

# Leitlinien



## **Leitlinien 02/2021 zu virtuellen Sprachassistenten**

**Version 2.0**

**Angenommen am 7. Juli 2021**

Translations proofread by EDPB Members.  
This language version has not yet been proofread.

## Versionsverlauf

Version 2.0	7. Juli 2021	Annahme der Leitlinien nach öffentlicher Konsultation
Version 1.0	9. März 2021	Annahme der Leitlinien für die öffentliche Konsultation

## ZUSAMMENFASSUNG

Ein virtueller Sprachassistent (Virtual Voice Assistant, VVA) ist ein Dienst, der Sprachbefehle versteht und ausführt oder gegebenenfalls an andere IT-Systeme vermittelt. VVAs sind heute auf den meisten Smartphones und Tablets, auf herkömmlichen Computern und seit einigen Jahren auch als eigenständige Geräte wie Smart Speaker verfügbar.

VVAs fungieren als Schnittstelle zwischen Nutzerinnen und Nutzern (im Folgenden „Nutzer“) und ihren Rechnern und Online-Diensten wie Suchmaschinen oder Online-Shops. Aufgrund ihrer Funktion haben VVAs Zugang zu einer riesigen Menge personenbezogener Daten, einschließlich der Befehle aller Nutzer (z. B. Browser- oder Suchverlauf) und der Antworten (z. B. Termine im Kalender).

Die allermeisten VVA-Dienste wurden von wenigen VVA-Designern konzipiert. VVAs können aber in Verbindung mit Anwendungen arbeiten, die von Dritten (VVA-Anwendungsentwicklern) programmiert wurden, und dadurch anspruchsvollere Befehle ermöglichen.

Für den ordnungsgemäßen Betrieb benötigt ein virtueller Sprachassistent ein Endgerät mit Mikrofonen und Lautsprechern. Das Gerät speichert Sprache und andere Daten, die heutige VVAs an entfernte VVA-Server übertragen.

Verantwortliche, die VVA-Dienste anbieten, und ihre Auftragsverarbeiter müssen sich daher sowohl an die DSGVO<sup>1</sup> als auch an die ePrivacy-Richtlinie<sup>2</sup> halten.

In den vorliegenden Leitlinien werden einige der wichtigsten Herausforderungen in Bezug auf die Einhaltung der Vorschriften aufgezeigt und für maßgebliche Interessenträger Empfehlungen formuliert, wie die Herausforderungen bewältigt werden können.

Auch wenn Verantwortliche VVA-Dienste über bildschirmlose Endgeräte anbieten, müssen sie die Nutzer gemäß der DSGVO informieren, wenn der Sprachassistent eingerichtet oder eine VVA-App installiert oder erstmalig genutzt wird. Deshalb empfehlen wir Anbietern/Designern und Entwicklern von VVAs, sprachbasierte Schnittstellen zu entwickeln, um die vorgeschriebene Information zu erleichtern.

Derzeit verlangen alle VVAs, dass sich mindestens ein Nutzer bei dem Dienst anmeldet. Entsprechend der Verpflichtung zum Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen sollten Anbieter/Designer und Entwickler von VVAs prüfen, ob für jede ihrer Funktionalitäten ein registrierter Nutzer erforderlich ist.

Im Benutzerkonto vieler VVA-Designer ist der VVA-Dienst mit anderen Diensten wie E-Mail oder Videostreaming gebündelt. Nach Auffassung des EDSA sollten Verantwortliche auf diese Praxis verzichten, weil sie umfassende und komplexe Datenschutzregelungen erfordert, die mit dem Transparenzgrundsatz der DSGVO nicht vereinbar sind.

In den Leitlinien werden vier der häufigsten Zwecke berücksichtigt, für die VVAs personenbezogene Daten verarbeiten: Erledigung von Anfragen, Verbesserung des Modells maschinellen Lernens für den

---

<sup>1</sup> Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung) (im Folgenden „DSGVO“).

<sup>2</sup> Richtlinie 2002/58/EG des Europäischen Parlaments und des Rates vom 12. Juli 2002 über die Verarbeitung personenbezogener Daten und den Schutz der Privatsphäre in der elektronischen Kommunikation (Datenschutzrichtlinie für elektronische Kommunikation), geändert durch die Richtlinie 2006/24/EG und die Richtlinie 2009/136/EG (im Folgenden „ePrivacy-Richtlinie“).

Sprachassistenten, biometrische Identifizierung und Profilerstellung für personalisierte Inhalte oder Werbung.

Soweit VVA-Daten verarbeitet werden, um Anfragen des Nutzers auszuführen, d. h. soweit die Verarbeitung für die Erbringung eines vom Nutzer gewünschten Dienstes unbedingt erforderlich ist, sind Verantwortliche von der Pflicht befreit, gemäß Artikel 5 Absatz 3 der ePrivacy-Richtlinie vorher die Einwilligung des Nutzers einzuholen. Diese Einwilligung nach Artikel 5 Absatz 3 der ePrivacy-Richtlinie muss jedoch eingeholt werden, wenn Daten für andere Zwecke als zur Erledigung von Nutzeranfragen gespeichert oder erhoben werden sollen.

Einige VVA-Dienste speichern personenbezogene Daten, bis die Nutzer ihre Löschung verlangen. Dies steht nicht im Einklang mit dem Grundsatz der Speicherbegrenzung. VVAs dürfen personenbezogene Daten nur so lange speichern, wie es für die Zwecke, für die diese Daten verarbeitet werden, erforderlich ist.

Erlangt ein Verantwortlicher (z. B. im Rahmen einer Qualitätsprüfung) Kenntnis von der versehentlichen Erfassung personenbezogener Daten, muss er feststellen, ob für jeden Zweck der Verarbeitung dieser Daten eine gültige Rechtsgrundlage besteht. Andernfalls sind die versehentlich erfassten Daten zu löschen.

Ein VVA kann Daten von mehreren betroffenen Personen verarbeiten. Deshalb sollten die Anbieter/Designer Zugangskontrollmechanismen einführen, um die Vertraulichkeit, die Integrität und die Verfügbarkeit personenbezogener Daten zu gewährleisten. Manche herkömmliche Zugangskontrollmechanismen wie Passwörter sind für Sprachassistenten allerdings ungeeignet, da sie laut ausgesprochen werden müssen. Die Leitlinien enthalten hierzu einige Überlegungen, u. a. im Abschnitt über die Verarbeitung besonderer Datenkategorien für die biometrische Identifizierung.

VVA-Anbieter/-Designer sollten berücksichtigen, dass bei der Aufzeichnung der Stimme des Nutzers Stimmen anderer Personen oder Daten wie Hintergrundgeräusche zu hören sein können, die für den Dienst nicht erforderlich sind. Deshalb sollten VVA-Designer nach Möglichkeit Technologien einsetzen, die nicht benötigte Daten herausfiltern und gewährleisten, dass nur die Stimme des Nutzers aufgezeichnet wird.

Hinsichtlich der Notwendigkeit einer Datenschutz-Folgenabschätzung (DSFA) ist nach Auffassung des EDSA davon auszugehen, dass für VVA-Dienste höchstwahrscheinlich die Kategorien und Bedingungen gelten, die eine DSFA erforderlich machen.

Verantwortliche, die VVA-Dienste erbringen, sollten sicherstellen, dass die Nutzer ihre Rechte als betroffene Personen durch unkomplizierte Sprachbefehle ausüben können. VVA-Anbieter/-Designer und auch App-Entwickler sollten die Nutzer am Ende des Prozesses per Sprachnachricht oder durch schriftliche Benachrichtigung an das Mobiltelefon, das Nutzerkonto oder in anderer, vom Nutzer gewählter Form darüber informieren, dass ihre Rechte ordnungsgemäß berücksichtigt wurden.

## Inhaltsverzeichnis

<b>ZUSAMMENFASSUNG</b> .....	3
<b>1 ALLGEMEINES</b> .....	7
<b>2 TECHNOLOGISCHER HINTERGRUND</b> .....	8
2.1 Basismerkmale von virtuellen Sprachassistenten .....	8
2.2 Akteure im VVA-Ökosystem .....	9
2.3 Schrittweise Ausführung .....	10
2.4 Aktivierungswörter .....	11
2.5 Voice Snippets und maschinelles Lernen .....	12
<b>3 ELEMENTE DES DATENSCHUTZES</b> .....	12
3.1 Rechtsrahmen .....	12
3.2 Datenverarbeitung und Interessenträger .....	15
3.2.1 Verarbeitung personenbezogener Daten.....	15
3.2.2 Verarbeitung durch Verantwortliche und Auftragsverarbeiter .....	17
3.3 Transparenz.....	19
3.4 Zweckbegrenzung und Rechtsgrundlage .....	23
3.4.1 Ausführung von Nutzeranfragen .....	24
3.4.2 Verbesserung des VVA durch Training der ML-Systeme und manuelle Überprüfung von Sprache und Transkripten .....	25
3.4.3 Nutzeridentifizierung (anhand von Sprachdaten).....	26
3.4.4 Erstellung von Nutzerprofilen für personalisierte Inhalte oder Werbung .....	26
3.5 Verarbeitung personenbezogener Daten von Kindern .....	28
3.6 Vorratsdatenspeicherung.....	28
3.7 Sicherheit.....	31
3.8 Verarbeitung besonderer Datenkategorien.....	33
3.8.1 Allgemeine Überlegungen zur Verarbeitung besonderer Datenkategorien .....	33
3.8.2 Spezielle Überlegungen zur Verarbeitung biometrischer Daten.....	34
3.9 Datenminimierung .....	36
3.10 Rechenschaftspflicht .....	36
3.11 Datenschutz durch Design und durch standardmäßige Voreinstellung.....	37
<b>4 Mechanismen zur Ausübung der Rechte betroffener Personen</b> .....	37
4.1 Recht auf Auskunft .....	38
4.2 Recht auf Berichtigung .....	39
4.3 Recht auf Löschung .....	39
4.4 Recht auf Datenübertragbarkeit .....	41

5	Anhang: Automatische Spracherkennung, Sprachsynthese und Verarbeitung natürlicher Sprache	42
5.1	Automatische Spracherkennung (ASR) .....	42
5.2	Verarbeitung natürlicher Sprache (NLP) .....	43
5.3	Sprachsynthese .....	43

## Der Europäische Datenschutzausschuss –

gestützt auf Artikel 70 Absatz 1 Buchstabe j und Buchstabe e der Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung) (im Folgenden „DSGVO“),

gestützt auf das Abkommen über den Europäischen Wirtschaftsraum, insbesondere auf Anhang XI und Protokoll 37, in der durch den Beschluss Nr. 154/2018 des Gemeinsamen EWR-Ausschusses vom 6. Juli 2018 geänderten Fassung<sup>3</sup>,

gestützt auf Artikel 12 und Artikel 22 seiner Geschäftsordnung –

### HAT FOLGENDE LEITLINIEN ANGENOMMEN:

## 1 ALLGEMEINES

1. Die jüngsten technologischen Fortschritte haben dazu geführt, dass virtuelle Sprachassistenten (VVA) sehr viel präziser geworden sind und beträchtlich an Popularität gewonnen haben. So werden Sprachassistenten unter anderem in Smartphones, vernetzte Fahrzeuge, smarte Lautsprecher und intelligente Fernsehgeräte integriert. Dadurch erhalten die VVAs Zugang zu intimen Informationen, was, wenn sie nicht ordnungsgemäß betrieben werden, die Rechte des Einzelnen auf Datenschutz und Privatsphäre beeinträchtigen könnte. Deshalb unterliegen VVAs und Geräte, in die sie integriert sind, der Kontrolle verschiedener Datenschutzbehörden.
2. Die Anwendung sprachbasierter Interaktionen hat verschiedene Vorteile wie die Natürlichkeit der Interaktion, die kein spezielles Lernen seitens der Nutzer erfordert, die Geschwindigkeit, mit der Befehle ausgeführt werden, oder die Ausweitung des Aktionsbereichs, was den Zugang zu Informationen beschleunigen kann. Die Verwendung von Sprache kann aber auch die korrekte Interpretation der Botschaft erschweren, bedingt durch die Variabilität des Sprachsignals verschiedener Sprecher, die akustische Umgebung, die Mehrdeutigkeit von Sprache usw.
3. In der Praxis ist die flüssige Erledigung oder Vereinfachung von Aufgaben nach wie vor das Hauptmotiv für die Anschaffung eines virtuellen Sprachassistenten. Solche Aufgaben sind beispielsweise das Tätigen/Beantworten eines Anrufs oder das Einschalten einer Zeituhr usw., vor allem, wenn der Nutzer keine Hand frei hat. Hausautomation ist die wichtigste Anwendung, die von VVA-Designern vorangetrieben wird. Sprachassistenten eignen sich als Vermittler im häuslichen Bereich, da sich mit ihrer Hilfe Aufgaben leichter ausführen lassen (Licht anstellen, Heizung regulieren, Rollläden herunterlassen usw.), die mit nur einem, aus der Ferne einfach zu aktivierenden Werkzeug zentralisiert gesteuert werden können. Außer zu persönlichen Zwecken oder im Haushalt lassen sich Sprachbefehle auch in einem beruflichen

---

<sup>3</sup> Soweit hierin auf „Mitgliedstaaten“ Bezug genommen wird, ist dies als Bezugnahme auf „EWR-Mitgliedstaaten“ zu verstehen.

Umfeld anwenden, in dem es schwierig ist, Computer-Tools zu handhaben oder schriftliche Anordnungen zu erteilen (z. B. in der Fertigung).

4. Rein theoretisch könnten vor allem Menschen mit einer Beeinträchtigung und betreuungsbedürftige Personen, denen die Nutzung herkömmlicher Schnittstellen Probleme bereitet, von sprachgesteuerten Schnittstellen profitieren. Virtuelle Sprachassistenten kann den Zugang zu Informationen und Computerressourcen erleichtern und so eine inklusive Logik fördern, da der Einsatz der Stimme es ermöglicht, die mit geschriebenem Text verbundenen Schwierigkeiten mancher Nutzergruppen zu überwinden.
5. Und schließlich gibt es auch im Gesundheitsbereich viele Anwendungsfälle, in denen sprechende oder andere Gesprächsagenten eingesetzt werden können. Während der COVID-19-Pandemie wurden beispielsweise verschiedene Callbots eingesetzt, um Anrufenden eine Vordiagnose bieten zu können. Teilweise wird davon ausgegangen, dass langfristig die gesamte Patientenbetreuung durch Interaktionen zwischen Mensch und Assistent erfolgen könnte, nicht nur in Fragen des Wohlbefindens und der Prävention, sondern auch zur Behandlung und Betreuung.
6. Derzeit gibt es mehr als 3 Milliarden Smartphones, und alle verfügen über integrierte Sprachassistenten, von denen die meisten standardmäßig eingeschaltet sind. Einige der am weitesten verbreiteten Betriebssysteme in Personal Computern und Laptops enthalten ebenfalls VVAs. Durch die zuletzt stark gestiegene Zahl von intelligenten Lautsprechern (im Jahr 2019 wurden 147 Millionen Stück verkauft<sup>4</sup>) sind Sprachassistenten in Millionen von Wohnungen und Büros gelangt. Bisher bietet das VVA-Design jedoch keine standardmäßigen Authentifizierungs- oder Zugangskontrollmechanismen.
7. Das vorliegende Dokument soll eine Orientierungshilfe für die Anwendung der DSGVO im Zusammenhang mit virtuellen Sprachassistenten geben.

## 2 TECHNOLOGISCHER HINTERGRUND

### 2.1 Basismerkmale von virtuellen Sprachassistenten

8. Ein VVA kann definiert werden als eine Softwareanwendung, die fähig ist, in natürlicher Sprache einen mündlichen Dialog mit dem Nutzer zu führen.
9. Natürliche Sprache hat eine für die menschliche Sprache typische Semantik. Abhängig von den sprachlichen Merkmalen und der lexikalischen Vielfalt kann ein und dieselbe Anweisung auf unterschiedliche Weise formuliert werden, während manche Befehle zwar ähnlich klingen, sich aber auf zwei unterschiedliche Objekte beziehen. Inferenzmechanismen werden häufig genutzt, um solche Mehrdeutigkeiten aufzulösen, beispielsweise abhängig von dem zuvor Gesagten, vom Zeitpunkt der Anweisung, vom Standort, von den Interessen der Person usw.
10. Ein VVA kann in Module unterteilt werden und auf diese Weise verschiedene Aufgaben ausführen: Tonaufnahme und -wiedergabe, automatische Sprachumwandlung (Sprache in Text), automatische Sprachverarbeitung, Dialogstrategien, Zugang zu Ontologien (Datensätze und strukturierte Konzepte zu einem bestimmten Bereich) und externen Wissensquellen,

---

<sup>4</sup> Siehe z. B. eine Pressemitteilung des Hamburgischen Beauftragten für Datenschutz und Informationsfreiheit vom 1. August 2019: <https://datenschutz-hamburg.de/pressemitteilungen/2019/08/2019-08-01-google-assistant>.

Sprachgenerierung, Sprachsynthese (Text in Sprache) usw. Konkret sollte der Assistent eine Interaktion ermöglichen, um dann Handlungen auszuführen (z. B. „Stell das Radio an“, „Schalte das Licht aus“) oder auf Wissen zuzugreifen (z. B. „Wie wird das Wetter morgen?“, „Fährt der Frühzug um 7.43 Uhr?“). Damit spielt der VVA die Rolle eines Vermittlers und Orchestrators, der die Ausführung der Aufgaben des Nutzers erleichtern soll.

11. In der Praxis ist ein VVA kein smarterer Lautsprecher, aber ein smarterer Lautsprecher kann mit einem Sprachassistenten ausgestattet werden. Beides wird häufig verwechselt, aber ein smarterer Lautsprecher ist nur eine materielle Ausgestaltung des Sprachassistenten. Ein VVA kann in einem Smartphone, einem smarten Lautsprecher, einer Smartwatch, einem Fahrzeug, einer Haushaltsanwendung usw. enthalten sein.

12. Die Organisation der zugrunde liegenden Datenverarbeitung kann verschiedene Informationsflussmuster umfassen. Hier lassen sich drei Hauptkomponenten unterscheiden:

**die Hardware-Komponente:** der Teil, in den der Assistent eingebettet ist (Smartphone, Lautsprecher, smarter Fernseher usw.), mit Mikrofonen, Lautsprechern sowie Netz- und Rechnerkapazitäten (je nach Ausführung mehr oder weniger entwickelt);

**die Software-Komponente:** der Teil, der die eigentliche Mensch-Maschine-Interaktion umsetzt und die Module für die automatische Spracherkennung, die Verarbeitung natürlicher Sprache, Dialog und Sprachsynthese umfasst. Die Steuerung kann direkt über die Hardware vorgenommen werden, aber häufig erfolgt sie aus der Ferne;

**die Ressourcen:** externe Daten wie Datenbanken, Ontologien oder Geschäftsanwendungen, die Wissen bereitstellen (z. B. „Nenn mir die Zeit an der Westküste der Vereinigten Staaten“, „Lies meine E-Mails vor“) oder die konkrete Umsetzung eines Befehls ermöglichen (z. B. „Erhöhe die Temperatur um 1,5 °C“).

13. Zur Erweiterung der Kernfunktionen von VVAs können Komponenten oder Anwendungen von Drittanbietern installiert werden. Jeder Sprachassistent bezeichnet die Komponenten anders, aber bei allen geht es um den Austausch personenbezogener Nutzerdaten zwischen dem VVA-Designer und dem App-Entwickler.

14. Obwohl die meisten VVAs keine Voice Snippets mit den App-Entwicklern gemeinsam nutzen, verarbeiten diese Akteure personenbezogene Daten. Darüber hinaus erhält der App-Entwickler je nach Art der bereitgestellten Funktionalität Intentionen und Slots, die sensible Informationen wie Gesundheitsdaten enthalten könnten.

## 2.2 Akteure im VVA-Ökosystem

15. An der gesamten Ausführungskette eines VVA können zahlreiche Akteure und Vermittler beteiligt sein. In der Praxis sind bis zu fünf verschiedene Akteure auszumachen. Je nach Geschäftsmodell und gewählter Technologie können manche Akteure auch mehrere Rollen in unterschiedlicher Kombination wahrnehmen, beispielsweise als Designer und Integrator oder als Designer und Anwendungsentwickler:

- a. **VVA-Anbieter (oder Designer):** verantwortlich für die Entwicklung des VVA, gestaltet und definiert die Möglichkeiten und Standardfunktionen des VVA wie Aktivierungsmodalitäten, Auswahl der Architektur, Datenzugang, Aufzeichnungsmanagement, Hardware-Spezifikationen usw.;

- b. **VVA-Anwendungsentwickler:** wie bei mobilen Anwendungen entwickelt er Anwendungen, die die Standardfunktionen des VVA erweitern. Dabei sind die Entwicklungsvorgaben des Designers zu beachten;
- c. **Integrator:** Hersteller verbundener Objekte, die er mit einem VVA ausstatten möchte. Dabei sollten die Anforderungen des Designers eingehalten werden;
- d. **Eigentümer:** zuständig für konkrete Bereiche, in denen Personen verkehren (Unterkünfte, berufliche Umgebung, Mietfahrzeuge usw.), und möchte seiner Kundschaft einen VVA anbieten (eventuell mit speziellen Anwendungen);
- e. **Nutzer:** letztes Glied in der VVA-Wertschöpfungskette, kann den Sprachassistenten in verschiedenen Geräten nutzen (Lautsprecher, Fernseher, Smartphone, Uhr usw.), je nachdem, wie und wo er installiert und eingerichtet wurde.

### 2.3 Schrittweise Ausführung

16. Damit ein VVA eine Aktion ausführen oder auf Informationen zugreifen kann, werden nacheinander verschiedene Aufgaben ausgeführt:
  - 1) Als Teil eines Geräts (Smartphone, Lautsprecher, Fahrzeug) befindet sich der VVA im Bereitschaftsmodus. Genau genommen hört er ständig zu. Doch solange er kein Aktivierungswort erkennt, wird kein Geräusch aus dem Sprachempfangsgerät übertragen, und nur die Erkennung des Aktivierungsbefehls ist aktiviert. Dazu wird ein Puffer von wenigen Sekunden vorgehalten (hierauf wird im folgenden Abschnitt genauer eingegangen).
  - 2) Sobald der Nutzer das Aktivierungswort spricht, vergleicht der VVA das Gehörte lokal mit dem Aktivierungsbefehl. Wenn sie übereinstimmen, öffnet der VVA einen Hörkanal, und der Audioinhalt wird sofort übertragen.
  - 3) Wenn die Verarbeitung des Befehls weiter entfernt erfolgt, wird die Aussprache des Schlüsselworts auf der Serverseite häufig erneut überprüft, um unerwünschte Aktivierungen einzuschränken.
  - 4) Der Nutzer gibt seine Anfrage ein, die unverzüglich an den VVA-Anbieter übermittelt wird. Die Sprachsequenz wird dann automatisch umgewandelt (Sprache in Text).
  - 5) Mittels Verarbeitung natürlicher Sprache (NLP) wird der Befehl interpretiert. Die Intentionen der Nachricht werden extrahiert, und Informationsvariablen (Slots) werden identifiziert. Dann wird durch einen Dialogmanager festgelegt, welches Interaktionsszenario mit dem Nutzer umgesetzt werden soll, und das entsprechende Antwortschema bereitgestellt.
  - 6) Wenn der Befehl eine Funktionalität betrifft, die von der App eines Drittanbieters bereitgestellt wird (Skill, Aktion, Shortcut usw.), sendet der VVA-Anbieter dem App-Entwickler die Intentionen und Informationsvariablen (Slots) der Nachricht.
  - 7) Eine für die Anfrage des Nutzers (vermutlich) passende Antwort wird ermittelt. „Ich kann Ihre Frage nicht beantworten“ ist eine passende Antwort, wenn der VVA die Anfrage nicht richtig interpretieren konnte. Gegebenenfalls werden entfernte Ressourcen genutzt in

Form öffentlich zugänglicher Wissensdatenbanken (Online-Enzyklopädie usw.) oder durch Authentifizierung (Bankkonto, Musikanwendung, Kundenkonto für Online-Käufe usw.), und die Informationsvariablen (Slots) werden mit dem erworbenen Wissen versehen.

- 8) Es wird ein Antwortsatz erstellt und/oder eine Handlung identifiziert (Rolläden herablassen, Temperatur erhöhen, Musik abspielen, eine Frage beantworten usw.). Der Satz wird synthetisiert (Text in Sprache) und/oder der Befehl für die auszuführende Handlung wird an das entsprechende Gerät gesendet.
- 9) Der VVA kehrt in den Bereitschaftsmodus zurück.

Hierzu ist anzumerken, dass die meisten sprachbezogenen Verarbeitungsvorgänge derzeit auf entfernten Servern durchgeführt werden, dass aber einige VVA-Anbieter Systeme entwickeln, die einen Teil dieser Verarbeitung lokal durchführen können.<sup>5</sup>

## 2.4 Aktivierungswörter

17. Wenn der VVA genutzt werden soll, muss er „wach“ sein. Das bedeutet, dass der Assistent in einen aktiven Zuhörmodus umschaltet, um Anweisungen und Befehle vom Nutzer entgegenzunehmen. Dieses Aufwecken kann durch eine physische Handlung erreicht werden (z. B. durch Tastendruck, Druck auf den intelligenten Lautsprecher usw.), doch fast alle gängigen VVAs reagieren auf einen Aktivierungsbefehl oder ein Wort (auch Aufwachwort oder Hotword), um in einen aktiven Hörmodus zu wechseln.
18. Dazu nutzt der Assistent das Mikrofon und geringe Rechnerkapazitäten, um festzustellen, ob das Schlüsselwort genannt wurde. Diese Analyse, die von dem Moment an, in dem der VVA angeschaltet ist, laufend stattfindet, wird ausschließlich lokal durchgeführt. Erst wenn das Schlüsselwort erkannt wurde, werden die Tonaufzeichnungen für die Interpretation und Ausführung des Befehls verarbeitet; dazu werden sie in vielen Fällen über das Internet an einen entfernten Server gesendet. Die Erkennung von Schlüsselwörtern basiert auf maschinellem Lernen. Die größte Herausforderung bei der Anwendung solcher Methoden besteht darin, dass es sich um eine probabilistische Identifizierung handelt. So liefert das System für jedes gesprochene Wort bzw. jeden Begriff einen Zuverlässigkeitswert, der angibt, ob das Schlüsselwort tatsächlich gesprochen wurde. Wenn dieser Wert einen bestimmten Schwellenwert übersteigt, gilt das Wort als gesprochen. Daher ist das System nicht fehlerfrei, denn während die Aktivierung in manchen Fällen möglicherweise nicht erkannt wird, obwohl das Schlüsselwort genannt wurde (Falschrückweisung), kann andererseits eine Aktivierung erkannt werden, obwohl der Nutzer das Schlüsselwort gar nicht ausgesprochen hat (Falschakzeptanz).
19. In der Praxis muss ein annehmbarer Kompromiss zwischen diesen beiden Fehlschlüssen gefunden werden, um den Schwellenwert festzulegen. Da eine Falscherkennung des Schlüsselworts zur Folge haben kann, dass Tonaufzeichnungen gesendet werden, ist mit einer unerwarteten und unerwünschten Datenübermittlung zu rechnen. Sehr häufig wenden VVA-Anbieter, die Daten auf entfernten Servern verarbeiten, einen zweistufigen Mechanismus zur Erkennung an. Die erste Stufe befindet sich lokal im Gerät und die zweite auf dem Remote Server, wo die nächste Datenverarbeitung erfolgt. In dem Fall setzen Entwickler eher einen relativ niedrigen Schwellenwert an, um die Nutzererfahrung zu verbessern und

---

<sup>5</sup> Darüber wurde beispielsweise berichtet unter: <https://www.amazon.science/blog/alexa-new-speech-recognition-abilities-showcased-at-interspeech>.

sicherzustellen, dass das vom Nutzer gesprochene Schlüsselwort fast immer erkannt wird, selbst wenn es dadurch etwas zu häufig identifiziert werden sollte, und daran schließt sich eine zweite, restriktivere Erkennungsstufe auf der Serverseite an.

## 2.5 Voice Snippets und maschinelles Lernen

20. Sprachassistenten basieren auf Methoden des maschinellen Lernens, um ein breites Spektrum von Aufgaben zu erfüllen (Schlüsselwörtererkennung, automatische Spracherkennung, Verarbeitung natürlicher Sprache, Sprachsynthese usw.), und dafür müssen umfangreiche Datensätze gesammelt, ausgewählt, gekennzeichnet werden usw.
21. Die Über- oder Unterrepräsentation bestimmter statistischer Merkmale kann die Entwicklung maschinenlernbasierter Aufgaben beeinflussen und sich anschließend in ihren Berechnungen und damit in ihrer Funktionsweise widerspiegeln. Daher ist nicht nur die Quantität, sondern auch die Qualität der Daten von Bedeutung für die Feinheit und die Genauigkeit des Lernprozesses.
22. Um die Qualität des Sprachassistenten zu verbessern und die angewandten Methoden des maschinellen Lernens zu verbessern, würden VVA-Designer sicherlich gerne auf Informationen über die Nutzung des Geräts unter realen Bedingungen, d. h. auf Voice Snippets, zugreifen, um es weiter zu verbessern.
23. Zur Qualifizierung der Lerndatenbank wie auch zur Berichtigung von Fehlern bei der Anwendung des Algorithmus muss der Mensch in das Lernen und das Training von Systemen künstlicher Intelligenz eingreifen. Dieser Teil der Arbeit, das sogenannte Digitallabor, wirft Fragen hinsichtlich der Arbeitsbedingungen wie auch der Sicherheit auf. In diesem Zusammenhang wurde in den Medien von Datenübertragungen zwischen VVA-Designern und Unterauftragnehmern berichtet, die ohne die erforderlichen Garantien zum Schutz der Privatsphäre erfolgt sein sollen.

## 3 ELEMENTE DES DATENSCHUTZES

### 3.1 Rechtsrahmen

24. Der einschlägige EU-Rechtsrahmen für VVAs ist in erster Linie die DSGVO, da die Verarbeitung personenbezogener Daten zur Kernfunktion von Sprachassistenten gehört. Zusätzlich zur DSGVO enthält die ePrivacy-Richtlinie<sup>6</sup> spezielle Regeln für alle Akteure, die im Endgerät von Teilnehmenden oder Nutzer im EWR gespeicherte Informationen speichern oder darauf zugreifen wollen.
25. Beispiele für „Endgeräte“<sup>7</sup> gemäß der Definition sind Smartphones, Smart-Fernseher und ähnliche IoT-Geräte. Bei VVAs handelt es sich zwar um Software-Dienste, doch sie werden

---

<sup>6</sup> Richtlinie 2002/58/EG des Europäischen Parlaments und des Rates vom 12. Juli 2002 über die Verarbeitung personenbezogener Daten und den Schutz der Privatsphäre in der elektronischen Kommunikation (Datenschutzrichtlinie für elektronische Kommunikation), geändert durch die Richtlinie 2006/24/EG und die Richtlinie 2009/136/EG (im Folgenden „ePrivacy-Richtlinie“). Xx Richtlinie 2002/58/EG des Europäischen Parlaments und des Rates vom 12. Juli 2002 über die Verarbeitung personenbezogener Daten und den Schutz der Privatsphäre in der elektronischen Kommunikation (Datenschutzrichtlinie für elektronische Kommunikation), ABl. L 201 vom 31.7.2002, S. 37, geändert durch die Richtlinie 2006/24/EG und die Richtlinie 2009/136/EG (im Folgenden „ePrivacy-Richtlinie“).

<sup>7</sup> Nach Artikel 1 der Richtlinie 2008/63/EG der Kommission vom 20. Juni 2008 über den Wettbewerb auf dem Markt für Telekommunikationsendeinrichtungen sind „Endeinrichtungen“: a) direkt oder indirekt an die

immer über ein physisches Gerät wie einen intelligenten Lautsprecher oder ein Smart-TV betrieben. **VVAs nutzen elektronische Kommunikationsnetze, um auf diese physischen Geräte zuzugreifen, die „Endgeräte“ im Sinne der ePrivacy-Richtlinie sind. Deshalb kommen die Bestimmungen von Artikel 5 Absatz 3 der ePrivacy-Richtlinie zur Anwendung, sobald ein Sprachassistent Informationen auf dem mit ihm verbundenen physischen Gerät speichert oder abrufen.**<sup>8</sup>

26. Die Verarbeitung personenbezogener Daten im Anschluss an die oben genannten Verarbeitungsvorgänge, einschließlich der Verarbeitung personenbezogener Daten, die durch den Zugriff auf Informationen im Endgerät erlangt wurden, ist nur rechtmäßig, wenn eine Rechtsgrundlage nach Artikel 6 der DSGVO gegeben ist.<sup>9</sup>
27. Da der Verantwortliche, der gemäß Artikel 5 Absatz 3 der ePrivacy-Richtlinie die Einwilligung in das Speichern oder den Zugang zu Informationen erlangen will, die betroffene Person über alle Zwecke der Verarbeitung (d. h. die „anschließende Verarbeitung“) – einschließlich jeder Verarbeitung nach den oben genannten Vorgängen – informieren muss, wird sich die Einwilligung nach Artikel 6 der DSGVO im Allgemeinen am besten als Rechtsgrundlage für die anschließende Verarbeitung der personenbezogenen Daten eignen. Daher wird die Einwilligung wahrscheinlich die Rechtsgrundlage sowohl für das Speichern und den Zugriff auf bereits gespeicherte Informationen als auch für die Verarbeitung personenbezogener Daten im Anschluss an die oben genannten Verarbeitungsvorgänge bilden. Bei der Prüfung, ob Artikel 6 der DSGVO eingehalten wird, ist zu beachten, dass die gesamte Verarbeitung auch besondere Tätigkeiten umfasst, für die der Unionsgesetzgeber zusätzlichen Schutz sicherstellen wollte.<sup>10</sup> Zudem müssen Verantwortliche bei der Bestimmung der angemessenen Rechtsgrundlage unter anderem die Auswirkungen auf die Rechte der betroffenen Personen berücksichtigen, um den Grundsatz von Treu und Glauben zu wahren.<sup>11</sup> Grundsätzlich gilt, dass sich Verantwortliche nicht auf Artikel 6 DSGVO berufen können, um den zusätzlichen Schutz nach Artikel 5 Absatz 3 der ePrivacy-Richtlinie abzuschwächen.
28. Wie in Abschnitt 2.3 (Schritte 2 und 3) erläutert wurde, verlangen heutige VVAs den Zugriff auf die vom VVA-Gerät gespeicherten Sprachdaten.<sup>12</sup> Daher kommt Artikel 5 Absatz 3 der ePrivacy-Richtlinie zur Anwendung. Die Anwendbarkeit von Artikel 5 Absatz 3 der ePrivacy-Richtlinie bedeutet, dass das Speichern von Informationen und der Zugriff auf bereits in einem

---

*Schnittstelle eines öffentlichen Telekommunikationsnetzes angeschlossene Einrichtungen zum Aussenden, Verarbeiten oder Empfangen von Nachrichten; sowohl bei direkten als auch bei indirekten Anschlüssen kann die Verbindung über Draht, optische Faser oder elektromagnetisch hergestellt werden; bei einem indirekten Anschluss ist zwischen der Endeinrichtung und der Schnittstelle des öffentlichen Netzes ein Gerät geschaltet; b) Satellitenfunkanlagen mit ihren Einrichtungen“.*

<sup>8</sup> Siehe EDSA, Leitlinien 1/2020 (im Folgenden „EDSA-Leitlinien 1/2020“); unter Randnummer 12 finden sich ähnliche Erwägungen in Bezug auf vernetzte Fahrzeuge. Siehe auch EDSA, Stellungnahme 5/2019 zum Zusammenspiel zwischen der e-Datenschutz-Richtlinie und der DSGVO, insbesondere in Bezug auf die Zuständigkeiten, Aufgaben und Befugnisse von Datenschutzbehörden.

<sup>9</sup> Ebd., Randnummer 41.

<sup>10</sup> Stellungnahme 5/2019, Randnummer 41.

<sup>11</sup> EDSA, Leitlinien 2/2019 für die Verarbeitung personenbezogener Daten gemäß Artikel 6 Absatz 1 Buchstabe b DSGVO im Zusammenhang mit der Bereitstellung von Online-Diensten für betroffene Personen, Version 2.0, 8. Oktober 2019, Randnummer 1.

<sup>12</sup> Es ist durchaus möglich, dass künftige VVA-Geräte das Edge-Computing-Paradigma anwenden und in der Lage sein werden, bestimmte Dienste lokal zu erbringen. In dem Fall wird die Anwendbarkeit der ePrivacy-Richtlinie neu zu bewerten sein.

VVA gespeicherte Informationen in der Regel die vorherige Einwilligung<sup>13</sup> des Endnutzers erfordern. Hier sind allerdings zwei Ausnahmen vorgesehen: zum einen, wenn die Übertragung einer Kommunikation über ein elektronisches Kommunikationsnetzwerk vorgenommen oder erleichtert wird, und zum andern, wenn dies unbedingt erforderlich ist, um einen vom Teilnehmer oder Nutzer ausdrücklich gewünschten Dienst der Informationsgesellschaft bereitzustellen.

