

# Asylverfahren: BAMF weitet automatische Sprachanalyse aus

Originalartikel

Backup

<html> <p>84.583 Menschen stellten im ersten Halbjahr 2022 einen Asylerstantrag in Deutschland. 7.808 Mal f&#252;hrte das Bundesamt f&#252;r Migration und Fl&#252;chtlinge (BAMF) eine automatische Dialektanalyse durch. Seit 2017 nutzt das BAMF Software, um auf Basis kurzer Sprechproben Hinweise auf die mutma&#223;liche Muttersprache einer Person zu bekommen &#8211; und so die Plausibilit&#228;t des Asylantrags zu &#252;berpr&#252;fen. Doch Sprach-Expert:innen kritisieren das System: Sprache <a href=„<https://www.vice.com/de/article/a3q8wj/fluechtlinge-bamf-sprachanalyse-software-entscheidet-asyl>“>h&#228;lt sich nicht an L&#228;ndergrenzen</a> und sie <a href=„<https://www.zeit.de/digital/internet/2017-03/bamf-asylbewerber-sprach-analyse-software-compilerlinguistik/komplettansicht>“>ver&#228;ndert sich im Lauf eines Lebens</a>.</p><p>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 <a href=„<https://dip.bundestag.de/vorgang/einsatz-von-dialekterkennungssoftware-im-bundesamt-f%C3%BCr-migration-und-fl%C3%BCchtlinge/290504>“>auf eine Kleine Anfrage der Linken-Abgeordneten Clara B&#252;nger</a> hat das BAMF seine Modelle mittlerweile auf andere Sprachen ausgedehnt.</p><p>&#8222;Ende Juli 2022 wurde die Software um die Sprachen Dari, Persisch und Paschto erweitert&#8220;, hei&#223;t es in der Antwort. Man plane au&#223;erdem, das Dialekterkennungssystem um ein kurdisches Sprachmodell zu erweitern. Dari und Paschto sind Amtssprachen in Afghanistan, Persisch (Farsi) ist die Amtssprache im Iran.</p><h3>&#220;ber vier Millionen Euro seit 2017</h3><p>Seit dem Start der Dialekterkennung hat das BAMF etwas mehr als vier Millionen Euro f&#252;r das System ausgegeben, davon 201.357,55 Euro in diesem Jahr. Davon zahlt das BAMF unter anderem die Lizenzen f&#252;r die Software Nuance Speech Suite. Sie wird vom US-amerikanischen Hersteller Nuance Communications entwickelt, einem gro&#223;en Anbieter von Sprachverarbeitungstechnologie. Die Software verf&#252;gt &#252;ber Hintergrundsprachmodelle f&#252;r 19 Sprachen, darunter Deutsch, Türkisch, Tamilisch oder Hindi.</p><div class=„embed-privacy-container is-disabled embed-default“ id=„oembed\_ffe2f05b3af376dd5974384dca292d40“ data-embed-provider=„><div class=„embed-privacy-overlay“><div class=„embed-privacy-inner“><p>Hier klicken, um den Inhalt von datawrapper.dwdcdn.net anzuzeigen</p><p class=„embed-privacy-input-wrapper“><label for=„embed-privacy-store-ffe2f05b3af376dd5974384dca292d40“ class=„embed-privacy-label“ data-embed-provider=„>Inhalt von datawrapper.dwdcdn.net immer anzeigen</label></p></div></div></div><p>Das BAMF nutzte die Software aber vor allem bei Gefl&#252;chteten aus arabischsprachigen L&#228;ndern. Es gibt viele verschiedene arabische Dialekte, f&#252;nf davon soll das System unterscheiden k&#246;nnen: levantinisches, &#228;gyptisches, irakisches, magrebinisches und Golf-Arabisch. Daf&#252;r hat das Bundesamt zus&#228;tzliche Sprachproben in das System eingespeist. Anhand dieser Sprachproben kann das System trainiert werden, verschiedene Sprachen auseinanderzuhalten.</p><p>Einen Gro&#223;teil der Sprachproben habe die Beh&#246;rde vom Linguistic Data Consortium (LDC) bezogen, hei&#223;t es in der Antwort. Das LDC ist ein Zusammenschluss von Forschungseinrichtungen und Unternehmen und an der University of Pennsylvania angesiedelt. Zus&#228;tzlich hat die Asylbeh&#246;rde einige Beispiele f&#252;r die arabischen Dialekte aus &#8222;BAMF-eigenen anonymisierten Sprachproben&#8220; genutzt. Und: &#8222;Ein kleiner Teil wurde &#252;ber die Clickworker GmbH beschafft&#8220;, 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</a> zu verfügen. Die nicht-angestellten Arbeiter:innen können über eine App Aufgaben übernommen 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 palästinensischen 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 erreichte man mittlerweile 85 Prozent.</p><p>Bei den neu aufgenommenen Dialekten erkennt 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 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 „gyptisches 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 Herkunftsänder 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</a> gesprochen. