2. Kapitel: Funktion und Manipulation der algorithmenbasierten Personalisierung

Nach dieser Einleitung wird nun im *zweiten Kapitel* vertieft auf das Szenario dieser Untersuchung eingegangen, an dem die praktische Bedeutung der Resilienz als Rechtsbegriff für die Daten- und IT-Sicherheit später demonstriert werden soll. Dabei werden v.a. die Funktionsweise der *algorithmenbasierten Personalisierung* und die Möglichkeit zur *Manipulation* dargestellt.

Im Einzelnen wird zunächst eine Darstellung anhand des Daten-, Informations-, Wissensmodells (DIW-Modells) vorgenommen (A.), mithilfe dessen die grundlegende Funktionsweise aus einer informationsrechtlichen Sicht nachgezeichnet werden soll. Als Ergänzung sollen in einem weiteren Schritt zumindest grob die zugrundeliegenden technischen Aspekte erläutert werden (B.). Im letzten Abschnitt dieses Teils (C.) wird sodann die Manipulation der Informationen als Angriff in diesem Szenario dargestellt.

A. Ermittlung von Personenwissen nach dem DIW-Modell

In einem ersten Schritt soll hierfür die Verarbeitung von Daten für die Personalisierung näher dargestellt werden.

Dabei ist Personalisierung ein Ausdruck einer neuen Form der Informationsverarbeitung: So waren IT-Systeme früher meist eher statische Systeme in denen Informationen vorrangig nur aufbewahrt, organisiert und zu spezifisch festgelegten Zwecken verarbeitet wurden, z.B. die Verwaltung und Pflege einer Kund:innendatenbank. Eine neue Erscheinung sind hingegen sog. „lernende Systeme", die automatisiert und teilweise sogar autonom (dazu in Abschnitt B.) aus den Daten zunächst Informationen sowie Wissen erzeugen, auf Basis dessen sie dann eine Entscheidung treffen und eine Steuerungswirkung herbeiführen. Zu beachten ist, dass sich das hier beschriebene Modell nur auf elektronische Datenverarbeitung bezieht und somit keine allgemeingültigen Definitionen für „Informationen" und „Wissen" liefern kann, welche weit über den Gegenstand der hier vorliegenden Untersuchung hinausgehen und wohl bis in philosophische Fragestellungen

hinein reichen würde.[79] Siehe zur Illustration der nachfolgenden Ausführungen bereits die Abbildung, S. 43, Abb. 2.

I. Daten

Die erste Kategorie des Informationsmodells bilden die Daten. Entgegen dem Sprachgebrauch bilden die Daten selbst oft gar nicht den Anknüpfungspunkt für rechtliche Regelungen wie etwa das Datenschutzrecht.[80] Denn der Begriff Daten beschreibt lediglich Zeichen, die durch Speicherung auf einem Datenträger physisch verkörpert werden und als (potenzielle) Grundlage für Informationen[81] sowie letztlich auch für Wissen und Entscheidungen dienen.[82] Diese Zeichen sind als solche jedoch rein syntaktischer Natur und können daher als „global" und „neutral" angesehen werden.[83] Sie erschöpfen sich auf tiefster Computer-Ebene bekanntlich in sog. Bits, deren Wert nur 0 oder 1 betragen kann. Jeweils 8 Bits bilden ein Byte, was in der IT regelmäßig das kleinste adressierbare Element mit 256 (2 hoch 8) möglichen Zuständen darstellt. Eine logische Ebene höher werden in der Informatik verschiedene Datentypen unterschieden, so z.B. Ganze Zahlen (INTEGER), Zeichen/Buchstaben (Char) oder logische Werte, insbesondere true/false (BOOLEAN). Aus mehreren Einzeldaten (ggf. mit unterschiedlichen Datentypen) lassen sich schließlich komplexere Datenstrukturen wie ein Datensatz abbilden. Ein Datensatz könne etwa in einem Serverlog-Eintrag wie diesem bestehen:

> 203.0.113.195 - user [07/Oct/2024:10:43:00 +0200] "GET /index.html HTTP/2.0" 200 2326

Diese Datensätze tragen schon erste Kennzeichen einer Semantik, da im Rahmen der Programmierung die Auswahl und die Anordnung der einzelnen Daten bereits mit einer verwendungsspezifischen Intention erfolgte. Damit wird der an sich der Information zuzuordnende Verwendungskon-

79 Vgl. *Aamodt/Nygård*, Data & Knowledge Engineering, Vol. 16 (1995), 191 (193).
80 BT-Drs. 17/8999, Fünfter Zwischenbericht der Enquete-Kommission „Internet und digitale Gesellschaft", S. 21.
81 *Albers*, in: Spiecker gen. Döhmann/Collin, Generierung und Transfer staatlichen Wissens im System des Verwaltungsrechts, 50 (54); *Spiecker gen. Döhmann*, RW 2010, 247 (253); *Jandt*, in: Hornung/Schallbruch, IT-Sicherheitsrecht, 391 (400), Rn. 23.
82 *Kloepfer*, Informationsrecht, S. 26; *Ebsen*, DVBl 1997, 1039 (1039).
83 *Aamodt/Nygård*, Data & Knowledge Engineering, Vol. 16 (1995), 191 (203).

text bereits in die Datenerhebung hineingetragen (bei einem Server-Log z.B. die Erkennung von Angriffen oder aber die Reichweiten-Messung). Insofern kann man attestieren, dass die jeweils erzeugten Datenstrukturen bereits *durch den informatorischen Verwendungskontext* determiniert sind.[84] Oder anders formuliert: Die Semantik der erzeugten Datensätze ist so gestaltet, wie es zur Interpretation zumindest der primären Informationen dienlich ist.

Damit von „personenbezogenen Daten" gesprochen werden kann, müssen diese Daten bzw. die daraus interpretierbaren Informationen zumindest einer Person zugeordnet werden können (auch dazu sogleich ausführlicher).

II. (Persönliche) Information

Die nächsthöhere Kategorie im Verarbeitungsprozess stellen die Informationen dar; sie sind von den Daten zu unterscheiden, aus denen sie interpretiert wurden.[85] Auch durch die eben beschriebenen determinierten Datenstrukturen reduziert sich de facto lediglich das Maß der Interpretationsleistung, die erforderlich ist, damit ein Datensatz zu einem Sinnelement und damit zu einer Information werden kann.[86]

Zur Interpretation dieser Daten sind stets entsprechende Kontextinformationen bzw. Kontextwissen erforderlich,[87] etwa über den Datentyp: Andernfalls lässt sich beispielsweise nicht erkennen, ob die Zeichenkette „AFFE" als das gleichnamige Säugetier (Datentyp: char) oder die Zahl 45054 (hexadezimal) interpretiert werden soll.[88]

Einen Sonderfall bei der Informationsinterpretation aus Daten stellen sog. Big-Data-Anwendungen dar. Dort werden große Datenmengen analy-

84 Vgl. *Aamodt/Nygård*, Data & Knowledge Engineering, Vol. 16 (1995), 191 (203), die insofern dahingehend differenzieren, dass zwar die Daten an sich neutral sind, aber nicht die Art und Weise der Datenerzeugung.

85 *Jendrian/Weinmann*, DuD 2010, 108 (108); *Eckert*, IT-Sicherheit, S. 4.

86 *Albers*, in: Spiecker gen. Döhmann/Collin, Generierung und Transfer staatlichen Wissens im System des Verwaltungsrechts, 50 (54).