29. Die zweite Ausnahme („unbedingt erforderlich, um einen vom Teilnehmer oder Nutzer ausdrücklich gewünschten Dienst der Informationsgesellschaft bereitzustellen“) würde es einem VVA-Diensteanbieter ermöglichen, Nutzerdaten zu verarbeiten, um Nutzeranfragen auszuführen (siehe Randnummer 72 in Abschnitt 3.4.1), ohne die Einwilligung gemäß Artikel 5 Absatz 3 der ePrivacy-Richtlinie einzuholen. Dagegen **wäre eine Einwilligung nach Artikel 5 Absatz 3 der ePrivacy-Richtlinie erforderlich**, um Informationen **für andere Zwecke als für die Erledigung von Nutzeranfragen** (z. B. zur Erstellung eines Nutzerprofils) zu speichern oder Zugriff darauf zu erhalten. Verantwortliche müssten die Einwilligung bestimmten Nutzern zuordnen. Deshalb sollten sie Daten von nicht registrierten Nutzern nur verarbeiten, um deren Anfragen auszuführen.
30. Sprachassistenten können versehentlich Audioaufnahmen von Personen aufnehmen, die gar keinen VVA-Dienst nutzen wollten. Erstens kann der Aktivierungsbefehl in einem gewissen Umfang und abhängig vom jeweiligen Assistenten geändert werden. Wer von dieser Änderung nichts weiß, könnte versehentlich den aktualisierten Aktivierungsbefehl verwenden. Zweitens können Sprachassistenten versehentlich oder irrtümlich einen Aktivierungsbefehl erkennen. Es ist höchst unwahrscheinlich, dass eine der in Artikel 5 Absatz 3 der ePrivacy-Richtlinie vorgesehenen Ausnahmen im Fall einer unbeabsichtigten Aktivierung Anwendung findet. Zudem ist die Einwilligung im Sinne der DSGVO definiert als eine *„unmissverständlich abgegebene Willensbekundung“* *„der betroffenen Person“*. Es ist daher höchst unwahrscheinlich, dass eine versehentliche Aktivierung als gültige Einwilligung interpretiert werden könnte. Erlangt ein Verantwortlicher (z. B. bei einer automatischen oder manuellen Überprüfung) Kenntnis davon, dass der VVA-Dienst versehentlich personenbezogene Daten verarbeitet hat, sollte er prüfen, ob für jeden Zweck der Verarbeitung dieser Daten eine gültige Rechtsgrundlage besteht. Andernfalls sind die versehentlich erfassten Daten zu löschen.
31. Dabei ist auch zu bedenken, dass es sich bei den von Sprachassistenten verarbeiteten personenbezogenen Daten um hochsensible Informationen handeln kann. Sowohl ihr Inhalt (Bedeutung des gesprochenen Textes) als auch ihre Meta-Information (Geschlecht oder Alter der Sprechenden Person usw.) können persönlicher Natur sein. Der EDSA erinnert daran, dass es sich bei Sprachdaten grundsätzlich um biometrische personenbezogene Daten handelt.<sup>14</sup> Wenn solche Daten ausschließlich zum Zweck der eindeutigen Identifizierung einer natürlichen Person verarbeitet werden oder als besondere Kategorie personenbezogener Daten einzustufen sind, muss die Verarbeitung auf einer gültigen Rechtsgrundlage gemäß Artikel 6 erfolgen, ergänzt durch eine der Ausnahmeregelungen nach Artikel 9 der DSGVO (siehe Abschnitt 3.7).

---

<sup>13</sup> Siehe auch EDSA-Leitlinien 1/2020, Randnummer 14.

<sup>14</sup> Nach Artikel 4 Nummer 14 der DSGVO sind „biometrische Daten“ „mit speziellen technischen Verfahren gewonnene personenbezogene Daten zu den physischen, physiologischen oder verhaltenstypischen Merkmalen einer natürlichen Person, die die eindeutige Identifizierung dieser natürlichen Person ermöglichen oder bestätigen, wie Gesichtsbilder oder daktyloskopische Daten“.

## 3.2 Datenverarbeitung und Interessenträger

32. In Anbetracht der vielfältigen Arten von Unterstützung, die ein Sprachassistent einer betroffenen Person in ganz unterschiedlichen Bereichen des täglichen Lebens<sup>15</sup> bieten kann, ist darauf hinzuweisen, dass bei der Verarbeitung personenbezogener Daten, auf die verschiedene Interessenträger Einfluss nehmen können, mit Sorgfalt vorgegangen werden muss.

### 3.2.1 Verarbeitung personenbezogener Daten

33. Was den Schutz personenbezogener Daten betrifft, sind unabhängig von dem von der betroffenen Person verwendeten VVA-Typ (d. h. Art des Geräts, Funktionen, Dienste oder Kombination daraus) verschiedene Konstanten festzustellen. Sie betreffen die Vielfalt der personenbezogenen Daten, der betroffenen Personen und der jeweiligen Datenverarbeitung.

#### **Vielfalt der personenbezogenen Daten**

34. Die Definition des Begriffs „personenbezogene Daten“ in Artikel 4 Nummer 1 der DSGVO umfasst eine große Bandbreite unterschiedlicher Daten; sie gilt in einem technologieneutralen Kontext für alle Informationen, die sich auf „eine identifizierte oder identifizierbare natürliche Person“ beziehen.<sup>16</sup> Jede Interaktion einer betroffenen Person mit einem VVA kann in den Anwendungsbereich dieser Definition fallen. Sobald die Interaktion stattfindet, können verschiedene personenbezogene Daten während der gesamten aktiven Phase des VVA verarbeitet werden, wie in Abschnitt 2.4 erläutert wurde.
35. Von der ursprünglichen Anfrage bis zur entsprechenden Antwort, Aktion oder einem Follow-up (z. B. dem Einstellen eines wöchentlichen Alarms) generiert die erste Eingabe personenbezogener Daten daher weitere personenbezogene Daten. Dazu gehören Primärdaten (z. B. Kontodaten, Sprachaufzeichnungen, Verlauf der Anfragen), Beobachtungsdaten (z. B. Gerätedaten, die sich auf eine betroffene Person beziehen, Tätigkeitsprotokolle, Online-Aktivitäten) sowie gefolgerte oder abgeleitete Daten (z. B. Erstellung von Nutzerprofilen). VVAs verwenden Sprache, um zwischen Nutzern und allen verbundenen Diensten (z. B. einer Suchmaschine, einem Online-Shop oder einem Musikstreaming-Dienst) zu vermitteln, doch im Gegensatz zu anderen Vermittlern können VVAs uneingeschränkter Zugang zum Inhalt der Anfragen haben und so dem VVA-Designer je nach Zweck der Verarbeitung ganz unterschiedliche personenbezogene Daten liefern.
36. Die Vielfalt personenbezogener Daten, die bei der Nutzung eines Sprachassistenten verarbeitet werden, bedeutet auch vielfältige Kategorien personenbezogener Daten, die zu beachten sind (siehe Abschnitt 3.7). Hinsichtlich der Verarbeitung besonderer

---

<sup>15</sup> Beispielsweise in der Wohnung, in einem Fahrzeug, auf der Straße, am Arbeitsplatz oder in anderen privaten, öffentlichen oder beruflichen Räumen oder einer Kombination dieser Räume.

<sup>16</sup> Artikel 4 Nummer 1 DSGVO spezifiziert zudem: „als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind, identifiziert werden kann“.

Datenkategorien<sup>17</sup> weist der EDSA darauf hin, dass der Verantwortliche gemäß Artikel 9 der DSGVO verpflichtet ist, eine gültige Ausnahme vom Verarbeitungsverbot nach Artikel 9 Absatz 1 und eine gültige Rechtsgrundlage nach Artikel 6 Absatz 1 zu ermitteln, sofern einer der in Artikel 9 Absatz 2 aufgeführten Fälle zutrifft. Die ausdrückliche Einwilligung kann eine der geeigneten Ausnahmen sein, wenn die Einwilligung die Rechtsgrundlage nach Artikel 6 Absatz 1 darstellt. In Artikel 9 wird zudem (im Einzelnen) darauf hingewiesen, dass die Mitgliedstaaten zusätzliche Bedingungen für die Verarbeitung biometrischer Daten oder anderer besonderer Kategorien einführen können.

### ***Vielfalt der betroffenen Personen***

37. Bei der Nutzung eines Sprachassistenten werden personenbezogene Daten aus der ersten Interaktion mit dem Assistenten verarbeitet. Für manche betroffenen Personen bezieht sich dies auf den Erwerb eines VVA und/oder die Konfiguration eines Nutzerkontos (registrierte Nutzer). Für andere betroffene Personen bedeutet es das erste Mal, dass sie wissentlich mit dem Sprachassistenten einer anderen betroffenen Person interagieren, die diesen VVA gekauft und/oder konfiguriert hat (nicht registrierte Nutzer). Neben diesen beiden Kategorien betroffener Personen gibt es eine dritte Kategorie: zufällige – registrierte oder nicht registrierte – Nutzer, die unwissentlich Anfragen an einen VVA stellen (z. B. durch den richtigen Aktivierungsbefehl, ohne zu wissen, dass der VVA aktiv ist, oder durch andere Äußerungen, die der VVA irrtümlich als Weckbefehl identifiziert).
38. Die Vielfalt der betroffenen Personen bezieht sich auch auf mehrere Nutzer eines VVA (z. B. eines von registrierten und nicht registrierten Nutzern, von Kollegen, einer Familie oder in der Schule gemeinsam genutzten Geräts) und auf Nutzer mit unterschiedlichen Fähigkeiten (z. B. Erwachsene, Kinder, ältere Menschen oder Menschen mit einer Beeinträchtigung). Da ein VVA die Interaktion mit einem digitalen Instrument vereinfachen und für einige Kategorien betroffener Personen viele Vorteile bieten kann, ist es wichtig, die Besonderheiten jeder Kategorie betroffener Personen und den Kontext der Anwendung des Sprachassistenten zu berücksichtigen.

### ***Vielfalt der Datenverarbeitung***

39. Die zur Bereitstellung eines VVA verwendeten Technologien haben auch Einfluss auf die Menge der verarbeiteten Daten und die Art der Verarbeitung. Je mehr Dienste oder Features ein Sprachassistent anbietet und je mehr er mit anderen Geräten oder Diensten von Drittanbietern verbunden ist, umso größer wird die Menge der verarbeiteten personenbezogenen Daten, und die Verarbeitung zu anderen Zwecken nimmt zu. Dies führt zu einer Vielzahl automatisierter Verarbeitungen, die in Abschnitt 2 erläutert wurden. Die Verarbeitung kann nicht nur automatisiert, sondern auch manuell erfolgen. Das ist beispielsweise der Fall, wenn die angewandte Technologie menschliches Eingreifen erfordert, etwa die Überprüfung der Umwandlung von Sprache in Text oder die Eingabe von Annotationen zu personenbezogenen Daten, die verwendet werden können, um neue Modelle in eine Technologie des maschinellen Lernens aufzunehmen. Es ist auch der Fall, wenn

---

<sup>17</sup> Nach Artikel 9 Absatz 1 DSGVO gilt für besondere Kategorien personenbezogener Daten: „Die Verarbeitung personenbezogener Daten, aus denen die rassische oder ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder die Gewerkschaftszugehörigkeit hervorgehen, sowie die Verarbeitung von genetischen Daten, biometrischen Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung einer natürlichen Person ist untersagt.“

Menschen personenbezogene Daten (z. B. Metadaten) analysieren, um den Dienst eines Sprachassistenten zu verbessern.

### 3.2.2 Verarbeitung durch Verantwortliche und Auftragsverarbeiter

40. Für betroffene Personen muss es möglich sein, die Rolle des jeweiligen Interessenträgers zu verstehen und zu identifizieren und nach Maßgabe der DSGVO zu jedem von ihnen Kontakt aufzunehmen oder mit ihm zu agieren. Die Rollenaufteilung darf nicht zulasten der betroffenen Personen gehen, auch nicht, wenn das Szenario kompliziert ist oder sich verändern kann. Zur Bewertung ihrer Rolle werden die Interessenträger auf die EDSA-Leitlinien 7/2020 zu den Begriffen „Verantwortliche“ und „Auftragsverarbeiter“ in der DSGVO verwiesen.<sup>18</sup>
41. Wie unter Randnummer 15 ausgeführt wurde, können die wesentlichen Interessenträger Anbieter oder Designer, Anwendungsentwickler, Integrator, Eigentümer oder eine Kombination daraus sein. Verschiedene Szenarien sind möglich, je nachdem, wer in der Geschäftsbeziehung von Interessenträgern was tut, und abhängig von der Anfrage des Nutzers, den personenbezogenen Daten, der Datenverarbeitung und ihrem Zweck. Die Akteure müssen sich festlegen und die betroffenen Personen darüber informieren, unter welchen Bedingungen jeder von ihnen tätig wird und seine Rolle als Verantwortlicher, gemeinsam Verantwortlicher oder Auftragsverarbeiter nach Maßgabe der DSGVO wahrnimmt.<sup>19</sup> Jeder von ihnen kann eine oder mehrere Rollen übernehmen, da es sich um einen Verantwortlichen, einen gemeinsamen Verantwortlichen oder einen Auftragsverarbeiter für eine und dieselbe Datenverarbeitung handeln kann, während sie bei einer anderen Datenverarbeitung eine andere Rolle ausüben.
42. Aus übergeordneter Perspektive kann der Designer bei der Festlegung der Zwecke und Mittel einer Verarbeitung als Verantwortlicher agieren, aber er kann auch als Auftragsverarbeiter tätig werden, wenn er personenbezogene Daten im Auftrag einer anderen Partei, beispielsweise eines Anwendungsentwicklers, verarbeitet. Wer einen Sprachassistenten nutzt, hätte es daher mit mehreren Verantwortlichen zu tun, dem Anwendungsentwickler und dem Designer. Es ist auch möglich, dass der Designer, der Integrator und der Entwickler in einer einzigen Stelle zusammengefasst sind, die als einzige Verantwortliche fungiert. Die geltenden Anforderungen sind in jedem einzelnen Fall zu ermitteln.

#### Beispiel 1:

Der VVA-Designer verarbeitet Nutzerdaten für viele Zwecke, u. a. zur Verbesserung der Fähigkeiten des VVA in Bezug auf das Sprachverständnis und die korrekte Beantwortung von Anfragen. Deshalb und obwohl für diesen Zweck Daten verarbeitet werden können, die aus der Nutzung von Anwendungen von Drittanbietern stammen, gibt es nur einen Verantwortlichen, den Designer des VVA, in dessen Namen und für dessen Zwecke die Verarbeitung erfolgt.

<sup>18</sup> EDPB Guidelines 07/2020 on the concepts of controller and processor in the GDPR, Version 2.0, angenommen am 7. Juli 2021 (ESDA, Leitlinien 07/2020 zu den Begriffen „Verantwortlicher“ und „Auftragsverarbeiter“ in der DSGVO, im Folgenden „Leitlinien 7/2020“).

<sup>19</sup> Artikel 12 bis 14 und Artikel 26 DSGVO.

Beispiel 2:

Eine Bank bietet ihrer Kundschaft eine Anwendung an, die die Verwaltung der eigenen Konten direkt über den Sprachassistenten ermöglicht.

An der Verarbeitung personenbezogener Daten sind zwei Akteure beteiligt, der Designer des VVA und der Entwickler der Bankanwendung.

In diesem Szenario ist die Bank die Verantwortliche für die Erbringung des Dienstes, da sie sowohl die Zwecke als auch die wesentlichen Mittel zur Verarbeitung im Zusammenhang mit der Anwendung bestimmt, die eine Interaktion mit dem Assistenten ermöglicht. Sie bietet eine spezielle Anwendung an, die es Kundinnen und Kunden der Bank ermöglicht, die eigenen Konten aus der Ferne zu verwalten. Außerdem entscheidet sie über die Mittel der Verarbeitung, indem sie einen geeigneten Auftragsverarbeiter wählt, den Designer des VVA, der durch seine Expertise wichtige Unterstützung bei der Bestimmung dieser Mittel leisten kann (er kann beispielsweise die Entwicklungsplattform betreiben, die die Integration der Anwendungen von Drittanbietern in den VVA ermöglicht, und legt daher den Rahmen und die Bedingungen fest, die von Anwendungsentwicklern einzuhalten sind).

43. Hinsichtlich der betroffenen Person ist darauf hinzuweisen, dass mehrere Interessenträger dieselben personenbezogenen Daten verarbeiten können, auch wenn die betroffene Person eigentlich nicht davon ausgeht, dass andere Parteien als der VVA-Anbieter an der Verarbeitungskette beteiligt sind. Wenn also eine betroffene Person in Bezug auf ihre personenbezogenen Daten mit dem VVA-Anbieter agiert (z. B. durch Ausübung ihrer Rechte als betroffene Person), bedeutet das nicht automatisch, dass für diese personenbezogenen Daten dasselbe gilt, wenn sie von einem anderen Interessenträger verarbeitet werden. Wenn es sich bei diesen Akteuren um unabhängige Verantwortliche handelt, ist es wichtig, dass die betroffenen Personen ganz klar davon in Kenntnis gesetzt werden und ihnen die verschiedenen Phasen und Akteure der Verarbeitung erläutert werden. Darüber hinaus ist bei gemeinsamer Verantwortlichkeit klarzustellen, ob jeder Verantwortliche dafür zuständig ist, dass alle Rechte der betroffenen Person berücksichtigt werden, oder welcher Verantwortliche für welches Recht zuständig ist.<sup>20</sup>

Beispiel 3:

In diesem Szenario möchte der VVA-Designer die für den von der Bank bereitgestellten Dienst erhobenen und verarbeiteten Daten nutzen, um das Spracherkennungssystem zu verbessern. Der VVA-Designer, der die Daten für seine eigenen Zwecke verarbeitet, wird damit zum Verantwortlichen für diese spezielle Verarbeitung.

44. Da viele Interessenträger und dementsprechend viele Mitarbeitende in die Verarbeitungskette eingebunden sein können, kann es zu riskanten Situationen kommen, falls keine geeigneten Maßnahmen und Sicherheitsvorkehrungen getroffen werden. Da die Verantwortlichen rechenschaftspflichtig sind, sollten sie sich auf den Schutz personenbezogener Daten konzentrieren und dafür insbesondere geeignete Geschäftspartner und Auftragsverarbeiter auswählen, den Datenschutz durch entsprechende Voreinstellungen und konzeptionelle Grundsätze gewährleisten<sup>21</sup> sowie angemessene Sicherheitsinstrumente und andere in der

<sup>20</sup> Leitlinien 07/2020, Randnummer 165.

<sup>21</sup> Siehe EDSA, Leitlinien 4/2019 zu Artikel 25 – Datenschutz durch Technikgestaltung und durch datenschutzrechtliche Voreinstellungen, Version 2.0, angenommen am 20. Oktober 2020.

DSGVO vorgesehene Instrumente anwenden wie Audits und rechtliche Vereinbarungen (z. B. nach Artikel 26 der DSGVO für gemeinsam Verantwortliche oder Artikel 28 für Auftragsverarbeiter).

45. In dem komplexen Ökosystem eines Sprachassistenten könnten zahlreiche Akteure als Verantwortliche oder Auftragsverarbeiter personenbezogene Daten untereinander austauschen und verarbeiten. Deshalb ist es besonders wichtig, dass die Rolle jedes Akteurs in Bezug auf jede Verarbeitung geklärt und der Grundsatz der Datenminimierung auch in Bezug auf den Datenaustausch eingehalten wird.
46. Darüber hinaus müssen die Verantwortlichen bei der Übermittlung personenbezogener Daten wachsam sein und den geforderten Schutz in der gesamten Verarbeitungskette gewährleisten, vor allem, wenn sie Dienste außerhalb des EWR in Anspruch nehmen.

