MATTHIAS LEISTNER
LUCIE ANTOINE
THOMAS SAGSTETTER

Big Data

Geistiges Eigentum
und Wettbewerbsrecht

Mohr Siebeck

Geistiges Eigentum und Wettbewerbsrecht

herausgegeben von

Peter Heermann, Diethelm Klippel,
Ansgar Ohly und Olaf Sosnitza

162

Matthias Leistner, Lucie Antoine
und Thomas Sagstetter

Big Data

Rahmenbedingungen im europäischen Datenschutz- und Immaterialgüterrecht und übergreifende Reformperspektive

Mohr Siebeck

Matthias Leistner ist Inhaber des Lehrstuhls für Bürgerliches Recht und Recht des Geistigen Eigentums mit Informations- und IT-Recht (GRUR-Lehrstuhl) an der LMU München.

Lucie Antoine ist wissenschaftliche Mitarbeiterin am Lehrstuhl für Bürgerliches Recht und Recht des Geistigen Eigentums mit Informationsrecht und IT-Recht (GRUR-Lehrstuhl) an der LMU München.

Thomas Sagstetter ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Bürgerliches Recht und Recht des Geistigen Eigentums mit Informationsrecht und IT-Recht (GRUR-Lehrstuhl) an der LMU München.

ISBN 978-3-16-160145-3 / eISBN 978-3-16-160198-9
DOI 10.1628/978-3-16-160198-9

ISSN 1860-7306 / eISSN 2569-3956 (Geistiges Eigentum und Wettbewerbsrecht)

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind über *http://dnb.dnb.de* abrufbar.

Das Buch wurde von Gulde Druck in Tübingen aus der Times gesetzt, auf alterungsbeständiges Werkdruckpapier gedruckt und gebunden.

Printed in Germany.

Vorwort

Die regulatorischen Herausforderungen der Datenökonomie sind in aller Munde. Dabei sind voreilige Forderungen nach der Schaffung neuer Exklusivrechte an Daten („Dateneigentum“) zwischenzeitlich zu Recht verhallt. Die aktuelle Diskussion wird eher vom Aspekt des Datenzugangs und insoweit von zunehmend konkreteren, präziseren Forschungsarbeiten und rechtspolitischen Entwicklungen bestimmt.

Hier reiht sich die vorliegende Studie ein, die aus immaterialgüter- und datenschutzrechtlicher Perspektive einen Beitrag zur Konkretisierung und Präzisierung der Diskussion liefern will. Einerseits wird Handwerksarbeit geleistet und der Stand der lex lata sowie aktueller Reformprojekte umfassend aufbereitet. Andererseits werden auf dieser Basis weitere Reformperspektiven aufgezeigt, die von den Leitgedanken vielfältiger Innovation durch wirksamen Wettbewerb sowie materieller Privatautonomie (einschließlich des Schutzes der Privatsphäre) auf der Basis der einschlägigen Grundrechtspositionen bestimmt sind. Ein besonderes Anliegen war es uns, den aus unserer Sicht für den immaterialgüterrechtlichen Rahmen der Datenökonomie zentral wesentlichen Geschäftsgeheimnisschutz mit einzubeziehen; entsprechend werden das GeschGehG und die EU-Geschäftsgeheimnis-Richtlinie in ihren für Daten als Schutzgegenstand wesentlichen Elementen ausführlich mit behandelt. Ebenso wichtig war es uns, den Einfluss der Datenschutzgrundverordnung auf die Entwicklung eines freien, unverzerrten und innovativen Wettbewerbs in den europäischen Datenmärkten praxisorientiert und differenziert zu diskutieren. Auch konnten wir zuletzt noch die aktuellen Vorschläge der Kommission für den Data Governance Act, den Digital Markets Act und den Digital Services Act mit einbeziehen.

Die Studie ist ein echtes Lehrstuhlprojekt und Gemeinschaftswerk. Auf der Grundlage einzelner Vorarbeiten haben die Autoren die einzelnen Teile unter der Federführung von *Leistner* gemeinsam bearbeitet, wobei *Antoine* schwerpunktmäßig den Teil zum Datenschutzrecht entworfen hat, *Sagstetter* schwerpunktmäßig den Teil zum Geschäftsgeheimnisschutz (und auch Abschnitte zum sui generis-Recht und zu den Grundlagen) und *Leistner* alle verbleibenden Teile, wobei ihm seine Mitarbeiter wiederum ihrerseits tatkräftig geholfen haben.