87 *Freimuth,* Die Gewährleistung der IT-Sicherheit Kritischer Infrastrukturen, S. 66; *Albers*, in: Hoffmann-Riem/Schmidt-Aßmann/Voßkuhle, Grundlagen des Verwaltungsrechts, 107, § 22, Rn. 12 f.; zur Abgrenzung zwischen Informationen und Wissen sogleich unter III.

88 *Jendrian/Weinmann*, DuD 2010, 108 (108); weiterführend zu den Schwierigkeiten bei Datensignatur und Informationsinterpretation: *Fox*, DuD 1997, 386 (387).

siert, die häufig ursprünglich zu anderen Zwecken erhoben wurden.[89] Aus diesen Daten soll folglich dann eine andere, als die ursprünglich vorgesehene Information interpretiert werden, so dass die jeweilige Semantik der Datensätze nicht mehr zwingend hilfreich ist, sondern ganz im Gegenteil sogar erhöhten Aufwand hervorrufen und größeren Raum für Fehlinterpretationen schaffen kann.

Eine durch Interpretation gewonnene oder auch jede sonstige Information enthält bei logischer Betrachtung[90] zumindest ein Subjekt[91] sowie einen Wert, eine Handlung oder eine Eigenschaft (nachfolgend: Parameter), den sie diesem Subjekt zuweist. Nur eine Information bietet somit auch für das Datenschutzrecht hinreichende persönlichkeitsrechtlich relevante Anknüpfungspunkte, die ihre Verarbeitung unter Aspekten des „Datenschutzes" daher entweder als legitim oder als verboten erscheinen lassen können. Insbesondere knüpfen an die Information die Interessen des Trägers des Grundrechts auf informationelle Selbstbestimmung, ob und inwieweit staatliche Stellen oder private Dritte diese Information erhalten und verarbeiten können sollen.[92] Das dem Schutz der informationellen Selbstbestimmung dienende Datenschutzrecht müsste daher streng genommen auch „Informationsschutzrecht" heißen.[93] Bezüglich der „Datensicherheit" ist der Begriff hingegen sachgerecht, da durch ein entsprechendes Ereignis wie einen Hacker-Angriff stets zunächst die Daten betroffen sind.

Eine Information kann und wird regelmäßig aus mehreren Parametern, also Einzelinformationen bestehen, die sich aber zweckgerichtet als eine Information zusammenfassen lassen.[94]

89 Vgl. *S. Schulz*, in: Gola/Heckmann, DS-GVO, 3. Auflage 2022, Art. 6, Rn. 151.

90 Siehe zur syntaktischen, semantischen und pragmatischen Dimension von Information: *Kloepfer*, Informationsrecht, S. 24, Rn. 53 ff.

91 Im Datenschutzrecht stets eine Person; in Abgrenzung dazu eine Sache, wenn es sich um eine Sachinformation handelt.

92 Fn. 80.

93 *Spiecker gen. Döhmann*, RW 2010, 247 (255); *Albers*, in: Hoffmann-Riem/Schmidt-Aßmann/Voßkuhle, Grundlagen des Verwaltungsrechts, 107 (114), Rn. 8 f.; auch die DSGVO knüpft in der insoweit missverständlichen Definition „personenbezogener Daten" im Kern an Informationen an, die sich auf eine betroffene Person beziehen (Art. 4 Nr. 1 DSGVO): *Veil*, NVwZ 2018, 686 (687).

94 Vgl. *Steinmüller*, Informationstechnologie und Gesellschaft, S. 197.

Beispiele für persönliche Informationen aus o.g. Auflistung sind daher:

– Bei seinem Webseitenaufruf am 01.03.2024 um 15:00 nutzte X das Endgerät Y.
– Am [Datum, Uhrzeit] suchte X nach das Produkt Y.

Aus der Interpretation von Daten ergeben sich indes nicht zwingend sinnstiftende Informationen. Dies ist zwar der Regelfall und insbesondere dann zu erwarten, wenn Informationen von Hand in ein IT-System eingegeben werden, also z.B. in das Bestellformular eines Webshops und damit quasi erst durch die Eingabe in Daten überführt werden.

Es mag allerdings auch andere Fälle geben. Im Normalfall wird bei o.g. Beispiel auf programmtechnischer Ebene des Webshops höchstwahrscheinlich eine MySQL-Datenbank gefüllt, wenn der/die Nutzer:in das Bestellformular ausfüllt. In dieser Datenbank, die sich letztlich als Tabelle verstehen lässt, wird eine neue Zeile angelegt und entsprechend der Spalten (z.B. „Vor- und Nachname" „Anschrift", „Zahlungsdaten") befüllt. Mithin wurden persönliche Informationen eingegeben, die im Idealfall nun durch die Daten als Träger repräsentiert werden.[95]

Gehen wir aber davon aus, der Programmierer des Webshops beherrschte sein Handwerk nicht und die Eingabe-Funktion ist grob fehlerhaft: Sie legt nicht für jede(n) Kund:in eine neue Zeile an, sondern verteilt die einzelnen Angaben eines bzw. einer Kund:in nach dem Zufallsprinzip über die Zeilen der Tabelle. In diesem Fall wurden unstreitig Daten geschaffen, eine Information zu einer einzelnen Person werden sie aus diesen Daten indes kaum erkennen können. Besteht die gewünschte Information dagegen darin, aus welchen PLZ-Bereichen die Kund:innen kommen, so lässt sich diese Information gleichwohl auch aus einer solch „fehlerhaft" gefüllten Datenbank interpretieren.[96]

Insofern verdeutlicht sich an diesem Beispiel die Subjektivität der Interpretation, d.h. die Gewinnung von Informationen durch die Interpretation

95 *M. Wagner,* Datenökonomie und Selbstdatenschutz, S. 33; *Börding et al.*, CR 2017, 134 (134).

96 Jenseits dessen gibt es außerdem Fälle, etwa in Folge von Angriffsszenarien oder auch bei technischen Defekten, bei denen sog. Stör- oder Zufallsdaten, die sich auch bei noch so ambitionierter Interpretation überhaupt nicht zu einer Information zusammenführen lassen. Dies müsste im Idealfall auch auf dieser Ebene der „Datenanalyse" erkannt werden, bevor hieraus unzutreffende Informationen interpretiert werden; *Tremmel*, Neue Schadsoftware möchte IoT-Geräte zerstören, golem.de vom 26.06.2019.

von Daten ist stark von dem individuellen Zweck- und Verwendungskontext abhängig[97] und jede Information ist damit „zweckrelativ".[98] Werden Informationen an Dritte weitergegeben, werden sie zunächst (wieder) in Daten überführt und müssen vom Empfänger neu als Informationen interpretiert werden. Diese Informationen sollten zwar zumeist, müssen aber nicht zwingend mit der ursprünglichen Information übereinstimmen.

III. Wissen

Aus Informationen kann in einem weiteren Schritt „Wissen" generiert werden. Auch wenn es kaum eine allgemeingültige Abgrenzung zwischen Information und Wissen geben kann,[99] lässt sich zumindest attestieren, dass Wissen durch die lernende Verknüpfung von Einzelinformationen entsteht.[100] Die Einzelinformationen werden hierbei „organisiert und systematisiert",[101] um daraus einen über die quantitative Informationsmenge hinausgehenden Erkenntnisgewinn zu erzielen.