### 3.3 Transparenz

47. Da VVAs personenbezogene Daten verarbeiten (z. B. die Stimme des Nutzers sowie Ort oder Inhalt der Kommunikation), sind die Transparenzanforderungen der DSGVO gemäß Artikel 5 Absatz 1 Buchstabe a sowie Artikel 12 und Artikel 13 (erläutert durch Erwägungsgrund 58) einzuhalten. Verantwortliche sind verpflichtet, die Nutzer in präziser, transparenter, verständlicher und leicht zugänglicher Form über die Verarbeitung ihrer personenbezogenen Daten zu informieren.
48. Unterbleibt die Bereitstellung der erforderlichen Information, so stellt das eine Pflichtverletzung dar, wodurch die Rechtmäßigkeit der Datenverarbeitung in Frage gestellt sein könnte. Die Transparenzanforderungen sind zwingend einzuhalten, weil damit ein Kontrollmechanismus für die Datenverarbeitung besteht, der es Nutzern ermöglicht, ihre Rechte wahrzunehmen. Die ordnungsgemäße Information der Nutzer über die Verwendung ihrer personenbezogenen Daten macht es den Verantwortlichen schwerer, den Sprachassistenten für Zwecke zu missbrauchen, die weit über das hinausgehen, womit Nutzer rechnen können. Beispielsweise sind patentierte Technologien darauf ausgelegt, den Gesundheits- und Gefühlszustand aus der Stimme eines Nutzers abzuleiten und die angebotenen Dienste entsprechend anzupassen.
49. Es kann für den VVA-Diensteanbieter und jede andere Stelle, die als Verantwortliche fungiert, besonders schwierig sein, die Transparenzanforderungen einzuhalten. Die Besonderheiten von Sprachassistenten bereiten Verantwortlichen bei der Einhaltung der Transparenzanforderungen der DSGVO einige Schwierigkeiten:
  - ) **Mehrere Nutzer:** Verantwortliche müssen alle (registrierten, nicht registrierten und zufälligen) Nutzer informieren und nicht nur die Person, die den Sprachassistenten eingerichtet hat.
  - ) **Komplexität des Ökosystems:** Wie im Abschnitt zu den technologischen Hintergründen erläutert wird, sind die Identitäten und Rollen derjenigen, die bei der Nutzung eines VVA personenbezogene Daten verarbeiten, für die Nutzer alles andere als offensichtlich.
  - ) **Besonderheiten der Sprachschnittstelle:** Digitalsysteme eignen sich noch nicht für rein sprachliche Interaktionen, was sich daran zeigt, dass fast immer ein ergänzender Bildschirm verwendet wird. Es ist unerlässlich, dass eine Anpassung an die Sprachschnittstelle vorgenommen und der Nutzer auf diesem Weg klar und korrekt informiert wird.

50. VVAs können als endliche Automaten angesehen werden, die während ihres normalen Betriebs verschiedene Zustände annehmen. Sie können lokal zuhören, um Aktivierungsbefehle zu erkennen, oder mit einem entfernten Server interagieren, um einen Befehl auszuführen, aber auch viele andere Zustände annehmen je nach Kontext (z. B. bei Hintergrundgeräuschen) oder mit ihnen kommunizierendem Nutzer (sie können z. B. mit einem identifizierten oder unbekanntem Nutzer sprechen). Leider besteht eine beträchtliche Informationsasymmetrie für den Nutzer, der kaum weiß, ob das Gerät zuhört, und noch weniger, in welchem Zustand es sich befindet.
51. Designern und Entwicklern von VVAs wird dringend empfohlen, geeignete Schritte zu unternehmen, um solche Asymmetrien abzubauen und die Sprachassistenten noch interaktiver zu gestalten. Nutzer sollten wissen, in welchem Zustand sich das Gerät befindet. Mehr Transparenz in dieser Hinsicht wäre durch eine interaktivere Gestaltung des Mensch-Maschine-Dialogs (das Gerät könnte z. B. den Eingang eines Sprachbefehls in irgendeiner Weise bestätigen) oder durch Meldung des Zustands der Maschine mittels bestimmter Signale zu erreichen. Hier kommen viele Möglichkeiten in Betracht, von gesprochenen Bestätigungen und sichtbaren Bildsymbolen oder Leuchtsignalen bis zu Anzeigen am Gerät.
52. Besondere Relevanz hat dieser Aspekt in Anbetracht der Vielfalt der Nutzer, zu denen auch schutzbedürftige Gruppen wie Kinder, ältere Menschen oder Personen mit einer Hör- oder Sehbehinderung gehören.
53. Zwei wichtige Fragen ergeben sich aus den genannten Aspekten: Wie können Nutzer am besten informiert werden, und zu welchem Zeitpunkt sollten sie informiert werden? Diese Fragen sollten in Bezug auf zwei unterschiedliche Situationen weiter untersucht werden, je nachdem, ob der Sprachassistent nur einen Nutzer (z. B. ein persönliches Smartphone) oder möglicherweise mehrere Nutzer (z. B. Smart-Home-Geräte) hat. Die Anwendung der VVA-Technologie ermöglicht auch eine Unter-Version dieser beiden Grundeinstellungen, etwa wenn ein Nutzer sein persönliches Smartphone mit einem Auto verbindet. Der VVA des Smartphones, das aller Wahrscheinlichkeit nach nur von diesem Nutzer verwendet wird, wird damit auf die anderen Fahrzeuginsassen „ausgeweitet“.
54. Alle derzeitigen VVAs sind mit einem Benutzerkonto verbunden und/oder werden durch eine Anwendung eingerichtet, die ein Benutzerkonto verlangt. Wie Verantwortliche die Nutzer bei der Einrichtung des VVA über die Datenschutzbestimmungen informieren könnten, lässt sich den Leitlinien der Artikel-29-Datenschutzgruppe zur Transparenz entnehmen. Bevor Apps heruntergeladen werden, sollten die erforderlichen Informationen in einem Online-Shop zur Verfügung gestellt werden.<sup>22</sup> Auf diese Weise erfolgt die Information zum frühestmöglichen Zeitpunkt, spätestens jedoch, wenn die personenbezogenen Daten eingehen. Manche VVA-Anbieter nehmen Apps von Drittanbietern in die Standardeinstellung des Sprachassistenten auf, damit diese Apps auf ein bestimmtes Aktivierungswort hin ausgeführt werden können. Wenn VVAs auf diese Weise die Apps von Drittanbietern aktivieren, muss sichergestellt sein, dass die Nutzer auch über die von diesen Dritten vorgenommene Verarbeitung hinreichend informiert werden.
55. Viele VVA-Designer verlangen VVA-Nutzerkonten, die den VVA-Dienst mit mehreren anderen Diensten wie E-Mail, Videostreaming oder Einkauf bündeln, um nur einige zu nennen. Die

---

<sup>22</sup> Artikel 29-Gruppe, Leitlinien für Transparenz gemäß der Verordnung 2016/679, WP 260 rev. 01, vom EDSA gebilligt, (im Folgenden „Leitlinien der Artikel 29-Gruppe WP260“), Randnummer 11.

Entscheidung des VVA-Designers, das Konto mit vielen verschiedenen Diensten zu verknüpfen, zieht langwierige und komplexe Datenschutzregelungen nach sich. Der Umfang und die Komplexität solcher Datenschutzregelungen sind ein beträchtliches Hindernis für die Einhaltung des Transparenzgrundsatzes.

Beispiel 4:

Ein VVA-Designer verlangt von seinen Nutzern ein Konto für den Zugang zum VVA-Dienst. Dieses Nutzerkonto gilt nicht ausschließlich für den VVA-Dienst, sondern kann auch für andere vom VVA-Designer angebotene Dienste wie E-Mail, Cloud-Speicherung und soziale Medien genutzt werden. Nutzer müssen vor dem Einrichten des Kontos eine 30-seitige Datenschutzerklärung lesen und akzeptieren. Sie enthält Informationen über die Verarbeitung personenbezogener Daten durch alle Dienste, die mit dem Nutzerkonto verknüpft sein können.

Die vom VVA-Designer in diesem Fall bereitgestellten Informationen sind sicherlich nicht zu knapp, und ihre Komplexität geht zulasten der geforderten Transparenz. Damit würde der VVA-Designer die Transparenzanforderungen gemäß Artikel 12 und 13 der DSGVO nicht erfüllen.

56. Es ist zwar allgemein üblich, die geforderten Informationen in schriftlicher Form bereitzustellen, aber nach Maßgabe der DSGVO sind auch „andere Mittel“ zulässig. In Erwägungsgrund 58 der Verordnung heißt es ausdrücklich, dass die Informationen in elektronischer Form, beispielsweise auf einer Website, bereitgestellt werden können. Darüber hinaus sollte bei der Wahl der geeigneten Methode zur Information der betroffenen Personen den jeweiligen Umständen Rechnung getragen werden, etwa der Form, in der der Verantwortliche und die betroffene Person sonst miteinander interagieren.<sup>23</sup> Bei bildschirmlosen Geräten könnte beispielsweise ein leicht verständlicher Link direkt oder per E-Mail bereitgestellt werden. Bereits bestehende Lösungen könnten als Beispiel für die Information dienen, z. B. die Praxis von Callcentern, die anrufende Person über die Aufzeichnung ihres Telefonanrufs zu informieren und sie zu ihren Datenschutzbestimmungen zu leiten. Die bei einem bildschirmlosen Sprachassistenten bestehenden Einschränkungen befreien den Verantwortlichen nicht von der Verpflichtung, die gemäß der DSGVO erforderlichen Informationen bei der Einrichtung des VVA oder der Installation oder Nutzung einer VVA-Anwendung bereitzustellen. Anbieter und Entwickler von VVAs sollten sprachbasierte Schnittstellen entwickeln, um die vorgeschriebene Information zu erleichtern.
57. Für sehbehinderte Nutzer können VVAs von großem Interesse sein, da sie eine Alternative zur Interaktion mit den IT-Diensten darstellen, die üblicherweise auf visuellen Informationen basieren. Gemäß Artikel 12 Absatz 1 der DSGVO kann die Information auch mündlich erteilt werden, jedoch nur auf Verlangen der betroffenen Person und nicht standardmäßig. Wegen der Einschränkungen bildschirmloser VVAs wäre eine automatisierte mündliche Information erforderlich, die in schriftlicher Form ergänzt werden könnte. Wenn die betroffenen Personen mündlich informiert werden, sollten die Verantwortlichen die erforderliche Information in knapper und klarer Form bereitstellen. Außerdem sollten die betroffenen Personen die Möglichkeit haben, die Information erneut abzuhören.<sup>24</sup>

<sup>23</sup> Leitlinien der Artikel 29-Gruppe WP260, Randnummer 19.

<sup>24</sup> Leitlinien der Artikel 29-Gruppe WP 260, Randnummer 21.

58. Wenn außer dem Eigentümer des Geräts noch andere Personen den Sprachassistenten nutzen, ist die Wahl geeigneter Maßnahmen zur Einhaltung der Transparenzanforderungen der DSGVO komplizierter. VVA-Designer müssen prüfen, wie nicht registrierte und zufällige Nutzer ordnungsgemäß darüber informiert werden können, dass ihre personenbezogenen Daten verarbeitet werden. Wenn die Einwilligung der betroffenen Personen die Rechtsgrundlage für die Verarbeitung von Nutzerdaten bildet, müssen die Nutzer ordnungsgemäß informiert werden, damit die Einwilligung gültig ist.<sup>25</sup>
59. Um die Bestimmungen der DSGVO einzuhalten, müssen Verantwortliche einen Weg finden, nicht nur registrierte, sondern auch nicht registrierte und zufällige VVA-Nutzer zu informieren. Diese Nutzer müssen zum frühestmöglichen Zeitpunkt, **spätestens jedoch zum Zeitpunkt der Verarbeitung**, informiert werden. Diese Anforderung könnte ganz besonders schwierig in die Praxis umzusetzen sein.
60. Aber auch unternehmensbedingte Besonderheiten dürfen sich nicht nachteilig auf die betroffenen Personen auswirken. Da viele Interessenträger als Unternehmen weltweit agieren oder für eine bestimmte Geschäftstätigkeit (z. B. Telekommunikation, elektronischen Geschäftsverkehr, Informationstechnologien, Webaktivitäten) bekannt sind, sollte klar sein, auf welche Weise sie einen VVA-Dienst anbieten. Durch angemessene Information müssen die betroffenen Personen erfahren, ob der von ihnen genutzte Sprachassistent mit anderen Verarbeitungsaktivitäten des VVA-Diensteanbieters (z. B. Telekommunikation, elektronischem Geschäftsverkehr, Informationstechnologien oder Webaktivitäten) über die eigentliche Nutzung des Sprachassistenten hinaus verbunden ist.

Beispiel 5:

Zur Nutzung seines Assistenten verlangt ein VVA-Designer, der auch eine Social-Media-Plattform und eine Suchmaschine betreibt, dass der Nutzer ihr Nutzerkonto mit dem Assistenten verknüpft. Durch die Verknüpfung ihres Kontos mit der Nutzung des VVA kann der Designer ihr Profil durch die Nutzung des Assistenten, der installierten Anwendungen (oder Skills), der getätigten Bestellungen usw. verbessern. Die Interaktionen des Assistenten sind somit eine neue Informationsquelle in Verbindung mit einem Nutzer. Der VVA-Designer muss die Nutzer klar darüber informieren, wie ihre Daten für jeden einzelnen Dienst verarbeitet werden, und ihnen die Möglichkeit geben zu entscheiden, ob die Daten für die Erstellung von Profilen verwendet werden dürfen.

## Empfehlungen

61. Wenn Nutzer über die Verarbeitung personenbezogener Daten gemäß der Datenschutzerklärung zum Nutzerkonto informiert werden und das Konto mit anderen unabhängigen Diensten (z. B. E-Mail oder Online-Käufe) verknüpft ist, empfiehlt der EDSA, einen gesonderten Abschnitt über die Verarbeitung personenbezogener Daten durch den VVA in die Datenschutzerklärung aufzunehmen.
62. Die Information für die Nutzer muss der durchgeführten Erhebung und Verarbeitung von Daten genau entsprechen. Jede Sprachprobe enthält Meta-Informationen (z. B. Stressniveau der Sprechenden Person), doch das bedeutet nicht automatisch, dass eine entsprechende

---

<sup>25</sup> Artikel 4 Nummer 11 DSGVO.

Analyse durchgeführt wird. Wichtig ist, dass Verantwortliche die spezifischen Aspekte der von ihnen verarbeiteten Rohdaten transparent machen.

63. Zudem sollte jederzeit erkennbar sein, in welchem Zustand sich der Sprachassistent befindet. Nutzer müssen feststellen können, ob ein VVA in seinem geschlossenen Regelkreis zuhört, und insbesondere, ob er Informationen an sein Backend weiterleitet. Diese Information muss auch für Menschen mit einer Beeinträchtigung wie Farbenblindheit (Daltonismus) oder Taubheit (Anakusis) zugänglich sein. Dabei ist besonders darauf zu achten, dass der VVA ein Nutzungsszenario anbietet, in dem kein Augenkontakt mit dem Gerät erforderlich ist. Sämtliche Rückmeldungen einschließlich Zustandsänderungen sollten zumindest visuell und akustisch verfügbar sein.
64. Besondere Aufmerksamkeit ist geboten, wenn ein Gerät durch Funktionalitäten von Drittanbietern ergänzt werden kann („Apps“ für VVAs). Der Nutzer kann zwar ganz allgemeine Informationen erhalten, wenn sie eine solche Funktionalität hinzufügen will (sofern sie sich selbst dafür entschieden hat), doch bei normaler Nutzung des Geräts können die Grenzen zwischen den verschiedenen Verantwortlichen sehr viel weniger klar sein, d. h. der Nutzer ist möglicherweise gar nicht ausreichend darüber informiert, wie und von wem (und in welchem Umfang) seine Daten bei einer bestimmten Abfrage verarbeitet werden.
65. Alle Informationen über die Verarbeitung auf der Grundlage von erhobenen und aus der Verarbeitung gespeicherter Sprache stammenden Daten müssen für Nutzer gemäß Artikel 12 der DSGVO ebenfalls verfügbar sein.
66. VVA-Designer müssen transparent machen, welche Art von Information ein VVA über seine Umgebung ableiten kann, beispielsweise in Bezug auf andere Personen im Raum, Musik im Hintergrund, die Verarbeitung der Stimme für medizinische Zwecke oder andere Marketinggründe, vorhandene Haustiere usw.

### 3.4 Zweckbegrenzung und Rechtsgrundlage

67. Die Verarbeitung von Sprachanfragen durch VVAs hat einen offensichtlichen Zweck, nämlich die Erledigung der Anfrage. Häufig gibt es jedoch weitere Zwecke, die nicht so offensichtlich sind, wie die Verbesserung des Sprachverständnisses des VVA durch Training des VVA-Modells mit Techniken des maschinellen Lernens. Zu den häufigsten Zwecken der Verarbeitung personenbezogener Daten durch VVAs gehören:
  - ) die Ausführung der Anfragen von Nutzern;
  - ) die Verbesserung des VVA durch Training des Modells maschinellen Lernens sowie die manuelle Überprüfung und Markierung von Transkriptionen;
  - ) die Identifizierung des Nutzers (anhand von Sprachdaten);
  - ) das Erstellen von Nutzerprofilen für personalisierte Inhalte oder Werbung.
68. Aufgrund ihrer Vermittlerrolle und ihrer Konzeption verarbeiten VVAs eine große Bandbreite personenbezogener und anderer Daten. Das ermöglicht die Verarbeitung personenbezogener Daten für viele Zwecke, die über die Beantwortung der Anfragen von Nutzern hinausgehen und die völlig unbemerkt vonstatten gehen können. Durch die Auswertung der über einen Sprachassistenten gesammelten Daten lassen sich Interessen, Zeitpläne, Fahrwege oder Gewohnheiten der Nutzer erkennen oder ableiten. Das könnte die Verarbeitung personenbezogener Daten für unvorhergesehene Zwecke ermöglichen (z. B.

Stimmungsanalyse oder Beurteilung des Gesundheitszustands<sup>26</sup>), was weit über die berechtigten Erwartungen der Nutzer hinausginge.

69. Verantwortliche sollten ihren Zweck bzw. die Zwecke in Bezug auf den Kontext, in dem der VVA genutzt wird, eindeutig angeben, damit die betroffenen Personen es genau verstehen (z. B. durch Einordnung der Zwecke in Kategorien). Nach Artikel 5 Absatz 1 der DSGVO sollen personenbezogene Daten „für festgelegte, eindeutige und legitime Zwecke erhoben werden und dürfen nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden“.

#### 3.4.1 Ausführung von Nutzeranfragen

70. Hauptsächlich werden VVAs für die Umsetzung von Sprachbefehlen genutzt, die von dem Assistenten oder einer dazugehörigen App oder einem Dienst (z. B. einem Musikstreaming-Dienst, einem Kartendienst oder einem elektronischen Schloss) ausgeführt werden sollen. Die Stimme des Nutzers und eventuell weitere Daten (z. B. ihr Standort, wenn sie nach der Route zu einem bestimmten Ziel fragt) könnten daher verarbeitet werden.

Beispiel 6:

Der Fahrer eines intelligenten Fahrzeugs mit einem VVA sucht den Weg zur nächstgelegenen Tankstelle. Der Sprachassistent verarbeitet die Stimme des Nutzers, um den Befehl zu verstehen und die Position des Fahrzeugs zu ermitteln und so die Route zu finden, und sendet diese Information an die intelligente Komponente, damit sie auf dem Bildschirm des Fahrzeugs angezeigt wird.

71. Soweit die Verarbeitung von Sprachbefehlen das Speichern von Informationen oder den Zugriff auf Daten umfasst, die auf Endgeräten des Endnutzers gespeichert sind, sind die Bestimmungen des Artikels 5 Absatz 3 der ePrivacy-Richtlinie einzuhalten. Nach Artikel 5 Absatz 3 darf das Speichern oder der Zugriff grundsätzlich nur mit vorheriger Einwilligung des Endnutzers erfolgen, doch eine Ausnahme von der Einwilligungspflicht ist vorgesehen, „wenn dies „unbedingt erforderlich ist, damit der Anbieter eines Dienstes der Informationsgesellschaft, der vom Teilnehmer oder Nutzer ausdrücklich gewünscht wurde, diesen Dienst zur Verfügung stellen kann“. Soweit die Sprachdaten verarbeitet werden, um Anfragen des Nutzers auszuführen, ist ausnahmsweise keine vorherige Einwilligung erforderlich.
72. Wie bereits erläutert wurde, muss jede Verarbeitung personenbezogener Daten nach dem Speichern oder dem Zugriff auf Daten im Endgerät der Endnutzern eine Rechtsgrundlage gemäß Artikel 6 der DSGVO haben, um rechtmäßig zu sein.
73. Im VVA finden zwei aufeinanderfolgende Verarbeitungsschritte statt. Wie oben ausgeführt wurde, verlangt der erste Schritt Zugang zum VVA (deshalb sind die Bestimmungen des Artikels 5 Absatz 3 der ePrivacy-Richtlinie einzuhalten). Außer den Voraussetzungen nach Artikel 5 Absatz 3 der ePrivacy-Richtlinie verlangt der zweite Schritt auch eine Rechtsgrundlage nach Artikel 6 der DSGVO.
74. Wenn jemand einen Sprachassistenten verwenden möchte, bedeutet das in der Regel, dass der Erstnutzer zunächst ein Nutzerkonto anlegen muss, um den Assistenten aktivieren zu

---

<sup>26</sup> Eoghan Furey, Juanita Blue, „Alexa, Emotion, Privacy and GDPR“, Conference paper, Human Computer Interaction Conference, Juli [2018].

können. Damit entsteht ein Vertragsverhältnis<sup>27</sup> zwischen dem registrierten Nutzer und dem VVA-Verantwortlichen. Inhaltlich und grundsätzlich besteht der Hauptzweck dieses Vertrags in der Nutzung des Assistenten, der Anfragen des Nutzers ausführen soll.

75. Jede Verarbeitung von personenbezogenen Daten, die zur Ausführung einer Nutzeranfrage erforderlich ist, kann sich daher auf die Rechtsgrundlage für die Ausführung des Vertrags stützen.<sup>28</sup> Diese Verarbeitung umfasst insbesondere die Erfassung der Sprachanfrage des Nutzers, ihre Umwandlung in Text, ihre Interpretation, die mit Wissensquellen ausgetauschte Information zur Vorbereitung der Antwort und die anschließende Umwandlung in die endgültige gesprochene Antwort, mit der die Anfrage des Nutzers abgeschlossen wird.
76. Ein Vertrag kann eine Rechtsgrundlage für die Verarbeitung personenbezogener Daten durch maschinelles Lernen (ML) sein, wenn diese für die Erbringung des Dienstes erforderlich ist. Die Verarbeitung personenbezogener Daten durch maschinelles Lernen für andere, nicht notwendige Zwecke wie eine Verbesserung des Dienstes darf sich nicht auf diese Rechtsgrundlage stützen.
77. Was nicht verwechselt werden darf, sind die Rechtsgrundlagen für die Erfüllung des Vertrags und die Einwilligung nach Maßgabe der DSGVO. Die für den Vertragsabschluss erteilte Einwilligung, d. h. die Zustimmung zum Vertrag, ist Teil der Gültigkeit dieses Vertrags und bezieht sich nicht auf die spezifische Bedeutung der Einwilligung gemäß der DSGVO.<sup>29</sup>
78. Wenn ein VVA ohne vorherige Konfigurierung eines Nutzerkontos genutzt werden kann, könnte die Einwilligung eine mögliche Rechtsgrundlage sein.

#### 3.4.2 Verbesserung des VVA durch Training der ML-Systeme und manuelle Überprüfung von Sprache und Transkripten

79. Es gibt unzählige Akzente und Varianten der menschlichen Sprache. Zwar sind alle VVAs direkt einsatzfähig, doch ihre Leistung kann durch Anpassung an die besonderen sprachlichen Merkmale des Nutzers noch verbessert werden. Wie in Abschnitt 2.6 ausgeführt wurde, erfolgt diese Anpassung mit Methoden des maschinellen Lernens und in zwei Prozessen, der Ergänzung des VVA-Trainingsdatensatzes durch neue, von seinen Nutzern stammende Daten und der manuellen Überprüfung der Daten, die für die Ausführung eines Teils der Anfragen verarbeitet wurden.

##### Beispiel 7:

Ein VVA-Nutzer muss dreimal denselben Sprachbefehl geben, weil der Assistent ihn nicht versteht. Die drei Sprachbefehle und die dazugehörigen Transkriptionen werden an menschliche Bearbeiter weitergeleitet, die die Transkriptionen überprüfen und korrigieren. Die Sprachbefehle und die überprüften Transkriptionen werden in den VVA-Trainingsdatensatz aufgenommen, um die Leistung des Assistenten zu verbessern.

---

<sup>27</sup> Sofern „der Vertrag nach dem geltenden nationalen Vertragsrecht gültig ist“, wie es formuliert ist in den Leitlinien 2/2019 für die Verarbeitung personenbezogener Daten gemäß Artikel 6 Absatz 1 Buchstabe b DSGVO im Zusammenhang mit der Erbringung von Online-Diensten für betroffene Personen („Leitlinien 2/2019“), Randnummer 26.

<sup>28</sup> Im Einklang mit den Leitlinien 2/2019, in denen außerdem darauf hingewiesen wird, dass die Stellungnahme 06/2014 nach wie vor relevant ist in Bezug auf Artikel 6 Absatz 1 Buchstabe b der DSGVO (siehe insbesondere Stellungnahme 06/2014, S. 11, 16, 17, 18 und 55).

<sup>29</sup> Siehe Leitlinien 2/2019, Randnummern 18, 19, 20, 21 und 27.

80. Da die im Beispiel beschriebenen Verarbeitungsaktivitäten nicht als (unbedingt) „erforderlich“ „für die Erfüllung eines Vertrags“ im Sinne von Artikel 6 Absatz 1 Buchstabe b der DSGVO anzusehen sind, erfordern sie eine andere Rechtsgrundlage gemäß Artikel 6 der DSGVO. Das ist vor allem darin begründet, dass VVAs von Anfang an funktionstüchtig sind und bereits das ausführen können, was für die Erfüllung des Vertrags (unbedingt) erforderlich ist. Der EDSA sieht in Artikel 6 Absatz 1 Buchstabe b im Allgemeinen keine geeignete Rechtsgrundlage für die Verarbeitung von Daten zur Verbesserung eines Dienstes oder zur Entwicklung neuer Funktionen innerhalb eines bestehenden Dienstes. In den meisten Fällen schließen Nutzer einen Vertrag, um einen bestehenden Dienst in Anspruch nehmen zu können. Zwar können Verbesserungen und Änderungen an einem Dienst routinemäßig in die Vertragsklauseln aufgenommen werden, doch normalerweise ist eine solche Verarbeitung nicht als objektiv notwendig für die Erfüllung des Vertrags mit dem Nutzer anzusehen.

#### 3.4.3 Nutzeridentifizierung<sup>30</sup> (anhand von Sprachdaten)

81. Um Sprachdaten zur Identifizierung der Nutzer verwenden zu können, müssen biometrische Daten im Sinne von Artikel 4 Nummer 14 der DSGVO verarbeitet werden. Daher muss der Verantwortliche außer einer Rechtsgrundlage gemäß Artikel 6 der DSGVO feststellen, welche Ausnahme nach Artikel 9 der DSGVO in Betracht kommt.<sup>31</sup>
82. Von den in Artikel 9 der DSGVO aufgeführten Ausnahmen scheint nur die ausdrückliche Einwilligung der betroffenen Personen für diesen spezifischen Zweck in Betracht zu kommen.
83. Da dieser Zweck die Anwendung der besonderen rechtlichen Regelung nach Artikel 9 der DSGVO erfordert, wird in Abschnitt 3.8 genauer auf die Verarbeitung von besonderen Datenkategorien eingegangen.

#### 3.4.4 Erstellung von Nutzerprofilen für personalisierte Inhalte oder Werbung

84. Wie bereits ausgeführt wurde, haben VVAs Zugang zum Inhalt aller Sprachbefehle, auch wenn diese sich an Dienste von Drittanbietern richten. Durch diesen Zugang hätte der VVA-Designer die Möglichkeit, sehr genaue Nutzerprofile zu erstellen, die für personalisierte Dienste oder Werbung verwendet werden könnten.

Beispiel 8:

Jedes Mal, wenn ein VVA-Nutzer eine Internetsuche durchführt, fügt der Sprachassistent VVA Markierungen ein, die für das Nutzerprofil interessante Themen signalisieren. Die Ergebnisse jeder neuen Suche werden dem bestellenden Nutzer anhand dieser Markierungen präsentiert.

---

<sup>30</sup> Technisch ist zwischen Identifizierung und Verifizierung (Authentifizierung) zu unterscheiden. Die Identifizierung stützt sich auf eine Eins-zu-Viele-Beziehung (1:n) und braucht für die Suche und den Vergleich im Prinzip eine Datenbank, in der mehrere Personen aufgeführt sind. Dagegen handelt es sich bei der Verarbeitung zum Zweck der Verifizierung um einen Eins-zu-Eins-Vergleich (1:1), bei dem durch biometrischen Abgleich überprüft und bestätigt wird, ob es sich um die Person handelt, von der die biometrischen Daten stammen. Nach Kenntnis des EDSA basieren die gängigen VVAs ausschließlich auf Techniken der Sprecheridentifizierung.

<sup>31</sup> Im Sinne der DSGVO reicht die Art der Daten allein nicht immer aus, um festzustellen, ob es sich um besondere Kategorien handelt, „da Lichtbilder nur dann von der Definition des Begriffs ‚biometrische Daten‘ erfasst werden, wenn sie mit speziellen technischen Mitteln verarbeitet werden, die die eindeutige Identifizierung oder Authentifizierung einer natürlichen Person ermöglichen“ (Erwägungsgrund 51). Das Gleiche gilt für Sprache.

Beispiel 9:

Jedes Mal, wenn ein VVA-Nutzer bei einem Online-Anbieter einkauft, speichert der VVA eine Aufzeichnung der Bestellung. Der VVA-Anbieter ermöglicht es Dritten, den VVA-Nutzer mit gezielter Werbung auf der Grundlage früherer Käufe anzusprechen.

85. Die Personalisierung von Inhalten stellt möglicherweise (aber nicht immer) ein inhärentes und zu erwartendes Element eines VVA dar. Ob diese Verarbeitung als untrennbarer Bestandteil eines VVA-Dienstes anzusehen ist, hängt von der Art des erbrachten Dienstes und den Erwartungen der durchschnittlichen betroffenen Person ab, nicht nur in Bezug auf die Nutzungsbedingungen, sondern auch darauf, wie der Dienst bei Nutzern beworben wird und ob er ohne Personalisierung erbracht werden kann.<sup>32</sup>
86. Wenn die Personalisierung im Rahmen eines Vertragsverhältnisses und als Teil eines Dienstes erfolgt, der ausdrücklich vom Endnutzer angefordert wird (und die Verarbeitung auf das für die Erbringung dieses Dienstes unbedingt erforderliche Maß beschränkt ist), kann diese Verarbeitung auf der Grundlage von Artikel 6 Absatz 1 Buchstabe b der DSGVO erfolgen.
87. Wenn die Verarbeitung nicht unbedingt „erforderlich“ „für die Erfüllung eines Vertrags“ im Sinne von Artikel 6 Absatz 1 Buchstabe b der DSGVO ist, muss der VVA-Anbieter grundsätzlich die Einwilligung der betroffenen Person einholen. Da nach Artikel 5 Absatz 3 der ePrivacy-Richtlinie die Speicherung oder der Zugang zu Informationen eine Einwilligung voraussetzt (siehe dazu Randnummern 28 und 29), wird die Einwilligung nach Artikel 6 Absatz 1 Buchstabe a der DSGVO grundsätzlich auch die geeignete Rechtsgrundlage für die Verarbeitung personenbezogener Daten im Anschluss an diese Vorgänge sein, denn das Vertrauen auf berechnete Interessen könnte in manchen Fällen dazu führen, dass das zusätzliche Schutzniveau nach Artikel 5 Absatz 3 der ePrivacy-Richtlinie nicht gewahrt wird.
88. Hinsichtlich der Erstellung von Nutzerprofilen für Werbezwecke ist darauf hinzuweisen, dass dieser Zweck niemals als ein von den Endnutzern ausdrücklich gewünschter Dienst anzusehen ist. Deshalb ist bei einer Verarbeitung von Daten für diesen Zweck die Einwilligung der Nutzer systematisch einzuholen.

## Empfehlungen

89. Nutzer müssen über den Zweck der Verarbeitung personenbezogener Daten informiert werden, und dieser Zweck sollte ihren Erwartungen an das von ihnen gekaufte Gerät entsprechen. Der Zweck eines Sprachassistenten besteht aus Sicht des Nutzers eindeutig darin, dass seine Stimme ausschließlich verarbeitet wird, um seine Anfragen zu interpretieren und entsprechend zu reagieren (durch Beantwortung einer Frage oder andere Reaktionen wie die Fernsteuerung eines Lichtschalters).
90. Wenn die Verarbeitung personenbezogener Daten eine Einwilligung voraussetzt, bedeutet das, dass diese Einwilligung „für ,einen oder mehrere bestimmte‘ Zwecke erteilt werden muss und dass eine betroffene Person in Bezug auf jeden dieser Zwecke eine Wahlmöglichkeit haben muss“. Und weiter, „dass ein Verantwortlicher, der die Einwilligung für verschiedene unterschiedliche Zwecke einholen möchte, für jeden Zweck ein gesondertes Opt-in bereitstellen sollte, damit die Nutzer für bestimmte Zwecke eine konkrete Einwilligung erteilen können“.<sup>33</sup>

<sup>32</sup> Siehe auch Leitlinien 2/2019, Randnummer 57.

<sup>33</sup> Siehe EDSA, Leitlinien 05/2020 zur Einwilligung gemäß Verordnung 2016/679, angenommen am 4. Mai 2020, Abschnitt 3.2.

Beispielsweise sollten Nutzer die Möglichkeit haben, für die manuelle Überprüfung und Markierung von Transkriptionen oder für die Verwendung ihrer Sprachdaten zur Identifizierung/Authentifizierung jeweils gesondert ihre Einwilligung zu erteilen oder zu verweigern (siehe Abschnitt 3.7).

### 3.5 Verarbeitung personenbezogener Daten von Kindern

91. Auch Kinder können auch mit Sprachassistenten interagieren oder eigene Profile erstellen, die mit denen der Erwachsenen verbunden sind. Manche VVAs sind in speziell für Kinder bestimmte Geräte integriert.
92. Wenn ein Vertrag die Rechtsgrundlage für die Verarbeitung bildet, sind die Bedingungen für die Verarbeitung personenbezogener Daten von Kindern durch das nationale Vertragsrecht geregelt.
93. Wenn die Einwilligung die Rechtsgrundlage für die Verarbeitung bildet, ist nach Artikel 8 Absatz 1 der DSGVO die Verarbeitung personenbezogener Daten von Kindern nur rechtmäßig, *„wenn das Kind das sechzehnte Lebensjahr vollendet hat. Hat das Kind noch nicht das sechzehnte Lebensjahr vollendet, so ist diese Verarbeitung nur rechtmäßig, sofern und soweit diese Einwilligung durch den Träger der elterlichen Verantwortung für das Kind oder mit dessen Zustimmung erteilt wird“*. In diesem Fall muss daher zur Einhaltung der DSGVO die ausdrückliche Erlaubnis von Eltern oder Vormündern eingeholt werden, um die Daten von Kindern erheben, verarbeiten und speichern zu dürfen (Sprache, Transkripte usw.).
94. Kinderschutz-Apps sind bis zu einem gewissen Grad verfügbar, aber in ihrer derzeitigen Form sind sie nicht benutzerfreundlich (beispielsweise ist die Anmeldung zu einem weiteren Dienst erforderlich), oder sie haben nur begrenzte Kapazitäten. Die Verantwortlichen müssen in die Entwicklung von Instrumenten investieren, die es Eltern und Vormündern ermöglichen, die Nutzung von VVAs durch ihre Kinder zu kontrollieren.

### 3.6 Vorratsdatenspeicherung

95. VVAs verarbeiten und generieren ganz unterschiedliche personenbezogene Daten wie Sprache, Transkriptionen, Metadaten oder Systemprotokolle. Derartige Daten können für vielfältige Zwecke wie die Bereitstellung eines Dienstes, eine bessere Verarbeitung natürlicher Sprache (NLP), Personalisierung oder wissenschaftliche Forschung verarbeitet werden. Nach dem in der DSGVO verankerten Grundsatz der Speicherbegrenzung dürfen personenbezogene Daten nur so lange gespeichert werden, wie es für die Zwecke, für die sie verarbeitet werden, erforderlich ist. Deshalb sollten die Fristen für die Datenspeicherung an den jeweiligen Zweck gebunden sein. Anbieter von VVA-Diensten oder Dritte, die über den Sprachassistenten Dienste anbieten, sollten die maximale Speicherdauer für jeden Datensatz und jeden Zweck bewerten.
96. Der Grundsatz der Datenminimierung ist eng verbunden mit dem Grundsatz der Speicherbegrenzung. Verantwortliche müssen nicht nur die Dauer der Datenspeicherung, sondern auch Art und Menge der gespeicherten Daten begrenzen.
97. Verantwortliche sollten sich unter anderem folgende Fragen stellen: Müssen alle Sprachaufzeichnungen oder Transkriptionen gespeichert werden, um Zweck X zu erreichen? Müssen die Sprachdaten auch noch gespeichert werden, nachdem die Transkription gespeichert ist? Und wenn ja, zu welchem Zweck? Wie lange werden Sprach- oder Transkriptionsdaten für den jeweiligen Zweck benötigt? Die Antworten auf diese und ähnliche

Fragen bestimmen die Speicherfristen, die in den Informationen für betroffene Personen genannt werden sollten.

