Peter Hense

Hi Alexa, can I trust you? Technologie, Wirtschaft und Rechtsentwicklungen bei Virtual Private Assistants

Alexa, Google Assistant, Cortana, Siri oder Bixby: Kaum ein relevanter Technologieanbieter kommt noch ohne eigenen Sprachassistenten und die notwendige Hardware aus. Massive Werbung und eine hohe Technologieakzeptanz in der Bevölkerung haben dazu geführt, dass insbesondere Endgeräte für die Amazon-Alexa-Plattform sich binnen weniger Jahre mehr als 100 Millionen Mal verkauften. Privathaushalte genießen die erleichterte Bedienung, Unternehmen ringen um plattformbezogene Geschäftsmodelle. Seit der Markteinführung 2015 von Technologieexperten kritisch begleitet, beginnen sich mittlerweile auch Wissenschaftler, Medien, Aufsichtsbehörden und Gerichte mit den sicherheits- und datenschutzrechtlichen Rahmenbedingungen von Virtual Private Assistants zu befassen.

Alexa, Google & Smart Home

Ein System, sie zu binden

Da, wo früher in Singlewohnungen die Lavalampe stand, thront heute eine Alexa. Die „smarte" Lautsprecherbox ist seit 2016 in Deutschland verfügbar und enthält sieben Mikrofone, mit denen sie ihre Umgebung abhört. Die Zahl sieben nimmt seit babylonischen Zeiten eine Sonderstellung in unserem Leben ein. Sie steht traditionell für göttliche Vollkommenheit und spannt aufgrund dieser überirdischen Assoziation einen Bogen zur Selbstwahrnehmung heutiger Repräsentanten bekannter Technologiekonzerne. Alexa, Ring und Nest thört zu, was Kunden sagen, und beantwortet alle Fragen, auf die das Internet eine Antwort weiß. Amazon Echo lernt, speichert und analysiert, was im Haushalt vor sich geht. Alexa ist stets bereit und willens, unser Haus zu steuern, Anrufe zu tätigen, Nachrichten vorzulesen und natürlich alle Konsumwünsche zu erfüllen, solange und soweit der Verfügungsrahmen der Kreditkarte reicht. Als Dank für die Nutzung erhält Amazon umfangreiche Informationen über alle Gespräche in der Umgebung des Geräts, ganz wie ein Mensch. Ein Mensch, der nie schläft und immer lauscht. Damit gelegentliche Zweifel an der Seriosität der technischen Errungenschaft bereits im Keim erstickt werden, spricht Alexa künftig auch mit der vertrauenserweckenden Stimme von Prominenten wie Samuel L. Jackson. Und wer würde Prominenten nicht vertrauen, wenn es um Privatsphäre und Datensicherheit geht?

Vor der Tür: Neue Patente von Google und Amazon

Wir stehen noch immer erst am Anfang einer großen Entwicklung, denn auf dem Weg zum „Smart Home" sind smarte Lautsprecher nur ein erstes Durchgangsstadium. *„The next data mine is your bedroom"* formulierte es Sidney Fussell treffend (*„The Atlantic"*, November 2018). Sowohl Amazon als auch Google haben bereits Lösungen marktreif verfügbar, die umfassende Sensortechnik in jedes Haus bringen und neben dem gesprochenen Wort auch Bildmaterial und sonstige sensible Sensordaten (Ultraschall, Infrarot, Bluetooth) aus dem Privatleben von Nutzern verarbeitet. Die Technologien zur Aufbereitung dieser Rohdaten sind in der Lage, Ableitungen („inferences") aus diesen Daten herzustellen und Nutzer anhand gewichteter Wahrscheinlichkeiten zu analysieren, je nachdem, was die installierten Smart-Home-Sensoren in ihrem Umfeld wahrnehmen. Objekterkennung und Raumvermessung sind bereits von autonomen Staubsaugerrobotern bekannt, nicht zuletzt, seit der (geplante) Verkauf der Grundrisse der gestaubsaugten Wohnungen an Werbetreibende durch Roomba im Jahr 2017 für einige Aufregung sorgte.