Wissensgenerierung ist damit, wie zuvor die Interpretation von Daten zur Informationserlangung, ein subjektiver Prozess, der abhängig von dem Verarbeitungskontext und der Zweckrichtung des Verarbeiters ist (dazu sogleich mit einem Beispiel). Erfolgt diese Interpretation im Wege einer automatisierten Verarbeitung durch Algorithmen, so treffen diese diverse Zwischenentscheidungen im Umgang mit den verfügbaren Informationen. Diese können insbesondere in die vier Kategorien *Priorisierung, Klassifikation, Assoziation* und *Filterung* unterteilt werden.[102]

Für das zu erlangende Wissen ist für das hier betrachtete Szenario zwischen zwei Kategorien zu unterscheiden. Zum einen das Wissen über eine spezifische Person, das z.B. in einem Profil mit entsprechenden übergreifenden Eigenschaften zusammengefasst wird und am Ende z.B. in der Zahlungsbereitschaft für ein bestimmtes Produkt ausgedrückt wird. Es wird in diesem Kontext als *Personenwissen* bezeichnet.

---

97 Vgl. *M. Wagner,* Datenökonomie und Selbstdatenschutz, S. 32

98 *Steinmüller*, Informationstechnologie und Gesellschaft, S. 199 f.; *Aamodt/Nygård*, Data & Knowledge Engineering, Vol. 16 (1995), 191 (198).

99 *M. Wagner,* Datenökonomie und Selbstdatenschutz, S. 30.

100 Vgl. *Aamodt/Nygård*, Data & Knowledge Engineering, Vol. 16 (1995), 191 (200); als „subjektive […] Vernetzung von Informationen": *Picot/Neuburger*, ZfCM 2005, 76 (76).

101 *Spiecker gen. Döhmann*, RW 2010, 247 (253).

102 *Diakopoulos*, Digital Journalism 2015, 398 (400 ff.).

Zum anderen gibt es Wissen aus der vergleichenden Betrachtung von Informationen vieler Personen, etwa um bestimmte Muster zu erkennen (z.B. dass sich Nutzer:innen die sich für das Produkt A (oder auch zugleich für B und C) interessieren mit hoher Wahrscheinlichkeit auch für Produkt D interessant finden werden). Dies kann man auch als *abstraktes Lernwissen oder Wissensbasis*[103] bezeichnen.

Diese Wissensgenerierung kann auch weiter verschachtelt sein, so dass aus einzelnen Wissenskategorien eine übergeordnete Erkenntnis erlangt werden kann.[104] Im Fall personalisierter Preise setzt sich das Wissen über die Zahlungsbereitschaft beispielsweise aus zwei Unterkategorien zusammen, wie dem Wissen über die Produktinteressen und dem Wissen über die Preissensibilität.

Schließlich ist zu berücksichtigen, dass auch die Kategorisierung in Informationen und Wissen subjektiver Natur ist. Für denjenigen, der die Verarbeitung vornimmt, werden aus Daten Informationen interpretiert und hieraus Wissen generiert. Wird das Wissen z.B. über ein Produktinteresse hingegen an Dritte weitergegeben, kann es sich für diesen wiederrum zunächst nur als aus Daten interpretierte Information darstellen.

IV. Entscheidung und Verhaltenssteuerung

Die zuvor genannten Wissenskategorien werden im Rahmen des Entscheidungsprozesses unterschiedlich einbezogen. Zum einen muss situationsabhängig untersucht werden, welches Personenwissen für eine bestimmte Entscheidung relevant ist. Hinsichtlich des abstrakten Lernwissens muss darüber hinaus noch geprüft werden, ob dieses auf die konkrete Person anwendbar ist, z.B. indem eine entsprechende Gruppenzugehörigkeit festgestellt wird.

Am Ende einer solchen Verarbeitung steht eine Entscheidung, bei der das gewonnene Wissen angewendet und in eine Interaktion mit dem Nutzer überführt wird. Die o.g. Wissenskategorien wirken sich dabei wie folgt aus: Zum einen wird verglichen, wie sich andere Personen in der Situation verhalten haben - das Profil des Betroffenen wird insoweit verwendet, als dass eine Zuordnung zu einer bestimmten Gruppe vorgenommen wird.

103 Siehe zu dem Begriff Wissensbasis: *Beierle/Kern-Isberner*, Methoden wissensbasierter Systeme, S. 11.

104 Vgl. *Aamodt/Nygård*, Data & Knowledge Engineering, Vol. 16 (1995), 191 (200).

Daneben wird das Personenwissen, also z.B. das Wissen über Produktinteressen aus Informationen wie der individuellen Bestellhistorie oder bisherige Reaktionen in sozialen Netzwerken abgeleitet.

Beides zusammen führt dann zu einer *algorithmenbasierten Entscheidung*[105] wie etwa dem/der Nutzer:in einen bestimmen Inhalt, ein Produkt oder einen ermittelten Preis für ein Produkt anzubieten. Damit findet eine indirekte *Steuerung* des/der Nutzer:in statt, d.h. im Erfolgsfalle wird diese(r) aufgrund der (passenden) Entscheidung des Algorithmus' zur Annahme des Angebots motiviert.

Die Steuerungswirkung auf den/die Nutzer:in liegt mithin entweder in der Auswahl, d.h. Filterung und der Präsentation von bestimmten Inhalten oder in der Vorgabe eines (vermeintlich) attraktiven Preises. Der/die Nutzer:in, welche(r) ein solches Angebot annimmt, hat diese Steuerungswirkung gebilligt und sich insoweit -idealtypisch freiwillig und wissentlich- aus Sicht des Dienstanbieters steuern lassen. Die Steuerung beschreibt im Falle personalisierter Dienste die Übernahme eines Entscheidungsprozesses, d.h. der/die Nutzer:in verzichtet im Vertrauen auf das Dienstangebot auf eine eigene, persönliche Entscheidung.

Schlägt die Steuerungswirkung fehl, reagiert der/die Nutzer:in also nicht wie beabsichtigt, so kann aus dessen/deren Reaktion, die wiederrum eine Information darstellt, aber zumindest neues individuelles Wissen „erlernt" werden. Etwa wenn der/die Nutzer:in eine Produktseite mit einem personalisierten Preis aufruft und sodann unter „Ähnliche Artikel" ein günstigeres Produkt aussucht. Das spräche möglicherweise dafür, dass die Zahlungsbereitschaft zu hoch angesetzt wurde, was bei künftigen Preisentscheidungen berücksichtigt werden kann. Im Ergebnis führt auch dies zu einer stetigen Präzisierung auch des Wissens über jede einzelne Person und damit grundsätzlich zu immer besseren Entscheidungen und erfolgreicherer Verhaltenssteuerung.

105 Es ist insofern darauf hinzuweisen, dass die Datenethikkommission ein abweichendes Begriffsverständnis verwendet, welches teilweise auch in der Kommentarliteratur zu Art. 22 DSGVO rezipiert wird: Hier meinen algorithmenbasierte Entscheidungen solche, bei denen für eine Entscheidung immer noch ein Mensch verantwortlich ist, der aber durch Algorithmen unterstützt wird; Was hier als algorithmenbasierten Entscheidung beschrieben wird, entspricht nach der Begrifflichkeit der Datenethikkommission dem algorithmendeterminierten Entscheidung, siehe: DEK, Gutachten der DEK, Oktober 2019, S. 17; *Buchner*, in: Kühling/Buchner, Datenschutz-Grundverordnung/BDSG, 4. Auflage 2024, Art. 22, Rn. 14.