98. Einige VVAs speichern standardmäßig personenbezogene Daten wie Voice Snippets oder Transkriptionen für einen unbestimmten Zeitraum und geben den Nutzern die Möglichkeit, solche Daten zu löschen. Das zeitlich unbegrenzte Speichern personenbezogener Daten widerspricht dem Grundsatz der Speicherbegrenzung. Wenn der Verantwortliche betroffenen Personen die Möglichkeit gibt, ihre personenbezogenen Daten zu löschen, entbindet ihn das nicht von der Pflicht zur Festlegung und Durchsetzung einer Strategie hinsichtlich der Vorratsdatenspeicherung.
99. Beim VVA-Design sind Kontrollmöglichkeiten für die Nutzer zur Löschung ihrer personenbezogenen Daten in ihren Geräten und in allen entfernten Speichersystemen zu berücksichtigen. Solche Kontrollen sind möglicherweise erforderlich, um verschiedene Nutzeranfragen auszuführen, z. B. einen Antrag auf Löschung oder den Widerruf einer zuvor erteilten Einwilligung. Bei manchen VVAs wurde diese Anforderung im Design nicht berücksichtigt.<sup>34</sup>
100. Wie in anderen Kontexten sind Verantwortliche möglicherweise gesetzlich zur Speicherung personenbezogener Daten verpflichtet als Nachweis für einen für einen Nutzer bereitgestellten Dienst. Auf dieser Grundlage darf der Verantwortliche personenbezogene Daten speichern. Die Speicherung muss sich aber auf die für die Einhaltung der gesetzlichen Verpflichtung unbedingt notwendigen Daten und den notwendigen Zeitraum beschränken. Natürlich dürfen die zur Erfüllung einer rechtlichen Verpflichtung gespeicherten Daten ohne eine Rechtsgrundlage gemäß Artikel 6 der DSGVO nicht für andere Zwecke verwendet werden.

Beispiel 10:

Ein Nutzer kauft mittels eines Sprachassistenten online ein Fernsehgerät. Selbst wenn der Nutzer später die Löschung seiner Daten beantragt, könnte der Anbieter oder Entwickler des VVA einige Daten aufgrund seiner steuerrechtlichen Pflicht zur Aufbewahrung von Kaufbelegen speichern. Die zu diesem Zweck gespeicherten Daten dürfen jedoch nicht über das zur Erfüllung der rechtlichen Verpflichtung erforderliche Mindestmaß hinausgehen, und ohne eine Rechtsgrundlage gemäß Artikel 6 der DSGVO dürfen sie nicht für andere Zwecke verarbeitet werden.

101. Wie in Abschnitt 2 ausgeführt wurde, verbessert sich die Spracherkennungsfähigkeit von VVAs durch das Training von Systemen maschinellen Lernens anhand von Nutzerdaten. Wenn Nutzer keine Einwilligung zur Verwendung ihrer Daten für diesen Zweck erteilen oder ihre Einwilligung widerrufen, dürfen ihre Daten nicht für weitere Modelle verwendet werden und müssen vom Verantwortlichen gelöscht werden, sofern kein anderer Zweck eine Vorratspeicherung rechtfertigt. Es gibt jedoch Hinweise auf mögliche Re-Identifizierungsrisiken bei einigen Modellen maschinellen Lernens.<sup>35</sup>

---

<sup>34</sup> Siehe schriftliche Antwort von Amazon vom 28. Juni 2019 auf das Schreiben von US-Senator Christopher Coons: [https://www.coons.senate.gov/imo/media/doc/Amazon%20Senator%20Coons Response%20Letter 6.28.19 \[3\].pdf](https://www.coons.senate.gov/imo/media/doc/Amazon%20Senator%20Coons%20Response%20Letter%206.28.19%20[3].pdf).

<sup>35</sup> Veale Michael, Binns Reuben, Edwards Lilian (2018), „[Algorithms that remember: model inversion attacks and data protection law](#)“, Philosophical Transactions of the Royal Society A, 37620180083, doi: 10.1098/rsta.2018.0083.

102. Verantwortliche und Auftragsverarbeiter sollten Modelle verwenden, die sie nicht daran hindern, die Verarbeitung einzustellen, wenn eine Person ihre Einwilligung widerruft, und sie sollten keine Modelle verwenden, die sie daran hindern, betroffenen Personen die Wahrnehmung ihrer Rechte zu erleichtern. Verantwortliche und Auftragsverarbeiter sollten Maßnahmen zur Risikominderung ergreifen, um das Risiko einer Re-Identifizierung auf einen akzeptablen Schwellenwert zu senken.
103. Falls der Nutzer seine Einwilligung widerruft, dürfen seine Daten für das weitere Training des Modells nicht mehr verwendet werden. Das bis dahin unter Verwendung dieser Daten trainierte Modell muss jedoch nicht gelöscht werden. Nach Kenntnis des EDSA gibt es aber Hinweise darauf, dass bei manchen Modellen maschinellen Lernens personenbezogene Daten einfließen könnten. Insbesondere haben verschiedene Studien gezeigt, dass sowohl Rekonstruktionen als auch Inferenzangriffe auf Mitgliederdaten möglich sind, durch die sich die Angreifer Informationen über Einzelpersonen verschaffen könnten.<sup>36</sup> Verantwortliche und Auftragsverarbeiter sollten daher Maßnahmen zur Risikominderung ergreifen, um das Risiko einer Re-Identifizierung auf einen akzeptablen Schwellenwert zu senken und um sicherzustellen, dass die verwendeten Modelle keine personenbezogenen Daten enthalten.
104. Betroffene Personen dürfen nicht gedrängt werden, einer unbegrenzten Speicherung ihrer Daten zuzustimmen. Die Löschung gespeicherter Sprachdaten oder Transkriptionen könnte zwar Auswirkungen auf die Dienstleistung haben, doch solche Auswirkungen sollten den Nutzern verständlich und überprüfbar erklärt werden. Die Anbieter von VVA-Diensten sollten allgemeine Aussagen vermeiden, wonach mit einer Verschlechterung des Dienstes zu rechnen sei, wenn personenbezogene Daten gelöscht werden.
105. Die Anonymisierung von Sprachaufzeichnungen stellt eine besondere Herausforderung dar, da Nutzer anhand des Inhalts der Nachricht und der Merkmale ihrer Stimme identifiziert werden können. Es wird an Techniken geforscht<sup>37</sup>, die es ermöglichen, situationsbezogene Informationen wie Hintergrundgeräusche zu entfernen und die Stimme zu anonymisieren.

## Empfehlungen

106. Aus Sicht der Nutzer besteht der Hauptzweck der Verarbeitung ihrer Daten darin, dass sie Anfragen stellen können und Antworten bekommen und/oder Aktionen wie das Abspielen von Musik oder das Ein- oder Ausschalten von Lampen ausgelöst werden. Nachdem eine Anfrage beantwortet oder ein Befehl ausgeführt wurde, müssen die personenbezogenen Daten gelöscht werden, außer wenn der VVA-Designer oder -Entwickler über eine gültige Rechtsgrundlage verfügt, um diese Daten für einen bestimmten Zweck zu speichern.
107. Bevor über eine Anonymisierung zur Einhaltung des Grundsatzes der Speicherbegrenzung nachgedacht wird, sollten VVA-Anbieter und -Entwickler prüfen, ob die Stimme durch das Anonymisierungsverfahren tatsächlich unidentifizierbar wird.
108. Standardkonfigurationen sollten diesen Anforderungen Rechnung tragen durch Beschränkung auf das absolute Minimum an gespeicherten Nutzerinformationen. Wenn diese Optionen als Teil eines Setup-Assistenten angeboten werden, muss sich dies in der Standardeinstellung

---

<sup>36</sup> N. Carlini et al., „Extracting Training Data from Large Language Models“, Dezember 2020.

<sup>37</sup> Siehe beispielsweise VoicePrivacy (<https://www.voiceprivacychallenge.org>), eine Initiative zur Entwicklung von Lösungen zum Schutz der Privatsphäre in der Sprachtechnologie.

Siehe auch Open-Source-Instrumente zur Stimmenanonymisierung, entwickelt vom H2020 Forschungs- und Innovationsprojekt COMPRISE: [https://gitlab.inria.fr/comprise/voice\\_transformation](https://gitlab.inria.fr/comprise/voice_transformation).

wiederfinden, und alle Optionen sollten gleichwertig ohne visuelle Diskriminierung dargestellt werden.

109. Stellt der VVA-Anbieter oder -Entwickler bei der Überprüfung fest, dass infolge einer fehlerhaften Aktivierung eine Aufzeichnung vorgenommen wurde, so sind die Aufzeichnung und alle damit verbundenen Daten unverzüglich zu löschen, und sie dürfen für keinen Zweck verwendet werden.

### 3.7 Sicherheit

110. Damit personenbezogene Daten sicher verarbeitet werden, muss der Sprachassistent ihre Vertraulichkeit, Integrität und Verfügbarkeit schützen. Abgesehen von den Risiken, die sich aus Elementen des VVA-Ökosystems ergeben, kommt es durch die Benutzung von Sprache als Kommunikationsmittel zu weiteren Sicherheitsrisiken.
111. VVAs sind Mehrbenutzer. Sie können mehr als einen registrierten Nutzer zulassen, und jeder in ihrer Umgebung kann Befehle ausgeben und ihre Dienste nutzen. Jeder VVA-Dienst, der Vertraulichkeit verlangt, verfügt über einen Zugangskontrollmechanismus und eine Nutzer-Authentifizierung. Ohne Zugangskontrolle könnte jede Person, die dem VVA Sprachbefehle erteilen kann, auf die personenbezogenen Daten der Nutzer zugreifen, sie ändern oder löschen (z. B. eingegangene Nachrichten, die Adresse oder Termine der Nutzer abfragen). Sprachbefehle können einem VVA auch erteilt werden, ohne dass sich die Person selbst in der Nähe befindet, da er manipulierbar ist, beispielsweise durch Signalübertragung<sup>38</sup> (z. B. durch Radio oder Fernsehen). Manche der bekannten Methoden zur ferngesteuerten Erteilung von Befehlen an VVAs wie Laser<sup>39</sup> oder (unhörbare) Ultraschallwellen<sup>40</sup> sind für menschliche Sinnesorgane gar nicht wahrnehmbar.
112. Zur Nutzerauthentifizierung können einer oder mehrere der folgenden Faktoren verwendet werden: etwas Ihnen Bekanntes (z. B. ein Passwort), etwas in Ihrem Besitz (z. B. eine Chipkarte) oder ein persönliches Merkmal (z. B. ein Stimmabdruck). Eine genauere Betrachtung dieser Authentifizierungsfaktoren im Zusammenhang mit dem Sprachassistenten ergibt Folgendes:
  - ) Die Authentifizierung durch etwas, was dem Nutzer bekannt ist, ist problematisch. Da das geheime Kennwort, mit dem Nutzer ihre Identität beweisen sollen, laut gesprochen werden muss, kann jeder in der Nähe mithören. Der Kommunikationskanal des Sprachassistenten ist die Umgebungsluft, d. h. er kann nicht verstärkt werden wie herkömmliche Kanäle (z. B. durch Beschränkung des Zugangs zum Kanal oder Verschlüsselung seines Inhalts).
  - ) Die Authentifizierung durch einen im Besitz des Nutzers befindlichen Gegenstand würde von den VVA-Diensteanbietern verlangen, „Token“ zu erstellen, zu verbreiten und zu verwalten, die als Identitätsnachweis verwendet werden können.

---

<sup>38</sup> X. Yuan et al., „All Your Alexa Are Belong to Us: A Remote Voice Control Attack against Echo“, 2018, IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, Vereinigte Arabische Emirate, 2018, S. 1-6, doi: 10.1109/GLOCOM.2018.8647762.

<sup>39</sup> Siehe beispielsweise: <https://lightcommands.com>.

<sup>40</sup> Siehe beispielsweise: <https://surfingattack.github.io>.

- ) Die Authentifizierung anhand eines persönlichen Merkmals des Nutzers impliziert die Verwendung biometrischer Daten zur eindeutigen Identifizierung einer natürlichen Person (siehe Abschnitt 3.7).
113. VVA-Nutzerkonten sind mit den Geräten verbunden, in denen der Dienst erbracht wird. Häufig wird das Konto für die Verwaltung des VVA auch für die Verwaltung anderer Dienste genutzt. Beispielsweise können Eigentümer eines Android-Mobiltelefons und eines Home Speaker von Google ihr Google-Konto mit beiden Geräten verbinden, was die meisten vermutlich auch tun. Die meisten VVAs verlangen oder bieten keinen Identifizierungs- oder Authentifizierungsmechanismus, wenn ein Gerät, das einen VVA-Dienst bereitstellt, lediglich ein Nutzerkonto hat.
  114. Wenn mehrere Nutzerkonten mit dem Gerät verbunden sind, bieten einige VVAs eine einfache optionale Zugangskontrolle in Form einer PIN ohne wirkliche Benutzerauthentifizierung an. Bei anderen VVAs besteht die Möglichkeit, Stimmabdrücke zur Identifizierung zu verwenden.
  115. Die Identifizierung oder Authentifizierung der Nutzer wird zwar nicht für den Zugang zu allen VVA-Diensten benötigt, aber für einige Dienste ist sie auf jeden Fall erforderlich. Ohne einen Identifizierungs- oder Authentifizierungsmechanismus könnte jeder auf die Daten von anderen Nutzern zugreifen und sie nach Belieben ändern oder löschen. Beispielsweise könnte jeder in der Nähe eines intelligenten Lautsprechers Wiedergabelisten von anderen Nutzern aus dem Musikstreaming-Dienst, Befehle aus dem Befehlsverlauf oder Kontakte aus der Kontaktliste löschen.
  116. Die meisten VVAs vertrauen blind auf ihre lokalen Netzwerke. Jedes falsch verwendete Gerät im selben Netzwerk könnte ohne Wissen oder Zustimmung des Nutzers die Einstellungen des intelligenten Lautsprechers verändern oder die Installation von Schadsoftware oder von gefälschten Apps/Skills erlauben.<sup>41</sup>
  117. VVAs haben wie jede andere Software Schwachstellen. Aufgrund der Konzentration im VVA-Markt<sup>42</sup> kann sich jede Schwachstelle auf Millionen von VVA-Nutzer auswirken. Wie VVAs bisher konzipiert sind, senden sie erst auf den Aktivierungsbefehl hin Informationen an den Cloud-Dienst zur Spracherkennung. Schwachstellen in der Software könnten es jedoch einem Angreifer ermöglichen, die Vorkehrungen und Sicherheitsmaßnahmen des VVA zu umgehen. So wäre es beispielsweise möglich, eine Kopie aller an die VVA-Cloud gesendeten Daten zu erhalten und an einen vom Angreifer kontrollierten Server weiterzuleiten.
  118. Anhand von Daten, die rechtmäßig von VVAs verarbeitet oder abgeleitet werden, lässt sich ein recht genaues Profil ihrer Nutzer erstellen, da VVAs den Standort, die Beziehungen und die Interessen ihrer Nutzer kennen oder herleiten können. Immer mehr VVAs sind in Wohnungen und Smartphones der Nutzer vorhanden. Dadurch erhöht sich das Risiko einer massenhaften Überwachung und Profilerstellung. Die Sicherheitsmaßnahmen zum Schutz der Daten sowohl bei der Übertragung als auch im Ruhezustand in den Geräten und in der Cloud müssen diesen Risiken Rechnung tragen.

---

<sup>41</sup> Siehe beispielsweise Deepak Kumar et al., *Skill Squatting Attacks on Amazon Alexa*, USENIX Security Symposium, August 2018, <https://www.usenix.org/conference/usenixsecurity18/presentation/kumar>; Security Research Labs, *Smart Spies: Alexa and Google Home expose users to vishing and eavesdropping*, November 2019, <https://srlabs.de/bites/smart-spies>.

<sup>42</sup> Den VVA-Markt teilen sich zurzeit weniger als ein Dutzend Diensteanbieter.

119. Wenn die zunehmende Verwendung von Sprachassistenten auf ungleichgewichtige Zugriffsrechte von Strafverfolgungsbehörden stößt, könnte das einen Abschreckungseffekt haben mit nachteiligen Folgen für Grundrechte wie die Meinungsfreiheit.
120. Strafverfolgungsbehörden sowohl innerhalb<sup>43</sup> als auch außerhalb<sup>44</sup> der EU haben bereits ihr Interesse an einem Zugriff auf von VVAs erfasste Voice Snippets bekundet. Der Zugriff auf von VVAs verarbeitete oder abgeleitete Daten in der EU muss mit den geltenden europäischen Rechtsvorschriften für den Datenschutz und den Schutz der Privatsphäre in Einklang stehen. Wenn Mitgliedstaaten besondere Rechtsvorschriften zur Beschränkung der Grundrechte auf Privatsphäre und Datenschutz erlassen wollen, müssen diese Vorschriften mit Artikel 23 der DSGVO in Einklang stehen.<sup>45</sup>
121. Die manuelle Überprüfung von Sprachaufzeichnungen und dazugehörigen Daten zur Verbesserung der Qualität des VVA-Dienstes ist bei VVA-Anbietern gängige Praxis. Da es sich um sensible Daten handelt, die von Menschen verarbeitet werden, und dieser Vorgang häufig an einen Auftragsverarbeiter vergeben wird, müssen unbedingt angemessene Sicherheitsmaßnahmen getroffen werden.

## Empfehlungen

122. VVA-Designer und Anwendungsentwickler sollten den Nutzern sichere Authentifizierungsverfahren nach dem neuesten Stand der Technik bieten.
123. Menschliche Überprüfer dürfen nur die unbedingt erforderlichen pseudonymisierten Daten erhalten. In den rechtlichen Vereinbarungen über die Überprüfung muss jede Verarbeitung, die zur Identifizierung der betroffenen Person führen könnte, ausdrücklich verboten sein.
124. Wenn ein Notruf als Dienst über den VVA bereitgestellt wird, muss eine stabile Uptime<sup>46</sup> gewährleistet sein.

## 3.8 Verarbeitung besonderer Datenkategorien

125. Wie bereits ausgeführt wurde, haben VVAs Zugang zu intimen Informationen, die nach Artikel 9 der DSGVO geschützt sein können (siehe Abschnitt 3.7.1), ebenso wie biometrische Daten (siehe Abschnitt 3.7.2). Deshalb müssen Designer und Entwickler von VVAs sorgfältig feststellen, in welchen Fällen es um die Verarbeitung besonderer Datenkategorien geht.

### 3.8.1 Allgemeine Überlegungen zur Verarbeitung besonderer Datenkategorien

126. VVAs können besondere Datenkategorien unter verschiedenen Voraussetzungen verarbeiten:
  - ) im Rahmen ihrer eigenen Dienste, z. B. bei der Verwaltung von Arztterminen im Terminkalender von Nutzern;
  - ) in ihrer Funktion als Schnittstelle für Dienste von Dritten. In dem Fall verarbeiten VVA-Anbieter den Inhalt der Befehle. Je nach Art des vom Nutzer angeforderten Dienstes können VVA-Anbieter besondere Datenkategorien verarbeiten. Eine Nutzerin könnte

<sup>43</sup> Siehe beispielsweise <https://www.ft.com/content/ad765972-87a2-11e9-a028-86cea8523dc2>.

<sup>44</sup> Siehe beispielsweise <https://cdt.org/insights/alexa-is-law-enforcement-listening>.

<sup>45</sup> Siehe auch EDPB, Guidelines 10/2020 on restrictions under Article 23 GDPR (Leitlinien zu den Beschränkungen gemäß Artikel 23 DSGVO).

<sup>46</sup> Die Zeit, in der ein Gerät oder ein Dienst unbeaufsichtigt laufen kann, ohne abzustürzen oder aus administrativen Gründen oder zur Wartung einen Neustart zu verlangen.

beispielsweise dem Sprachassistenten befehlen, eine Drittanwendung zu nutzen, um ihren Eispung zu verfolgen<sup>47</sup>;

- ) wenn Sprachdaten nur verwendet werden, um den Nutzer eindeutig zu identifizieren (siehe unten).

