



## Beyond Big Data

Themenkurzprofil Nr. 34 | Simone Ehrenberg-Silies | November 2019

Der Anspruch auf die wirtschaftliche und wissenschaftliche Führungsposition im Bereich der künstlichen Intelligenz (KI) scheint bislang mit den USA und China sowie ihren Tech-Konzernen verbunden. Diese sind weltweit einzigartig darin, große Mengen an Nutzerdaten zu sammeln, die KI-Algorithmen vom Grundsatz her zum Lernen benötigen. Die Verfügbarkeit großer Volumina potenzieller Lerndaten in Verbindung mit vergleichsweise niedrigen Datenschutzstandards bildet vermeintlich die Basis für erfolgreiche Produktentwicklungen und einen uneinholbaren Vorteil im Wettrennen um die bestmöglichen KI-Anwendungen gegenüber Deutschland und zahlreichen anderen europäischen Ländern.

Neuere Entwicklungen im Bereich der Datensynthetisierung sowie der KI-Forschung mit Small Data, wie das bestärkende Lernen (Reinforcement Learning), die Methode des One-Shot-Lernens oder auch Zero-Shot-Lernens sowie das Transferlernen, erwecken jedoch die Hoffnung, dass KI-Innovationen auch unter Einhaltung der europäischen Datenschutznormen möglich sind. Dies bietet dem deutschen Mittelstand, der meist nur über geringe Datenmengen verfügt, die Aussicht, im Bereich der künstlichen Intelligenz den Anschluss zu finden.

Zurzeit wird die Methode der Datensynthetisierung jedoch fast ausschließlich von öffentlichen Stellen wie Statistikbehörden angewendet, um im Sinne von Open Data sensible Mikrodatsätze mit personenbezogenen Daten für die wissenschaftliche Forschung zugänglich zu machen. Die KI-Forschung mit Small Data steckt ihrerseits noch in den Kinderschuhen. Beide Aspekte spiegeln sich in der geringen Anzahl von Start-ups wider, die Dienstleistungen in den Bereichen synthetischer Daten und Small Data anbieten. Es wird

also noch einige Zeit benötigen, bis sich das Potenzial der beiden Entwicklungen im KI-Bereich tatsächlich abschätzen und auch wirtschaftlich ausschöpfen lässt. Die Forschungs- und Innovationsentwicklung wird hier unter anderem stark davon abhängen, ob durch zielgerichtete Schwerpunktsetzung in der Förderprogrammatische die Grundlage für eine gewisse Gründungsdynamik geschaffen werden kann.

### Hintergrund und Entwicklung

Der Wettlauf der Industrienationen um die Vormachtstellung im Bereich der künstlichen Intelligenz hat seit einigen Jahren durch „Innovationen in der Prozessor- und Speichertechnologie, im Cloud Computing, der Sensorik, dem Internet der Dinge und der Robotik“ (Bitkom e.V./ Deutsches Forschungszentrum für Künstliche Intelligenz GmbH 2017, S.8) beträchtlich an Geschwindigkeit gewonnen. Die Publikation zahlreicher nationaler und supranationaler KI-Strategien – z.B. China, Japan, Kanada 2017; Deutschland, Frankreich, Großbritannien, USA, Europäische Union 2018 – (Dutton 2018) belegt den großen politischen Stellenwert. Insbesondere die europäischen Nationen und auch die Europäische Union selbst sind dabei auf der Suche nach einem goldenen Mittelweg, bei dem sowohl der Anschluss an die Innovationsdynamik der USA und Chinas nicht verloren geht als auch ethische Standards, wie Datenschutz, Transparenz oder Nichtdiskriminierung, gewahrt bleiben (EK 2019). Mit Blick auf die künstliche Intelligenz resultiert daraus jedoch ein Dilemma, denn für die Entwicklung von KI-Anwendungen gilt die Verfügbarkeit großer Datenmengen bei gleichzeitig niedrigen Datenschutzanforderungen als ideale Ausgangsbasis. Selbst für die USA mit 328 Mio. Einwohnern stellt sich die Frage, wie sie



mit China erfolgreich konkurrieren können, das mit 1,4 Mrd. Menschen nicht nur über wesentlich mehr potenzielle Datenproduzenten verfügt, sondern auch über ein vergleichsweise geringes Datenschutzniveau, das die Nutzung dieser Daten kaum einschränkt (Hao 2019a).

