

## **Kleine Anfrage**

**der Abgeordneten Clara Bünger, Nicole Gohlke, Gökay Akbulut, Anke Domscheit-Berg, Dr. André Hahn, Ina Latendorf, Cornelia Möhring, Petra Pau, Sören Pellmann, Heidi Reichinnek, Martina Renner, Dr. Petra Sitte, Kathrin Vogler und der Fraktion DIE LINKE.**

### **Einsatz von Dialekterkennungssoftware im Bundesamt für Migration und Flüchtlinge**

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/>; <https://www.heise.de/newsticker/meldung/Sherlock-Bamf-Asyltaeuschern-auf-der-Spur-4257898.html>).

Wir fragen die Bundesregierung:

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?
2. In wie vielen Fällen wurden die Angaben der Asylsuchenden bestätigt bzw. widerlegt (bitte nach Herkunftsländern und Jahren aufschlüsseln)?
3. Für welche Sprachen bzw. Dialekte kommt die Software aktuell zum Einsatz, und für welche weiteren Sprachen bzw. Dialekte ist ggf. eine Erweiterung 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)?
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/tmzbdl416/files/docs/cdr\\_slides\\_tks\\_dias\\_common\\_language\\_analysis.pdf](https://migrationnetwork.un.org/sites/g/files/tmzbdl416/files/docs/cdr_slides_tks_dias_common_language_analysis.pdf), S. 16)?
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)?
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)?
8. Wo und wie lange werden Sprachaufzeichnungen der Antragstellenden gespeichert?
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?
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?
11. Welches Unternehmen hat die Software, die für den Sprach- bzw. Dialekterkennungsassistenten verwendet wird, entwickelt und ist momentan deren Eigentümer?

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?
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?
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)?
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?
- b) Welche Art von Sprachaufnahmen bzw. Audiodateien sind gemeint?
- Geht es dabei um Sprachaufnahmen aus abgeschlossenen oder laufenden Asylverfahren?

Berlin, den 11. August 2022

**Amira Mohamed Ali, Dr. Dietmar Bartsch und Fraktion**