### 3.8.2 Spezielle Überlegungen zur Verarbeitung biometrischer Daten

127. Manche VVAs können die Nutzer allein anhand ihrer Stimme eindeutig identifizieren. Dabei handelt es sich um ein Modell zur Spracherkennung. In der Anmeldephase der Spracherkennung verarbeitet der VVA die Stimme des Nutzers, um ein Sprachmodell (oder Voiceprint) zu erstellen. Während der regelmäßigen Nutzung kann der VVA auch das Sprachmodell von anderen Nutzern berechnen und es mit den angemeldeten Modellen vergleichen, um den Nutzer, der einen Befehl ausgeführt hat, eindeutig zu identifizieren.

**Beispiel 11:**

Eine Gruppe von Nutzern richtet einen VVA ein, um ein Modell zur Spracherkennung zu verwenden. Anschließend meldet jede Person ihr Sprachmodell an.

Später verlangt einer der Nutzer vom VVA Zugang zu den Sitzungsterminen in seinem Kalender. Da der Zugriff auf den Kalender eine Nutzeridentifizierung erfordert, extrahiert der VVA das Modell aus der Stimme der Anfrage, berechnet ihr Sprachmodell und prüft, ob es mit einer der angemeldeten Personen übereinstimmt und ob diese Person auf den Kalender zugreifen darf.

128. In dem genannten Beispiel müssen zur Erkennung der Stimme eines Nutzers auf der Grundlage eines Sprachmodells besondere Kategorien personenbezogener Daten im Sinne von Artikel 9 der DSGVO (Verarbeitung biometrischer Daten zur eindeutigen Identifizierung einer natürlichen Person) verarbeitet werden.<sup>48</sup> Die im Beispiel erforderliche Verarbeitung biometrischer Daten zur Nutzeridentifizierung setzt die ausdrückliche Einwilligung der betroffenen Person/en voraus (Artikel 9 Absatz 2 Buchstabe a DSGVO). Daher müssen Verantwortliche, wenn sie die Einwilligung der Nutzer einholen, die Bedingungen des Artikels 7 gemäß den Erläuterungen in Erwägungsgrund 32 der DSGVO einhalten und im Hinblick auf die Freiwilligkeit der Einwilligung eine alternative Methode zur Identifizierung anhand biometrischer Daten anbieten.

129. Bei der Verwendung von Sprachdaten zur biometrischen Identifizierung oder Authentifizierung müssen die Verantwortlichen transparent machen, wo eine biometrische Identifizierung erfolgt und wie Voiceprints (biometrische Modelle) gespeichert und geräteübergreifend verbreitet werden. Um diese Transparenzanforderung zu erfüllen, empfiehlt der EDSA die Beantwortung folgender Fragen:

- ) Wird durch die Aktivierung der Spracherkennung auf einem Gerät diese Funktion automatisch auch auf allen anderen Geräten aktiviert, die mit demselben Konto verbunden sind?

---

<sup>47</sup> Ein Beispiel für ein verfügbares Produkt: <https://www.amazon.com/Ethan-Fan-Ovulation-Period-Tracker/dp/B07CRLSHKY>.

- ) Überträgt sich die Aktivierung der Spracherkennung über die Infrastruktur des VVA-Verantwortlichen auf Geräte anderer Nutzer?
  - ) Wo werden biometrische Modelle erzeugt, gespeichert und verglichen?
  - ) Sind biometrische Modelle für VVA-Anbieter, Entwickler oder andere Interessierte zugänglich?
130. Wenn ein registrierter Nutzer die VVAs zur Identifizierung der Stimmen seiner Nutzer konfiguriert, werden auch die Stimmen von nicht registrierten und zufälligen Nutzern zur eindeutigen Identifizierung verarbeitet.
131. Damit die Stimme des registrierten Sprechers erkannt wird, muss sie natürlich mit den Stimmen anderer Personen in der Nähe des Assistenten verglichen werden. Deshalb wird die in Sprachassistenten implementierte Funktionalität der Spracherkennung möglicherweise verlangen, dass die Stimmbiometrie von Personen, deren Stimme in dem Haushalt zu hören ist, aufgezeichnet wird, damit die stimmlichen Merkmale des Nutzers von denen der Person, die erkannt werden möchte, unterschieden werden können. Durch die biometrische Identifizierung werden daher möglicherweise die Daten einer nicht informierten Person biometrisch verarbeitet, weil ihr Modell registriert und mit dem des Nutzers verglichen wird, der erkannt werden möchte.
132. Um eine Erfassung biometrischer Daten ohne Wissen der betroffenen Personen zu vermeiden und gleichzeitig die Erkennung des Nutzers durch den Sprachassistenten zu ermöglichen, sollten vorrangig Lösungen angestrebt werden, die sich allein auf die Daten des Nutzers stützen. Konkret bedeutet das, dass die biometrische Erkennung nur auf Initiative des Nutzers und nicht durch die laufende Analyse der vom Sprachassistenten gehörten Stimmen aktiviert wird. So könnte beispielsweise durch ein spezielles Stichwort der anwesenden Personen oder eine an sie gerichtete Frage ihre Zustimmung eingeholt werden, bevor mit der Verarbeitung biometrischer Daten begonnen wird. Der Nutzer könnte „Identifizierung“ sagen, oder der Sprachassistent könnte fragen „Möchten Sie identifiziert werden?“ und auf Zustimmung warten, bevor die biometrische Verarbeitung aktiviert wird.

Beispiel 12:

Wenn der Nutzer eine biometrische Authentifizierung für den Zugriff auf bestimmte geschützte Daten wie ihr Bankkonto wünscht, könnte der Sprachassistent die Sprecherverifizierung beispielsweise erst aktivieren, wenn die Anwendung gestartet wird, und die Identität auf diese Weise verifizieren.

## Empfehlungen

133. Sprachmodelle sollten ausschließlich auf dem lokalen Gerät und nicht auf entfernten Servern erzeugt, gespeichert und verglichen werden.
134. In Anbetracht der Sensitivität von Voiceprints sollten Normen wie die ISO/IEC 24745 und Techniken zum Schutz biometrischer Modelle<sup>49</sup> umfassend angewandt werden.

<sup>49</sup> Siehe beispielsweise:

Jain, Anil & Nandakumar, Karthik & Nagar, Abhishek (2008), „*Biometric Template Security*“, EURASIP Journal on Advances in Signal Processing, 2008, 10.1155/2008/579416.

135. Wenn ein VVA die sprachbasierte biometrische Identifizierung anwendet, sollte der VVA-Anbieter:
- ) eine so präzise Identifizierung gewährleisten, dass personenbezogene Daten zuverlässig den richtigen Personen zugeordnet werden können;
  - ) für alle Gruppen von Nutzern die gleiche Präzision gewährleisten und sich vergewissern, dass es keine nennenswerte Verzerrung gegenüber anderen Bevölkerungsgruppen gibt.

### 3.9 Datenminimierung

136. Verantwortliche müssen die Menge der direkt oder indirekt erhobenen und durch Verarbeitung und Analyse gewonnenen Daten minimieren und dürfen beispielsweise keine Analyse der Stimme oder anderer akustischer Informationen des Nutzers durchführen, um Informationen über ihren psychischen Zustand, eventuelle Krankheiten oder ihre Lebensumstände abzuleiten.
137. Standardmäßige Einstellungen sind einzurichten, um die Erhebung und/oder Verarbeitung von Daten auf das für die Erbringung des Dienstes erforderliche Minimum zu begrenzen.
138. Je nach Standort, Nutzungskontext und Empfindlichkeit des Mikrofons könnte der Sprachassistent Sprachdaten von Dritten als Teil des Hintergrundgeräuschs erfassen, wenn er die Stimme des Nutzers aufnimmt. Selbst Hintergrundgeräusche, die keine Sprachdaten enthalten, können Informationen über die Umgebung enthalten, die sich verarbeiten lassen, um etwas über die betroffene Person (z. B. ihren Standort) zu erfahren.

### Empfehlungen

139. VVA-Designer sollten Technologien zum Ausschalten von Hintergrundgeräuschen in Erwägung ziehen, um die Aufzeichnung und Verarbeitung von Stimmen im Hintergrund und Informationen über den Standort zu vermeiden.

### 3.10 Rechenschaftspflicht

140. Für jede Verarbeitung, die eine Einwilligung voraussetzt, muss der Verantwortliche gemäß Artikel 7 Absatz 1 der DSGVO nachweisen können, dass die Einwilligung der betroffenen Personen vorliegt. Die Rechenschaftspflicht kann anhand von Sprachdaten erfüllt werden (z. B. zum Nachweis der Einwilligung). Die Verpflichtung zur Aufbewahrung solcher Sprachdaten wäre dann durch die Rechenschaftspflicht nach Maßgabe der betreffenden Rechtsvorschriften gegeben.
141. Zur Bewertung, ob eine Datenschutz-Folgenabschätzung (DSFA) erforderlich ist, hat der EDSA Kriterien<sup>50</sup> zusammengestellt, die die Datenschutzbehörden anwenden können, wenn sie Listen der Verarbeitungsvorgänge erstellen, die zwingend eine DSFA erfordern, und er nennt Beispiele für Verarbeitungen, für die wahrscheinlich eine DSFA erforderlich ist. Für VVA-Dienste werden höchstwahrscheinlich die Kategorien und Bedingungen gelten, die eine DSFA erforderlich machen. Dazu gehören auch die Frage, ob das Gerät möglicherweise betroffene Personen beobachtet, überwacht oder kontrolliert oder eine systematische umfangreiche

---

S. K. Jami, S. R. Chalamala, A. K. Jindal, „*Biometric Template Protection Through Adversarial Learning*“, 2019, *IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, 2019, S. 1-6, doi: 10.1109/ICCE.2019.8661905.

<sup>50</sup> Artikel 29-Gruppe, Leitlinien zur Datenschutz-Folgenabschätzung (DSFA), WP 248 Rev. 01, vom EDSA gebilligt.

Überwachung gemäß Artikel 35 Absatz 3 Buchstabe c der DSGVO erfolgt, sowie der Einsatz „neuer Technologien“ oder die Verarbeitung von sensiblen Daten und Daten schutzbedürftiger Personen.

142. Alle Datenerhebungs- und Verarbeitungstätigkeiten sind gemäß Artikel 30 der DSGVO zu dokumentieren. Darunter fällt auch jede Verarbeitung von Sprachdaten.

## Empfehlungen

143. Wenn Sprachnachrichten zur Information der Nutzer gemäß Artikel 13 verwendet werden sollen, müssen die Verantwortlichen diese Mitteilungen auf ihrer Website veröffentlichen, um sie Nutzern und den Datenschutzbehörden zugänglich zu machen.

### 3.11 Datenschutz durch Design und durch standardmäßige Voreinstellung

144. VVA-Anbieter und -Entwickler müssen prüfen, ob für jede ihrer Funktionalitäten ein registrierter Nutzer erforderlich ist. Klar ist, dass ein registrierter Nutzer für die Verwaltung eines Kalenders oder Adressbuchs erforderlich ist, doch weniger klar ist, dass der VVA für einen Telefonanruf oder eine Internetsuche einen registrierten Nutzer benötigt.
145. Standardmäßig dürfen Dienste, für die kein identifizierter Nutzer erforderlich ist, keinen der identifizierten VVA-Nutzer mit den Befehlen in Verbindung bringen. Ein datenschutzfreundliches Standard-VVA würde Nutzerdaten nur für die Ausführung von Nutzeranfragen verarbeiten und weder Sprachdaten speichern noch ein Register der ausgeführten Befehle anlegen.
146. Während einige Geräte nur einen VVA betreiben können, können andere unter verschiedenen VVAs wählen. VVA-Anbieter sollten Industriestandards entwickeln, die die Datenübertragbarkeit gemäß Artikel 20 der DSGVO ermöglichen.
147. Einige VVA-Anbieter haben behauptet, dass ihre VVAs nicht alle Nutzerdaten löschen könnten, selbst wenn dies von der betroffenen Person verlangt wird. VVA-Anbieter müssen sicherstellen, dass alle Nutzerdaten auf Verlangen des Nutzers gemäß Artikel 17 der DSGVO gelöscht werden können.

## 4 MECHANISMEN ZUR AUSÜBUNG DER RECHTE BETROFFENER PERSONEN

148. Im Einklang mit der DSGVO müssen Verantwortliche, die VVA-Dienste erbringen, allen registrierten und nicht registrierten Nutzern die Ausübung ihrer Rechte als betroffene Personen ermöglichen.
149. Anbieter und Entwickler von VVA müssen betroffenen Personen die Kontrolle über ihre Daten während des gesamten Verarbeitungszeitraums erleichtern, insbesondere ihr Recht auf Auskunft, auf Berichtigung, auf Löschung, auf Einschränkung der Verarbeitung und, je nach Rechtsgrundlage der Verarbeitung, ihr Recht auf Datenübertragbarkeit und ihr Widerspruchsrecht.
150. Der Verantwortliche muss Informationen über die Rechte der betroffenen Person zu dem Zeitpunkt bereitstellen, zu dem die betroffene Person einen Sprachassistenten anschaltet, spätestens jedoch, wenn die Sprachanfrage des ersten Nutzers verarbeitet wird.

151. Da Sprache das wichtigste Interaktionsmittel für VVAs ist, müssen VVA-Designer sicherstellen, dass registrierte und nicht registrierte Nutzer ihre Rechte als betroffene Personen mittels einfach anzuwendender Sprachbefehle ausüben können. VVA-Designer und, sofern sie Teil der Lösung sind, App-Entwickler müssen die Nutzer am Ende des Vorgangs darüber informieren, dass ihre Rechte ordnungsgemäß berücksichtigt wurden, und sie durch eine gesprochene Mitteilung oder durch schriftliche Benachrichtigung an das Mobiltelefon, das Konto des Nutzers oder auf andere, vom Nutzer gewählte Weise informieren.
152. VVA-Designer und insbesondere App-Entwickler sollten spezielle Instrumente einsetzen, die eine wirksame und effiziente Ausübung dieser Rechte ermöglichen. Sie sollten für ihre Geräte eine Möglichkeit zur Ausübung der Rechte betroffener Personen vorsehen und diesen Personen Self-Service-Tools zur Verwaltung des Nutzerprofils als Profilverwaltungssystem<sup>51</sup> zur Verfügung stellen. Das könnte die wirksame und zeitnahe Wahrnehmung der Rechte betroffener Personen erleichtern und es dem Verantwortlichen ermöglichen, den Identifizierungsmechanismus in das Self-Service-Tool zu integrieren.
153. Wenn es bei der Ausübung der Rechte betroffener Personen um mehrere Nutzer geht, muss jede dieser registrierten oder nicht registrierten Personen jedes ihrer Rechte unbeschadet etwaiger anderer Nutzerrechte wahrnehmen können. Alle registrierten und nicht registrierten Nutzer können ihre Rechte ausüben, solange der Verantwortliche die Daten noch verarbeitet. Verantwortliche sollten ein Verfahren einrichten, mit dem sichergestellt ist, dass betroffene Personen ihre Rechte wahrnehmen können.

#### 4.1 Recht auf Auskunft

154. Nach Artikel 12 Absatz 1 der DSGVO erfolgt die Kommunikation gemäß Artikel 15 schriftlich oder in anderer Form, gegebenenfalls auch elektronisch. Hinsichtlich des Rechts auf Auskunft über die personenbezogenen Daten, die Gegenstand der Verarbeitung sind, regelt Artikel 15 Absatz 3: „Stellt die betroffene Person den Antrag elektronisch, so sind die Informationen in einem gängigen elektronischen Format zur Verfügung zu stellen, sofern sie nichts anderes angibt.“ Was als gängiges elektronisches Format anzusehen ist, bestimmt sich nach den berechtigten Erwartungen der betroffenen Personen und nicht nach dem Format, das der Verantwortliche normalerweise verwendet. Die betroffene Person darf nicht gezwungen sein, eine bestimmte Software oder Hardware anzuschaffen, um Zugang zu den Informationen zu bekommen.
155. Auf Verlangen müssen Verantwortliche deshalb eine Kopie der personenbezogenen Daten und insbesondere der Audiodaten (auch Sprachaufzeichnungen und Transkripte) in einem gängigen, für die betroffene Person lesbaren Format übermitteln.
156. Bei der Entscheidung, in welchem Format die Informationen gemäß Artikel 15 bereitgestellt werden sollen, muss der Verantwortliche beachten, dass die Informationen in dem Format sowohl verständlich dargestellt als auch leicht zugänglich sein sollen. Verantwortliche sollten die Information auch an die besondere Situation der betroffenen Person, die den Antrag stellt, anpassen.

---

<sup>51</sup> Das Profilverwaltungssystem ist ein Ort im VVA-System, wo Nutzer jederzeit ihre Präferenzen speichern, Änderungen vornehmen und ihre Datenschutzeinstellungen ganz einfach ändern können.

#### Beispiel 13:

Ein Verantwortlicher, der einen VVA-Dienst bereitstellt, erhält von einem Nutzer einen Antrag auf Auskunft und einen Antrag auf Datenübermittlung. Der Verantwortliche will die Informationen gemäß Artikel 15 und gemäß Artikel 20 als PDF-Datei bereitstellen. In dem Fall kann nicht davon ausgegangen werden, dass er beide Anträge korrekt bearbeitet. Eine PDF-Datei erfüllt in technischer Hinsicht zwar die Pflichten des Verantwortlichen nach Artikel 15, jedoch nicht nach Artikel 20.<sup>52</sup>

Es ist darauf hinzuweisen, dass der Verantwortliche seiner Pflicht hinsichtlich des Auskunftsrechts nicht nachkommt, indem er den Nutzer einfach auf den Verlauf seiner Interaktion mit dem Sprachassistenten hinweist, da die zugänglichen Daten in der Regel nur einen Teil der Informationen ausmachen, die im Zusammenhang mit der Erbringung des Dienstes verarbeitet werden.

157. Das Auskunftsrecht darf nicht dazu genutzt werden, den Grundsatz der Datenminimierung und den Grundsatz der Datenspeicherung zu verletzen bzw. zu umgehen.

## 4.2 Recht auf Berichtigung

158. Um die Berichtigung von Daten zu erleichtern, müssen registrierte und nicht registrierte Nutzer ihre Daten, wie oben beschrieben, jederzeit direkt über das VVA-Gerät per Sprache verwalten und aktualisieren können. Darüber hinaus sollte das Self-Service-Tool in dem Gerät oder einer Anwendung implementiert sein, damit die personenbezogenen Daten leichter berichtigt werden können. Nutzer sind per Sprache oder schriftlich über die Aktualisierung zu benachrichtigen.
159. Generell gilt das Recht auf Berichtigung für alle Stellungnahmen und Inferenzen<sup>53</sup> des Verantwortlichen, einschließlich Profilbildung, wobei es sich in den allermeisten Fällen um sehr subjektive Daten handeln wird.<sup>54</sup>

## 4.3 Recht auf Löschung

160. Sowohl registrierte als auch nicht registrierte Nutzer müssen jederzeit per Sprache über das VVA-Gerät oder über ein Self-Service-Tool, das in ein mit dem VVA verbundenes Gerät integriert ist, sie betreffende Daten löschen können. Insoweit können die personenbezogenen Daten von einer betroffenen Person ebenso einfach gelöscht werden, wie sie übermittelt wurden. Aufgrund der schwierigen Anonymisierung von Sprachdaten und der Vielfalt personenbezogener Daten zu der betroffenen Person, die gesammelt, beobachtet und

---

<sup>52</sup> Artikel 29-Gruppe, Leitlinien zum Recht auf Datenübertragbarkeit, vom EDSA gebilligt, S. 18.