Auch hierzulande wird thematisiert, dass Deutschland durch das hohe Datenschutzniveau auf dem Feld der künstlichen Intelligenz ins Hintertreffen geraten könnte. So wies Iris Plöger, Mitglied der Geschäftsleitung des BDI, auf dem Symposium des Bundesbeauftragten für den Datenschutz und die Informationssicherheit zum Thema „Chancen und Risiken für den datenschutzgerechten Einsatz von Künstlicher Intelligenz“ im September 2019 auf die Notwendigkeit hin, „eine wettbewerbsfähige Datenwirtschaft [zu] schaffen“ (Krempel 2019). Diese würde durch einige rechtliche Vorgaben der Datenschutz-Grundverordnung, wie das prinzipielle Verbot der Verarbeitung personenbezogener Daten mit Erlaubnisvorbehalt, den jederzeit möglichen Widerspruch zu bereits erteilten Einwilligungen, die enge Zweckbindung der erhobenen Daten und die Vorgaben hinsichtlich der Voraussagbarkeit der Ergebnisse der Datenverarbeitung, eingeschränkt. Da die Forschung im Bereich der künstlichen Intelligenz große Datenmengen erfordert, würden unter den gegebenen rechtlichen Voraussetzungen zahlreiche deutsche Konzerne ihre Forschungs- und Entwicklungsaktivitäten bereits im Ausland durchführen (Krempel 2019). Während Industrievertreter die Vorschriften der Datenschutz-Grundverordnung daher als Hemmnis für die Entwicklung innovativer KI-Anwendungen in Deutschland sehen, betrachten Datenschützer, darunter der Bundesdatenschutzbeauftragte Ulrich Kelber, sie als eigentliche „Innovationsmotoren“ (Krempel 2019). Wissenschaftliche und wirtschaftliche Führerschaft im Bereich der künstlichen Intelligenz und ein hohes Datenschutzniveau stehen nach dieser Lesart nicht im Widerspruch. Die Bundesregierung sieht in ihrer „Strategie Künstliche In-

telligenz“ vom November 2018 zwei strategische Ansatzpunkte, um beides in Einklang zu bringen: Zum einen sollen Methoden gefördert werden, auf deren Basis KI-Systeme auch mit kleineren Datenmengen (Small Data) trainiert und angewendet werden können (Bundesregierung 2018, S.36). Zum anderen soll auf die Forschung zur Erzeugung synthetischer, also quasi künstlicher Daten fokussiert werden, die als Alternative zu realen Daten für die Trainingszwecke einer KI genutzt werden können (Bundesregierung 2018, S.35).

### Small Data

Die aktuelle Blüte der künstlichen Intelligenz geht auf das tiefe Lernen (Deep Learning) mit künstlichen neuronalen Netzen zurück (Kirste/Schürholz 2019, S.29). In der Tat benötigt diese spezielle Methode des maschinellen Lernens (ML) sehr große Datenmengen. Beim ML leiten Lernalgorithmen auf der Grundlage von Erfahrungen komplexe Modelle ab, die sie auf unbekannte Daten gleicher Erfahrungszusammenhänge übertragen können. ML benötigt also keine manuelle Wissensangabe oder Programmierung eines Lösungswegs. Künstliche neuronale Netze sind Modelle, die an die neuronale Struktur des menschlichen Gehirns erinnern. Die künstlichen Neuronen werden als Knoten in mehreren Schichten in der Software angelegt. Beim Lernen passt der Algorithmus immer wieder die Gewichtungen zwischen den Knoten an, bis er seine Aufgabenstellung ausreichend erfüllen kann (FhG 2018, S.44). Tiefes Lernen meint „das Lernen in Künstlichen Neuronalen Netzen mit mehreren bis sehr vielen inneren Schichten“, also das Lernen in sehr komplexen neuronalen Netzen (FhG 2018, S.43).

Soll ein künstliches neuronales Netzwerk beispielsweise in der Objekterkennung trainiert werden, benötigt es rund 15 Mio. Bilder der jeweiligen Objektart als Trainingsinput (IBM 2017). Je nachdem, wie komplex und vielfältig die Objektart ist, kann die Anzahl der notwendigen Lernbilder nach unten oder oben variieren. Unter der Voraussetzung, tendenziell große Datenmengen zu benötigen, ist es nicht verwunderlich, dass auf dem internationalen Markt für ML vor allem große Technologieunternehmen aus den USA (Amazon, Facebook, Google, IBM, Intel, Microsoft, NVIDIA, Tesla) und China (Baidu, Alibaba, Tencent) dominieren. Diese können nicht nur auf sehr große Mengen potenzieller Lerndaten zurückgreifen, sondern verfügen für das Training von ML-Algorithmen auch über rechenstarke Hardware, eigene ML-Plattformen und den Zugriff auf renommierte Forscherteams (FhG 2018, S.24).