Das Unternehmen hat gelernt und ist heute einen großen Schritt weiter: Man kann seinen neuesten Roomba auch über Alexa und Google steuern. Das immer engmaschigere Smart Home, gepowert durch unzählige Endpunkte des Internet of Things ("IoT") wird immer unausweichlicher, je mehr Amazon und Google Industriestandards (er)setzen. Praktisch, dass die iRobot-Cloud von Roomba auf Amazon Web Services (AWS) aufsetzt, dann ist der Weg für die Daten später nicht allzu weit. Von dieser Vernetzung hat jeder etwas: Geräte wie "Echo Look" und "Echo Frames" nutzen bereits heute nach eigenen Angaben Objekterkennung von Bekleidung, um (schlechten) Modegeschmack festzustellen und (bessere) Mode zum Kauf zu empfehlen.

Daneben können die Sensoren auch aus den optisch oder über Bluetooth erkannten elektronischen Geräten im Umfeld aber auch mühelos Annahmen zum verfügbaren Einkommen erstellen, sowie anhand von Audiosignaturen Personen im Geräteumfeld, nicht notwendig den "Nutzer", erkennen und ohne Mühe einordnen, z. B. anhand der Klangfarbe hinsichtlich Geschlecht und Alter.

Technologieanbieter können und wollen auf diese Weise neben dem Mode-, Musik- und Technikgeschmack auch Gesundheitsdaten erfassen, z. B. indem sie anhand der Sensorik erkennen, dass ein Nutzer mit ungewöhnlich rau-

er Stimme erkältet ist, wegen Wortwahl oder Tonlage depressiv ist oder aufgrund dem menschlichen Sinn kaum wahrnehmbarer Veränderungen in der Wortbildung und Aussprache gegebenenfalls an Alzheimer leidet und ihm entsprechende Lebensoptimierungs-, Gesundheits- und Konsumvorschläge unterbreitet. Details zu dieser Sensorik für den Haushalt findet der interessierte Datenschutzexperte in einigen in den Jahren 2017 und 2018 publizierten Patenten und denen auch jenseits der Fachkreise mediale Aufmerksamkeit zuteil wurde.

Zur Erkennung von Stimmungen und Gesundheitszustand lohnt ein Blick auf Amazons Patent No.: US 10,276,188 B2, *„Systems and Methods for identifying human emotions and/or mental health states based on analyses of audio inputs and/or behavioral data collected from computing devices."* sowie auf Googles Patent No.: US 10,096,31 B1, *„Voice-based determination of physical and emotional characteristics of users"*, wen die Personalisierung von Produktempfehlungen und Medieninhalten interessiert, der wird mit der Beschreibung zu Googles Patent Application Publication: US 2016/0260135 A1 *„Privacy-aware personalized content for the smart home"* relevante Lektüre erhalten.

Google denkt Smart Home noch einen Schritt weiter: Der gleichfalls patentierte *„Household Policy Manager"* verarbeitet in mit Sensorik angereicherten Wohnobjekten sämtliche Sensordaten, steuert den gesamten Haushalt, inklusive der Beschaffung aller relevanten Waren und Dienstleistungen und optimiert, ganz nebenbei, auch noch das Berufs- und Privatleben der Verbraucher (Patent No.: US 20160259308A1). Während in Deutschland über „Smart Metering" und nach Messstellenbetriebsgesetz regulierte Datenverarbeitung hitzig debattiert wird, entsteht im Schatten der großer Versorgungsunternehmen der Energiewirtschaft die Informationstechnik einer neuen Ära, deren Herrscher sich keine Sorgen drohenden Normenvollzug und Sanktionen machen müssen. Datenschutzregulierung in Deutschland ist auch heute noch im Wesentlichen ein Papiertiger.

