrittshürden sind, wird sich wohl keine definitive Antwort finden lassen. Wie für jede andere Ressource gilt, ist

---

1443 "Essentially arbitrary": *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 27.

1444 *D. Tucker/Wellford*, Big Mistakes Regarding Big Data, Antitrust Source, Dezember 2014, S. 6ff.

1445 *D. Tucker/Wellford*, Big Mistakes Regarding Big Data, Antitrust Source, Dezember 2014, S. 7.

1446 *Schweitzer/Peitz*, Datenmärkte in der digitalisierten Wirtschaft, S. 88.

1447 *Rubinfeld/M. Gal*, Access Barriers to Big Data, S. 17.

1448 *Stucke/Grunes*, Big Data and Competition Policy, S. 337.

1449 Z. B. *D. Tucker/Wellford*, Big Mistakes Regarding Big Data, Antitrust Source, Dezember 2014, S. 8f.

es nicht auszuschließen, dass in bestimmten Konstellationen die Unverfügbarkeit Marktzutritte verhindert. Die Heterogenität von Daten und die Diversität der Anwendungsmöglichkeiten erfordern eine komplexe Einzelfallbeurteilung. Tatsächlich genießen bereits in ein System integrierte Unternehmen privilegierten Zugang zu relevanten Datensets, um daran anknüpfende Anwendungen zu entwickeln und ihre Geschäftsmodelle im Sinne von Economies of Scope auszudehnen.[1450]

Stattdessen bietet sich die Bezeichnung als „Kaltstart-Problematik" an.[1451] Fehlt einem neuen Marktteilnehmer der Zugang zu mindestoptimalen Datenmengen, kann ihn dies an dem zügigen Fortschritt bei der Entwicklung seines selbstlernenden Systems hindern. Ähnlich wie ein Kaltstart beim Auto den Motor in höherem Maße belastet und zu Verschleiß führt, ist in der Informatik der „Cold Start" bekannt, der auftritt, wenn datengetriebene Empfehlungssysteme zu wenig trainiert sind, um akkurate Ausgaben zu produzieren. Mit zu wenig Daten ist das System „noch nicht auf Betriebstemperatur". Die Analyse einer größeren Menge spezifischer Nutzungsdaten erlaubt ebenso wie eine höhere Temperatur eines Motors bessere Bedingungen, um das System einwandfrei laufen zu lassen und präzise Ergebnisse auszugeben. Für gewisse selbstlernende Systeme kommen zur Überwindung einer Kaltstart-Problematik synthetische Trainingsdaten infrage.[1452]

Im Ergebnis dürfte für solche Unternehmen, die ohne Zugang zu relevanten Datensets selbstlernende Systeme entwickeln, in der Regel ein Kaltstartproblem zu bejahen sein. Eine Marktzutrittsschranke darf aber auf keinen Fall ohne Einzelfallprüfung angenommen werden. Daten konstituieren nur dann eine Marktzutrittsschranke, wenn sie kumulativ wesentlich[1453], einzigartig, exklusiv und rival sind.[1454] In der Regel sollte dabei das Volumen des Datensets anderen Faktoren wie der Qualität, der Repli-

1450 Zu Economies of Scope: *BMWi*, Wettbewerbsrecht 4.0, S. 14, mit unrichtigem Bezug zu „datengetriebenen indirekten Netzwerkeffekten"; *Crémer/de Montjoye/Schweitzer*, Competition policy for the digital era, S. 33.

1451 Z. B. *Gefferie,* The Cold Start Problem with Artificial Intelligence, 6. März 2018; Lösungen für ähnliche Situationen werden auch als „Bootstrapping" bezeichnet.

1452 Siehe S. 239; sowie *Nisselson*, Deep Learning with Synthetic Data Will Democratize the Tech Industry, TechCrunch.

1453 Zu diesem Kriterium: *FTC*, Statement Concerning Google/DoubleClick, FTC File No. 071–0170, S. 12.

1454 Vgl. *Manne/Morris/Stout/Auer*, Comments of the ICLE, 7. Januar 2019, S. 30: „What seems to be required in order that data may be treated as a potential entry barrier is that the data at issue be some combination of essential, unique,

zierbarkeit, der Eingliederung in den Kontext und dem Entwicklungsziel gleichgestellt werden.

