

Antwort

der Bundesregierung

**auf die Kleine Anfrage der Abgeordneten Clara Bünger, Nicole Gohlke, Gökyak Akbulut, weiterer Abgeordneter und der Fraktion DIE LINKE.
– Drucksache 20/3133 –**

Einsatz von Dialekterkennungssoftware im Bundesamt für Migration und Flüchtlinge

Vorbemerkung der Fragesteller

Seit 2017 setzt das Bundesamt für Migration und Flüchtlinge (BAMF) verschiedene IT-Assistenzsysteme ein, die zur Klärung der Herkunft und Identität von Asylsuchenden beitragen sollen. Dabei werden u. a. mobile Datenträger von Geflüchteten ausgewertet oder Sprachproben von Asylsuchenden mittels einer Sprach- bzw. Dialekterkennungssoftware analysiert (vgl. u. a. die Antwort der Bundesregierung auf die Kleine Anfrage auf Bundestagsdrucksache 19/1663).

Beim Einsatz der Sprachbiometrie müssen Asylsuchende ca. zwei Minuten in ihrer Muttersprache ein Bild beschreiben und dabei in ein Telefon sprechen. Was sie sagen, wird aufgezeichnet. Die Sprechproben werden anschließend automatisch mit einem zugrunde liegenden Sprachmodell abgeglichen. So werden Wahrscheinlichkeiten für den gesprochenen Dialekt berechnet und in einem Ergebnisbericht festgehalten. Die Dialekterkennungssoftware kann bislang die wichtigsten arabischen Dialekte erkennen und wird folglich in erster Linie bei arabischsprachigen Asylsuchenden angewendet. Von Januar bis November 2019 kam das Spracherkennungstool bei rund 4 000 Personen zum Einsatz. Gut ein Viertel von ihnen kam aus Syrien, weitere wichtige Staatsbürgerschaften waren Algerien, Marokko und Sudan. Allerdings ist aus Sicht der Fragestellerinnen und Fragesteller unklar, ob mit der Spracherkennungssoftware überhaupt geeignete Hinweise auf die Herkunft von Geflüchteten gewonnen werden können. Zum einen wurde diese mithilfe von Experten und Expertinnen aus dem Bereich der Künstlichen Intelligenz, nicht aber der Sprachwissenschaften entwickelt. Linguistinnen und Linguisten betonen, dass sich die Sprechweise von Menschen über die Lebensspanne verändern kann und deshalb nicht unbedingt Rückschlüsse auf die Herkunft erlaubt. Auch stimmen Sprach- und Landesgrenzen nur teilweise überein. Zum anderen ist die Spracherkennungssoftware nach Ansicht der Fragestellerinnen und Fragesteller fehleranfällig, und es mangelt an Transparenz darüber, wie viele Sprachproben pro Dialekt vorliegen und wie die Fehlerquoten, die das BAMF mit 15 Prozent angibt, im Einzelnen zustande kommen (<https://www.fiff.de/publikationen/fiff-kommunikation/fk-2022/fk-2022-1/fk-1-2022-content/fk-1-22-p45.pdf>; <https://netzpolitik.org/2020/automatisiertes-misstrauen/>; [---

Die Antwort wurde namens der Bundesregierung mit Schreiben des Bundesministeriums des Innern und für Heimat vom 31. August 2022 übermittelt.](https://ww</p></div><div data-bbox=)

w.heise.de/newsticker/meldung/Sherlock-Bamf-Asyltaeuschern-auf-der-Spur-4257898.html).

1. In wie vielen Fällen wurde die Dialekterkennungssoftware des BAMF in den Jahren 2020, 2021 und im ersten Halbjahr 2022 eingesetzt (bitte nach Jahren und nach BAMF-Standorten aufschlüsseln), und was waren die Herkunftsländer der betroffenen Asylsuchenden?

Die Antwort kann der nachfolgenden Übersicht der in den Jahren 2020, 2021 und dem ersten Halbjahr 2022 durchgeführten Dialekterkennungen entnommen werden:

| Zeitraum | Dialekterkennung* |
|---------------------|-------------------|
| 01.01. – 31.12.2020 | 9.923 |
| 01.01. – 31.12.2021 | 15.052 |
| 01.01. – 30.06.2022 | 7.808 |

* Durch nachträgliche Änderungen im Erfassungssystem, z. B. die Einreichung von Pass/Passersatz-Dokumenten, kann es zu Abweichungen der Daten im vgl. zu vorherigen Angaben kommen. Es werden diejenigen Fälle ausgewertet, bei denen aus Prozesssicht alle erforderlichen Angaben im Kerndatensystem korrekt und valide hinterlegt sind.

Nach Organisationseinheiten/Standorten des Bundesamts für Migration und Flüchtlinge (BAMF):

| 01.01. – 31.12.2021 | |
|-------------------------------|---|
| Organisationseinheit/Standort | Anteil des Einsatzes der Dialekterkennung |
| AS Heidelberg im AZ | 10,9 % |
| AS Gießen im AZ, LAS | 9,3 % |
| AS Bad Fallingbostal im AZ | 4,8 % |
| AS Halberstadt im AZ, LAS | 4,6 % |
| AS Trier, LAS | 4,4 % |
| AS Berlin im AZ | 3,8 % |
| AS Chemnitz im AZ, LAS | 3,5 % |
| AS Bonn im AZ | 3,5 % |
| AS Regensburg in AnKER | 4,1 % |
| AS Ellwangen | 3,2 % |
| AS Lebach in AnKER, LAS | 3,2 % |
| AS Schwerin im AZ | 3,2 % |
| AS Suhl im AZ | 3,2 % |
| AS München | 3,1 % |
| AS Dresden in AnKER | 3,0 % |
| AS Schweinfurt in AnKER | 2,9 % |
| AS Mönchsgladbach im AZ | 2,8 % |
| AS Zirndorf in AnKER | 2,5 % |
| AS Neumünster, LAS | 2,5 % |
| AS Bamberg in AnKER | 2,3 % |
| AS Bielefeld im AZ | 2,3 % |
| AS Deggendorf in AnKER | 2,2 % |
| AS Augsburg in AnKER | 2,0 % |
| AS Leipzig im AZ | 1,9 % |
| AS Bramsche im AZ | 1,7 % |
| AS Eisenhüttenstadt, LAS | 1,6 % |

| 01.01. – 31.12.2021 | |
|--------------------------------------|--|
| Organisationseinheit/Standort | Anteil des Einsatzes der Dialekterkennung |
| AS Speyer | 1,5 % |
| AS Sigmaringen | 1,5 % |
| AS Hamburg im AZ, LAS | 1,5 % |
| AS Bremen im AZ, LAS | 0,9 % |
| AS Büdingen | 0,5 % |
| AS Oldenburg | 0,3 % |
| AS Nostorf-Horst, LAS | 0,2 % |
| 32F Dublinzentrum Bayreuth | 0,2 % |
| AS Unna im AZ | 0,1 % |
| 32D Dublinzentrum Berlin | 0,1 % |
| AS Essen | 0,1 % |
| 32E Dublinzentrum Bochum | 0,1 % |
| AS Manching in AnKER | 0,1 % |
| AS Braunschweig | 0,1 % |
| AS Freiburg | 0,1 % |
| AS Friedland, LAS | 0,0 % |
| AS Jena/Hermsdorf, LAS | 0,0 % |
| AS Neustadt | 0,0 % |
| Unbekannt | 0,0 % |