Erwerbsentscheidungen durch Algorithmen: Das Ende klassischer Werbung

Tobias Haberkorn beschreibt in einem Interview mit dem Autor des Buches „Platform Capitalism", Nick Srnicek, das überraschende Ziel dieses Trends zur massenhaften Verwendung von Sensorik durch Technologieunternehmen treffend: *„What tech firms are now pushing for are personal assistants at every instant in the chain of consummation, a type of service where the wish is fulfilled at the very moment it is formed, so that there is no need for advertising anymore."*. Das Ziel von „Smart Home" sind demnach weder ein „besseres Leben" für den Nutzer, noch allein die Monetarisierung von dessen privater Konversationen und Bewegungen, sondern das Ausschalten der mit lästigen Streueffekten verbundenen klassischen Werbung. Wenn Sensoren statt Menschen fühlen und eine Maschine statt des eigenen Verstandes kalkuliert und entscheidet, dann ist im Beschaffungsprozess mangels Emotionen und sinnlicher Wahrnehmung für klassische Werbung kein Raum mehr.

Technologie & Recht

Alexa, anatomisch

Wer als Datenschutzexperte mit Technologieinteresse näher erfahren will, wie das System Alexa funktioniert, kann sich auf das Projekt *„Anatomy of an AI system"* unter „https://anatomyof.ai" stützen, welches bis auf die Prozessorebene herunter Aufbau und Funktionsweise des technischen und menschlichen Ökosystems rund um Amazon Alexa dokumentiert und erläutert. Die nicht unbekannten Forscher Kate Crawford (Microsoft Research, AI Now Institute) und Prof. Vladan Joler (University of Novi Sad) haben mit ihrem Projekt *„The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources"* eine bemerkenswerte Wissens- und Kontemplationsquelle bereitgestellt, deren Wert Datenschutzbeauftragte nicht zuletzt bei der Erstellung von Datenschutz-folgeabschätzungen (DSFA) schätzen werden.

Joint Controllership: Alexa Skills Kit

Amazon Echo stellte für Entwickler bereits seit Juni 2015 diverse Programmierschnittstellen (APIs) bereit, die zum Anschluss und über das „Alexa Skills Kit" zur Entwicklung eigener sprachgesteuerter Applikationen namens „Skills" („Fähigkeiten") freigegeben ist und durch die Entwickler gezielt Nutzer der Plattform werblich ansprechen können. Diese Skills werden von Amazon geprüft und im Alexa Skills Store bereitgestellt, derzeit sind mehr als 60.000 Skills verfügbar.

Da die Anbieter von Skills Daten von Nutzern verarbeiten (müssen), darunter transkribiertes Audio, userID, Standortdaten sowie z. B. nutzerspezifische Einkaufslisten, sind sie auch gezwungen, die Vorgaben der DSGVO einzuhalten und transparente Informationen zur Datenverarbeitung (IDV) nach Art. 12, 13 und 14 DSGVO bereitzustellen. In der Praxis bestehen bei der Erfüllung der Informationspflichten erhebliche Defizite. Bereits 2017 stellte eine Studie von Alhadlaq, Tang, Almaymoni, Korolova: *„Privacy in the Amazon Alexa Skills Ecosystem"* fest, dass für 75 % aller Skills noch nicht einmal irgendwelche Datenschutzinformationen vorlagen, geschweige denn solche, die konform mit europäischem Recht wären. Daran hat sich bis heute nichts geändert, man kann insoweit den Begriff des „normvollzugsfreien Raumes" verwenden.