Nicht in allen Anwendungsbereichen ist es möglich, auf solch große Datenmengen zuzugreifen. Beispielsweise umfassen klinische Studien in der Medizin oftmals nur vierstellige Probandenzahlen (IBM 2017). Oder falls die Digitalisierung in der Produktion eines Unternehmens noch nicht besonders weit fortgeschritten ist, ist zumeist auch das verfügbare Datenpotenzial gering (WIK 2019, S.10).

Insgesamt sind die für Deutschland relevantesten KI-Anwendungsbereiche – Robotik, industrielle Produktion, Medizin und Sicherheit von Infrastrukturen – dadurch charakterisiert, dass nur geringe repräsentative Datenmengen anfallen respektive nicht ausreichend brauchbare Daten verfügbar sind (FhG 2018, S.29).

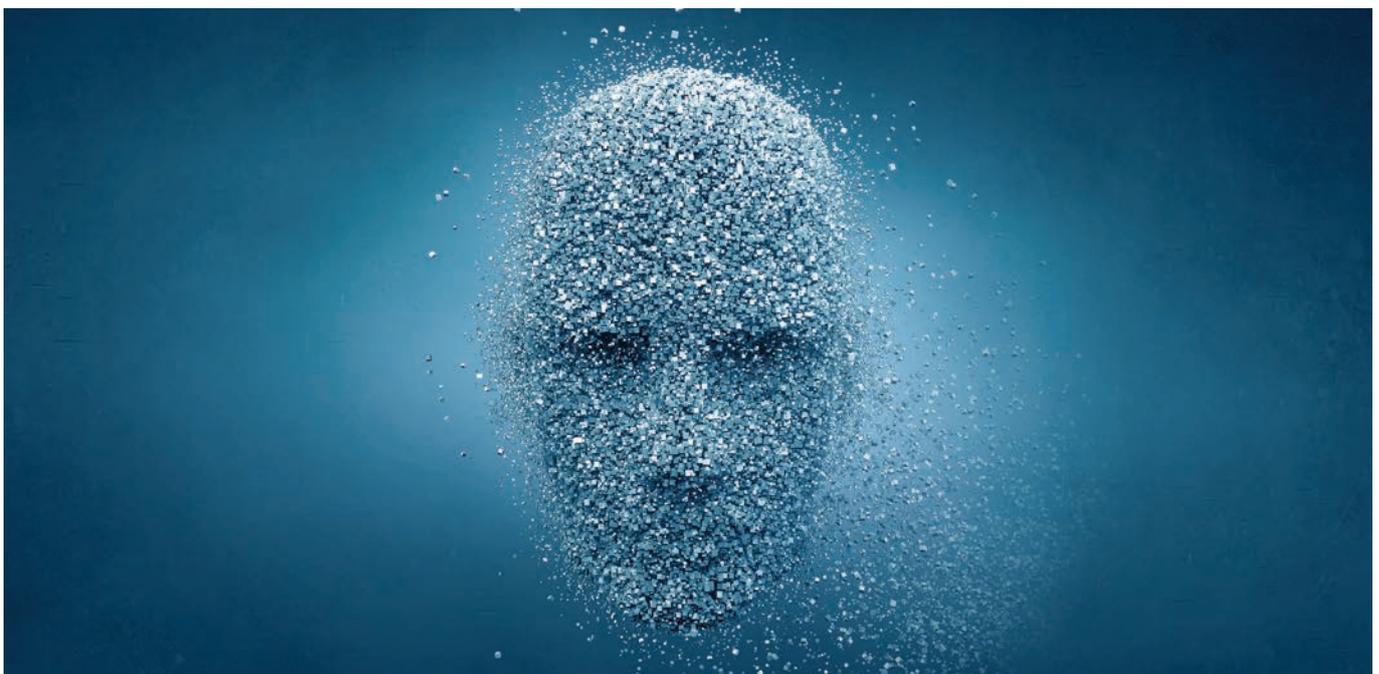
Zusätzlich zum regulativen Rahmen der Datenerhebung und -verwendung ist es also schlicht die Tatsache, dass nur sehr wenige Unternehmen über große Datenmengen verfügen (Greco et al. 2019, S.3), die Small Data zu einem der wichtigsten Innovationsfelder der Zukunft im Bereich der künstlichen Intelligenz werden lässt. Eine Auswertung von 16.625 Vorabdrucken aus den Wissenschaftsdisziplinen Physik, Mathematik, Informatik, Statistik, Finanzmathematik und Biologie des Dokumentenservers arXiv.org zeigt, dass Methoden der künstlichen Intelligenz, die keine großen Datensätze benötigen, wie das bestärkende Lernen, in der Forschung mehr und mehr an Bedeutung gewinnen (Hao 2019b). Beim bestärkenden Lernen lernt der Algorithmus aus der direkten Interaktion mit seiner Umgebung, indem das Programm – ähnlich der Konditionierung eines Tieres – für richtige Ergebnisse belohnt wird (Hao 2019b; Kirste/Schürholz 2019, S.29). Durch das Feedback ist der Lernalgorithmus in der Lage, die Erfolgsaussichten seiner Aktionen besser abzuschätzen (FhG 2018, S.43). Potenzielle Anwendungen für das bestärkende Lernen sind z.B. das Greifen-Lernen unterschiedlicher Objekte durch einen Roboter oder das Erlernen erwünschter und unerwünschter Aktionen beim autonomen Fahren (FhG 2018, S.33).

Auch die Methode des One-Shot- oder auch Zero-Shot-Lernens wird als geeignetes Verfahren des ML mit kleinen Datenmengen diskutiert (IBM 2017). Beim One-Shot-Lernen wird menschliches Lernen nachgeahmt. Der Algorithmus

lernt auf Basis „weniger Beispiele einer Klasse oder durch die Kombination bekannter Beispiele anderer Klassen“ (FhG 2018, S.45). Mögliche Anwendungen liegen in den Bereichen Medizin und Gesundheit, der Gefahrenabwehr (z.B. Produktionsprozesse, Finanzaktivitäten, Videoüberwachung, kritische Infrastrukturen) sowie der industriellen Produktion (FhG 2018, S.37).

Großes Potenzial wird ebenfalls der Methode des Transferlernens zugeschrieben. Beim Transferlernen geht es darum, geeignete Teilbereiche eines künstlichen neuronalen Netzes zur Bewältigung ähnlicher Aufgabenstellungen zu verwenden (FhG 2018, S.46). Der Vorteil besteht darin, KI-Lösungen, die durch das Training mit großen Datenmengen entwickelt werden konnten, auf ähnliche Aufgabengebiete zu übertragen, in denen potenziell weniger Trainingsdaten vorhanden sind bzw. deren Beschaffung sehr kostenintensiv wäre (t3n – digital pioneers 2019, S.1). So hat beispielsweise das Start-up audEERING GmbH große Fortschritte in der Emotionserkennung menschlicher Stimmen erzielen können, indem es auf künstliche neuronale Netze zurückgriff, die auf Emotionserkennung in der Musik und auf menschliche Stimmungen bei Geräuschen spezialisiert waren (t3n – digital pioneers 2019, S.1). Die Anwendungsgebiete für das Transferlernen sind weitestgehend mit denen des One-Shot-Lernens identisch. Zusätzlich zu diesen kann das Transferlernen allerdings auch zur Gestaltung von flexibleren Robotern, beispielsweise im Servicebereich, eingesetzt werden (FhG 2018, S.37).

Alle drei Methoden des ML auf Basis kleinerer Datenmengen konnten in ihren Anwendungsbereichen bereits den Technologiereifegrad eines Demonstrators erreichen (Nachweis der generellen Machbarkeit einer Lösung vor der tatsächlichen Umsetzung) (Innovationsprozesse.com 2011).





für Arbeitsmarktforschung (IAB) in Deutschland. Auch Institutionen und Behörden in Schottland, England, Kanada, Australien und Neuseeland stellen synthetische Datensätze zur Verfügung (SNV 2018, S. 2 u. 6).

Am Beispiel der LBD des U.S. Census Bureau soll im Folgenden kurz dargestellt werden, wie bei der (Teil-)Synthetisierung des Datensatzes vorgegangen wird: Der Ursprungsdatensatz umfasst Realdaten zu Betriebsgründung und -schließung, Wirtschaftszweigen, Umsätzen sowie Beschäftigtenzahlen seit 1976. Diese Variablen korrelieren auf eine bestimmte Art und Weise miteinander. Ziel war es, hieraus einen synthetischen Datensatz zu schaffen, der die gleichen statistischen Eigenschaften wie die LBD aufweist. Vor diesem Hintergrund wurde zunächst unter Rückgriff auf den Wirtschaftszweig die Lebensdauer der Betriebe synthetisiert. Dadurch entstanden quasi synthetische Betriebe, die es nie gegeben hat, mit synthetischen Zeiten der aktiven Geschäftstätigkeit. Im nächsten Schritt wurde die Betriebsgröße der „künstlichen“ Betriebe für die Jahre ihrer gesamten Lebensdauer modelliert. Sollte sich das entwickelte Synthetisierungsmodell auf Betriebsdatensätze anderer Länder, wie z.B. das IAB-Betriebspanel, übertragen lassen, wären erstmals datenschutzkonforme länderübergreifende Vergleichsstudien möglich, indem die US-amerikanischen und deutschen synthetischen Betriebsdatensätze zu Analyse Zwecken gemeinsam auf einem Server angeboten werden (Stiftung Neue Verantwortung 2018, S.16).

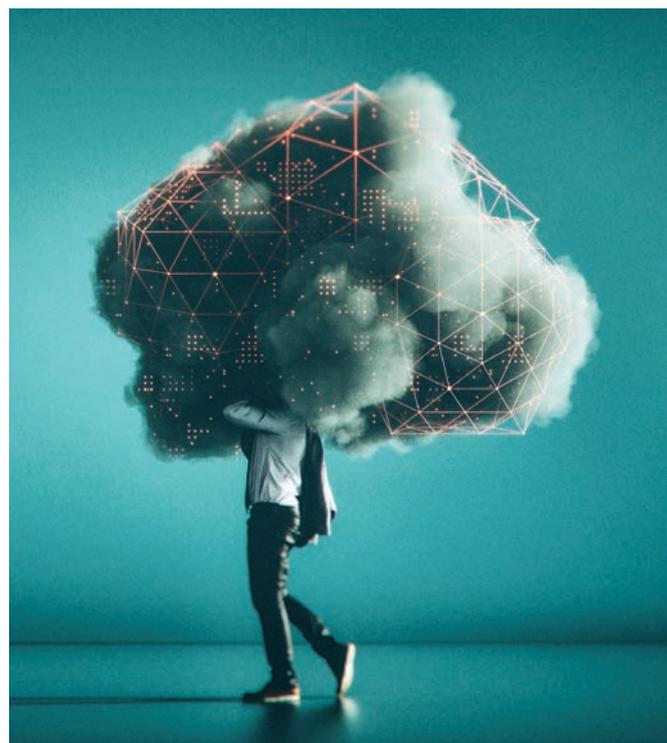
Wenngleich die Methode der Datensynthetisierung zurzeit vor allem im öffentlichen Sektor verwendet wird, gibt es auch erste kommerzielle Akteure in diesem Bereich wie das Berliner Software-Start-up Stative. Stative bietet seinen Kunden an, deren sensible Datensätze datenschutzkonform in synthetische Datensätze umzuwandeln, zu analysieren und KI-Anwendungen zu trainieren. Bisher nutzen vor allem Krankenhäuser und Pharmaunternehmen die Dienstleistung des Unternehmens für ihre Gesundheitsdaten. Die im April geflossene Seedfinanzierung (Bereitstellung von Wagniskapital in der Phase der Vorgründung und Gründung eines Unternehmens) der Wagniskapitalgeber Capnamic und Westtech soll es allerdings ermöglichen, das Geschäft auf die Finanz- und Versicherungsbranche auszuweiten (Schlenk 2019). Die Twenty Billion Neurons GmbH, ebenfalls mit Unternehmenssitz in Berlin, bietet eine eigene Video-Dataset-Collection an (<https://20bn.com/products/datasets>), die Videos für das Training von KI-Anwendungen synthetisiert (Hedemann 2018).