V. Zwischenergebnis

Einige datenreiche Unternehmen innovieren permanent, andere entwickeln zwar neue Produkte, aber geben dabei die Richtung der Entwicklung vor, und wieder andere schaffen keine Innovationen.[1455] Innovationen ist grundsätzlich das Potential inhärent, Monopolisierungstendenzen zu realisieren – ebendieses Potential setzt Investitionsanreize. Die Konzentration von Innovationsressourcen auf wenige Anbieter senkt theoretisch die Anreize, weiter Innovationen zu schaffen, wenn zu erwarten ist, dass andere Unternehmen nicht die Mittel haben, aufzuholen. Dass sich diese Einstellung unter datenreichen Unternehmen durchsetzt, ist wegen des dynamischen Charakters der von Digitalisierung geprägten Sektoren zu bezweifeln. Nicht nur ein im Hintergrund arbeitender Algorithmus und seine Trainingserfahrung machen ein gutes Produkt aus: Verschiedene Faktoren wirken zusammen, sodass selten die Verfügbarkeit von Daten als entscheidender Erfolgsfaktor isoliert werden kann. Insgesamt werden mehr und qualitativ bessere Daten generiert,[1456] was zu einer höheren Zahl datenbasierter Innovationen führen dürfte.

Es wirkt beliebig, solche Monopolisierungstendenzen, für die bis vor kurzem Netzwerkeffekte und Plattformmärkte verantwortlich gemacht wurden, nun Datennetzwerkeffekten zuzuschreiben. Das Bestehen von Datennetzwerkeffekten erscheint zwar intuitiv logisch, aber es darf nicht außer Acht gelassen werden, dass die theoretische und empirische Literatur dazu aktuell recht dünn ist und sich noch nicht den gegenläufigen Effekten gewidmet hat. Trotzdem ist es richtig, dass für dieses prognostizierte Phänomen eine Bezeichnung gefunden wurde.[1457] Die Beobachtung und

---

exclusive, and rivalrous."; ähnlich *Nuys*, WuW 2016, 512 (514): „einzelfallbezogen zu prüfen".

1455 *Stucke*, Georgetown Law Technology Review, Vol. 2.2, S. 275–324, 304 (2018).

1456 *Lundqvist*, Big Data, Open Data, Privacy Regulations, Intellectual Propery and Competition Law in an Internet of Things World – The Issue of Access S. 26; *Rusche*, Intereconomics, Vol. 54, No. 2, S. 114–119 (2019).

1457 Dagegen: *Varian*, Artificial Intelligence, Economics, and Industrial Organization, S. 15; sowie *Varian/Dolmans/Baird/Senges*, Digitale Herausforderungen für die Wettbewerbspolitik, in: Wirtschaftsrat der CDU (Hrsg.), Soziale Marktwirtschaft im digitalen Zeitalter, S. 75–92 (75ff).

Untersuchung wird dadurch erleichtert und die Problematik wird greifbarer. Viel zu kurz kommt in Diskussionen allerdings die Heterogenität von Daten sowie die Tatsache, dass es ultimativ auf eine Vielfalt und Masse von Informationen und nicht auf einzelne Signale ankommt. Zudem ist ein „überragendes" selbstlernendes System nicht in überragende Marktanteile zu übersetzen. Selten ist das selbstlernende System das Produkt, sondern es stützt ein Produkt, beispielsweise ein Empfehlungssystem als Bestandteil eines Online-Shops. Es ist von außen nicht erkennbar, ob und zur Erfüllung welcher Aufgaben ein Unternehmen ein selbstlernendes System implementiert hat; dies erschwert den Vergleich des Fortschritts von Wettbewerbern.[1458]

F. Zwischenergebnis – Auswirkungen selbstlernender Systeme auf den Innovationswettbewerb

Wenn Schumpeters Thesen zutreffen, ist die potentielle Monopolstellung eines datenreichen Unternehmens kein Grund zur Besorgnis, weil es selbst aus Angst vor dem Verlust seiner Marktstellung weiter innoviert. Gilt das allerdings auch, wenn ihm klar ist, dass anderen der Zugang zu einem dafür nötigen „Rohstoff" fehlt? Dann wären die Angst vor Überholung neutralisiert und Innovationsanreize verringert. In der Zukunft werden sich alle Wirtschaftssektoren dahingehend entwickeln, dass sie entweder datengetrieben oder datengestützt sind. Eine ausschließlich datengetriebene Marktmacht wäre dann vergänglich.