Zur datenschutzrechtlichen Einordnung der Beteiligten: Dass Amazon selbst nicht etwa ein schlichter Auftragsverarbeiter i. S. v. Art. 28 DSGVO sondern ein eigener Verantwortlicher mit dem vollen datenschutzrechtlichen Pflich-

tenkatalog ist, liegt angesichts der Nutzung der Daten durch Amazon für eigen Zwecke, darunter Training der Hauseigenen Machine-Learning-Systeme („KI") sowie Amazons Hoheit über sämtliche Mittel der Datenverarbeitung auf der Hand. Die gemeinsame Nutzung der Plattforminfrastruktur und der intensive Datenaustausch zwischen den Beteiligten jeweils zu eigenen Zwecken macht Entwickler jedoch neben Amazon zu gemeinsamen Verantwortlichen i. S. v. Art. 26 DSGVO, was neben der gemeinsamen und gesamtschuldnerischen Haftung nach Art. 26 Abs. 3 DSGVO auch die Pflicht zum Abschluss eines Joint-Controller-Agreement und die Publikation dessen wesentlicher Inhalte nach Art. 26 Abs. 2 S. 2 DSGVO mit sich bringt.

Zu beachten ist, dass sich die Verantwortlichkeit beider Parteien, Entwickler wie Amazon, auch angesichts der Fashion-ID-Entscheidung des EuGH (C-40/17) auf alle Verarbeitungsvorgänge bezieht, die im Zusammenhang mit der Plattformnutzung durch den Entwickler stehen, da dieser mit Amazon gemeinsam die Verarbeitung der Sprachdaten ermöglicht, erleichtert und intensiviert. Ob und inwieweit Amazon notwendige Mitwirkungshandlungen zur Aufbereitung von IDV seiner Entwickler oder gar den Abschluss von Art. 26 – Verträgen unternimmt, darüber liegen bisher keine öffentlich verfügbaren Informationen vor. Vergleichbare juristische Herausforderungen stellen sich für den, der Alexa Voice Services über die „Connected Speaker APIs" in der eigenen Hardware implementiert, um den Anschluss an den erhofften Boom-Markt „SmartHome" nicht zu verpassen. Schade nur, dass dieser Branchenstandard per Default bereits die Datenschutzrechtswidrigkeit in sich trägt.

„Echo Dot Kids Edition": Stein des Anstoßes

Amazon wendet sich mit Erfolg beim arglosen Verbraucher nicht nur an Erwachsene, sondern auch an diejenigen Mitglieder der Rechtsgemeinschaft, die am Verwundbarsten sind: Kinder. Die in bei Freunden von Disneyfilmen beliebten bunten Farben „rainbow" und „frost blue" vertriebene *„Amazon Dot 3 Kids Edition"* wurde 2019 noch einmal überarbeitet und spielt die von Amazon lizenzierte Musik im kostenpflichtigen Abonnement nunmehr in erster Linie lauter, beantwortet Fragen in „kindgerechterer Sprache" und bietet „kindgerechte Spiele" neben ebenso „kindgerechten" Produktempfehlungen. Ein derartiges Eindringen in Kinder- und Jugendzimmer stellt nicht zuletzt das Recht vor enorme Herausforderungen.

Die bereits 2015 geäußerte Kritik der US-amerikanischen *Campaign for a Commercial-Free Childhood* (CCFC) an smarten Spielzeugen wie „Hello Barbie" ist vor dem Hintergrund der Marktmacht, mit der Amazon seine Echo-Plattform für Kinder nutzbar macht gut nachvollziehbar: *„Kids using ‚Hello Barbie' aren't only talking to a doll, they are talking directly to a toy conglomerate whose only interest in them is financial."*. Es überrascht wenig, dass bei dem von Amazon gewählten Markteintrittsgeschwindigkeit („Blitz Scaling") das „Echo Dot Kids Edition" gegen geltendes US-Bundesrecht verstößt. Zumindest verletzt es wohl den *Children Online Privacy Protection Act* (COPPA), weil es Daten a) ohne adäquate Information sammelt, b) diese Sammlung zudem ohne Einwilligung der Eltern „parental consent" erfolgt und c) für die erfassten und verarbeiteten Daten keine Möglichkeit der Löschung vorgesehen ist (öffentlicher *„Request for Investigation of Amazon Inc."* bei der FTC vom 9. Mai 2019). Es bedarf keiner besonderen Fantasie, um diese drei ausgewählten Verstöße gegen US-Recht auch als Verstöße gegen europäisches Datenschutzrecht zu subsumieren, insbesondere als Verletzung von zwingenden Informationspflichten (Art. 12 ff. DSGVO), rechtsgrundlose Verarbeitung von Daten von Kindern (Art. 5, 6 DSGVO) sowie als Verletzung der Grundsätze von Fairness, Datenminimierung sowie Datensicherheitsstandards durch die Unmöglichkeit der Löschung von Aufzeichnungen (Art. 5, 32 DSGVO).