### Gesellschaftliche und politische Relevanz

Die Methoden der Datensynthetisierung und des ML mit kleinen Datenmengen sind vielversprechende Ansätze, Entwicklungen im Bereich der künstlichen Intelligenz zu ermöglichen, die den hohen europäischen Anforderungen an den Datenschutz gerecht werden. Zudem könnten sie

dazu beitragen, das Innovationspotenzial des deutschen Mittelstands bei der Entwicklung und Implementierung von KI-Anwendungen besser auszuschöpfen. Dies wird zurzeit noch durch das mangelnde Datenvolumen der kleinen und mittleren Unternehmen eingeschränkt, welches bei entsprechender Technologiereife der genannten Methoden allerdings keine zentrale Bedeutung mehr haben sollte. Selbst das US-amerikanische IT- und Beratungsunternehmen IBM ging schon 2017 davon aus, dass kleine Datensätze große Datenvolumina als Treiber für KI-Innovationen ablösen könnten (IBM 2017). Möglicherweise könnte sich also der entscheidende strategische Vorteil der Tech-Giganten in den USA und in China, der in der massenhaften Verfügbarkeit von Nutzerdaten besteht, abschwächen und dadurch die Dominanz US-amerikanischer und chinesischer Forschungseinrichtungen sowie Unternehmen im Bereich der künstlichen Intelligenz ansatzweise ins Wanken gebracht werden. Dass die Entwicklung allerdings noch eher am Anfang steht, belegen die wenigen Suchergebnisse, die die Internetrecherche zu den Themen „Start-ups“, „Synthetische Daten“ sowie „Small Data“ hervorgebracht hat. Bisher fehlt in Deutschland eine spezifische Förderprogrammatik, die Forschung in den Bereichen KI und Small Data sowie KI und Datensynthetisierung unterstützt sowie speziell auf Gründungen aus diesen Forschungsfeldern heraus abhebt. Es wird von Interesse sein zu beobachten, wie die Bundesregierung ihre KI-Strategie in den kommenden Jahren in Bezug auf synthetische Daten und Small Data konkret umsetzen wird.

Bei aller Zuversicht in das Potenzial der drei Methoden darf nicht übersehen werden, dass zumindest die Datensynthetisierung und das Transferlernen auch ethische Risiken in



sich bergen. So können synthetische Daten missbraucht werden, um sensible Anwendungen zu korrumpieren, wie biometrische Authentifizierungsverfahren, die auf der Echtheit von Daten basieren (SNV 2018, S. 20). Ebenso ist es möglich, dass Datenanalysten in synthetischen Datensätzen bzw. zusammengeführten unterschiedlichen synthetischen Datensätzen sensible, ethisch problematische Zusammenhänge identifizieren, die dann zu Prognosezwecken – z.B. von Krankheiten – auf tatsächliche Menschen im Originaldatensatz angewendet werden können. Anders als nichtanonyme Datensätze unterliegen anonyme synthetische Datensätze nämlich nicht den Vorschriften der Datenschutz-Grundverordnung hinsichtlich Zweckbindung sowie Dokumentations- und Auskunftspflichten (SNV 2018, S. 19f.). Mit dem Ursprungsdatensatz wären Prognosen dieser Art nur zulässig, wenn die betroffenen Personen zuvor in die Verarbeitung ihrer Daten zu genau diesem Prognosezweck (Vorhersagbarkeit von Krankheiten) eingewilligt hätten.

Dem Transferlernen wird unterstellt, dass es die Gefahr von z.B. rassistischer und/oder sexistischer Diskriminierung aufgrund verzerrter Trainingsdaten, die im Allgemeinen mit künstlicher Intelligenz assoziiert wird, noch einmal steigern könnte. Der Grund hierfür ist, dass der Anwender von vortrainierten Modellen häufig in Unkenntnis über die Eigenschaften des ursprünglichen Trainingsdatensatzes ist und deshalb nicht abschätzen kann, ob sich hieraus möglicherweise problematische Diskriminierungen für seinen Anwendungsfall ergeben könnten, die er dann unbewusst in seinem neuen Modell reproduziert (t3n – digital pioneers 2019, S.2).