| 01.01. – 30.06.2022 | |
|--------------------------------------|--|
| Organisationseinheit/Standort | Anteil des Einsatzes der Dialekterkennung |
| AS Heidelberg im AZ | 9,8 % |
| AS Gießen im AZ, LAS | 7,5 % |
| AS Dresden in AnKER | 5,1 % |
| AS Trier, LAS | 5,1 % |
| AS Bonn im AZ | 4,8 % |
| AS Bad Fallingb. im AZ | 4,6 % |
| AS Chemnitz im AZ, LAS | 4,5 % |
| AS Bielefeld im AZ | 3,9 % |
| AS Schwerin im AZ | 3,8 % |
| AS Berlin im AZ | 3,7 % |
| AS Halberstadt im AZ, LAS | 3,5 % |
| AS Suhl im AZ | 3,5 % |
| AS Neumünster, LAS | 3,5 % |
| AS Mönchengladbach im AZ | 3,2 % |
| AS Regensburg in AnKER | 3,7 % |
| AS Speyer | 2,8 % |
| AS Essen | 2,8 % |
| AS Deggendorf in AnKER | 2,3 % |
| AS Lebach in AnKER, LAS | 2,3 % |
| AS Unna im AZ | 2,4 % |
| AS Bramsche im AZ | 2,1 % |
| AS Hamburg im AZ, LAS | 1,8 % |
| AS Bamberg in AnKER | 1,7 % |
| AS Zirndorf in AnKER | 1,7 % |
| AS Eisenhüttenstadt, LAS | 1,7 % |
| AS Schweinfurt in AnKER | 1,4 % |
| AS Bremen im AZ, LAS | 1,2 % |
| AS München | 1,2 % |

| 01.01. – 30.06.2022 | |
|--------------------------------------|--|
| Organisationseinheit/Standort | Anteil des Einsatzes der Dialekterkennung |
| AS Leipzig im AZ | 0,9 % |
| AS Büdingen | 0,9 % |
| AS Oldenburg | 0,7 % |
| AS Nostorf-Horst, LAS | 0,4 % |
| AS Sigmaringen | 0,4 % |
| AS Ellwangen | 0,3 % |
| AS Augsburg in AnKER | 0,2 % |
| AS Braunschweig | 0,2 % |
| AS Jena/Hermsdorf, LAS | 0,2 % |
| AS Freiburg | 0,1 % |
| AS Karlsruhe, LAS | 0,1 % |
| 32F Dublinzentrum Bayreuth | 0,0 % |
| AS Bochum, LAS | 0,0 % |

Nach Staatsangehörigkeiten (HKL):

| 01.01. – 31.12.2020 | |
|--|--|
| Staatsangehörigkeit | Anteil des Einsatzes der Dialekterkennung |
| Syrien | 44,3 % |
| Algerien | 13,3 % |
| Irak | 12,7 % |
| Ungeklärt | 9,6 % |
| Marokko | 7,2 % |
| Tunesien | 3,7 % |
| Libyen | 1,9 % |
| Sudan | 1,9 % |
| Ägypten | 1,4 % |
| Libanon | 1,2 % |
| Pers. aus palästinensischen Gebieten (nicht als Staat anerkannt) | 0,6 % |
| Staatenlos | 0,7 % |
| Jemen | 0,5 % |
| Kuwait | 0,2 % |
| Tschad | 0,3 % |
| Jordanien | 0,1 % |
| Iran | 0,1 % |
| Südsudan | 0,1 % |
| Eritrea | 0,1 % |
| Somalia | 0,1 % |
| Äthiopien | 0,0 % |
| Mauretanien | 0,0 % |
| Saudi-Arabien | 0,0 % |

| 01.01. – 31.12.2021 | |
|--|--|
| Staatsangehörigkeit | Anteil des Einsatzes der Dialekterkennung |
| Syrien | 52,7 % |
| Algerien | 10,2 % |
| Irak | 8,3 % |
| Ungeklärt | 7,6 % |
| Marokko | 6,1 % |
| Tunesien | 4,9 % |
| Jemen | 3,1 % |
| Ägypten | 1,2 % |
| Libanon | 1,2 % |
| Libyen | 1,1 % |
| Sudan | 1,0 % |
| Pers. aus palästinensischen Gebieten (nicht als Staat anerkannt) | 1,0 % |
| Staatenlos | 0,7 % |
| Kuwait | 0,2 % |
| Tschad | 0,1 % |
| Eritrea | 0,1 % |
| Jordanien | 0,1 % |
| Iran | 0,1 % |
| Südsudan | 0,1 % |
| Somalia | 0,1 % |
| Mauretanien | 0,1 % |
| Afghanistan | 0,0 % |
| Bahrain | 0,0 % |
| Myanmar | 0,0 % |
| ohne Angabe | 0,0 % |

| 01.01. – 30.06.2022 | |
|--|--|
| Staatsangehörigkeit | Anteil des Einsatzes der Dialekterkennung |
| Syrien, Arabische Republik | 50,0 % |
| Algerien | 9,8 % |
| Ungeklärt | 9,2 % |
| Irak | 6,8 % |
| Marokko | 6,2 % |
| Tunesien | 4,9 % |
| Libanon | 3,9 % |
| Jemen | 2,4 % |
| Ägypten | 1,9 % |
| Libyen | 1,6 % |
| Sudan | 1,1 % |
| Staatenlos | 0,7 % |
| Pers. aus palästinensischen Gebieten (nicht als Staat anerkannt) | 0,4 % |
| Kuwait | 0,2 % |
| Iran | 0,1 % |
| Somalia | 0,1 % |
| Jordanien | 0,1 % |
| Mauretanien | 0,1 % |
| Südsudan | 0,1 % |
| Eritrea | 0,1 % |

| 01.01. – 30.06.2022 | |
|----------------------------|--|
| Staatsangehörigkeit | Anteil des Einsatzes der Dialekterkennung |
| Tschad | 0,1 % |
| Niger | 0,0 % |
| Venezuela | 0,0 % |

2. In wie vielen Fällen wurden die Angaben der Asylsuchenden bestätigt bzw. widerlegt (bitte nach Herkunftsländern und Jahren aufschlüsseln)?

Die Ergebnisse der Dialekterkennung können die Angaben zur Herkunft im Sinne der Fragestellung weder bestätigen noch widerlegen. Sie stellen einen Hinweis bzgl. der Herkunft des Antragstellers dar, welcher im Rahmen der Anhörung, die unter anderem auch der Klärung der Identität und Staatsangehörigkeit dient, berücksichtigt wird. Sie haben keinen Beweischarakter.

Im Jahr 2020 führte das Ergebnis der Dialekterkennung-Auswertungen von persönlichen Erstantragstellenden ohne Pass/Passersatz dazu, dass in rund 79 Prozent der Fälle die Angaben zur Identität der Antragstellenden gestützt und in rund 21 Prozent der Fälle die Angaben zur Identität nicht gestützt werden konnten.

Im Jahr 2021 führte das Ergebnis der Dialekterkennung-Auswertungen von persönlichen Erstantragstellenden ohne Pass/Passersatz dazu, dass in rund 78 Prozent der Fälle die Angaben zur Identität der Antragstellenden gestützt und in rund 22 Prozent der Fälle die Angaben zur Identität nicht gestützt werden konnten.

Im ersten Halbjahr 2022 führte das Ergebnis der Dialekterkennung-Auswertungen von persönlichen Erstantragstellenden ohne Pass/Passersatz dazu, dass in rund 76 Prozent der Fälle die Angaben zur Identität der Antragstellenden gestützt und in rund 24 Prozent der Fälle Angaben zur Identität nicht gestützt werden konnten.

Die Ergebnisse sind im Einzelnen den nachfolgenden Tabellen zu entnehmen.

Nach Staatsangehörigkeiten (HKL):

| 01.01. – 31.12.2020 | | |
|----------------------------|---------------------------------------|---|
| Gesamt | 78,7 % | 21,3 % |
| Staatsangehörigkeit | Report stützt Ast-Angaben in % | Report stützt Ast-Angaben nicht in % |
| Syrien | 73,0 % | 27,0 % |
| Algerien | 90,9 % | 9,1 % |
| Irak | 87,1 % | 12,9 % |
| Ungeklärt | 66,7 % | 33,3 % |
| Marokko | 89,1 % | 10,9 % |
| Tunesien | 93,8 % | 6,2 % |
| Libyen | 76,3 % | 23,7 % |
| Ägypten | 81,8 % | 18,2 % |
| Libanon | 83,3 % | 16,7 % |
| Sudan | 64,0 % | 36,0 % |
| Jemen | k. A. | k. A. |
| Staatenlos | k. A. | k. A. |
| Tschad | k. A. | k. A. |
| Eritrea | k. A. | k. A. |

| 01.01. – 31.12.2020 | | |
|--|-----------------------------------|--|
| Gesamt | 78,7 % | 21,3 % |
| Staatsangehörigkeit | Report stützt Ast-Angaben in % | Report stützt Ast-Angaben nicht in % |
| Pers. aus palästinensischen Gebieten (nicht als Staat anerkannt) | k. A.* | k. A. |
| Jordanien | k. A. | k. A. |
| Südsudan | k. A. | k. A. |
| Äthiopien | k. A. | k. A. |
| Iran | k. A. | k. A. |