Zahlreiche Anregungen und auch Kritik, die uns in einzelnen Bereichen geholfen hat, die Gedanken schärfer zu konturieren, verdanken wir der kontinuierlichen Diskussion mit KollegInnen in Wissenschaft, Politik und Praxis. So zahlreich waren die Diskussionen in den vergangenen Jahren, so intensiv und ertragreich der Austausch, dass wir nicht einmal annähernd allen gebührend danken können. Immer wieder überaus fruchtbringend waren etwa der Dialog mit *Josef Drexl, Reto Hilty* und der gesamten „Datengruppe" des Max-Planck-Instituts für Innovation und Wettbewerb in München sowie der kontinuierliche inspirierende Austausch zu einzelnen Themen mit *Tanya Aplin, Jeanne Fromer, Jane Ginsburg, Pamela Samuelson, Heike Schweitzer, Thomas Ackermann, Barton Beebe, Lionel Bently, Henning Große Ruse-Khan, Bernt Hugenholtz, Wolfgang Kerber, Kung-Chung Liu, Axel Metzger, Ansgar Ohly, Rupprecht Podszun, Jule Sigall, Gerald Spindler, Christopher Sprigman, Yoshiuki Tamura, Tatsuhiro Ueno, Jacob Victor, Michael Weinberg, Peter Yu* und *Herbert Zech.* Auch allen anderen KollegInnen und Studierenden, mit denen wir uns in den letzten Jahren in ganz unterschiedlichen nationalen und internationalen Foren zu Einzelaspekten ausgetauscht haben, sei von Herzen für diese faszinierenden, zunehmend immer genaueren, tiefer gehenden Diskussionen zur Regulierung der Datenökonomie gedankt.

Großer Dank gilt schließlich allen wissenschaftlichen und studentischen MitarbeiterInnen des Lehrstuhls für Bürgerliches Recht und Recht des Geistigen Eigentums mit Informationsrecht und IT-Recht (GRUR-Lehrstuhl) an der LMU. Ohne die lebhafte, auch kontroverse, dabei stets anregende und freundschaftliche Diskussionskultur am Lehrstuhl (und darüber hinaus an der Juristischen Fakultät der LMU und am Münchener Max-Planck-Institut) und ohne die einsatzfreudige Mithilfe der LehrstuhlmitarbeiterInnen in Recherche- und Redaktionsangelegenheiten wäre dieses Buch nicht in seiner jetzigen Form zustande gekommen.

Abschließend sei den Herausgebern der Schriftenreihe „Geistiges Eigentum und Wettbewerbsrecht" für die Aufnahme in die Reihe und allen MitarbeiterInnen des Mohr Siebeck-Verlags für die sorgsame verlegerische Betreuung unseres Werks gedankt. Es ist für uns eine bewusste Entscheidung und besondere Freude gewesen, das Buch in einer so gediegenen, genuin wissenschaftlich geprägten Reihe zu veröffentlichen.

München, Januar 2021

Matthias Leistner
Lucie Antoine
Thomas Sagstetter

Inhaltsübersicht

Inhaltsverzeichnis

A. Einführung

Der angemessene und sichere *Zugang zu Daten* in hinreichender *Qualität* steht im Mittelpunkt der von Daten vorangetriebenen Innovation. Zugleich wirft der digitale Strukturwandel in Gestalt von big data diverse ethisch-philosophische, sozialwissenschaftliche, ökonomische, politische und rechtliche Fragen auf.

Die zugrundeliegenden Zielsetzungen sind zum Teil *komplementär*. So entspricht es dem heutigen Stand der Forschung, dass die Zielsetzung *freier Datenmärkte mit funktionierendem Wettbewerb* zur Zielsetzung der *Innovationsförderung*, idealerweise in einem *Komplementaritätsverhältnis* steht.[1] Hier sind mit Blick auf die Datenökonomie in erster Linie *funktionale Rahmenbedingungen* zu schaffen, die dieses Komplementaritätsverhältnis angemessen ausgestalten. Diese Aufgabe ist insbesondere für das Immaterialgüterrecht nicht neu – es geht also eher um eine *Anpassung des rechtlichen Rahmens* an neue, so bisher nicht dagewesene regulatorische Herausforderungen der Datenökonomie. Das betrifft insbesondere die Sicherstellung eines angemessenen *Zugangs zu notwendigen Daten* bei gleichzeitig hinreichender Wahrung tatsächlich notwendiger *Anreize*, um Innovation optimal zu ermöglichen, insbesondere aber auch um der Ausprägung marktübergreifender Machtpositionen vorzubeugen oder jedenfalls deren Missbrauch zu unterbinden.[2]

Darüber hinaus sind zweifellos auch regelrecht *divergierende Rechte und Interessen* zu angemessenem, verhältnismäßigem Ausgleich zu bringen. Das heißt, es ist anzuerkennen, dass bestimmte weitere verfassungsrechtlich abgesicherte, ethische und politische Zielsetzungen in einem Spannungsverhältnis zu Wettbewerbsfreiheit und angemessenen Innovationsanreizen stehen können. So hat ein

[1] Die Zielsetzung der *Innovationsförderung* kann es auch erforderlich machen, bestimmte Marktchancen mit Ausschließlichkeitscharakter (ausschließliche Rechte) oder auch in Form bestimmter Abwehrrechte (bloße rechtlich abgesicherte Freiheiten) einzelnen Marktteilnehmern zuzuweisen.