### Mögliche vertiefte Bearbeitung des Themas

Das Thema „Beyond Big Data“ eignet sich grundsätzlich für eine zukünftige Bearbeitung in Form einer TAB-Kurzstudie. Allerdings wird empfohlen, mit der Untersuchung frühestens in 2 bis 3 Jahren zu beginnen. Dann könnten auf den FuE-Feldern Datensynthesierung und Small Data bereits tiefer gehende Erkenntnisse vorliegen, die den technologischen Reifegrad der Methoden im Bereich KI erhöhen und eine gewisse Gründungsdynamik auslösen könnten. Erst auf dieser Basis lässt sich einschätzen, welches Potenzial diese Ansätze hinsichtlich Funktionsweise und Anwendungsbreite tatsächlich aufweisen und wie sich Deutschland in diesem Technologiebereich im Vergleich zu anderen Nationen mittelfristig entwickeln wird. In der Kurzstudie sollte auch die Forschungsförderung hinsichtlich ihres Beitrags zur Stärkung einer datenschutzkonformen KI untersucht werden. Gegebenenfalls müsste der Fokus der Kurzstudie neu ausgerichtet werden, wenn sich neben Datensynthesierung und Small-Data-Anwendungen andere vielversprechende Verfahren abzeichnen, die eine datenschutzkonforme KI-Entwicklung ermöglichen.

Zudem sollte betrachtet werden, wie die großen Technologiekonzerne sich des Themas „Small Data“ annehmen,



insbesondere wenn sich abzeichnet, dass die Grenzen des Big-Data-Geschäfts tatsächlich erreicht werden, wie von IBM vermutet.

### Literatur

- ▶ Bitkom e. V.; Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (2017): Künstliche Intelligenz. Wirtschaftliche Bedeutung, gesellschaftliche Herausforderungen, menschliche Verantwortung (Autoren: Buschbacher, F.; Weber, M.; Burchardt, A.). [https://www.dfki.de/fileadmin/user\\_upload/import/9744\\_171012-KI-Gipfelpapier-online.pdf](https://www.dfki.de/fileadmin/user_upload/import/9744_171012-KI-Gipfelpapier-online.pdf) (10.10.2019)
- ▶ Bundesregierung (2018): Strategie Künstliche Intelligenz der Bundesregierung. [https://www.bmbf.de/files/Nationale\\_KI-Strategie.pdf](https://www.bmbf.de/files/Nationale_KI-Strategie.pdf) (1.10.2019)
- ▶ Degeler, A. (2019): French „small data“ machine learning startup MyDataModels raises €1 million. Tech.eu, 31.1.2019, <https://tech.eu/brief/french-small-data-machine-learning-startup-mydatamodels-raises-e1-million/> (11.10.2019)
- ▶ Dutton, T. (2018): An Overview of National AI Strategies. Medium, 28.6.2018, <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd> (10.10.2019)
- ▶ EK (Europäische Kommission) (2019): Künstliche Intelligenz. Kommission treibt Arbeit an Ethikleitlinien weiter voran. [https://ec.europa.eu/commission/news/artificial-intelligence-2019-apr-08\\_de](https://ec.europa.eu/commission/news/artificial-intelligence-2019-apr-08_de) (11.10.2019)
- ▶ FhG (Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e. V.) (2018): Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung (Autoren: Döbel, I.; Leis, M.; Vogelsang, M. Molina; Neustroev, D.; Petzka, H.; Riemer, A.; Rüping, S.; Voss, A.; Wegele, M.; Welz, J.). <https://www.bigdata.fraunhofer.de/>

content/dam/bigdata/de/documents/Publikationen/Fraunhofer\_Studie\_ML\_201809.pdf (10.10.2019)