* k. A. (keine Angaben): bei diesen HKL liegt eine absolute, statistische Bezugsgröße von weniger als zehn Fällen vor. Hier können daher keine aussagekräftigen relativen Angaben abgeleitet werden.

| 01.01. – 31.12.2021 | | |
|--|-----------------------------------|--|
| Gesamt | 77,5 % | 22,5 % |
| Staatsangehörigkeit | Report stützt Ast-Angaben in % | Report stützt Ast-Angaben nicht in % |
| Syrien | 69,1 % | 30,9 % |
| Algerien | 92,3 % | 7,7 % |
| Ungeklärt | 73,3 % | 26,7 % |
| Irak | 88,4 % | 11,6 % |
| Tunesien | 94,4 % | 5,6 % |
| Marokko | 95,5 % | 4,5 % |
| Jemen | 91,3 % | 8,7 % |
| Libyen | 71,4 % | 28,6 % |
| Libanon | 79,6 % | 20,4 % |
| Ägypten | 66,7 % | 33,3 % |
| Sudan | 53,8 % | 46,2 % |
| Pers. aus palästinensischen Gebieten (nicht als Staat anerkannt) | 70,8 % | 29,2 % |
| Staatenlos | k. A.* | k. A. |
| Eritrea | k. A. | k. A. |
| Jordanien | k. A. | k. A. |
| Somalia | k. A. | k. A. |
| Afghanistan | k. A. | k. A. |
| Südsudan | k. A. | k. A. |
| Tschad | k. A. | k. A. |
| Iran | k. A. | k. A. |
| Kuwait | k. A. | k. A. |
| ohne Angabe | k. A. | k. A. |

* k. A. (keine Angaben): bei diesen HKL liegt eine absolute, statistische Bezugsgröße von weniger als zehn Fällen vor. Hier können daher keine aussagekräftigen relativen Angaben abgeleitet werden.

| 01.01. – 30.06.2022 | | |
|----------------------------|-----------------------------------|---|
| Gesamt | 76,3 % | 23,7 % |
| Staatsangehörigkeit | Report stützt Ast-Angaben in % | Report stützt Ast-Angaben nicht in % |
| Syrien | 70,7 % | 29,3 % |
| Ungeklärt | 71,3 % | 28,7 % |

| 01.01. – 30.06.2022 | | |
|--|--------------------------------|--------------------------------------|
| Gesamt | 76,3 % | 23,7 % |
| Staatsangehörigkeit | Report stützt Ast-Angaben in % | Report stützt Ast-Angaben nicht in % |
| Algerien | 95,8 % | 4,2 % |
| Marokko | 97,0 % | 3,0 % |
| Tunesien | 92,9 % | 7,1 % |
| Irak | 84,9 % | 15,1 % |
| Libanon | 90,5 % | 9,5 % |
| Libyen | 78,0 % | 22,0 % |
| Ägypten | 59,4 % | 40,6 % |
| Jemen | 27,8 % | 72,2 % |
| Sudan | 29,4 % | 70,6 % |
| Staatenlos | 72,7 % | 27,3 % |
| Pers. aus palästinensischen Gebieten (nicht als Staat anerkannt) | k. A.* | k. A. |
| Iran | k. A. | k. A. |
| Somalia | k. A. | k. A. |
| Eritrea | k. A. | k. A. |
| Jordanien | k. A. | k. A. |
| Südsudan | k. A. | k. A. |

* k. A. (keine Angaben): bei diesen HKL liegt eine absolute, statistische Bezugsgröße von weniger als zehn Fällen vor. Hier können daher keine aussagekräftigen relativen Angaben abgeleitet werden.