[2] An dieser Stelle ist naturgemäß ganz zentral auch das Kartellrecht angesprochen. Angesichts des in diesem Bereich auf deutscher und europäischer Ebene vergleichsweise schon überaus fortgeschrittenen Forschungsstandes wird das Kartellrecht in dieser Untersuchung als eigenständiger Gegenstand allerdings ausgespart, dennoch informiert die diesbezügliche Forschung natürlich ganz konkret auch viele der hier angestellten Überlegungen und entwickelten Vorschläge insbesondere im Immaterialgüterrecht, vgl. unten S. 35.

falsch verstandener *Datenschutz* durchaus das Potential, die von Daten vorangetriebene Innovation nachhaltig zu hemmen. Einerseits muss es hier wiederum darum gehen, die diesbezüglich dennoch vorhandenen Potentiale für mögliche *Konvergenzen* dieser Zielsetzungen durch entsprechend funktional ausgestaltete institutionelle Rahmenbedingungen zunächst einmal auszuschöpfen. Zum Beispiel kann ein angemessen und praxistauglich ausgestalteter Schutz persönlicher Daten zweifellos das Vertrauen der Datensubjekte in den rechtlichen Rahmen stärken und damit zur Bereitschaft beitragen, persönliche Daten für bestimmte Innovationszwecke überhaupt erst zur Verfügung zu stellen. Andererseits ist aber zugleich auch anzuerkennen, dass bestimmte verfassungsrechtlich abgesicherte Rechtspositionen gegebenenfalls auch dazu führen können, dass nicht allein ein *Maximum* an datengetriebener, effizienter Innovation anzustreben ist, sondern vielmehr ein *Optimum* angemessenen Ausgleichs zwischen den unterschiedlichen involvierten Rechten, Interessen und Werten im Sinne praktischer Konkordanz. Hier sind zugleich ersichtlich auch zahlreiche *weitere private und öffentliche Interessen* betroffen, wie etwa der Schutz vor *Diskriminierung*, *Informationsfreiheit* als Funktionsvoraussetzung demokratischer Willensbildungsprozesse, *Fairness* und *Vertrauen*.

Angesichts dieser komplexen Gemengelage herrscht über die richtigen regulatorischen Rahmenbedingungen derzeit in weiten Bereichen noch *Unsicherheit*. Diese beruht teilweise auf *fehlenden Informationen* über die Funktionsbedingungen neuer datengetriebener Innovationsräume und Geschäftsmodelle. Teilweise sind auch schlichtweg die *politischen Wertentscheidungen* noch nicht getroffen, die notwendig sind, um divergierende Rechte und Interessen in einen angemessenem Ausgleich zu bringen und dabei zugleich hinreichende *Handlungsspielräume* für Wettbewerb und Innovation zu sichern, die dann ihrerseits garantieren, dass die resultierenden Zielsetzungen auch effizient erreicht werden können. Hinzu kommt die Tatsache, dass sich die aufgeworfenen Rechtsfragen naturgemäß nicht *abstrakt allgemeingültig,* sondern vielmehr nur *konkret bereichsspezifisch* beantworten lassen werden.

Die derzeit demnach teilweise bestehende *Unsicherheit* über Chancen und Risiken der Datenökonomie und ihre zutreffende *bereichsspezifische Regulierung* ist kein rechtspolitisches Zukunftsproblem. Vielmehr spiegelt sie sich bereits konkret in erheblicher *Rechtsunsicherheit* auf Grundlage der lex lata überall dort, wo europäische Regelungsinstrumente auf dem Wege von unbestimmten Rechtsbegriffen und Generalklauseln der Rechtsprechung erhebliche Spielräume zur Rechtskonkretisierung und -fortentwicklung einräumen. Das betrifft insbesondere, aber nicht nur den datenschutzrechtlichen Rahmen für big data in Europa. Die Eigenheiten der Rechtsauslegung und Rechtsfortbildung im europäischen Mehrebenensystem führen dabei dazu, dass eine verlässliche Klärung der wesentlichs-

ten offenen Fragen unter diesen Bedingungen erhebliche Zeit beansprucht und häufig zudem – auf europäischer Ebene – eher lediglich allgemeine Leitlinien vorgegeben werden, die die Rechtsanwendung im konkreten Einzelfall unter Hinweis auf die allgemeinen Rechtsgrundsätze des Unionsrechts wiederum den mitgliedstaatlichen Gerichten überlassen. Tatsächlich kann angesichts bestehender informationeller Unsicherheiten eine derartige offene Regulierung im Wege flexibler *standards* statt konkreter und bestimmter *rules* auch in vielen Bereichen ersichtlich gerade der richtige regulatorische Ansatz sein. Zugleich ist aber im Interesse der dynamischen Entwicklung der datengetriebenen Innovation dafür zu sorgen, dass entsprechende Unsicherheiten nicht zu derart erheblichen *chilling effects* führen, dass diese – selbst im Verhältnis zu möglichen zukünftigen Effizienzgewinnen aufgrund solcherart abgesicherter Spielräume für die iterative Rechtsentwicklung – nicht mehr zu rechtfertigen wären.

Vor diesem Hintergrund setzt sich die vorliegende Untersuchung im Wesentlichen zwei Ziele. *Erstens* soll auf Grundlage der *lex lata* der regulatorische Rahmen für big data im europäischen Datenschutz- und Immaterialgüterrecht umfassend dargestellt und diskutiert werden, so dass eine entsprechend *verlässliche Informationsgrundlage bezüglich des bestehenden gesetzlichen Rahmens für big data* mit seinen schon de lege lata existierenden erheblichen *Gestaltungsspielräumen* zur Verfügung steht. Das soll auch dazu beitragen, durch Gewährleistung einer gewissen minimalen Berechenbarkeit die Möglichkeit für risikoaffine Marktteilnehmer zu stärken, bestimmte innovative Ansätze zum Ausgleich der Interessen in der Rechtsprechung „testen" zu lassen. Nur so können sich schließlich auf iterativem Wege in der Rechtsprechung neue, angemessene Lösungsansätze kristallisieren.

Auf der solcherart erarbeiteten Basis soll, *zweitens*, ein *Abgleich* der zwischenzeitlich in der Forschung in bestimmten Szenarien als besonders wesentlich für die Datenökonomie identifizierten und strukturierten Funktionsbedingungen eines angemessenen *Zugangs zu und der Teilung von Daten in hinreichender Qualität* mit den diesbezüglich relevanten datenschutz- und immaterialgüterrechtlichen Rahmenbedingungen erfolgen. Hinsichtlich dieser zweitgenannten Zielsetzung geht es um einen ganzheitlichen Ansatz im Sinne einer Art *„Hin- und Herwandern des Blicks"*: Einerseits werden Defizite des bestehenden Datenschutz- und Immaterialgüterrechts identifiziert und entsprechend *konkret dringliche Reformvorschläge* unterbreitet. Andererseits ist zugleich anzuerkennen, dass die künftige Regulierung der Datenökonomie Wertentscheidungen erfordert, die ihrerseits unvermeidlich von der objektiven Werteordnung des bestehenden Wettbewerbsrechts, des Vertragsrechts, aber auch des Datenschutz- und Immaterialgüterrechts und weiterer Rechtsgebiete informiert und beeinflusst werden. Das heißt, das das Datenschutz- und insbesondere das Immaterialgüter-

recht halten nach der hier zugrunde gelegten Untersuchungsprämisse schon heute bestimmte gesetzgeberische Wertentscheidungen, Strukturen und Instrumente bereit stellt, die ihrerseits für bestimmte Aspekte der künftigen Regulierung der Datenökonomie relevant sind, ja teilweise Vorbildcharakter haben oder zumindest bestimmte konkrete building blocks für bestimmte regelungsbedürftige Einzelszenarien bereitstellen könnten.

Der Schwerpunkt der Untersuchung wird dabei hinsichtlich des identifizierten kurzfristigen Reformbedarfs wesentlich beim sui generis-Schutzrecht für Datenbanken, bei bestimmten Aspekten des Schutzes von Geschäftsgeheimnissen sowie bei Einzelheiten des datenschutzrechtlichen Rahmens liegen. Demgegenüber erscheinen das allgemeine Urheberrecht sowie das Patentrecht in Europa (mit kleineren Einschränkungen) vergleichsweise fit für die Datenökonomie. Hier werden eher bestimmte Details angesprochen und angemahnt, die zum Teil auch in der Praxis der Patentämter und der Rechtsprechung hinreichend gelöst werden könnten.

B. Grundlagen

I. Strukturwandel durch big data und AI

1. Big data

Eine *allgemeingültige Definition* des Begriffs „big data" existiert nicht. Typischerweise wird auf die drei *Strukturmerkmale volume, velocity und variety* abgestellt.[1] Teilweise wird diese Definition noch um das Element der *veracity*, also Richtigkeit bzw. Zuverlässigkeit der Daten[2], sowie um weitere Elemente ergänzt.[3]

Die Strukturmerkmale des Umfangs der Datenmengen und der Verarbeitungsgeschwindigkeit verdeutlichen, dass es bei den neuen big data-Anwendungsmöglichkeiten in erster Linie um die Ableitung von Erkenntnisgewinn aufgrund der heute praktisch unbegrenzt *skalierbaren* Auswertung großer Datenmengen geht.[4] Die insoweit neue Qualität beruht in erster Linie auf dieser *Skalierbarkeit* der zugrundeliegenden Datenmengen aufgrund heute praktisch unbegrenzter Speicherkapazitäten in der Cloud und der Geschwindigkeit der Verarbeitung dieser Daten – gerade soweit der Datenauswertung klassische kybernetische Methoden der sechziger Jahre zugrunde liegen, aber auch im Bereich der auf mathematischen Weiterentwicklungen seit den achtziger Jahren beruhenden neueren Methoden der AI und insbesondere des machine learning (ML).[5] Das Element der

