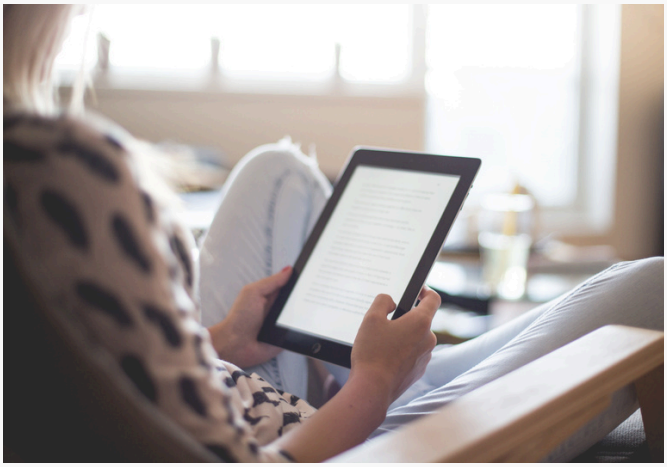
FORUM WISSEN, DAS ANKOMMT.


Leseprobe

Unsere Fachinhalte bieten Ihnen praxisnahe Lösungen, wertvolle Tipps und direkt anwendbares Wissen für Ihre täglichen Herausforderungen.

- **Praxisnah und sofort umsetzbar**: Entwickelt für Fach- und Führungskräfte, die schnelle und effektive Lösungen benötigen.
- **Fachwissen aus erster Hand**: Inhalte von erfahrenen Expertinnen und Experten aus der Berufspraxis, die genau wissen, worauf es ankommt.
- **Immer aktuell und verlässlich**: Basierend auf über 30 Jahren Erfahrung und ständigem Austausch mit der Praxis.

Blättern Sie jetzt durch die Leseprobe und überzeugen Sie sich selbst von der Qualität und dem Mehrwert unseres Angebots!

Data Science im Spannungsfeld der Datenschutz-Compliance

Der technologische Fortschritt ist auf Data Science angewiesen. Das gilt insbesondere für die Entwicklung von Systemen Künstlicher Intelligenz (KI). In dieser Entwicklung dient der Datenschutz als Korrektiv, damit die Rechte und Freiheiten der Menschen nicht „unter die Räder geraten“.

Bildquelle: NicoElNino – stock.adobe.de

Data Science befasst sich mit enormen Datenmengen, woraus die verschiedensten Informationen gewonnen werden können.

Gibt es Wege, damit der Datenschutz nicht allein den Fortschritt ausbremst, sondern ihn konstruktiv begleitet? Welche Risiken entstehen Unternehmen in der Praxis, wenn Data Science am Datenschutz vorbei Anwendung findet?

Was ist Data Science?

Bei Data Science handelt es sich um eine interdisziplinäre Wissenschaft, die sich mit der Analyse von – i. d. R. sehr großen – Datenmengen befasst. Sie bildet dabei eine Schnittstelle aus Informatik, Mathematik und dem Know-how der zu untersuchenden Branchen oder Bereiche. In der praktischen Anwendung sollen dabei aus den Datenmengen Informationen gewonnen werden, die

- zur (unternehmerischen) Entscheidungsfindung dienen,
- Zusammenhänge innerhalb der sehr großen Datenmengen aufzeigen oder
- Wahrscheinlichkeiten für zukünftiges (Kunden-) Verhalten oder andere Entwicklungen bestimmen.

Da die Datenmengen oft sehr groß sind, wird in diesem Zusammenhang auch von Big Data gesprochen. Da Systeme Künstlicher Intelligenz für ihre Lernprozesse große Datenmengen benötigen, liefert Data Science dafür meist die notwendige Grundlage.

Solange Data Science rein anonyme Daten verarbeitet, besteht keine Anwendung der DSGVO. Das ergibt sich aus Art. 2 Abs. 1 DSGVO. Unglücklicherweise ist dies nur sehr selten per se der Fall. Daher muss Data Science oft direkt oder indirekt auf personenbezogene Daten zurückgreifen. Dabei entsteht unweigerlich ein Spanungsfeld mit allen Vorschriften, die dem Schutz personenbezogener Daten dienen. Laut der bitkom-Studie von 2021 behaupten rund 36 % aller befragten Unternehmen, dass die DSGVO die Entwicklung innovativer Projekte im Bereich KI oder Big Data verhindert hätte.

Mit Data-Science-Methoden möchten Unternehmen Hypothesen über die unternehmerische Lebens- oder Marktwirklichkeit verifizieren bzw. falsifizieren oder Aussagen über das zukünftige Verhalten von Menschen treffen. Insbesondere das Verhalten von Kunden und Mitarbeitern steht dabei im Fokus. Wie müssen Werbung oder Arbeitsplätze für die optimale Erreichung der unternehmerischen Ziele beschaffen sein? Diese und andere Fragen sind aus Unternehmenssicht berechtigt. Demgegenüber stehen jedoch auch die Grundrechte der Kunden und Mitarbeiter auf den Schutz ihrer Daten (Art. 8 EU-GrCh).

Künstliche Intelligenz und ihre aktuellen Grenzen

Unabhängig vom Datenschutz-Aspekt sind die sehr hohen Erwartungen an Data Science und die Künstliche Intelligenz oft unbegründet hoch, da die statistischen Effekte falsch eingeschätzt werden.

Ein Beispiel: Im Juli 2018 wurde in einer Pressemitteilung der Test der Gesichtserkennung mittels KI am Bahnhof Berlin-Südkreuz mit einer Trefferrate von über 80 % und einer Falschalarmrate von 0,1 % als erfolgreich in der Presse verkündet. Wie sich jedoch mit einfachen statistischen Überlegungen zeigt, würde dieses System zu 99,3 % falsche Ergebnisse liefern[1].

Den hohen Erwartungen an die KI versucht man nun oft, anstatt das Systemdesign zu hinterfragen, mit noch größeren Datenmengen als „Lernfutter“ für die Systeme zu begegnen. Nicht immer befasst sich daher Data Science mit diesen Themen, die aber aus Sicht des Datenschutzes obligatorisch sind.

Die Verarbeitung personenbezogener Daten

Einkaufsverhalten, Gesichtserkennung, Mitarbeiterstatistiken – das sind nur einige sehr offen-

1 Vgl. https://dstatg.de/unstatistik-des-monats-oktober (zuletzt aufgerufen: 18.10.2022).

sichtliche Beispiele dafür, dass Data Science i. d. R. auf Daten zurückgreifen muss, die einen Personenbezug haben. Die Auslegung des Personenbezugs in der DSGVO ist bekanntermaßen hoch, sodass auch scheinbar rein technische Daten wie die IP-Adresse oder die Fahrgestellnummer eines Pkw als personenbezogene Daten gelten, weil sich anhand dieser eine Person ggf. auch unter Hinzuziehung weiterer Informationen identifizieren lässt.

Daher stellen sich in Data Science folgende grundlegenden Fragen zum Datenschutz:

- Auf welcher Rechtsgrundlage ist die Verarbeitung der personenbezogenen Daten erlaubt?
- Wie kann und muss über die Verarbeitung nach Art. 13 oder 14 DSGVO informiert werden?
- Welche Schutzmaßnahmen sind zu treffen, damit insbesondere die Vertraulichkeit der Daten gewahrt bleibt?

Diese Fragen sind nicht nur dem Grunde, sondern auch besonders in der prozessualen Umsetzung nach zu beantworten.

Rechtsgrundlagen

In der Liste der möglichen Erlaubnistatbestände zur Verarbeitung nach Art. 6 DSGVO, insbesondere der Verarbeitung besonderer Daten nach Art. 9 DSGVO, ist zunächst die Einwilligung zu prüfen. Diese ist mit besonderen Problemen verbunden. Zum einen sind speziell im Beschäftigtenkontext an die Freiwilligkeit der Einwilligung besonders hohe Anforderungen geknüpft (§ 26 Abs. 2 BDSG). Auf der anderen Seite bildet die Einwilligung bei der Verarbeitung von Daten nach Art. 9 DSGVO die vermutlich einzig mögliche Rechtsgrundlage. Die Ausnahme für wissenschaftliche oder historische Forschungszwecke und zu statistischen Zwecken (§ 27 BDSG) scheidet im Unternehmenskontext regelmäßig wegen der immanenten wirtschaftlichen Interessen des Verantwortlichen an der Verarbeitung aus. Erschwerend kommt hinzu, dass die Bewertung, ob Daten nach Art. 9 DSGVO vorliegen, nach der jüngsten Rechtsprechung des EuGH weit auszulegen ist.[2] Damit wird bereits der Vorname des Ehepartners schnell zu einem Datum, das nach Art. 9 DSGVO geschützt ist, da sich daraus Rückschlüsse auf die mögliche sexuelle Orientierung ziehen lassen.

Aber auch methodisch werden die Ergebnisse von Statistiken, die nur auf Daten basieren, die mit einer Einwilligung gewonnen werden, anzuzweifeln sein. Freiwillig eingebrachte Daten verfälschen die statistische Signifikanz. Hinzu kommt die prozessuale Herausforderung möglicher Widerrufe.