3. Für welche Sprachen bzw. Dialekte kommt die Software aktuell zum Einsatz, und für welche weiteren Sprachen bzw. Dialekte ist ggf. eine Erweiterung geplant?

Die Software kommt aktuell für die arabischen Großdialekte Irakisch, Maghrebisch, Levantinisch, Golf und Ägyptisch zum Einsatz. Ende Juli 2022 wurde die Software um die Sprachen Dari, Persisch und Paschto erweitert. Eine Erweiterung um ein kurdisches Sprachmodell ist geplant.

4. Welche Kosten sind mit aktuellem Stand seit der Pilotierung im Jahr 2017 für die Anschaffung und den Einsatz der biometrischen Sprach- bzw. Dialektanalysesoftware entstanden und im laufenden Haushalt vorgesehen (bitte nach Jahren aufschlüsseln)?

Für die Sprach- bzw. Dialektanalysesoftware sind seit der Pilotierung folgende Kosten entstanden:

2017: 170 715,00 Euro

2018: 953 139,00 Euro

2019: 1.039 960,00 Euro

2020: 822 810,00 Euro

2021: 831 060,00 Euro

2022:

Im laufenden Haushalt 2022 wurden bislang 201 357,55 Euro in Anspruch genommen.

5. Um welche Software handelt es sich bei der lizenzierten Third-Party-Software, auf der das Sprachanalyse-System einer Präsentation des BAMF aus dem Jahr 2020 zufolge basiert (vgl. https://migrationnetwork.un.org/sites/g/files/tmzbd1416/files/docs/cdr_slides_tks_dias_common_language_analysis.pdf, S. 16)?

Es handelt sich um die Software Nuance Speech Suite.

6. Aus welchen Quellen stammen die für das Sprach- bzw. Dialekterkennungssystem verwendeten „existierenden Hintergrundsprachmodelle“ und die arabischen Dialektmodelle, die laut der genannten Präsentation vom BAMF trainiert wurden (vgl. ebd.)?

Für welche Sprachen liegen Modelle vor, und aus wie vielen Sprachbeispielen setzen sich die Trainingsdatensätze jeweils zusammen (bitte nach jeweiliger Sprache bzw. jeweiligem Dialekt aufschlüsseln)?

Der größte Teil der Trainingsdaten wurde vom Linguistic Data Consortium (LDC) bezogen. Ein kleiner Teil wurde über die Clickworker GmbH beschafft. Die arabischen Modelle sind darüber hinaus auch mit BAMF-eigenen anonymisierten Sprachproben trainiert worden.

Die Anzahl der Sprachproben können der folgenden Tabelle entnommen werden.

Sprachmodelle im Einsatz:

| Sprache | Anzahl Trainingsaudios |
|------------------------|-------------------------------|
| Arabisch-Levantinisch | 6.680 |
| Arabisch-Maghrebinisch | 1.823 |
| Arabisch-Irakisch | 1.158 |
| Arabisch-Ägyptisch | 1.968 |
| Arabisch-Golf | 1.812 |
| Persisch (Farsi) | 1.759 |
| Dari | 614 |
| Paschto | 5.404 |

Für weitere 19 Sprachen liegen Hintergrundmodelle vor. Diese dienen der Unterscheidung und Differenzierung der produktiven Sprachmodelle (arabische Dialekte sowie Dari, Persisch und Paschto).

Hintergrundmodelle durch Hersteller bereitgestellt:

| | |
|----------------|-------|
| Bulgarisch | 707 |
| Chinesisch | 1.053 |
| Deutsch | 435 |
| Englisch | 1.239 |
| Französisch | 327 |
| Hebräisch | 722 |
| Hindi | 461 |
| Italienisch | 908 |
| Japanisch | 660 |
| Koreanisch | 544 |
| Kroatisch | 745 |
| Niederländisch | 741 |
| Portugiesisch | 699 |
| Rumänisch | 746 |
| Russisch | 1.904 |

| | |
|---------------|-------|
| Spanisch | 762 |
| Tamilisch | 495 |
| Türkisch | 3.170 |
| Vietnamesisch | 489 |

7. Welche Fehlerquoten hat die Sprach- bzw. Dialekterkennungssoftware aktuell (bitte nach jeweiliger Sprache bzw. jeweiligem Dialekt aufschlüsseln), und wie haben sich die Fehlerquoten seit 2017 entwickelt (bitte nach jeweiliger Sprache bzw. jeweiligem Dialekt sowie nach Jahren aufschlüsseln)?