---

[1] Vgl. *Fasel/Meier*, in: Fasel/Meier (Hrsg.), Big Data. Grundlagen, Systeme und Nutzungspotenziale, S. 5–6; *Schütze/Hänold/Forgó*, in: Kolany-Raiser/Heil/Orwat et al. (Hrsg.), Big Data und Gesellschaft, Eine multidisziplinäre Annäherung, S. 233, 237 jeweils m. w. N.

[2] Zu den teilweise unklaren (Rechts-)Begriffen „Daten" und „Informationen" unten Fn. 33.

[3] Vgl. zum dem Element der *Veracity* sowie weiteren Begriffsergänzungen *Gervais*, JIPITEC 2019, 22, 23 m. w. N.

[4] Vgl. dazu *Sagstetter*, in: Husemann/Korves/Rosenkranz et al. (Hrsg.), Strukturwandel und Privatrecht, S. 249 Fn. 1 m. w. N. Zur Menge der weltweit produzierten Daten (33 Zettabyte im Jahr 2018; voraussichtlich 175 Zettabyte im Jahr 2025) jüngst etwa Mitteilung der Kommission v. 19.2.2020, Eine europäische Datenstrategie (im Folgenden Datenstrategie), COM(2020) 66 final, 2 m. w. N.

[5] Vgl. dazu *Sagstetter*, in: Husemann/Korves/Rosenkranz et al. (Hrsg.), Strukturwandel und Privatrecht, S. 251–252 m. w. N.

variety verdeutlicht, dass die Datengrundlage aus einer *Vielfalt vollkommen unterschiedlicher strukturierter, semi-strukturierter oder unstrukturierter Daten aus einer Vielzahl unterschiedlicher Quellen* stammen kann, die mit Hilfe teils hergebrachter, teils neuerer Methoden und Algorithmen ausgewertet wird.[6]

2. Soziale und ökonomische Funktion entscheidend für die rechtliche Strukturierung

Für die juristische Annäherung an das Phänomen big data sind dabei weniger die technischen Grundlagen, als vielmehr die *soziale und ökonomische Funktion* der technologischen Entwicklungen entscheidend[7]: Durch die Modellierung und Analyse großer und vielfältiger Datenmengen aus einer Vielzahl kreativ kombinierter Quellen sowie Auswertungsinstrumente und -parameter sollen *neue Erkenntnisse gewonnen* werden, die sich in Ansehung einzelner Datensätze so nicht erschlossen hätten.[8] Diese neuen Erkenntnisse können Ausgangspunkt der Entwicklung *neuer Dienste* und *Produkte* sein, die teils Effizienzgewinne ermöglichen (gezieltes Marketing, Bild- und Mustererkennungsverfahren in der Medizin und vielen anderen Bereichen), teils bestehende Bedürfnisse auf vollkommen neuartige, innovative Art und Weise befriedigen (smart grids, smart farming), damit neue Märkte erschließen und sich entsprechend *disruptiv auf bestehende Märkte* auswirken können.[9] Wegen des Strukturmerkmals der *variety*, der marktübergreifenden Vielfalt zugrundeliegender Daten und des – teils unvorhersehbaren – marktübergreifenden oder markteröffnenden Werts der gewonnenen neuen Informationen, können derartig gewonnene Erkenntnisse überdies Ausgangspunkt *neuartiger Verbundeffekte* sein, die *marktübergreifend* neue Räume für Innovation überhaupt erst eröffnen.[10]

Daten und Informationen sowie die *Instrumente zu ihrer Auswertung* werden damit in höherem Maße als bisher zu einem eigenständigen Marktfaktor; ein Marktfaktor, der wegen der hierdurch ermöglichten werthaltigen Einsatz- und Innovationsmöglichkeiten einen *eigenständigen signifikanten Marktwert* über die Generierung marktbezogenen Wissens hinaus aufweist.

---

[6] Vgl. dazu *Sagstetter*, in*:* Husemann/Korves/Rosenkranz et al. (Hrsg.), Strukturwandel und Privatrecht, S. 251–252; *Fasel/Meier*, in: Fasel/Meier (Hrsg.), Big Data. Grundlagen, Systeme und Nutzungspotenziale, S. 3, 5–6 jeweils m. w. N.

[7] S. dazu *Surblytė-Namavičienė*, S. 8 ff.

[8] Vgl. dazu *Sagstetter*, in: Husemann/Korves/Rosenkranz et al. (Hrsg.), Strukturwandel und Privatrecht, S. 249 Fn. 1 m. w. N.