<sup>53</sup> Dass Stellungnahmen und Inferenzen als personenbezogene Daten eingestuft werden können, wurde vom EuGH bestätigt. Er stellte fest, dass der Ausdruck „alle Informationen“ in der Bestimmung des Begriffs „personenbezogene Daten“ potenziell alle Arten von Informationen „sowohl objektiver als auch subjektiver Natur in Form von Stellungnahmen oder Beurteilungen“ umfasst, „unter der Voraussetzung, dass es sich um Informationen ‚über‘ die in Rede stehende Person handelt“. Urteil des Gerichtshofs vom 20. Dezember 2017, Peter Nowak / Data Protection Commissioner, C-434/16, ECLI:EU:C:2017:994, Rn. 34.

<sup>54</sup> Getting Data Subject Rights Right: A submission to the European Data Protection Board from international data rights academics, to inform regulatory guidance, November 2019.

abgeleitet werden,<sup>55</sup> lässt sich das Recht auf Löschung durch die Anonymisierung personenbezogener Datensätze in diesem Zusammenhang kaum verwirklichen. Da die DSGVO technologieneutral ist und die Technologie sich rasch weiterentwickelt, ist dennoch nicht auszuschließen, dass das Recht auf Löschung durch Anonymisierung wirksam werden kann.

161. In manchen Fällen ist es ohne einen Bildschirm von Dritten oder die Möglichkeit, gespeicherte Daten anzuzeigen (z. B. über eine mobile Anwendung oder ein Tablet), nicht ohne Weiteres möglich, eine Übersicht über die aufgezeichneten Bewegungen zu erhalten, um die Relevanz der Vorschläge beurteilen zu können. Ein Dashboard (oder eine Anwendung), das Nutzern allgemein zugänglich ist, um die Nutzung zu erleichtern, sollte den Sprachassistenten ergänzen, um den Verlauf der gestellten Anfragen löschen und das Tool an die Erfordernisse des Nutzers anpassen zu können.<sup>56</sup>
162. Bei jeder Datenverarbeitung und insbesondere, wenn registrierte Personen mit der Transkription von Sprachaufzeichnungen einverstanden sind, die der Anbieter zur Verbesserung seiner Dienste verwendet, müssen VVA-Anbieter auf Verlangen des Nutzers die ursprüngliche Sprachaufzeichnung sowie jede damit verbundene Transkription von personenbezogenen Daten löschen können.
163. Der Verantwortliche muss sicherstellen, dass nach Ausübung des Rechts auf Löschung keine weitere Verarbeitung mehr erfolgen kann. In Bezug auf frühere Vorgänge kann das Recht auf Löschung allerdings an rechtliche und vor allem technische Grenzen stoßen.

**Beispiel 14:**

Wenn ein Nutzer vor dem Löschungsersuchen einen Online-Kauf über seinen VVA getätigt hat, kann der VVA-Anbieter die Sprachaufzeichnung im Zusammenhang mit dem Online-Kauf löschen und sicherstellen, dass keine weitere Nutzung mehr erfolgt. Der Kauf bleibt jedoch wirksam, ebenso wie die mündliche Bestellung oder Umwandlung in geschriebenen Text, die über die E-Commerce-Website verarbeitet wird (hier ist die Ausnahme in der rechtlichen Verpflichtung der E-Commerce-Website begründet).

Gleiches gilt, wenn der Nutzer vor dem Löschungsersuchen über seinen Sprachassistenten einen bestimmten Song zu seiner Playlist hinzugefügt hat. Der VVA-Anbieter kann die mündliche Anfrage löschen, nicht jedoch ihre bisherigen Folgen (die Löschung wirkt sich nicht auf die Playlist des Nutzers aus).

164. Nach den obigen Ausführungen ist hinsichtlich der Verarbeitung derselben personenbezogenen Daten für unterschiedliche Verarbeitungszwecke festzustellen, dass Verantwortliche Löschungsersuchen als klares Signal auslegen sollten, die Verarbeitung der Daten für alle Zwecke, die nicht aus rechtlichen Gründen ausgenommen sind, zu beenden.

Im Einklang mit den Bedingungen nach Artikel 21 Absatz 1 der DSGVO dürfen Daten, die auf der Grundlage berechtigter Interessen der VVA-Anbieter verarbeitet werden, nicht vom Recht auf Löschung ausgenommen werden, zumal die betroffenen Personen nicht damit rechnen können, dass ihre personenbezogenen Daten weiterverarbeitet werden.

---

<sup>55</sup> Artikel-29-Datenschutzgruppe, Stellungnahme 5/2014 zu Anonymisierungstechniken, WP216, angenommen am 10. April 2014.

<sup>56</sup> Conseil Supérieur de l'Audiovisuel, „Assistants vocaux et enceintes connectées, l'impact de la voix sur l'offre et les usages culturels et médias“, Mai 2019.

#### 4.4 Recht auf Datenübertragbarkeit

165. Die Datenverarbeitung durch die VVA-Anbieter fällt in den Bereich der Datenübertragbarkeit, da die Verarbeitungsvorgänge hauptsächlich auf der Einwilligung der betroffenen Person beruhen (Artikel 6 Absatz 1 Buchstabe a oder Artikel 9 Absatz 2 Buchstabe a, wenn es sich um besondere Kategorien personenbezogener Daten handelt) oder auf einem Vertrag, bei dem die betroffene Person gemäß Artikel 6 Absatz 1 Buchstabe b Vertragspartei ist.
166. In der Praxis sollte das Recht auf Datenübertragbarkeit den Wechsel zwischen verschiedenen VVA-Anbietern erleichtern. Vor allem, wenn VVAs in einem digitalen Umfeld betrieben werden und die Stimme der betroffenen Person in einer Anwendung oder auf einer Plattform aufgezeichnet wird, muss das Recht auf Datenübertragbarkeit für alle von der betroffenen Person stammenden personenbezogenen Daten gelten. Außerdem sollte der Verantwortliche Nutzern die Möglichkeit bieten, ihre personenbezogenen Daten direkt aus ihrem Nutzerbereich mittels eines Self-Service-Tools abzurufen. Nutzer sollten dieses Recht auch per Sprachbefehl ausüben können.
167. VVA-Anbieter und Entwickler müssen den betroffenen Personen eine umfassende Kontrolle über die sie betreffenden personenbezogenen Daten ermöglichen, damit sie personenbezogene Daten von einem VVA-Anbieter an einen anderen übermitteln können. Betroffene Personen sollten daher ihre personenbezogenen Daten, die dem Verantwortlichen zur Verfügung gestellt wurden, in einem strukturierten, gängigen und maschinenlesbaren Format und durch Werkzeuge erhalten<sup>57</sup>, die die Beantwortung von Anfragen zur Datenübertragbarkeit unterstützen (z. B. Download-Tools oder Anwendungsschnittstellen)<sup>58</sup>. Wie in den Leitlinien zum Recht auf Datenübertragbarkeit dargelegt wird, muss der Verantwortliche im Falle einer umfangreichen oder komplexen Erhebung personenbezogener Daten, wie es hier der Fall sein könnte, einen Überblick „in präziser, transparenter, verständlicher und leicht zugänglicher Form in einer klaren und einfachen Sprache“ (siehe Artikel 12 Absatz 1 DSGVO) geben, damit die betroffenen Personen stets klare Informationen darüber haben, welche Daten für einen bestimmten Zweck an einen anderen Verantwortlichen heruntergeladen oder übermittelt werden können. Beispielsweise sollten betroffene Personen Softwareanwendungen nutzen können, um bestimmte Daten leicht identifizieren, erkennen und verarbeiten zu können.

---

<sup>57</sup> Siehe beispielsweise die Begründung der Artikel-29-Datenschutzgruppe in den Leitlinien zum Recht auf Datenübertragbarkeit, vom EDSA gebilligt, S. 19:

*„Verantwortliche sollten auf technischer Ebene zwei unterschiedliche, einander ergänzende Verfahren analysieren und bewerten, durch die portable Daten an betroffene Personen oder andere Verantwortliche übermittelt werden können:*

- direkte Übermittlung des vollständigen Datensatzes (oder mehrerer Auszüge von Teilen des Datensatzes);
- Einsatz eines automatisierten Werkzeugs, das die Extrahierung der relevanten Daten ermöglicht.

*Das zweite Verfahren dürfte vom Verantwortlichen bei komplexen und umfangreichen Datensätzen vorzuziehen sein, denn es ermöglicht ihm, genau jene Teile des Datensatzes zu extrahieren, die für die betroffene Person laut deren Anfrage relevant sind. Dies kann zur Risikominimierung beitragen und auch die Nutzung von Datensynchronisierungsverfahren (beispielsweise im Rahmen der regelmäßigen Kommunikation zwischen Verantwortlichen) ermöglichen. Dieses Verfahren eignet sich vielleicht besser, um die Vorschrifteneinhaltung des ‚neuen‘ Verantwortlichen sicherzustellen und kann als bewährtes Verfahren zur Verringerung von Datenschutzrisiken bei der Datenverarbeitung aufseiten des ursprünglichen Verantwortlichen gelten.“*

<sup>58</sup> Siehe dazu: Artikel-29-Datenschutzgruppe, Leitlinien zum Recht auf Datenübertragbarkeit, vom EDSA gebilligt, S. 3.

168. Dieses Recht soll es dem Nutzer ermöglichen, die Daten, die er mit seiner Stimme übermittelt hat (z. B. den Verlauf der sprachlichen Interaktionen), für seine persönliche Verwendung und insbesondere im Rahmen der Einrichtung seines Nutzerkontos (z. B. Name und Vorname) abzurufen.
169. Damit betroffene Personen dieses Recht in einem digitalen Binnenmarkt uneingeschränkt ausüben können, sollten VVA-Designer und insbesondere App-Entwickler einheitliche maschinenlesbare Formate entwickeln, die die Interoperabilität des Datenformats zwischen VVA-Systemen<sup>59</sup>, einschließlich der Standardformate für Sprachdaten, erleichtern. Technisch sollten sie so strukturiert sein, dass personenbezogene Daten, einschließlich Sprachdaten, von dem neuen Verantwortlichen einfach und vollständig wiederverwendet werden können.<sup>60</sup>
170. Was das Format betrifft, sollten VVA-Anbieter personenbezogene Daten in gängigen offenen Formaten (mp3, wav, csv, gsm usw.) zusammen mit geeigneten Metadaten bereitstellen, um die Bedeutung ausgetauschter Informationen genau zu beschreiben.<sup>61</sup>

## 5 ANHANG: AUTOMATISCHE SPRACHERKENNUNG, SPRACHSYNTHESE UND VERARBEITUNG NATÜRLICHER SPRACHE

171. Infolge der theoretischen Grundlagen der Signalverarbeitung, insbesondere Claude Shannons Informationstheorie und Abtasttheorem, ist die automatische Sprachverarbeitung zu einem grundlegenden Bestandteil der Ingenieurwissenschaften geworden. An der Schnittstelle zwischen Physik (Akustik, Wellenausbreitung), angewandter Mathematik (Modellierung, Statistik), Informatik (Algorithmen, Lerntechniken) und Humanwissenschaften (Perzeption, Argumentation) hat sich die Sprachverarbeitung sehr schnell in verschiedene Forschungsrichtungen entwickelt: Sprecheridentifizierung und -verifizierung, automatische Spracherkennung, Sprachsynthese, Stimmungserkennung usw. In den letzten rund 15 Jahren hat dieser gesamte Forschungsbereich erhebliche Fortschritte gemacht, wozu verschiedene Faktoren beigetragen haben: verbesserte Methoden, eine erhebliche Zunahme der Rechnerkapazitäten und größere verfügbare Datenvolumen.

### 5.1 Automatische Spracherkennung (ASR)

172. Die automatische Spracherkennung (Umwandlung von Sprache in Text) lief üblicherweise in drei Stufen ab, in denen es darum geht, 1. anhand eines akustischen Modells zu bestimmen, welche Phoneme gesprochen wurden; 2. anhand eines phonetischen Wörterbuchs zu bestimmen, welche Wörter gesprochen wurden; 3. anhand eines Sprachmodells die Sequenz von Wörtern (Satz), die höchstwahrscheinlich gesprochen wurde, umzuwandeln. Dank der

---

<sup>59</sup> Siehe dazu: Randnummer 68 der DSGVO; Leitlinien der Artikel-29-Datenschutzgruppe zum Recht auf Datenübertragbarkeit, vom EDSA gebilligt, S. 20.

<sup>60</sup> „Diesbezüglich sieht Erwägungsgrund 68 vor, dass die Verantwortlichen interoperable Formate entwickeln sollten, die die Datenübertragbarkeit ermöglichen, wobei die Verantwortlichen jedoch nicht dazu verpflichtet werden sollten, technisch kompatible Datenverarbeitungssysteme zu übernehmen oder beizubehalten. Die DSGVO verbietet Verantwortlichen, die Übermittlung zu behindern.“ Leitlinien der Artikel-29-Datenschutzgruppe zum Recht auf Datenübertragbarkeit, vom EDSA gebilligt, S. 5.

<sup>61</sup> Der EDSA unterstützt nachdrücklich eine Zusammenarbeit zwischen Interessenträgern von Industrie und Wirtschaftsverbänden zur Entwicklung gemeinsamer interoperabler Standards und Formate, um die Anforderungen des Rechts auf Datenübertragbarkeit zu erfüllen.

Fortschritte durch Deep Learning (einer maschinellen Lerntechnik) bieten viele Systeme heute eine „End-to-End“-ASR an. Das erspart das komplexe Training von drei verschiedenen Modellen, während gleichzeitig bessere Ergebnisse und Verarbeitungszeiten erreicht werden. Fast alle großen digitalen Akteure bieten mittlerweile eigene ASR-Implementierungen an, die von API-Systemen problemlos genutzt werden können, aber es gibt auch Open-Source-Systeme (z. B. DeepSpeech<sup>62</sup> oder Kaldi<sup>63</sup>).

## 5.2 Verarbeitung natürlicher Sprache (NLP)

173. Die Verarbeitung natürlicher Sprache ist ein multidisziplinärer Bereich, der Linguistik, Informatik und künstliche Intelligenz einschließt und darauf abzielt, Instrumente für die Verarbeitung natürlicher Sprache für vielfältige Anwendungen zu entwickeln. Es gibt eine Vielzahl von Forschungsfeldern und Anwendungen: syntaktische Analyse, maschinelle Übersetzung, automatische Texterstellung und Textzusammenfassung, Rechtschreibprüfung, Frage-Antwort-Systeme, Text Mining, Eigennamenerkennung, Sentiment-Analyse usw. Konkret sollen Computer durch die Verarbeitung natürlicher Sprache in die Lage versetzt werden, menschliche Sprachen zu lesen, zu verstehen und ihre Bedeutung zu erkennen. Die Entwicklung von NLP-Anwendungen stellt eine Herausforderung dar, da Computerwerkzeuge normalerweise verlangen, dass Menschen mit ihnen in einer Programmiersprache interagieren, die formal, d. h. präzise, eindeutig und hochgradig strukturiert ist. Menschliche Sprache ist aber nicht immer präzise. Sie ist oft mehrdeutig, und die Sprachstruktur kann von vielen komplexen Variablen wie Slang, regionalen Dialekten und sozialem Kontext abhängen.
174. Syntaxanalyse und semantische Analyse sind zwei wesentliche Techniken in der Verarbeitung natürlicher Sprache. Syntax ist die Anordnung der Wörter in einem Satz, so dass sich grammatikalisch ein Sinn ergibt. Die NLP nutzt die Syntax, um die Bedeutung einer Sprache anhand von grammatischen Regeln zu erkennen. Die Syntaxanalyse umfasst Parsing (grammatikalische Analyse des einzelnen Satzes), Wortsegmentierung (Zerlegung eines größeren Textteils in Einheiten), Satzumbruch (Satzgrenzen in einem langen Text erkennen), morphologische Segmentierung (Gruppierung von Wörtern) und Stemming (führt flektierte Wörter auf ihre Grundform zurück). Die Semantik befasst sich mit der Verwendung und der Bedeutung von Wörtern. Die NLP verwendet Algorithmen, um die Bedeutung und die Struktur von Sätzen zu verstehen. Die semantische Analyse der NLP umfasst die Word Sense Disambiguation (erklärt die Bedeutung eines Wortes abhängig vom Kontext), die Eigennamenerkennung (bestimmt Begriffe, die Gruppen zugeordnet werden können) und die Erzeugung natürlicher Sprache (verwendet eine Datenbank zur Bestimmung der Semantik von Wörtern). Während sich die NLP früher auf regelbasierte Ansätze stützte, bei denen einfachen Algorithmen für maschinelles Lernen erklärt wurde, welche Wörter und Wortfolgen in einem Text gesucht werden sollten, und spezifische Antworten gegeben wurden, wenn diese Wortfolgen erschienen, basiert die NLP mittlerweile auf dem Deep Learning, einer Form von künstlicher Intelligenz, bei der Muster in Daten untersucht und verwendet werden, um die Verständnisfähigkeit eines Programms zu verbessern.

## 5.3 Sprachsynthese

175. Sprachsynthese ist die künstliche Generierung menschlicher Sprache. Die Sprachsynthese wird hauptsächlich durch eine Verkettung von Worteinheiten umgesetzt, die in einer Datenbank gespeichert sind. Bei dieser Technik werden aus allen Aufzeichnungen eines Akteurs, die zuvor

---

<sup>62</sup> <https://github.com/mozilla/DeepSpeech>

<sup>63</sup> <https://github.com/kaldi-asr/kaldi>

in Phoneme, Silben und Wörter umgewandelt wurden, die Tonbausteine ausgewählt, die den Wörtern entsprechen, die der VVA aussprechen soll, und zu einem verständlichen Satz mit natürlicher Diktion aneinandergereiht. Alternativ dazu kann ein Sprachsynthesizer ein Modell des Sprechapparates und andere Merkmale menschlicher Sprache umfassen, um die Parameter einer Stimme wie Intonation, Rhythmus und Klang durch generative statistische Modelle (wie WaveNet<sup>64</sup>, Tacotron<sup>65</sup> oder DeepVoice<sup>66</sup>) zu modellieren und einen rein synthetischen Sprachausgang zu erzeugen.

---

<sup>64</sup> Aäron van den Oord, Sander Dieleman, *WaveNet: A generative model for raw audio*, Deepmind Blog, September 2016, <https://deepmind.com/blog/article/wavenet-generative-model-raw-audio>

<sup>65</sup> Yuxuan Wang, *Expressive Speech Synthesis with Tacotron*, Google AI Blog, März 2018, <https://ai.googleblog.com/2018/03/expressive-speech-synthesis-with.html>

<sup>66</sup> *Deep Voice 3: 2000-Speaker Neural Text-to-Speech*, Baidu Research Blog, Oktober 2017, <http://research.baidu.com/Blog/index-view?id=91>