V. Zusammenfassung

Die Daten stellen die technische Grundlage dar, die durch Algorithmen verarbeitet werden. Aus ihnen werden im ersten Interpretationsschritt Informationen und aus diesen wiederrum in einem zweiten Interpretationsschritt Wissen gewonnen. Hierzu zählt sowohl individuelles *Personenwissen* (Person A interessiert sich für X) als auch *abstraktes Lernwissen*, z.B. über bestimmte Gruppeneigenschaften (Gruppen mit Eigenschaft B, C, D interessieren sich mit hoher Wahrscheinlichkeit für X). Beide Wissenskategorien dienen als Grundlage einer automatisierten, personalisierten Entscheidung und diese verursacht bei positivem Verlauf eine Steuerungswirkung bei dem von dieser Entscheidung Betroffenen.

B. Technische Grundlagen

Im nun folgenden Schritt werden die technischen Grundlagen für den zuvor dargestellten Veredlungsprozess von Daten über Informationen und Wissen bis hin zu einer Entscheidung kurz erläutert. Grundlegend findet dieser Prozess in Form einer Verarbeitung statt. Eine solche kann sowohl automatisiert (I.) als auch autonom (II.) erfolgen. Unter III. wird dann noch spezifisch auf die Verarbeitung im Rahmen personalisierter Dienste eingegangen.

I. Automatisierte Verarbeitung

Von einer Automatisierung i.S.d. Art. 4 Nr. 2 DSGVO kann bereits dann gesprochen werden, wenn eine Verarbeitung mit technischen Hilfsmitteln vorgenommen wird.[106] Dies umfasst insbesondere die elektronische Datenverarbeitung (EDV),[107] bei der die Daten anhand einer programmtechnisch, d.h. durch Algorithmen vorgegebene Logik verarbeitet werden.

Denkbar sind damit etwa Empfehlungen anhand sachlicher Verknüpfungen, wie die Empfehlung entsprechender Zubehör- zu einem Hauptartikel. Bei dieser Form der Verarbeitung wurde programmtechnisch eine entsprechende Funktion implementiert, mit der in einer Datenbank einem Haupt-

106 *Herbst*, in: Kühling/Buchner, Datenschutz-Grundverordnung/BDSG, 4. Auflage 2024, Art. 4, Rn. 17.

107 *Schild*, in: BeckOK DatenschutzR, 47. Edition 2024, Art. 4 DSGVO, Rn. 34.

artikel ein oder mehrere Zubehörartikel zugeordnet werden können und diese daher dem/der Kundin, wenn er oder sie den Hauptartikel kauft, angeboten werden. Auch automatisierte Empfehlungssysteme (dazu sogleich ausführlich), die z.B. Artikel mit ähnlichen Eigenschaften zu den bereits ausgewählten vorschlagen, fallen unter diese automatisierte Verarbeitung.

Entscheidend für die Abgrenzung zur nachfolgenden autonomen Verarbeitung ist insoweit, dass die Logik und die Entscheidung der Zuordnung und Empfehlung der Zubehörartikel final von einem menschlichen Entscheider festgelegt bzw. getroffen wird (*regelbasierte Programmierung*).[108] Das Sachwissen über diese Zuordnung (Zubehör-/Hauptartikel) oder zumindest das Regelwissen zur Programmierung, z.B. eines automatisierten Empfehlungssystems ist mithin bei diesem schon vorhanden und wird von ihm in das informationstechnische System eingebracht.

II. Autonome Verarbeitung durch maschinelles Lernen

Anders verhält es sich bei autonomer Verarbeitung. Hier beruhen die Empfehlungen des Systems auf Entscheidungen, denen ein maschinelles Lernverfahren vorangegangen ist. Wie die Bezeichnung des *maschinellen Lernens (ML)* bereits nahelegt, ist es hier nicht mehr (allein) ein menschlicher Entscheider, der durch die Programmierung auf Basis bereits vorhanden Wissens den Verarbeitungsprozess abschließend vorgibt.[109] Vielmehr kann ein ML-System nach einer entsprechenden Implementierung selbstständig lernen, d.h. Wissen erwerben,[110] indem es Muster oder Korrelationen erkennt[111] und auf dieser Basis auch eigenständige Entscheidungen trifft.[112] Es handelt sich mithin immer noch um eine algorithmenbasierte Personalisierung. Allerdings werden die Algorithmen zur Entscheidungsfindung nicht mehr (ausschließlich) durch ein(e) Programmierer:in vorgegeben;

108 Vgl. *Staehelin*, GRUR 2022, 1569 (1569); *Steege*, MMR 2019, 715 (716).

109 Vgl. *Sattler*, in: Ebers/Steinrötter, Künstliche Intelligenz und smarte Robotik im IT-Sicherheitsrecht, 197 (220 f.).

110 *Portugal/Alencar/Cowan*, Expert Systems with Applications, Vol. 97 (2018), 205 (206).

111 *Buxmann/H. Schmidt*, in: Buxmann/Schmidt, Künstliche Intelligenz, 3 (11); *Sattler*, in: Ebers/Steinrötter, Künstliche Intelligenz und smarte Robotik im IT-Sicherheitsrecht, 197 (220 f.) m.w.N.

112 *Hof*, in: Tinnefeld/Buchner/Petri/Hof, Einführung in das Datenschutzrecht, 477 (485), Rn. 18; zu den unterschiedlichen Verfahren des MLs: *Buxmann/H. Schmidt*, in: Buxmann/Schmidt, Künstliche Intelligenz, 3 (10 ff.).

vielmehr verändert sich die Entscheidungsfindung eines ML-Systems beim Lernen autonom.[113]

Es wird somit z.B. möglich die Aktivitäten der Nutzer:innen auf einem Online-Marktplatz, einem sozialen Netzwerk oder einer Online-Suchmaschine zu erfassen und hierin bislang unbekannte Muster zu erkennen und auf Basis dessen neue Empfehlungen zu generieren. Sowohl die Entscheidungen einer autonomen Verarbeitung, als auch solche die (ausschließlich) auf einer automatisierten Verarbeitung beruhen, unterfallen grundsätzlich dem Tatbestand des Art. 22 Abs. 1 DSGVO, der aber bei den hiesigen personalisierten Diensten mangels rechtlicher oder in ähnlicher Weise erheblich beeinträchtigenden Wirkung dieser Entscheidungen nicht erfüllt sein dürfte.[114]

III. Verarbeitung in personalisierten Dienstangeboten

Entscheidungen werden bei den digitalen Diensten in der Form personalisierter Dienstangebote getroffen. Hierfür werden (ggf. auch ML-gestützte) Empfehlungssysteme verwendet. So werden sie u.a. von dem Onlinemarktplatz-Anbieter *Amazon* genutzt, um dem/der Nutzer:in entsprechend personalisierte Angebote machen zu können.[115] Im Falle einer personalisierten Suche (z.B. *Google Personalized Search)* können auf eine Suchanfrage hin die Ergebnisse an den jeweiligen Nutzer angepasst werden, sowohl in ihrer Auswahl (z.B. durch das Entfernen zumindest vermeintlich irrelevanter Ergebnisse) als auch bezüglich der angezeigten Reihenfolge der Suchergebnisse (sog. Re-Ranking).[116] Schließlich werden Empfehlungssysteme für die

113 Vgl. *Buxmann/H. Schmidt*, in: Buxmann/Schmidt, Künstliche Intelligenz, 3 (9 f.).

114 So etwa zu Preisdifferenzierungen und personalisierter Werbung: *Martini*, in: Paal/Pauly, DSGVO, BDSG, 3. Auflage 2021, Art. 22, Rn. 27 ff. Ob man bei den Feeds sozialer Netzwerke oder den Ergebnissen von einer Online-Suchmaschine das Erfordernis einer „nicht nur marginalen" Beeinflussung der persönlichen Entfaltungsfreiheit (*Martini*, ebd.) erreicht, und somit zu einem anderen Ergebnis kommt, wäre zumindest diskussionswürdig, soll aufgrund des Fokus' dieser Untersuchung auf die Datensicherheit aber nicht weiter verfolgt werden.