[9] Vgl. etwa auch die zahlreichen Beispiele in der Datenstrategie, COM(2020) 66 final, 2 ff.

[10] Vgl. *Schweitzer,* GRUR 2019, 569 f.

Hinsichtlich der *Auswertungsinstrumente* in big data-Sachverhalten steht in der heutigen Diskussion der, wiederum schillernde und wenig trennscharfe[11], dabei zugleich linguistisch aus deutscher Sicht in die Irre führende[12] Begriff der *artificial intelligence* (AI) im Mittelpunkt. Die hierdurch bezeichneten Phänomene reichen selbst innerhalb des derzeit wesentlichsten innovativen Unterbereichs des sogenannten *machine learning* (ML) von klassischen lediglich weiterentwickelten *regelbasierten statistisch-empirischen Methoden* der Mustererkennung bis hin zu mehr oder weniger autonomem machine learning im Rahmen *neuronaler Netze*.[13] Hinzu kommen neue Methoden, die ebenfalls im Bereich des machine learning eingesetzt werden können, dabei aber über ein weiteres Einsatzfeld verfügen und damit den angestammten Bereich des machine learning sprengen, wie insbesondere *evolutionary algorithms* und verwandte Forschungsansätze.[14]

3. Datenzugang und Datenqualität als Bottleneck

Unabhängig von dieser Diversität der zugrundeliegenden Methoden ist angesichts der weitgehenden Verfügbarkeit von Rechenleistung das derzeit wesentlichste Bottleneck für Innovationen im Bereich des machine learning unstreitig der *Zugang* zu und die Generierung von *Trainingsdaten* in hinreichender *Qualität, Quantität* und *Varietät* sowie die innovative Entwicklung und Kombination entsprechender *Algorithmen,* um zweckgerichtete Analyseinstrumente zur Verfügung zu stellen und *biases*[15] zu vermeiden.[16] Hinsichtlich des Zugangs zu *Analy-*

---

[11] Vgl. *Drexl/Hilty/Beneke et al.*, Max Planck Institute for Innovation & Competition Research Paper No. 19-13, S. 1, 3; *Fink*, ZGE 2017, 288, 297: „*AI is whatever hasn't been done yet*"; *Sagstetter*, in: Husemann/Korves/Rosenkranz et al. (Hrsg.), Strukturwandel und Privatrecht, S. 249 Fn. 1 m. w. N.

[12] Suggeriert doch schon der Begriff der *intelligence* in der deutschen Sprache unvermeidlich einen engen Zusammenhang mit „Intelligenz", während der Begriff im Englischen sehr viel neutraler auch für „Informationsgewinnung" steht.

[13] Für einen guten besonders bündigen Überblick über die technischen Grundlagen vgl. *Drexl/Hilty/Beneke et al.*, Max Planck Institute for Innovation & Competition Research Paper No. 19-13, 1 f.; *Stiemerling*/*Ehinger*, CR 2018, 761 f.; *Stiemerling*, CR 2015, 762 f. Vgl. allgemein zu *machine learning* auch COM(2018), 237 final, S. 10.

[14] Vgl. etwa *Drexl/Hilty/Beneke et al.*, Max Planck Institute for Innovation & Competition Research Paper No. 19-13, S. 11 m. w. N.

[15] Vgl. nur *Lemley*/*Caseys*, Fair Learning – Draft 23.3.2020, S. 42 ff. m.w.N: „[…] *providing ML systems with broader access to data actually helps to mitigate some of the very negative outcomes that critics of ML systems fear.* […]"; vgl. auch *Executive Office of the President*, Preparing for the future of Artificial Intelligence, S. 30: „*AI needs good data. If the data is incomplete or biased, AI can exacerbate problems of bias.*".

[16] Vgl. zur Bedeutung der *Trainingsdaten* statt vieler *Drexl/Hilty/Beneke et al.*, Max Planck

seinstrumenten ist insoweit allerdings Differenzierung geboten: Zahlreiche Analyseinstrumente stehen inzwischen in der Cloud oder sogar allgemeinzugänglich im Internet wie eine Art „Toolbox" zur Verfügung und werden auch entsprechend beworben und unter Open Source-Bedingungen angeboten.[17] Dabei wird ihre Zurverfügungstellung seitens der Anbieter sogar häufig gleichsam als „Angel" für die Gewinnung zusätzlicher Daten seitens der Nutzer dieser Analyseinstrumente eingesetzt. Lediglich die dahinterstehenden Kombinationen von Algorithmen und Gewichtungsfaktoren sowie der Sonderbereich besonders hochwertiger, innovativer und spezifischer Algorithmen oder Algorithmen, deren Aufdeckung zu Manipulationsmöglichkeiten im Markt führen würde, werden derzeit typischerweise geheim gehalten. Dennoch ist festzuhalten, dass gängige, auch ausreichende Analyseinstrumente für die Erkenntnisgewinnung im Rahmen von big data derzeit grundsätzlich in durchaus hinreichendem Maße in den Märkten zur Verfügung stehen.

Wesentlichstes Bottleneck ist damit nach dem heutigen Stand der Forschung zweifelsfrei der *Zugang zu* entsprechenden *Rohdaten* und für spezifische Modelle weiterentwickelten *Trainingsdaten*.[18] Im Zusammenhang des *Zugangs zu Daten* spielt dabei auch die *Qualität der Daten* eine entscheidende Rolle (daher auch das Strukturmerkmal der *veracity* in der Beschreibung von big data Anwendungen).[19] Werden Daten als Input genutzt bzw. als Bausteine für zur Verfügung stehende ML-Analyseverfahren eingesetzt, über deren *Qualität* Unsicherheit besteht, muss dies bei darauf basierender Entscheidungsfindung Berücksichtigung finden. Ansonsten besteht die Gefahr, dass die für sich genommen mathematisch

---

Institute for Innovation & Competition Research Paper No. 19-13, S. 1, 8: „*training data is the most valuable element of the machine learning process*."; *Fink*, ZGE 2017, 288, 295 ff.

[17] Vgl. etwa Datenstrategie, COM(2020) 66 final, 7 f.; *Drexl/Hilty/Beneke et al.*, Max Planck Institute for Innovation & Competition Research Paper No. 19-13, S. 1, 7; *Stiemerling/Ehinger*, CR 2018, 761 m. w. N.

[18] Im Übrigen besteht gewisse Einigkeit, dass im Zusammenhang von big data und AI auch eine Verlagerung der Wertschöpfung auf einzelne Forscher und Ingenieure stattgefunden hat, so dass etwa auch der Aspekt der Arbeitnehmermobilität im AI-Bereich eine besonders wesentliche Rolle für die optimale Strukturierung der Innovationsprozesse spielt. Vgl. zu dem Einsatz heuristischer und erfahrungsbasierter Verfahren im Kontext des machine learnings und der Abhängigkeit vom Know-how der Mitarbeiter etwa *Drexl/Hilty/Beneke et al.*, Max Planck Institute for Innovation & Competition Research Paper No. 19-13, S. 1, 6–7. Die vorliegende Studie greift dies etwa im Zusammenhang des Geschäftsgeheimnisschutzes am Rande auf, soweit es um die differenzierte Regelung des in diesem Zusammenhang besonders wesentlichen (und europaweit gleichermaßen von gewisser Rechtsunsicherheit charakterisierten) Bereichs nachvertraglicher Nutzung von Geschäftsgeheimnissen geht.

[19] Vgl. dazu aus jüngerer Zeit etwa *OECD*, Enhancing Access to and Sharing of Data, S. 94 ff.; *Tentative Draft No. 1* der ALI-ELI Principles for a Data Economy vom 22. Mai 2020, S. 126, online abrufbar unter: https://www.ali.org/projects/show/data-economy/.

richtigen Methoden der Algorithmen einer auf unsicherer Basis getroffenen, möglicherweise unzutreffenden Entscheidung bzw. Prognose den Anschein von Richtigkeit, Sicherheit bzw. hoher Wahrscheinlichkeit verleihen und diese dadurch unberechtigt legitimieren.[20] Aus fehlenden (standardisierten) Angaben zur Daten-/Analysequalität resultiert somit die Gefahr, dass Marktteilnehmer und Individuen unbemerkt zum Objekt nur vermeintlich verlässlicher big data-Analysen werden.[21] Neben dem *Datenzugang* und dem Zugang zu AI-Ergebnissen als solchen ist daher *Transparenz* im Hinblick auf verwendete Input-Daten und Analysemethoden von wesentlicher Bedeutung.[22] Die Bedeutung des Zugangs zu Daten mit entsprechender Qualität greift nunmehr zutreffend auch der Vorschlag der Kommission für einen Digital Markets Act auf, der in den vorgesehenen Regelungen zu Datenportabilität und Datenzugang explizit Bezug nimmt auf „effective, high quality, continuous and real-time access" (s. Art. 6 Abs. 1 lit. i).[23]

4. Angestammte und neuartige Geschäftsmodelle auf unterschiedlichen Marktebenen und mit marktübergreifender Relevanz

Der marktübergreifenden Vielfalt von Datenmengen und Datenverarbeitungsmethoden entspricht in der Datenökonomie eine entsprechende *Vielfalt erfolgversprechender Geschäftsmodelle auf ganz unterschiedlichen Marktebenen*, von denen nur wenige auf die Anreizwirkung eines bestehenden oder gar zusätzlichen rechtlichen Schutzes der immateriellen Basis angewiesen sind.