- ▶ Greco, C.; Polonioli, A.; Tagliabue, J. (2019): Less (Data) Is More. Why Small Data Holds the Key to the Future of Artificial Intelligence. <https://arxiv.org/ftp/arxiv/papers/1907/1907.10424.pdf> (22.10.2019)
- ▶ Hao, K. (2019a): China's masses of data give it an edge in AI—but they may not forever. MIT Technology Review, 5.3.2019, <https://www.technologyreview.com/f/613077/chinas-masses-of-data-give-it-an-edge-in-ai-but-they-may-not-forever/> (2.10.2019)
- ▶ Hao, K. (2019b): We analyzed 16,625 papers to figure out where AI is headed next. MIT Technology Review, 25.1.2019, <https://www.technologyreview.com/s/612768/we-analyzed-16625-papers-to-figure-out-where-ai-is-headed-next/> (2.10.2019)
- ▶ Hedemann, F. (2018): Synthetische Daten. Ein neues Geschäftsmodell? DMEXCO, 4.6.2018, <https://dmexco.com/de/stories/synthetische-daten-ein-neues-geschaeftsmodell/> (1.10.2019)
- ▶ IBM (2017): The new AI innovation equation. <https://www.ibm.com/watson/advantage-reports/future-of-artificial-intelligence/ai-innovation-equation.html> (7.10.2019)
- ▶ Innovationsprozesse.com (2011): Demonstrator. <http://www.innovationsprozesse.com/demonstrator.htm> (17.10.2019)
- ▶ Kirste, M.; Schürholz, M. (2019): Einleitung: Entwicklungswege zur KI. In: Wittpahl, V. (Hg.): Künstliche Intelligenz. Technologie, Anwendung, Gesellschaft. Berlin, Heidelberg, S.21 – 35
- ▶ Kreml, S. (2019): BDI: Datenschutzvorgaben sind „toxisch für Künstliche Intelligenz“. heise online, 25.9.2019, <https://www.heise.de/newsticker/meldung/BDI-Datenschutzvorgaben-sind-toxisch-fuer-Kuenstliche-Intelligenz-4538401.html> (1.10.2019)
- ▶ Schlenk, C. T. (2019): Millionenbetrag für Daten-Startup Stative. Gruenderszene.de, 16.4.2019, <https://www.gruenderszene.de/business/stative-millionen-cap-namic?interstitial> (11.10.2019)
- ▶ SNV (Stiftung Neue Verantwortung) (2018): Synthetische Daten. Innovationspotential und gesellschaftliche Herausforderungen (Autoren: Drechsler, J.; Jentzsch, N.). [https://www.stiftung-nv.de/sites/default/files/synthetische\\_daten.pdf](https://www.stiftung-nv.de/sites/default/files/synthetische_daten.pdf) (1.10.2019)
- ▶ T3n – digital pioneers (2019): Transfer Learning: Das Wundermittel für mehr Intelligenz? 12.3.2019, <https://t3n.de/magazin/transfer-learning-wundermittel-fuer-mehr-intelligenz-247734/> (22.10.2019)
- ▶ Weyer, S. (2019): Synthetische Daten. Datenschutz und Datennutzung in Zeiten von DSGVO. Computerwoche, 17.7.2019, <https://www.computerwoche.de/a/datenschutz-und-datennutzung-in-zeiten-von-dsgvo,3547346> (1.10.2019)
- ▶ WIK (Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste) (2019): Künstliche Intelligenz im Mittelstand. Relevanz, Anwendungen, Transfer (Autoren: Lundborg, M.; Märkel, C.). [https://www.mittelstand-digital.de/MD/Redaktion/DE/Publikationen/kuenstliche-intelligenz-im-mittelstand.pdf?\\_\\_blob=publicationFile&v=5](https://www.mittelstand-digital.de/MD/Redaktion/DE/Publikationen/kuenstliche-intelligenz-im-mittelstand.pdf?__blob=publicationFile&v=5) (7.10.2019)

Das Horizon-Scanning ist Teil des methodischen Spektrums der Technikfolgenabschätzung im TAB.

**Horizon**  
**SCANNING**

Mittels Horizon-Scanning werden wissenschaftlich-technische Trends und sozio-ökonomische Entwicklungen in frühen Entwicklungsstadien beobachtet und in den Kontext gesellschaftlicher Debatten eingeordnet. So sollen Innovationssignale möglichst früh erfasst und ihre technologischen, ökonomischen, ökologischen, sozialen und politischen Veränderungspotenziale beschrieben werden. Ziel des Horizon-Scannings ist es, einen Beitrag zur forschungs- und innovationspolitischen Orientierung und Meinungsbildung des Ausschusses für Bildung, Forschung und Technikfolgenabschätzung zu leisten.

In der praktischen Realisierung wird das Horizon-Scanning als Kombination softwaregestützter Such- und Analyse-schritte und eines expertenbasierten Validierungs- und Bewertungsprozesses durchgeführt.

Herausgeber: Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB)

Gestaltung: VDI/VDE Innovation + Technik GmbH

Bildnachweise: © peterschreiber.media/AdobeStock (S.1), Jakub Krechowicz/AdobeStock (S.2), Sergey Nivens/AdobeStock (S.3), elenabsl/AdobeStock (S.4), gremlin/iStock (S.5), peterhowell/iStock (S.6)

ISSN-Internet: 2629-2874