„The value isn't your algorithm, it's your data"[1459]: Was dieses und ähnliche Zitate zum Ausdruck bringen sollen, ist die Unabdingbarkeit und der überragende Wert von Daten bei der Entwicklung selbstlernender Systeme. Die Vielfalt von Herangehensweisen und Entwicklungszielen verbietet aber eine derart schwarz-weiße Einordnung. Der richtige Algorithmus (oder Data Scientist) kann aus schlechten oder kleinen Datensets den größtmöglichen Wert schöpfen, während ein ungeeigneter, unüberwachter Algorithmus mit großen und guten Datensets unzureichende Lernerfolge hervorbringt. Möglicherweise sind die derzeit erfassten Daten ungleich verteilt, nicht aber der Zugang zu Instrumenten der Datenerfassung

1458 *Mateos-Garcia*, The Complex Economics of Artificial Intelligence, S. 9.

1459 Z. B. *Staun-Olsen*, The Value Isn't Your Algorithm, It's Your Data, 3. November 2016; *Wertz*, Data, Not Algorithms Is Key to Machine Learning Success, 6. Januar 2016.

– praktisch jeder kann Sensoren erwerben, um Echtzeitdaten zu erfassen. Für viele Anwendungen liegt der Wert der Daten in ihrer Aktualität und sofortigen Verfügbarkeit.[1460] Schon nach kurzer Zeit verlieren diese Daten an Relevanz, weshalb historische Daten in der Regel von geringerem Wert sind als aktuelle Daten. Techniken, Produktionsformen und Maschinen der traditionellen Industrie altern ebenfalls, allerdings in vergleichsweise niedriger Geschwindigkeit. Während sich die industriellen Produktionsprozesse häufig sprunghaft („revolutionär") entwickeln, sollten die Rechtsetzung und -durchsetzung eine sprunghafte Entwicklung scheuen. Die Entwicklung zur Industrie 4.0 benötigt einerseits Rechtssicherheit sowie faire und belastbare Wettbewerbsbedingungen. Andererseits darf der diesen Entwicklungen bereits innewohnende Innovationsdruck nicht unterschätzt werden. Die Diversität der mit KI arbeitenden Unternehmen könnte nicht größer sein. Trotzdem ist in Zukunft sicherzustellen, dass sowohl große Unternehmen als auch potentielle Markteintreter Zugang zu den Daten haben, die ihnen KI-Innovationen ermöglichen.[1461]

Daten können – wie finanzielle Mittel und Talent – ein wichtiges Instrument im Innovationswettbewerb sein, aber sie sind jeweils nie allein ausreichend zur Entwicklung neuer Produkte oder Prozesse. Der Zugang zu Daten ist ein Wettbewerbsvorteil gegenüber fehlendem Zugang. Dass er quasi-automatisch eintritt, wie ein Datennetzwerkeffekt vorgibt, ist jedoch grundlegend falsch. Inwiefern dieser Wettbewerbsvorteil innovationshemmend oder regulierungsbedürftig ist, ist eine weitere Frage, die im nächsten Kapitel betrachtet wird.

Im Ergebnis orientiert sich Künstliche Intelligenz an menschlicher Intelligenz, ohne deren natürliche Begrenzungen aufzuweisen. Überdurchschnittlich kompetente Arbeitnehmer, moderne Fabriken oder zu große Lagerhallen müssen nicht mit anderen Unternehmen geteilt werden. Hinzukommen muss also eine spezielle, innovationsstimulierende Rechtfertigung: Diejenigen, die die Sensoren der Industrie 4.0 und die generierten Daten kontrollieren, entnehmen den Daten möglicherweise nicht sämtliche – den gesellschaftlichen Fortschritt stützende – Informationen, was ein Grund für eine Öffnung der Zugänge sei könnte.[1462] Auch aus volkswirt-

1460 *Duch-Brown/Martens/Müller-Langer*, The economics of ownership, access and trade in digital data, S. 14.

1461 *Himel/Seamans,* Artificial Intelligence, Incentives to Innovate, And Competition Policy, CPI Antitrust Chronicle December 2017, S. 2; *Rusche*, Intereconomics, Vol. 54, No. 2, S. 114–119, 119 (2019).

1462 *Goodman*, The Atomic Age of Data, S. 14.

schaftlichen Gründen gilt es, das Potential der Technologie vollumfänglich zu nutzen.