So hat sich im Rahmen der Datenökonomie ein weiter Bereich von *Infrastrukturdienstleistern in der Cloud* entwickelt, die für den Erfolg ihres Geschäftsmodells – der Zurverfügung-stellung von Rechenkapazität und teilweise auch „Werkzeugkästen" üblicher AI-Tools – keines zusätzlichen rechtlichen Schutzes bedürfen. Zusätzlicher Regelungsbedarf im Immaterialgüterrecht – sei es durch zusätzliche Anreize im Immaterialgüterrechtssystem oder (umgekehrt) Rückschnitt bestehenden immaterialgüterrechtlichen Schutzes – ist hier derzeit nicht erkennbar.

---

[20] In diese Richtung auch *Hoeren,* MMR 2016, 8, 11, der vor diesem Hintergrund eine Erneuerung der alten Datenqualitätsdiskussion im Lichte von big data fordert.

[21] Zu den Risiken von big data vgl. allgemein *Mayer-Schönberger/Cukier*, Big Data, S. 189 ff. m. w. N.; eine weitere interessante Frage ist, ob und inwieweit der Mensch durch die Anwendung häufig nicht nachvollziehbarer selbstlernender KI-Algorithmen zum Objekt degradiert wird, vgl. *Martini*, JZ 2017, 1017 ff. m. w. N.

[22] Vgl. allgemein zu dem Problem der mangelnden Offenlegung der Art und Weise, wie Daten gesammelt und verarbeitet werden (disclosure problem) *Mattioli*, BTJL 2017, 1 ff. m. w. N.

[23] Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act), COM(2020) 842 final. S. näher unten S. 25 ff., 345 ff.

Wiederum andere Unternehmen setzen die ihnen zuwachsenden Datenmengen in erster Linie ein, um die *Qualität ihrer eigenen Produkte oder Dienstleistungen* zu erhöhen, ohne die entsprechenden potentiellen Erkenntnisgewinne auf breiter Front marktübergreifend zu nutzen. Insoweit wäre etwa Amazon als Beispiel zu nennen, das wegen der dokumentierten Verkaufsprozesse über vergleichsweise sehr hoch qualitative Daten verfügt, diese aber praktisch bisher weder in größerem Umfang zum Gegenstand eigenständiger Dienstleistungen noch zum Gegenstand neuartiger marktübergreifender Geschäftsmodelle macht. Stattdessen werden die Daten vorwiegend zur Verbesserung der eigenen, angestammten Geschäftsmodelle verwendet. Auch weite Bereiche der gezielten Maintenance und anderer effizienzsteigernder Anwendungen im internet of things (smart maintenance, smart farming etc.) gehören in diesen Kontext. Hier spielen faktische Datenkontrolle und flankierend auch Geschäftsgeheimnisschutz und teilweise das Schutzrecht sui generis für Datenbankhersteller eine signifikante Rolle. Entsprechend könnten für die Zukunft tatsächlich mehr oder weniger weitreichende *Reformen im Rechtssystem* mit dem Ziel, die Bereitschaft zur institutionell strukturierten *Offenlegung, Teilung und Übertragung* von Daten zu erhöhen, geboten sein. An dieser Stelle setzt auch die Diskussion um mögliche neue *sektorspezifische Zugangsregimes* an, in deren Rahmen sorgsam nach Rechtsverhältnissen und nutzungsbezogenen Fallgruppen zu differenzieren ist.

Massive Wertschöpfung auf der Grundlage von immensen Datenmengen (ohne erkennbaren Bedarf nach *zusätzlichem* immaterialgüterrechtlichen Schutz) findet auch heute schon im Rahmen der *„Aufmerksamkeitsökonomie" im Internet*[24] statt, insbesondere soweit es um *gezieltes Marketing* auf der Grundlage von Suchmaschinen- oder sonstigen Plattformdaten sozialer Netzwerke und vergleichbarer Kommunikationsplattformen geht. An dieser Stelle drohen möglicherweise auch *neue datenbasierte, markttranszendierende Machtpositionen* einzelner Unternehmen, die nicht nur im eigentlich zugrundeliegenden Dienstleistungsmarkt bestehen. Vielmehr können sie wegen der faktischen Beherrschung des vorgelagerten Markts für Zugang zu den gesammelten Daten in marktübergreifender Form auch in vollkommen anders gelagerte oder gänzlich neuartige Innovationsmärkte vordringen.[25] Insoweit spielt allerdings wiederum auch die *Datenqualität* eine entscheidende mitigierende Rolle: Die immense Datenmenge entspricht in diesen Bereichen nicht selten einer hierzu umgekehrt proportional abnehmenden Datenqualität im Hinblick auf die eigentliche Funktion dieser Daten, soweit sie beispielsweise in unterschiedlichen Märkten für eine

[24] Zu dem Begriff der „Aufmerksamkeitsplattformen" ausführlich *Stieper*, ZUM 2017, 132, 133 m. w. N. Vgl. in diesem Kontext etwa auch *Wu*, The Attention Merchants.

[25] Vgl. statt vieler *Schweitzer,* GRUR 2019, 569, 575 ff. m. w. N.

Sachregister