115 *Linden/B. Smith/York*, IEEE Internet Computing 2003, 76 (76).

116 *Z. Ma/Pant/Sheng*, ACM TOIS, Vol. 25 (2007), Heft 1, 1 (5); zu Begriff und Funktionsweise des Rankings: *Lewandowski/Kerkmann/Sünkler*, in: Stark/Dörr/Aufenanger, Die Googleisierung der Informationssuche, 75 (83 f.).

Auswahl der angezeigten Beiträge in den Feeds sozialer Netzwerke verwendet.[117]

Um eine verlässliche technische Darstellung vornehmen zu können, ist die Personalisierung von Preisen noch zu wenig verbreitet und wissenschaftlich erforscht, wobei die nachfolgenden Mechanismen grundsätzlich auch hierfür verwendet werden können.[118] Deshalb wird im Nachfolgenden nur auf die Personalisierung bei den zuvor genannten personalisierten Diensten eingegangen, d.h. die Personalisierung von Suchergebnissen einerseits und das Geben bzw. Hervorheben von Empfehlungen unabhängig von einer konkreten Suchanfrage andererseits.

Empfehlungssysteme (en: Recommender Systems) bringen ihre Entscheidung in der Empfehlung eines bestimmten Elements (en: Item) zum Ausdruck.[119] Inzwischen besteht in Art. 3 lit s) DSA[120] auch eine gesetzliche Definition: ein „Empfehlungssystem" ist demnach

> *„ein vollständig oder teilweise automatisiertes System, das von einer Online-Plattform verwendet wird, um auf ihrer Online-Schnittstelle den Nutzern bestimmte Informationen vorzuschlagen oder diese Informationen zu priorisieren, auch infolge einer vom Nutzer veranlassten Suche, oder das auf andere Weise die relative Reihenfolge oder Hervorhebung der angezeigten Informationen bestimmt;"*

Ein „vollständig automatisiertes System" dürfte wie in Art. 22 Abs. 1 DSGVO „ausschließlich auf einer automatisierten Verarbeitung beruhend" verstanden werden, so dass auch (teil-)autonom agierende, also ML-gestützte Empfehlungssysteme adressiert werden.

Bei den für Empfehlungssysteme verwendeten, technischen Ansätzen ist insbesondere zwischen „Content-based Filtering" (CBF) und „Collaborative Filtering" (CF) zu unterscheiden. In diesen finden sich auch die im Informationsmodell abgeschichteten Formen unterschiedlicher Wissensgenerierung wieder.

---

117 *Ziegler/Loepp*, in: Kollmann, Handbuch digitale Wirtschaft, 717 (719).

118 *Kamishima/Akaho*, in: Cantador/Brusilovsky/Kuflik, Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems, 57 (57 ff.).

119 *Jürgens/Stark/Magin*, in: Stark/Dörr/Aufenanger, Die Googleisierung der Informationssuche, 98 (105); *Ziegler/Loepp*, in: Kollmann, Handbuch digitale Wirtschaft, 717 (719); *Ricci et al.*, Recommender systems handbook, S. 1 ff.

120 Digital Services Act (EU-VO 2022/2065 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Gesetz über digitale Dienste).

Beim CBF werden die (bisherigen) Reaktionen (Feedback) der einzelnen Nutzer:innen auf bestimmte Elemente wie z.B. einem bestimmten Produkt auf einer Webseite erfasst. Äußert er oder sie dieses Feedback ausdrücklich auf Nachfrage zu einem zuvor empfohlenen Element, spricht man von explizitem Feedback; werden die Reaktionen hingegen nur „stillschweigend" überwacht und untersucht, spricht man von implizitem Feedback.[121] Ein Beispiel für letzteres ist, wenn ein(e) Nutzer:in einen Artikel auf einem Online-Marktplatz kauft oder auch nur danach sucht, was als positives Feedback bezüglich dieses Artikels angesehen werden kann.[122] Am Ende werden die aus diesen Feedbacks abzuleitenden Informationen über das bisherige Verhalten des Nutzers oder der Nutzerin in Form eines Profils zusammengeführt.[123] Dieses Profil wird sodann mit den verfügbaren „Content-", d.h. Inhaltsinformationen (auch als „Features" bezeichnet) über neue Elemente abgeglichen und bei einer entsprechenden Ähnlichkeit bzw. Übereinstimmung wird dann das jeweilige Element empfohlen.[124] Hat der/die Nutzer:in z.B. zuvor einen oder mehrere Film(e) positiv bewertet, die dem Comedy-Genre angehören (Feature), so würde das System ihm oder ihr noch weitere Filme aus diesem Genre empfehlen.[125] Das für diesen Abgleich verwendete individuelle Profil entspricht dem Personenwissen nach dem Informationsmodell.

Der Ansatz des CF ist dagegen nicht auf einzelne Nutzer*innen fokussiert, sondern nutzt das gesammelte Feedback von mehreren Nutzer:innen in einer kollaborativen Weise. Im Detail bestehen unterschiedliche Verfahren, zunächst können etwa Kohorten aus Nutzer:innen erstellt, die in der Vergangenheit an ähnlichen Elementen interessiert waren (kohortenbasiertes Verfahren).[126] Daraus wird geschlossen, dass die Übereinstimmungen innerhalb der Kohorten aus der Vergangenheit dazu führen, dass die zuge-

121 *Mahmoud/John*, in: SAI Intelligent Systems Conference (IntelliSys), Enhanced content-based filtering algorithm using Artificial Bee Colony optimisation, 155 (156).

122 *Aggarwal*, Recommender Systems, S. 1. Ausführlich zu dem von Amazon.com genutzten Collaborative Filtering: *Linden/B. Smith/York*, IEEE Internet Computing 2003, 76 (76, 78 f.).

123 *Mahmoud/John*, in: SAI Intelligent Systems Conference (IntelliSys), Enhanced content-based filtering algorithm using Artificial Bee Colony optimisation, 155 (155).

124 *Aggarwal*, Recommender Systems, S. 14. *Badriyah et al.*, in: Seventh International Conference on Innovative Computing Technology (INTECH), A hybrid recommendation system for E-commerce based on product description and user profile, 95 (95 f.).

125 *Ricci et al.*, Recommender systems handbook, S. 11.

126 *Aggarwal*, Recommender Systems, S. 2.

hörigen Nutzer:innen auch auf künftige Empfehlungen gleich oder zumindest ähnlich ansprechen.[127] Im Falle eines Online-Shops kann eine Kohorte aus ähnlichen Nutzer:innen gebildet werden, die in der Vergangenheit bestimmte Artikel gekauft oder bewertet haben. Die daraus aggregierte Menge an Elementen wird dann dem oder der jeweiligen Nutzer*in vorgeschlagen, wenn er/sie aufgrund des bisherigen Feedbacks ebenfalls dieser Kohorte zugeordnet werden kann; natürlich abzüglich aller Elemente, die der/die jeweilige Nutzer:in schon gekauft oder bewertet hat.[128] Daneben existieren auch weniger komplexe Ansätze wie etwa das Co-Visitation-Verfahren[129], bei dem zwei Elemente betrachtet werden, die von vielen Nutzer:innen gemeinsam bzw. direkt aufeinander folgend besucht werden (z.B. Gartenschere und Arbeitshandschuhe in einem Online-Shop). Es wird insofern davon ausgegangen, dass sich dieser Präferenzzusammenhang auch in Zukunft fortsetzt und daher den künftigen „Gartenscherenkund:innen" entsprechend empfohlen auch Arbeitshandschuhe zu kaufen.