Diese Verhaltensweisen können neben Unterlassungsansprüchen auch harte Bußgelder nach sich ziehen, so denn die zuständigen Behörden den Sachverhalt erkennen, verstehen und adäquat reagieren. Das in Europa noch junge Recht zur Verbandsklage sowie die Musterfeststellungsklage könnten angesichts der verbreiteten Überlastung und auch dadurch bedingten Ineffizienz von Datenschutzbehörden vor diesem Hintergrund geeignete Instrumente sein, um Rechtsverletzungen zu sanktionieren, abzustellen und dadurch auch ein „level playing field" für rechtskonforme Anbieter zu schaffen. In den USA nehmen Verbraucher das Recht mittlerweile selbst in die Hand: Die Mutter einer zehnjährigen Tochter aus Massachusetts reichte wegen der bekannten Verstöße von Amazon Anfang Juni 2019 eine Privacy Class Action in Seattle ein (Case No.: 2:19-cv-910), wobei die Lektüre der Klage mit 15 frei verfügbaren Seiten für Nutzer und Fachpublikum erhellend ist. Die Klägerin kritisiert, soviel sei verraten, neben der Löschunwilligkeit von Amazon auch einen weiteren Punkt in der Verwertungskette des angegriffenen Konzerns, nämlich die Erstellung von "biometrischen Sprachsignaturen von Millionen von Kindern".

„Natural Language Processing (NLP)"

Das Training von Spracherkennungssystemen ist mühsam und teuer, „Natural Language Processing" („NLP") ist jedoch ein lukrativer Wachstumsmarkt. Statt der kostenintensiven Bereitstellung von Trainingsdaten durch bezahlte Testpersonen nutzen Unternehmen wie Amazon und Google die eigenen Kunden als Datenlieferanten oder gehen, wie im Fall des smarten Spielzeugpuppe *„My friend Cayla"* eine Partnerschaft mit einem (Spielzeug-)Hersteller ein, der entsprechend nah am Kunden ist. Der Markt für Spracherkennung hat Potenzial. Nuance Technologies aus Burlington, Massachusetts z. B. ist bekannt durch die Spra-

cherkennungssoftware „Dragon“, die in vielen Büros weltweit im Einsatz ist.

Das Unternehmen verfügt nach eigenen Angaben über eine Datenbank von mehr als 60 Millionen gespeicherten Voiceprints zu Menschen aus aller Welt: Erwachsene, Kinder, Ärzte, Rechtsanwälte. Das Unternehmen analysiert Stimmen und speichert die entsprechenden Charakteristika als digitalen „Fingerabdruck“. Nuance bietet auf www.nuance.com unter dem Schlagwort „Voice Biometrics“ an, dass Wirtschaftsunternehmen, aber auch die Öffentliche Hand, über Schnittstellen eigene Sprachaufzeichnungen mit den biometrischen Sprachprofilen („Samples“) bei Nuance abgleichen und auf diese Weise die zugehörigen Sprecher identifizieren können. Die Werbeaussage *„Every voice matters: Our system knows who is talking and why.“* lässt erahnen, dass man nicht bei der bloßen Identifizierung der Sprecher stehen bleibt, sondern auch dynamische Stimmanalysen in Echtzeit betreibt und auf diese Weise die jeweilige Stimmung des Sprechenden erkennt. Jenseits rechtlicher Bedenken, die sich aus der Verarbeitung biometrischer Daten zu Identifikationszwecken bei direkter und insbesondere indirekter Erhebung ergeben, blüht das Geschäft mit dem staatlichen Sicherheitssektor. Als *„new public security weapon“* stellt Nuance seine Identifikations- und Stimmungsanalysesysteme den Sicherheitsbehörden dieser Welt zur Verfügung.