Von übrigen Erlaubnistatbeständen verbleibt oft nur das berechtigte Interesse des Verantwortlichen (Art. 6 Abs. 1 lit. f DSGVO). Andere Rechtsgrundlagen, wie beispielsweise die Vertragserfüllung, scheiden meist wegen mangelnder, zwingender Erforderlichkeit aus. Daten nach Art. 9 DSGVO lassen sich hiermit nicht legitimieren. Aber abgesehen davon, lässt diese Rechtsgrundlage die Anwendung von Data Science prinzipiell zu. Ausschlaggebend ist dafür aber eine umfassende Interessenabwägung, die insbesondere die Erwartungshaltung und die Schutzbedürftigkeit der Betroffenen berücksichtigt.

Prozessual ist dabei ebenso das Widerspruchsrecht nach Art. 21 Abs. 1 DSGVO zu berücksichtigen. Auch dieses kann Ergebnisse durch seinen Opt-Out-Charakter verfälschen.

Zu berücksichtigen sind je nach Anwendungsfall auch das Thema Profiling und die Rechte der Betroffenen aus Art. 22 DSGVO. Unter Profiling versteht die DSGVO eine „automatisierte Verarbeitung per-

2 EuGH-Urteil vom 01.08.2022 (Rs. C-184/20) https://curia.europa.eu/juris/document/document.jsf?text=&docid=263721&pageIndex=0&doclang=DE&mode=lst&dir=&occ=first&part=1&cid=1391320 (zuletzt aufgerufen: 18.10.2022).

sonenbezogener Daten, die darin besteht, dass diese personenbezogenen Daten verwendet werden, um bestimmte persönliche Aspekte, die sich auf eine natürliche Person beziehen, zu bewerten“ (Art. 4 Nr. 4 DSGVO). Vor dem Profiling – oder besser seinen Folgen – sollen die Rechte aus Art. 22 DSGVO schützen. Aber genau das Profiling ist in sehr vielen Fällen das Ziel der unternehmerischen Anwendung von Data Science.

Pseudonymisierung und Anonymisierung

Um diesen Problemen zu begegnen, werden personenbezogene Daten pseudonymisiert oder anonymisiert. Im ersten Fall lässt sich der Personenbezug mittels eines Schlüssels wiederherstellen, wobei hingegen im zweiten Fall der Personenbezug (im Idealfall) dauerhaft verloren ist.

Im Fall der Pseudonymisierung bleiben die Regeln der DSGVO wegen der möglichen Identifizierbarkeit gültig. Nur die echte Anonymisierung löst die rechtlichen und methodischen Probleme.

Auch wenn die Anonymisierung nicht im Katalog der Beispiele nach Art. 4 Nr. 2 DSGVO genannt ist, handelt es sich dennoch um eine Datenverarbeitung i. S. d. DSGVO. Sie stellt gleichsam einen Sonderfall der Löschung oder Vernichtung von Daten dar. Die DSGVO findet demnach Anwendung auf die Anonymisierung als Verarbeitung, aber nicht auf die Daten, die Folge der Anonymisierung sind.

In der Praxis bedeutet dies, dass auch für diesen Vorgang eine Rechtsgrundlage gefunden und die Betroffenen nach Art. 13 oder 14 DSGVO informiert werden müssen, neben den üblichen weiteren Pflichten (z. B. Aufnahme ins Verzeichnis der Verarbeitungstätigkeiten). Da von einer erfolgreichen Anonymisierung nur geringe Risiken für die Rechte und Freiheiten der Betroffenen ausgehen, dürfte das oben angeführte berechtigte Interesse eine stabile Rechtsgrundlage bilden.

Bildquelle: Lazy_Bear

Bei einer echten Anonymisierung geht der Personenbezug idealerweise dauerhaft verloren.

Angriffsmöglichkeiten und Sicherheitsmaßnahmen

Genau hier liegt nun die Crux: Ist die Anonymisierung erfolgreich oder nicht? Und genau hier liegt auch die Herausforderung für die junge Wissenschaft Data Science.

Ein erfolgreiches Anonymisierungsverfahren muss derart durchgeführt werden, dass ein Personenbezug nicht mit angemessenen Mitteln wiederhergestellt werden kann. Dem risikobasierten Ansatz des Datenschutzrechts folgend, muss sich dabei die Qualität der Anonymisierung an dem für die Betroffenen bestehenden Risiko orientieren. Ein analoger Maßstab ist hier die DIN 66399 zur Datenvernichtung. Dabei geht es also um die Frage, über welche Mittel, Zeit und Kenntnisse ein potenzieller Angreifer verfügen muss, um aus dem anonymisierten Datensatz einen Personenbezug ganz oder in Teilen wiederherstellen zu können.