Das Wissen über solche Präferenzzusammenhänge zwischen den Elementen (in den jeweiligen Kohorten) gehört im DIW-Modell in die Kategorie des abstrakten *Lernwissens*, was dann bei Erkennung einer entsprechenden Kohortenzugehörigkeit oder bei der Auswahl des entsprechenden Elements auf den/die jeweilige(n) Nutzer:in angewendet wird.

Werden CBF, CF und ggf. auch noch weitere Ansätze[130] kombiniert um eine Entscheidung zu treffen, spricht man von „hybrid filtering". In der Praxis ist dies innerhalb der Erbringung eines Dienstes häufig der Fall, um die Nachteile der einzelnen Ansätze zu kompensieren: So hat etwa CF den Nachteil, dass es keine neuen Elemente empfehlen wird, solange diese noch nicht von anderen Nutzer:innen bewertet wurden.[131] Dieses Problem hat CBF hingegen nicht, da es auch neue Elemente (nur) anhand der Ähnlichkeit zu anderen, bereits von dem/der Nutzer:in bewerteten Elementen

127 *Ricci et al.*, Recommender systems handbook, S. 2.

128 *Linden/B. Smith/York*, IEEE Internet Computing 2003, 76 (76).

129 *Mahmood/Adnan*, in: 9th International Conference on Networking, Systems and Security (NSysS), Detecting Fake Co-visitation Injection Attack in Graph-based Recommendation Systems, 30 (30 f.); *G. Yang/Gong/Cai*, in: Proceedings 2017 Network and Distributed System Security Symposium, Fake Co-visitation Injection Attacks to Recommender Systems, S. 2 f.

130 Exemplarischer weiterer Ansatz: Demographic Filtering, basiert auf demografischen Informationen über die Nutzer:innen wie Alter, Geschlecht, Beruf oder Bildung. um entsprechende Empfehlungen abzuleiten, vgl. *Aggarwal*, Recommender Systems, S. 19; *Ricci et al.*, Recommender systems handbook, S. 8.

131 *Ricci et al.*, Recommender systems handbook, S. 13.

vorschlagen kann.[132] Bei allen Ansätzen kann auch ML eingesetzt werden, um die Ergebnisse weiter zu verbessern.[133]

C. Manipulation der Informationen

Nach der modellhaften Beschreibung der algorithmenbasierten Personalisierung sowie den zugehörigen technischen Grundlagen werden nun die Möglichkeiten der Manipulation durch die Einspeisung korrumpierter Daten dargestellt. Aus diesen werden sodann unrichtige Informationen interpretiert, die dann die Wissenserzeugung sowie die darauf basierenden Entscheidungen beeinträchtigen.

Nach einer allgemeinen Darstellung (I.) wird auf die Unterfälle der singulären (II.) und der pluralen (III.) Informationsmanipulation eingegangen. Dieses Kapitel schließt mit einem Fazit, in dem die Resilienz als mögliche Gegenmaßnahme auf diese Manipulationen positioniert wird (IV.).

I. Allgemeine Darstellung

Es sind zunächst zwei Angriffsvektoren zu unterscheiden, die sich in ihrer unmittelbaren Wirkung unterscheiden:

1. Es werden für einzelne Identitäten der Nutzer:innen falsche Informationen in das System eingebracht. Diese werden im Rahmen des MLs richtig verarbeitet, also etwa klassifiziert, aber da die Informationen objektiv unrichtig sind, wird im System falsches *Personenwissen* erzeugt und am Ende trifft der personalisierte Dienst bezüglich dieser Person auch eine falsche, unpassende Entscheidung. Dies wird nachfolgend als *singuläre Informationsmanipulation* bezeichnet.
2. Werden hingegen viele Identitäten manipuliert bzw. viele künstliche Identitäten geschaffen und etwa über ein sog. Bot-Netz zentral gesteuert, ist es auch möglich, dass insbesondere ein im Modus des Online Lernens operierendes System „umtrainiert" wird, d.h. es wird falsches *Lernwissen* erzeugt indem etwa bei der Co-Visitation alle korrumpierten Systeme

132 Wie zuvor.

133 *Portugal/Alencar/Cowan*, Expert Systems with Applications, Vol. 97 (2018), 205 (205); teilweise wird dies dann auch als computational intelligence-based (CI-based) bezeichnet: *Lu et al.*, Decision Support Systems, Vol. 74 (2015), 12 (14 f.).

positives Feedback bezüglich zweier Elemente abgeben, zwischen denen tatsächlich kein Präferenz-Zusammenhang besteht. Für die Verarbeitung nach dem Informationsmodell bedeutet dies, dass künftig auch von echten, manipulationsfreien Nutzer:innen stammende Daten bzw. Informationen falsch verarbeitet werden, da Verarbeitung auf dem manipulierten Lernwissen beruht, was im Ergebnis deshalb auch bei diesen Nutzer:innen bzw. generell zu falschen Dienstentscheidungen führt. Dies wird nachfolgend als *plurale Informationsmanipulation* bezeichnet.

Beide Angriffsvektoren werden in den folgenden Abschnitten näher dargestellt.

II. Singuläre Informationsmanipulation

Werden nun z.B. durch das manipulierte Endgerät eines/einer Kund:in „unrichtige Informationen" in das System eingebracht, so stellt sich die Frage, wie sich dies im Rahmen des DIW-Modells auswirkt (1.). Anschließend wird die technische Ausgestaltung dieser Manipulation dargestellt (2.).

1. Wirkung nach dem DIW-Modell

In diesem Fall werden Einzelinformationen wie z.B. eine Suchanfrage in das System eingebracht, die wie folgt aussehen könnte:

Von dem Account des A fand am 08.08.19 um 13:30 Uhr eine Suche nach dem Produkt Y statt.

Zu beachten ist, dass diese Information bei einer technisch-neutralen Betrachtung nicht unrichtig ist. Die Produktsuche vom Account des A fand gleichwohl statt. Allerdings ist die abgeleitete Information vor dem vorliegenden Zweckkontext unrichtig. Hier sollen die Produktinteressen des A ermittelt werden. In diesem Kontext kommt es darauf an, ob die Produktsuche tatsächlich ein entsprechendes Produktinteresse indiziert und nicht ob rein technisch betrachtet eine entsprechende „Suchfunktion" aktiviert wurde. Mithin ist die Information vor diesem Hintergrund unrichtig, weil sie als Indikator für die Produktinteressen des A falsch ist.

Unter Berücksichtigung des o.g. Informationsmodells ist weiterhin festzustellen, dass die Datenbasis, die durch die Webseite bei der Suchfunktion

angelegt wurde, im Vergleich zu einem „echten" Suchvorgang unverändert ist. Mithin zeigt sich auch hier, dass es für die Frage, ob eine unrichtige oder richtige Information vorliegt auf den jeweiligen Zweck- und Interpretationskontext ankommt. Aus den manipulierten Daten wird unter Berücksichtigung des vorliegenden Zweckkontexts (Ermittlung der individuellen Präferenzen) eine unrichtige Information interpretiert. Man könnte daher auch formulieren, dass die Störungsursache darin liegt, dass die Daten nicht in diesen Zweckkontext passen, sondern missbräuchlich in diesen eingebracht wurden.[134]

2. Technische Ausgestaltung

Solche Einzelangriffe sind v.a. aus dem Bereich der sozialen Netzwerke[135] bekannt geworden: Grundlage ist i.d.R. das Ausspähen von Passwörtern ggf. i.V.m. der Ausnutzung fehlender 2-Faktor-Authentifizierung. Über den Kontozugriff werden dann nicht nur bereits vorhandene Daten ausgespäht oder manipuliert, sondern es werden vielmehr neue Daten in ein System eingebracht. Sofern die Personalisierung des Dienstangebots nicht (nur) über den Server, sondern auch über im Endgerät gespeicherte Daten (Cookies) erfolgt,[136] ist z.T. nicht mal ein Kontozugriff erforderlich.

Allgemein wurde das Phänomen des Einbringens von falschen Informationen bereits 2005 beobachtet, als der Wurm Samy neue, wenn auch belanglose Informationen in den Online-Dienst Myspace einbrachte: Durch den entsprechenden Schadcode fügte der Browser der Betroffenen unbemerkt den Autor des Wurms „Samy" auf Myspace als Freund hinzu und platzierte in den Profilen der Betroffenen die Zeichenfolge „but most of all, samy is my hero".[137] Ähnlich verhielt sich auch der Wurm Mikeyy, der

134 Dieser Gedanke findet sich auch in Art. 5 Abs. 1 lit b) DSGVO. Zwar spricht die DSGVO aufgrund der definitorischen Gleichstellung von Daten und Informationen in Art. 4 Nr. 1 DSGVO von „Datenrichtigkeit", konkretisiert aber in der Sache zutreffend, dass die Frage der Unrichtigkeit der Daten „im Hinblick auf die Zwecke ihrer Verarbeitung" zu beantworten ist.

135 *Sahoo/Gupta*, Enterprise Information Systems 2019, 832 (833).

136 *Xing et al.*, in: Proceedings of the 22nd USENIX Security Symposium, Take This Personally: Pollution Attacks on Personalized Services, 671 (680 f.).

137 *Grossman,* Cross-Site Scripting Worms & Viruses, Juni 2007, S. 8; *Alcorn*, Network Security 2006, 7 (8).

auf Twitter 2008 mindestens 190 Konten kompromittierte und von diesen mindestens 10.000 Tweets verschickte.[138]

Das hier gegenständliche Einbringen von Informationen zum Zwecke der zielgerichteten Manipulation eines Profils und damit letztlich auch des personalisierten Dienstes wird im Englischen auch als „*pollution attack*" bezeichnet.[139] Insofern wird nur der Dienst für den betroffenen Nutzer manipuliert, nicht hingegen der Empfehlungsdienst generell (dazu sogleich). In einer Studie konnte die Möglichkeit der Manipulation des Rankings der *Google Suche* als auch der von *Amazon* vorgeschlagenen Produkte bereits nachgewiesen werden. Hierfür wurden falsche Informationen im Sinne von tatsächlich nicht vom User vorgenommenen Suchanfragen bzw. Seitenaufrufen von Produkten in das Profil eingebracht und dieses somit verfälscht,[140] mithin falsches Personenwissen erzeugt. Dabei wurde der Umstand ausgenutzt, dass Web Authentifizierung in der Regel nur absichern kann, dass eine Anfrage von dem Browser oder der App eines Nutzers ausgelöst wurde, aber nicht sicherstellen kann, dass der Nutzer diese auch tatsächlich durchgeführt oder autorisiert hat.[141] Der vermeintlich vertrauenswürdige Endpunkt (Browser/App) wird mithin dahingehend korrumpiert, dass er bei seiner Benutzung unbemerkt manipulierte Anfragen an die jeweiligen Server richtet.

Auf beiden Webseiten (*Google Suche, Amazon*) konnten hierdurch die personalisierten Rankings der Suchergebisse bzw. die Produktempfehlungen manipuliert werden.[142] Grundsätzlich können solche *Pollution Attacks* gegen alle personalisierten Dienste und somit insbesondere auch gegen

138 *Luo et al.*, in: Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing (DASC), An Analysis of Security in Social Networks, 648 (648); Teen claims responsibility for disrupting Twitter, CNN vom 13.04.2009.

139 *Xing et al.*, in: Proceedings of the 22nd USENIX Security Symposium, Take This Personally: Pollution Attacks on Personalized Services, 671 (671); *H. Huang et al.*, in: Proceedings 2021 Network and Distributed System Security Symposium, Data Poisoning Attacks to Deep Learning Based Recommender Systems, S. 3 f.; *M. Fang et al.*, in: Proceedings of the 34th Annual Computer Security Applications Conference, Poisoning Attacks to Graph-Based Recommender Systems, 381 (384); Technisch wird dies insbesondere durch sog. Cross-Site-Request-Forgery (kurz CSRF oder XSRF), umgesetzt, hierzu ausführlich: *Zeller/Felten,* Cross-Site Request Forgeries: Exploitation and Prevention S. 1 ff.

140 *Xing et al.*, in: Proceedings of the 22nd USENIX Security Symposium, Take This Personally: Pollution Attacks on Personalized Services, 671 (677 ff.).

141 *Zeller/Felten,* Cross-Site Request Forgeries: Exploitation and Prevention, S. 4.

142 Wie zuvor.

soziale Netzwerke wie *Facebook, Instagram* oder *Twitter* ausgeführt werden.[143]

III. Plurale Informationsmanipulation

1. Wirkung nach dem Informationsmodell

Bei der pluralen Informationsmanipulation werden anders als bei der singulären Informationsmanipulation unrichtige Informationen mit zahlreichen realen, aber korrumpierten oder gefakten Identitäten eingebracht. Dies kann zwar -bei realen Identitäten- wie zuvor zu einer Verfälschung des Personenwissens führen; vor allem wird aber durch diese Breite des Angriffs das *abstrakte Lernwissen* über die Beliebtheit von Items oder die Zusammenhänge zwischen mehreren Items beeinträchtigt.

2. Technische Gestaltung

Im Rahmen von Angriffen auf CF-Systeme spricht man hier auch von sog. *Shilling-Attacks:* Dabei wird eine große Anzahl von gefakten Identitäten erstellt oder es werden reale Identitäten manipuliert und mit diesen entweder positive (en: push-attack) oder negative Bewertungen (en: nuke-attack) abgegeben, um das System zur Abgabe falscher Empfehlungen zu zwingen.[144] So sind insbesondere sog. *Fake-Co-Visitation-Angriffe* möglich, d.h. es werden mit vielen Identitäten zwei Elemente (ein Zielelement und ein Ankerelement) aufgerufen, um zwischen diesen eine Präferenzverbindung *(Lernwissen)* herzustellen, so dass bei Auswahl des Ankerelements als nächstes das Zielelement empfohlen bzw. sogleich das bisherige Zielelement verdrängt wird.[145] Neben solchen gezielten Angriffen auf ein bestimmtes

143 *Xing et al.*, in: Proceedings of the 22nd USENIX Security Symposium, Take This Personally: Pollution Attacks on Personalized Services, 671 (684).

144 *Kaur/Goel*, in: 2016 International Conference on Inventive Computation Technologies (ICICT), Shilling attack models in recommender system, 1 (1 f.); *Sundar et al.*, IEEE Access, Vol. 8 (2020), 171703 (171704).