Bei Kauf eines jeden smarten Assistenten erhalten nicht nur Nuance, Amazon, Google oder andere Technologieanbieter neue unverwechselbare Sprachprofile, sondern auch deren Kunden bei Unternehmen und Regierungen rund um die Welt. Anonymität wird zunehmend ein Luxusgut und man erkauft es mit Verzicht.

Abwehransprüche und Sanktionen

Neben datenschutzrechtlich begründeten staatlichen Sanktionen und Individualansprüchen sollte nicht vergessen werden, dass das deutsche Recht bei unberechtigter Datenverarbeitung, die eine Verletzung des Allgemeinen Persönlichkeitsrechts mit sich bringt, ganz eigene Anspruchsgrundlagen bereithält, die mit den Ansprüchen aus der DSGVO nicht deckungsgleich sind, sondern konkurrieren (vgl. dazu den Beitrag von *Hense*, Datenschutz-Berater 2019, S. 204 ff. zu „Google Analytics“). Ein Vorgehen nach § 823 Abs. 1 i. V. m. § 1004 BGB analog steht den von unberechtigter Datenverarbeitung Betroffenen jederzeit offen und kann, aufgrund der besseren Geläufigkeit der Anspruchsgrundlagen bei deutschen Gerichten, prozesstaktisch das Mittel der Wahl sein, um eine unberechtigte Verarbeitung zu unterbinden oder gar, was im Fall der vorsätzlichen und rechtswidrigen Verarbeitung von Daten von Kindern naheliegt, adäquaten Schadensersatz für die Persönlichkeitsrechtsverletzungen als immaterielle Schäden einzufordern.

Regulierung in den USA

Die Faszination für Virtual Private Assistants scheint in den USA bereits etwas abgeklungen zu sein. Eine Reihe neuerer Publikationen beschäftigt sich intensiv mit der jungen Geschichte und der noch jüngeren Rechtsgeschichte der Systeme, insbesondere mit dem omnipräsenten Amazon Alexa. Hervorzuheben, weil instruktiv, ist ein längerer Aufsatz mit dem Titel *„Alexa, what should we do about privacy?“* von Anne Pfeifle (Washington Law Review, 2019). Für US-amerikanische Juristen endet der Spaß smarter Assistenten spätestens dann, wenn staatliche Institutionen, insbesondere Strafverfolger, sich der gesammelten Sprachaufzeichnungen bemächtigen und diese im Strafprozess verwenden wollen. Bedenken hinsichtlich der Verletzung des Rechts auf „Privacy“ äußern sich in den USA, anders als in Europa, häufiger in schneller sektor- und technologiespezifischer Ad-hoc-Gesetzgebung. Kalifornien, das gerade den Gesetzgebungsprozess für das umfassendste Datenschutzgesetz eines US-Bundesstaates (*„California Consumer Privacy Protection Act“*, CCPA, siehe hierzu *Hense/Fischer*, Datenschutz-Berater 2019, S. 27 ff.) erfolg- und folgenreich abgeschlossen hat, lässt die Technologiekonzerne nicht zur Ruhe kommen. In der Assembly Bill (AB) No. 1395 ist geplant, stand-alone „Smart speaker devices“ mit „integrated virtual assistants“ wesentlich schärfer zu regulieren, als dies bisher der Fall war. Ausgenommen sind lediglich integrierte Systeme z. B. in Telefonen, Tablets und Connected Cars. Das Gesetz verbietet unter anderem ausdrücklich das unfreiwillige „data sharing“ mit Dritten, nicht nur der Audiodateien selbst, sondern auch der Transkripte und fordert für die dauerhafte Speicherung beider Datenarten ein gesondertes „opt-in“, was Unternehmen in den USA angesichts der bisher vorherrschenden Datennutzungspraxis vor besondere Herausforderungen stellen dürfte.

Praxiseinsatz: Datenschutzfolgeabschätzung

User Control: Privacy by Default?

Hilfreich für das individuelle Assessment von Risiken beim Einsatz von Sprachassistenzsystemen ist eine Veröffentlichung des „Future of Privacy Forum (FPF)“ von 2016 namens *„Always On: Privacy Implications of Microphone-Enabled Devices“*. Das Arbeitspapier enthält insbesondere Hinweise zur Unterscheidung der verschiedenen Aktivierungsmodi („always on“, „speech activated“, „manually activated“), die in Deutschland nicht zuletzt vor dem Hintergrund telekommunikationsrechtlicher Regulierung von verdeckten Abhöreinrichtungen in § 90 TKG relevant sein können. Aufschlussreich auch für europäische Rechtsexperten sind zudem die Ausführungen zu den Einstellungsmöglichkeiten durch Nutzer (*„user control“*), der Mikrofontechnologie sowie zur Datenspeicherung.

Systemarchitektur

Neben dem eingangs erwähnten Projekt „Anatomy of an AI system“ ist bei einem vollverantwortlichen, also nicht aus-

schließlich im Rahmen der hier sehr schmalen Haushaltsausnahme oder zu wissenschaftlichen Zwecken erfolgenden Einsatzes von Alexa ein umfassendes Lagebild nebst Risikoeinschätzung zu erstellen. Zumindest für das erste Modell von Amazon Alexa existiert durch den Aufsatz von *Clinton, Cook und Banik* (2016): *„A Survey of Various Methods for Analyzing the Amazon Echo"* für den technikinteressierten Juristen eine intensive Analyse der eingesetzten Hardware und Software, in deren Rahmen über Reverse Engineering nach bekannten Möglichkeiten zur Entwicklung von Exploits gesucht wurde. Wem Debugging, Linux Kernel und die Vivisektion von Bootsequenzen zu langatmig sind, dem sei die Zusammenfassung der Autoren empfohlen: Alexa ist hardwareseitig angreifbar über den SD-Karten-Pinout, Hardware-Rooting sowie JTAG, was Alexa mit anderen Geräten aus dem Hause Amazon durchaus gemein hat.

Angriffe und Angriffsvektoren

Hilfreiche Informationen können neben der offiziellen, aber spärlichen Dokumentation zu Systemarchitektur und Datenflüssen seitens Amazons dem detaillierten Beitrag von Leong, *„Analyzing the Privacy Attack Landscape"* (2018), entnommen werden. Darunter finden sich wertvolle Hinweise, z. B. a) dass durch die Definition von spezifischen, leicht misszuverstehenden Steuerbefehlen in bösartigen Skills eine Übermittlung des Gesprächsinhalts an Dritte getriggert werden kann, b) auf das Risiko von Exploits für Alexa Firmware sowie c) auf die Risiken beim Einsatz von Hardware Dritter, bei der neben dem Abhören der gesamten Konversation im Alexa-Umfeld auch jede Manipulation der Steuerbefehle möglich ist. Weiterführende Lektüre stammt von Haack, Severance, Wallace und Wohlwend (2017): *„Security Analysis of the Amazon Echo"* und widmet sich der bemerkenswerten Erstellung einer Security Policy für das jeweilige Endgerät, was unter dem Gesichtspunkt „geeignete Abhilfemaßnahmen" für jede DSFA von Interesse ist.

