

Asylverfahren: BAMF weitet automatische Sprachanalyse aus

[Originalartikel](#)

[Backup](#)

84.583 Menschen stellten im ersten Halbjahr 2022 einen Asylerstantrag in Deutschland. 7.808 Mal führte das Bundesamt für Migration und Flüchtlinge (BAMF) eine automatische Dialektanalyse durch. Seit 2017 nutzt das BAMF Software, um auf Basis kurzer Sprechproben Hinweise auf die mutmaßliche Muttersprache einer Person zu bekommen und so die Plausibilität des Asylantrags zu überprüfen. Doch Sprach-Expert:innen kritisieren das System: Sprache kann sich nicht an Ländergrenzen und sie verändert sich im Lauf eines Lebens. Bisher wandte das BAMF seine Dialektanalyse nur bei Personen an, die einen arabischen Dialekt sprechen und ohne Pass nach Deutschland kommen. Doch laut der Antwort des Bundesinnenministeriums auf eine Kleine Anfrage der Linken-Abgeordneten Clara Banger hat das BAMF seine Modelle mittlerweile auf andere Sprachen ausgedehnt. Ende Juli 2022 wurde die Software um die Sprachen Dari, Persisch und Paschto erweitert, heißt es in der Antwort. Man plane außerdem, das Dialekterkennungssystem um ein kurdisches Sprachmodell zu erweitern. Dari und Paschto sind Amtssprachen in Afghanistan, Persisch (Farsi) ist die Amtssprache im Iran. Über vier Millionen Euro seit 2017 hat das BAMF etwas mehr als vier Millionen Euro für das System ausgegeben, davon 201.357,55 Euro in diesem Jahr. Davon zahlt das BAMF unter anderem die Lizenzen für die Software Nuance Speech Suite. Sie wird vom US-amerikanischen Hersteller Nuance Communications entwickelt, einem großen Anbieter von Sprachverarbeitungstechnologie. Die Software verfügt über Hintergrundsprachmodelle für 19 Sprachen, darunter Deutsch, Türkisch, Tamilisch oder Hindi.

Hier klicken, um den Inhalt von datawrapper.dwcdn.net anzuzeigen

Inhalt von datawrapper.dwcdn.net immer anzeigen

Das BAMF nutzte die Software aber vor allem bei Geflüchteten aus arabischsprachigen Ländern. Es gibt viele verschiedene arabische Dialekte, von denen soll das System unterscheiden können: levantinisches, ägyptisches, irakisches, maghrebinisches und Golf-Arabisch. Dafür hat das Bundesamt zusätzliche Sprachproben in das System eingespeist. Anhand dieser Sprachproben kann das System trainiert werden, verschiedene Sprachen auseinanderzuhalten.

Einen Großteil der Sprachproben habe die Behörde vom Linguistic Data Consortium (LDC) bezogen, heißt es in der Antwort. Das LDC ist ein Zusammenschluss von Forschungseinrichtungen und Unternehmen und an der University of Pennsylvania angesiedelt. Zusätzlich hat die Asylbehörde einige Beispiele für die arabischen Dialekte aus BAMF-eigenen anonymisierten Sprachproben genutzt. Und: Ein kleiner Teil wurde über die Clickworker GmbH beschafft, so die Antwort. Die Clickworker GmbH aus Essen wirbt auf