145 *G. Yang/Gong/Cai*, in: Proceedings 2017 Network and Distributed System Security Symposium, Fake Co-visitation Injection Attacks to Recommender Systems, S. 2, 10 ff.; *Y. Zhang et al.*, IEEE Trans. Inf. Forensics Secur., Vol. 15 (2020), 3807 (3809, 3813 f.); beispielsweise könnte im obigen Beispiel (S. 72) durch den Angriff bei Aus-

Empfehlungsverhalten sind außerdem ungezielte Angriffe möglich, die auf eine generelle Verschlechterung der Empfehlungsleistung abzielen.[146] In der Praxis ist beispielsweise der Online-Marktplatz *Amazon* nach einer wissenschaftlichen Untersuchung aus 2017 massiv von diesem Phänomen der falschen Bewertungen betroffen.[147] Im Recht benennt inzwischen auch der DSA dieses Phänomen in EG 57 mit Blick auf die „Einrichtung von Scheinkonten, die Nutzung von Bots und anderen automatisierten oder teilautomatisierten Verhaltensweisen".

Wird auch ML eingesetzt und soll ein dann zumeist online (auch: „inkrementell") lernendes ML-System manipuliert werden, spricht man auch von *Model Skewing* oder einer *Poisoning Attack* bzw. *Data Poisoning*.[148] Der entscheidende Angriffspfad liegt hier in dem Online-Training, d.h. in dem Umstand, dass regelmäßig neue Daten genutzt werden, um das Modell des ML-Systems zu aktualisieren.[149] Durch dieses fortlaufende Training mit neuen Daten kann es grundsätzlich ausreichen, nur geringe Datenmengen zu manipulieren um das Modell zu vergiften, d.h. es subtil „umzutrainieren", so dass es falsche Entscheidungen trifft.[150]

IV. Fazit und Ansatz für das Erfordernis der Resilienz

Wie voranstehend dargestellt können bei der Betrachtung von Angriffen zwei unterschiedliche Vektoren unterschieden werden. Zwar haben beide

---

wahl der Gartenschere (Ankerelement) statt der Arbeitshandschuhe eine Schreibtischlampe (Zielelement) empfohlen werden.

146 *Himeur et al.*, Computers & Security, Vol. 118 (2022), AS-Nr. 102746, S. 12; *Deldjoo/Di Noia/Merra*, ACM CSUR 2022, AS-Nr. 35, Heft 2, AS-Nr. 35, S. 8 ff.

147 *P. Liu et al.*, in: IEEE International Conference on Software Quality, Reliability and Security, Identifying Indicators of Fake Reviews Based on Spammer's Behavior Features, 396 (401 f.).

148 *Xue et al.*, IEEE Access, Vol. 8 (2020), 74720 (74723); diese Begriffe des Poisonings, also der Vergiftung, werden teilweise auch bei regelbasierten Empfehlungssystemen verwendet: *Chen et al.*, Trans Emerging Tel Tech 2021, AS-Nr. e3872, Heft 6, AS-Nr.: e3872.

149 *Heinemeyer/Herpig*, in: Ebers/Steinrötter, Künstliche Intelligenz und smarte Robotik im IT-Sicherheitsrecht, 65 (66); *Jagielski et al.*, in: 2018 IEEE Symposium on Security and Privacy (SP), Manipulating Machine Learning: Poisoning Attacks and Countermeasures for Regression Learning, 19 (19).

150 So reicht es beispielsweise in der personalisierten Medizin bereits aus bei den Sensoren der Medizingeräte weniger Patient:innen die Daten zu manipulieren, um die ML-Modelle, die für eine große Gruppe von Patient:innen genutzt werden, zu „vergiften": *Jagielski et al.,* wie zuvor.

Vektoren stark abstrahiert betrachtet den gleichen Auslöser (Manipulation des Datenflusses) sowie dieselbe informationstechnische Auswirkung (falsche Entscheidungen).

Allerdings werden bei einer *singulären Informationsmanipulation* lediglich bezüglich einzelner Identitäten falsche Informationen in das System eingebracht. Im Ergebnis ist damit insbesondere das Datenschutzgrundrecht betroffen, weil das persönliche Profil der jeweiligen Personen etwa durch die dargestellte *Pollution Attack* manipuliert wird. Es werden insoweit durch eine Sicherheitslücke unrichtige personenbezogene Daten in das System eingebracht und durch deren Verarbeitung wird im Ergebnis das *Personenwissen* in Gestalt des Profils beeinträchtigt.

Bei der *pluralen Informationsmanipulation* wird hingegen das *abstrakte Lernwissen* verändert, indem durch viele reale oder gefakte Accounts in großem Umfang falsche Informationen in das lernende System eingebracht werden (*Poisoning Attack*) und so die Entscheidungen des Systems generell beeinflusst werden. Ist dies erfolgreich, werden somit etwa Rahmen eines (ML-gestützten) CF-Empfehlungssystem auch für viele Personen falsche Entscheidungen getroffen, obwohl das individuelle Personenwissen auch weiterhin richtig sein kann.[151]

Die Ausgangslage für beide Angriffsvektoren besteht darin, dass von (vermeintlichen) Nutzer:innen stammende Informationen verwendet werden, die manipuliert sein können. Der Verantwortliche bzw. der Anbieter des jeweiligen Dienstes kann die Qualität der Informationen dabei nicht mit letzter Sicherheit garantieren, da diese Interaktionen beschreiben, die außerhalb seines Kontrollbereichs initiiert werden. So ist es ihm etwa nicht möglich mit letzter Sicherheit zu überprüfen, ob auf dem Endgerät, von welchem etwa Bewertungen oder Anfragen abgegeben werden, dieselben tatsächlich auch von der/dem jeweiligen Nutzer:in stammen. So könnte das Endgerät etwa kompromittiert sein; auf die Daten- und IT-Sicherheit desselben hat er regelmäßig keinen Einfluss.

Damit verbleibt eine erhebliche Ungewissheit über die Qualität der gesammelten Informationen und es kann insoweit als Hypothese vorweg gestellt werden, dass die *Resilienz* als besondere funktionale Anforderung solchen Ungewissheiten und den damit verbundenen Folgen entgegentreten muss.

151 Dies könnte grundsätzlich aber mit Art. 22 DSGVO angefochten werden, dazu sogleich.

Wichtig ist es, in diesem Zusammenhang darauf hinzuweisen, dass die weitere Untersuchung an dieser Stelle auf *die Resilienz als Erfordernis der Datensicherheit* ausgerichtet ist und insofern sich ausschließlich auf den Fall bezieht, dass personenbezogene Daten im Entscheidungsprozess manipuliert wurden. Nicht nachgegangen wird der Frage, ob und inwieweit eine automatisierte Entscheidung jenseits von Manipulationen der eigenen personenbezogenen Daten (etwa durch eine plurale Informationsmanipulation) nach Art. 22 DSGVO richtig sein muss. Nach Art. 22 Abs. 3 DSGVO kann eine automatisierte Entscheidung, die die betroffene Person als unbillig empfindet grundsätzlich im Rahmen des Art. 22 Abs. 3 DSGVO angefochten werden.[152]

152 Beachte allerdings zum Nicht-Vorliegen des Tatbestands des Art. 22 DSGVO bei personalisierten Inhaltsempfehlungen als auch bei der personalisierten Preisgestaltung bereits S. 69, Fn. 114.