Betroffenenrechte

Schwierigkeiten im Umgang mit Betroffenenrechten sind bei großen Technologieunternehmen weit verbreitet. Die Übersendung von 1.700 WAV-Audiofiles per ZIP-Datei nebst PDF der Transkriptionen durch Amazon an einen Dritten statt den nach Art. 15 DSGVO anfragenden Betroffenen im Jahr 2018 stellt nicht nur eine Verletzung der Betroffenenrechte dar, sondern legt strukturelle Defizite des Unternehmens im Bereich Art. 32 DSGVO offen. Vorfälle dieser Art sind nach Art. 33, 34 DSGVO mitteilungs- und meldepflichtig.

Wer den kommerziellen Einsatz von Alexa und Co., z. B. in Hotels oder im Retail plant, übernimmt damit auch Verantwortung für die Erfüllung von Betroffenenrechten durch Amazon. Sollten sich Verstöße gegen Betroffenenrechte nach Art. 12 ff. sowie Verletzungen der Vertraulichkeit von Daten, wie der eben geschilderte, nicht mehr als Ausnahme, sondern als Regel erweisen, dann wäre die Zusammenarbeit mit notorisch unzuverlässigen Diensteanbietern nicht nur ein Grund für ein intensiveres DPIA, sondern nach Art. 28, 26 sowie 32 in einem Kontrahierungsverbot resultieren.

Ein Recht auf Vergessenwerden: Cui bono?

Die dauerhafte Speicherung von Rohdaten („Audiodateien") ist ein Punkt, der bei einer Datenschutzfolgeabschätzung besondere Beachtung geschenkt werden sollte. Einerseits scheint es ja in Zeiten der Stärkung von Betroffenenrechten in der DSGVO eine Selbstverständlichkeit zu sein, dass Unternehmen Daten auch löschen können müssen. Der Löschvorgang ist jedoch in Machine-Learning-Systemen mitnichten trivial umzusetzen. Zwei Gründe stechen hierbei ins Auge. Erstens setzen die eingesetzten Machine-Learning-Modelle zur Spracherkennung quantitativ und qualitativ auf einer großen Menge unstrukturierter Rohdaten („Big Data") auf. Ein Entfall dieser „Rohdaten" duch Obfuszierung oder Löschung beeinträchtigt die Validität der erstellten „Predictive Models" und gefährdet dadurch deren Zweckerreichung, nämlich eine datenbasierte und gewichtete Vorhersage. Diesen Konflikt mag das Gesetz in Art. 17 DSGVO im Falle biometrischer und sensibler Daten im erforderlichen Abwägungsprozess zu Gunsten der Geltung starker Betroffenenrechte gelöst haben, wenngleich womöglich nicht interessengerecht.

Das Training von Machine Learning-Modellen führt jedoch in unvermeidbarer Weise zu einer disparaten Speicherung von personenbezogenen Daten (Bilder, Texte etc.) im modellierten System selbst. Diese Datenpartikel können vom Nutzer der Modelle insbesondere bei „Machine Learning as a Service" (MLaaS) ausgelesen werden, was unter dem Phänomen *„training data leakage"* bekannt ist (hierzu Ateniese, Mancini, Spognardi et. al. [2015]: *„Hacking smart machines with smarter ones: How to extract meaningful data frommachine learning classifiers."*). Eine gezielte „Suche" nach diesen Daten oder Datenfragmenten ist jedoch weder technisch noch wirtschaftlich sinnvoll zu erreichen, was zu erheblichen Compliance-Rückständen im Bereich von Forschung und Entwicklung im Bereich Machine Learning bzw. „Künstlicher Intelligenz" führen kann.

Autor: **Peter Hense ist Rechtsanwalt und Partner bei SPIRIT LEGAL LLP Rechtsanwälte im Bereich Technologie und Datennutzung in Leipzig.**