Grundsätzlich unterscheidet man folgende Verfahren der Anonymisierung:

- Aggregation: Einzelne Daten werden zu Gruppen zusammengefasst.
- zufallsbasierte Verfahren: Einzelne Daten werden zufallsbasiert verändert.
- synthesebasierte Verfahren: Anhand der statistischen Rahmenbedingungen des Ursprungsdatensatzes werden synthetische Daten erzeugt, die über die gleichen statistischen Parameter verfügen.

Eine Gemeinsamkeit aller Verfahren ist, dass sie unterschiedliche Stärken und Schwächen haben. Denn aus der Sicht von Data Science ist selbstverständlich nicht allein der Grad der Sicherheit von Bedeutung, sondern in besonderem Maß, ob die anonymen Daten auch noch in hinreichendem Umfang Informationen enthalten, die für den gewünschten Zweck nutzbar gemacht werden können. Tritt der Aspekt der Informationsgewinnung ungefiltert in den Vordergrund, leidet meist die Sicherheit.

Jeder anonymisierte Datensatz lässt sich, vorbehaltlich des Aufwands, auch angreifen. Diese Angriffe sind meistens nur in einem bestimmten Umfang erfolgreich. Aber hier schlagen die Risiken von Big Data zu: Wenn aus einem anonymisierten Datensatz von einer Million betroffenen Personen sich nur 0,1 % re-anonymisieren lassen, dann tritt ein Schaden für 1.000 Personen ein.

Die Wahrscheinlichkeit eines erfolgreichen Angriffs hängt zum einen von den getroffenen Schutzmaßnahmen insbesondere im Bereich der Vertraulichkeit ab. Das bedeutet auch, dass die anonymisierten Daten in der Praxis risikobasiert in dem Umfang zu schützen sind, wie es auch für die Ursprungsdaten gilt. An die Qualität der technischen und organisatorischen Maßnahmen, gekoppelt mit Maßnahmen der Informationssicherheit, sind bei Unternehmen, die es mit Data Science oder Big Data zu tun haben, besonders hohe Maßstäbe anzulegen.

Zum anderen ist ein Angriff umso wahrscheinlicher, je leichter der Angreifer über zusätzliche Informationen über die Ursprungsdaten verfügt. Angriffe auf anonymisierte Daten sind oft wie ein Puzzle, das einfacher wird, je mehr Teile man bereits gelegt hat. Daher müssen neben dem grundlegenden Schutz der personenbezogenen Daten auch besondere Maßnahmen zum Geheimnisschutz, z. B. zu den tatsächlich angewandten Anonymisierungsverfahren, getroffen werden.

Fazit

Die Regeln des Datenschutzes behindern die grundlegenden Ziele von Data Science im unternehmerischen Kontext durchaus, dies jedoch aus einem guten Grund und zwar um die Grundrechte der Betroffenen zu wahren. Eine Anwendung von Data Science, die diese Regeln ignoriert, löst für ein Unternehmen hohe Haftungsrisiken aus und kann in der Folge auch als Compliance-Versagen eine persönliche Haftung der Geschäftsleitung bedeuten.

Das muss jedoch nicht bedeuten, dass auf Data Science und Big Data verzichtet werden sollte. Die genannten Probleme lassen sich mit einem bestimmten Aufwand lösen.

Das Scheitern derartiger Projekte liegt nicht in einem angeblichen fortschrittsfeindlichen Datenschutzrecht, sondern eher daran, dass es an der Bereitstellung von Ressourcen zur legitimen Umsetzung fehlt. „Wasch mir den Pelz, aber mach mich nicht nass“ – so wird der Fortschritt in Wirklichkeit behindert.

Andreas Sutter
Andreas Sutter ist als Datenschutzbeauftragter, Dozent und Unternehmensberater tätig. Zudem berät er Mandanten aus den Branchen Finanzdienstleistung, Wohnungswirtschaft und Rechtsberatung.

WISSEN, DAS ANKOMMT.

Bestelloptionen

Infodienst KI-Recht.IT-Sicherheit.Datenschutz.

Sie haben Fragen zum Produkt oder benötigen Unterstützung bei der Bestellung? Unser Kundenservice ist für Sie da:

08233 / 381-123 (Mo - Do 7:30 - 17:00 Uhr, Fr 7:30 - 15:00 Uhr)

service@forum-verlag.com

Oder bestellen Sie bequem über unseren Online-Shop:

Jetzt bestellen

FORUM VERLAG HERKERT GMBH, Mandichostraße 18, 86504 Merching,
Tel.: (08233) 381 123, E-Mail: service@forum-verlag.com, Web: shop.forum-verlag.com