Eine Fehlerquote im Sinne der Fragestellung liegt nicht vor. Die Ergebnisse von DIAS stellen lediglich einen Hinweis bezüglich der Herkunft der Antragstellenden dar, der im Rahmen der Anhörung, die unter anderem auch der Klärung der Identität und Staatsangehörigkeit dient, berücksichtigt wird. Sie haben keinen Beweischarakter.

Die Erkennungsquote der arabischen Dialekte wird in Gesamttests gemessen. Von 2017 bis 2020 lag die Erkennungsquote für die arabischen Dialekte bei rund 80 Prozent. Durch erstmalige Trainings der Sprachmodelle in 2021 erreichen die arabischen Dialekte aktuell eine Erkennungsquote von rund 85 Prozent.

Die kürzlich neu eingeführten persischen Sprachmodelle (Dari und Persisch/Farsi) erreichen eine Erkennungsquote von 73,07 Prozent, und Paschto eine Erkennungsquote von 77,7 Prozent. Es ist davon auszugehen, dass sich die Erkennungsquote durch weitere geplante Trainings der Sprachmodelle noch erhöht.

8. Wo und wie lange werden Sprachaufzeichnungen der Antragstellenden gespeichert?

Die Aufbewahrung der Sprachproben erfolgt im BAMF. Die Löschung erfolgt entsprechend rechtlicher Aufbewahrungsvorgaben des Asylgesetzes (§ 16 Absatz 6 Asylgesetz) zehn Jahre nach unanfechtbarem Abschluss des Asylverfahrens.

9. Wurden die Sprachaufzeichnungen von Antragstellenden jemals an eine Institution außerhalb des BAMF und/oder an einen anderen Staat oder an sonstige Dritte weitergegeben, und wenn ja, an wen bzw. welche Institutionen und/oder Staaten, und aus welchem Grund?

DIAS-Sprachaufnahmen der Asylverfahren werden im Rahmen der Durchführung des Asylverfahrens nicht an Institutionen außerhalb des BAMF oder an einen anderen Staat gegeben. Vom Asylverfahren getrennte und anonymisierte BAMF-eigene Sprachproben wurden erstmals zur linguistischen Prüfung im Rahmen eines Pilotprojekts an die Niederlande übermittelt.

Im Übrigen wird auf die Antwort zu Frage 10 verwiesen.

10. Können Antragstellende gegen ein falsches Ergebnis der Spracherkennungssoftware Einspruch erheben, und wenn ja, welches Verfahren ist hierfür vorgesehen?

Ist es zum Beispiel möglich, die Durchführung eines weiteren Tests zu beantragen?

Können Antragstellende zusätzliche Sprachanalyse durch einen Linguisten oder eine Linguistin beantragen?

Antragstellende werden vor Erhebung der Sprachprobe über den Ablauf und über den Sinn und Zweck der Maßnahme informiert. Dies umfasst die Information darüber, dass das Ergebnis der Sprachprobe (Report) zur Akte genommen wird. Falls die Ergebnisse im Widerspruch zu den Angaben des Antragstellers oder der Antragstellerin stehen bzw. sich daraus Klärungsbedarf ergibt, erfolgt eine entsprechende Aufklärung des Sachverhalts in der Anhörung.

Sofern nach Durchführung der Anhörung Zweifel bezüglich der Herkunft oder Staatsangehörigkeit des Antragsstellenden verbleiben, besteht in Ausnahmefällen die Möglichkeit, ein separates Sprachgutachten zu veranlassen. Diesem Sprachgutachten liegt nicht die DIAS-Sprachaufnahme zugrunde. Die erforderliche Sprachprobe wird in einem eigens dafür anberaumten Termin aufgezeichnet.