seiner Website damit, über einen Pool aus <a href=„<https://www.clickworker.com/clickworker-crowd/>“>3,6 Millionen Clickworkern in 136 Ländern zu verfügen. Die nicht-angestellten Arbeiter:innen können über eine App Aufgaben lösen und werden dann für gelöste Tasks bezahlt.</p><p>Grob kann man sagen: Mit je mehr Beispielen ein System für die unterschiedlichen Sprachen und Dialekte trainiert wurde, desto besser kann es sie auseinanderhalten. Doch diese Beispiele sind im Fall der Dialekterkennung des BAMF noch ungleich verteilt. Während für levantinisches Arabisch 6.680 Trainingsaudiodateien hinterlegt sind, sind es für irakisches Arabisch nur 1.158. Ersteres sprechen etwa Menschen aus Syrien oder palästinischen Gebieten. 1.759 Proben kennt das System für Farsi, für das sehr ähnliche Dari jedoch lediglich 614. Die Antwort gibt jedoch keine Information darüber, wie lang die verwendeten Proben sind, deshalb ist es schwer, die reinen Zahlen zu vergleichen.</p><p>Doch wie zuverlässig sind die Analysen? Wie die Fehlerquoten für die einzelnen Dialekte aussehen, beantwortet das Ministerium nicht direkt. „Eine Fehlerquote im Sinne der Fragestellung liegt nicht vor“, so die Antwort. Stattdessen spricht das Ministerium lieber von einer „Erkennungsquote“. Die habe von 2017 bis 2020 für die arabischen Dialekte bei 80 Prozent gelegen. „Durch erstmalige Trainings der Sprachmodelle in 2021“ erreiche man mittlerweile 85 Prozent.</p><p>Bei den neu aufgenommenen Dialekten erkenne die Software bei Dari und Farsi in 73,07 Prozent der Fälle erfolgreich, bei Paschto in 77,7 Prozent. Bei den neu hinzugefügten Sprachen ist das System also offenbar weniger zuverlässig. Es seien jedoch weitere Trainings geplant und man gehe davon aus, dass sich dadurch Erkennungsquote erhöht.</p><p>Nimmt man die 7.808 durchgeführten Dialektanalysen aus dem ersten Halbjahr, ergeben sich aus der Erkennungsquote für arabische Dialekte 1.562 Antragsteller:innen, bei denen die Sprache nicht erfolgreich erkannt wurde. Ob die Ergebnisse in diesen Fällen offensichtlich und für Entscheider:innen leicht erkennbar falsch waren, geht aus der Antwort nicht hervor. Das wäre etwa der Fall, wenn bei einer arabischsprachigen Person „Deutsch“ als wahrscheinlichste Muttersprache herauskommt. Dann könnten die BAMF-Mitarbeitenden das leicht als unsinniges Ergebnis ausschließen.</p><h3>Fragezeichen bei sudanesischen Antragsteller:innen</h3><p>Kommt das System aber bei einem Geflüchteten aus Syrien zum Ergebnis, die wahrscheinlichste Muttersprache sei ägyptisches Arabisch, kann es kompliziert werden und ungerechtfertigte Zweifel an den Angaben des Asylsuchenden hervorrufen. Dann müssen die Entscheider:innen in der Anhörung versuchen, diese Widersprüche aufzuklären.</p><p>Das Ministerium gibt an, die Ergebnisse der Dialektanalyse hätten „in rund 76 Prozent der Fälle die Angaben zur Identität der Antragstellenden gestützt“, in 24 Prozent nicht. Schaut man sich die Aufteilung auf verschiedene Herkunftsländer an, fällt ein Land besonders auf: In 70,6 Prozent der Fälle sei die Angabe nicht bestätigt worden, wenn das Herkunftsland Sudan war. Das verwundert kaum. Im Sudan werden <a href=„<https://www.refworld.org/docid/3ae6ab7794.html>“>Dutzende Sprachen gesprochen. Das sudanesische Arabisch ist <a href=„<https://plc.sas.upenn.edu/sudanese-arabic>“>ein eigener Dialekt, der dem ägyptischen Dialekt ähnlich ist. In der Trainingsdatenaufstellung kommt er nicht gesondert vor. Ebenso wie <a href=„<https://de.wikipedia.org/wiki/Jemenitisch-Arabisch>“>der jemenitische Dialekt, hier werden in ähnlich vielen Fällen die Angaben nicht gestützt.</p><div class=„embed-privacy-container is-disabled embed-default“ id=„oembed_2f34bd3129fa5134fc906d5c59f77215“ data-embed-provider=„“><div class=„embed-privacy-overlay“><div class=„embed-privacy-inner“><p>Hier klicken, um den Inhalt von datawrapper.dwcdn.net anzuzeigen</p><p class=„embed-privacy-input-wrapper“><label for=„embed-privacy-store-2f34bd3129fa5134fc906d5c59f77215“ class=„embed-privacy-label“ data-embed-provider=„“>Inhalt von datawrapper.dwcdn.net immer

anzeigen

Wie geht das BAMF mit einer solchen Unsicherheit um, die die Ergebnisse mit sich bringen? Die Ergebnisse der Dialekterkennung seien lediglich Hinweise, hätten aber keinen Beweischarakter, so das Ministerium. Die Fragestellerin Clara Bünger überzeugt das nicht: „Gerade bei hoher Arbeitsbelastung, unter Zeitdruck oder bei mangelnder Schulung besteht die Gefahr, dass BAMF-Mitarbeiter:innen die Ergebnisse der Software letztlich doch wie unumstößliche Fakten behandeln, um schnell eine Entscheidung treffen zu können“, so die Bundestagsabgeordnete gegenüber netzpolitik.org <a href=„<https://algorithmwatch.org/en/bamf-dialect-recognition-expanding/>“>und AlgorithmWatch. Dann sei es für Asylsuchende „äußerst schwierig, daraus resultierende Fehleinschätzungen anzufechten“.

Bünger kritisiert auch die „Misstrauenskultur“ im BAMF, die ihrer Auffassung nach der Ursprung für die Einführung der technischen Assistenzsysteme ist. „Dabei wird unterstellt, dass Asylsuchende massenhaft falsche Angaben zu ihrer Identität und Herkunft machen würden, wofür es jedoch keinerlei Beweise gibt.“ Statt „fehleranfälliger und teurer technischer Lösungen“ wünscht sie sich „eine angemessene Schulung der Mitarbeiter:innen im BAMF, um faire Asylverfahren zu ermöglichen“.

Schon 2017 kündigte das BAMF an, die Dialektanalyse im Jahr 2018 <a href=„<https://netzpolitik.org/2018/die-it-tools-des-bamf-fehler-vorprogrammiert/>“>wissenschaftlich begleiten lassen zu wollen. Es sollten „technologische Fragestellungen“ und „weiterführende sprachwissenschaftliche Aspekte“ untersucht werden. Das könnte zusätzlich Aufschluss darüber geben, ob die computergenerierten Analysen überhaupt geeignet sein können, belastbare Hinweise für Asylverfahren zu geben. Eine wissenschaftliche Begleitung gab es noch nicht. Laut Innenministerium sei mittlerweile „die Einbindung einer deutschen Hochschule“ in Vorbereitung. <a href=„<https://www.bamf.de/SharedDocs/Meldungen/DE/2017/20170726-am-vorstellung-modellprojekt-bamberg.html?nn=282388>“>Mehr als fünf Jahre, nachdem die Dialekterkennung erstmals eingesetzt wurde, bei mittlerweile zehntausenden Menschen.

 src=„<https://vg03.met.vgwort.de/na/94c08d494327422ea26dba9acc4e0fb8>“ width=„1“ height=„1“ alt=„“ referrerpolicy=„no-referrer“ />

From:
<https://schnipsel.qgelm.de/> - Qgelm

Permanent link:
https://schnipsel.qgelm.de/doku.php?id=wallabag:wb2asylverfahren_bamf-weitet-automatische-sprachanalyse-aus

Last update: 2025/06/27 11:17