11. Welches Unternehmen hat die Software, die für den Sprach- bzw. Dialekterkennungsassistenten verwendet wird, entwickelt und ist momentan deren Eigentümer?

Der Dienstleister heißt „Nuance“. Die Software wird über Lizenzen bezogen.

12. Arbeitet das BAMF mit anderen EU-Staaten zusammen, um den Sprach- bzw. Dialekterkennungsassistenten gemeinsam zu entwickeln und zu nutzen, und wenn ja, mit welchen?

Wird die Software des BAMF auch in anderen EU-Mitgliedstaaten verwendet, und wenn ja, in welchen?

Im Rahmen eines europäischen Pilotprojekts wird derzeit eine generelle europäische Zusammenarbeit bei Sprachanalysen diskutiert und erprobt. An der Diskussion beteiligt sind Österreich, Finnland, Norwegen, Schweden, Litauen, Griechenland, Schweiz und die Niederlande. Im Übrigen wird auf die Antwort zu Frage 13 verwiesen.

Welche Staaten die Software von Nuance nutzen ist nicht bekannt.

13. Mit welchen anderen Staaten ist das BAMF zum Einsatz von Spracherkennungstools im Austausch?

Hat das BAMF den Sprach- bzw. Dialekterkennungsassistenten anderen EU- und Nicht-EU-Ländern zur Nutzung zur Verfügung gestellt?

In welchen Ländern haben Vertreterinnen und Vertreter des BAMF die Nutzung des Sprach- bzw. Dialekterkennungsassistenten demonstriert?

Das BAMF ist nur mit den in der Antwort zu Frage 12 genannten Ländern zum Einsatz von Spracherkennungstools im Austausch. Das Spracherkennungstool des BAMF selbst wurde dabei nicht zur Nutzung zur Verfügung gestellt. Die Anwendung des Programms erfolgte im Rahmen des Pilotprojekts BAMF-

intern. Das Spracherkennungstool DIAS wurde in den Ländern Norwegen und Schweiz demonstriert.

14. Findet mittlerweile eine wissenschaftliche Begleitung der Dialektanalyse statt, und wenn ja, mit welchen wissenschaftlichen Partnern bzw. Partnerinnen und Ergebnissen (vgl. Antwort zu Frage 11c auf Bundestagsdrucksache 19/190)?

Die Einbindung einer deutschen Hochschule befindet sich in Vorbereitung.

15. Was sind die Ziele eines vom BAMF koordinierten europäischen Projekts, mit dem u. a. untersucht werden soll, ob Agenturen innerhalb der EU technisch und rechtlich in der Lage sind, Audiodateien auszutauschen (<https://r.algorithmwatch.org/nl2/x3ysn/1ox7.html>)?

- a) Zu welchem Zweck wollen die Asyl- bzw. Migrationsbehörden der EU-Mitgliedstaaten, die in dieser Studie teilnehmen, Audiodateien austauschen?

Geht es dabei darum, mehr Trainingsdaten für die Dialekterkennungssoftware zu sammeln?

Die Fragen 15 und 15a werden gemeinsam beantwortet.

Bei dem Austausch von Audiodateien geht es um Möglichkeiten der gegenseitigen Unterstützung im Bereich der Sprachanalyse, da die beteiligten Staaten vor vergleichbaren Herausforderungen bei der Herkunftsklärung von Asylantragstellenden stehen. Insbesondere sollen der Prozess des Sprachprobenaustauschs erprobt und dabei rechtliche, technische und organisatorische Fragestellungen erörtert werden. Das Pilotprojekt hat dabei nicht zum Ziel, Trainingsdaten für die Spracherkennungssoftware zu sammeln.

- b) Welche Art von Sprachaufnahmen bzw. Audiodateien sind gemeint?

Geht es dabei um Sprachaufnahmen aus abgeschlossenen oder laufenden Asylverfahren?

Es wurden seitens BAMF vom Asylverfahren getrennte und anonymisierte Sprachproben verwendet. Der Verfahrensstand ist aufgrund der Anonymisierung nicht zu ermitteln. Informationen zum Verfahrensstand von Sprachproben anderer Staaten liegen nicht vor.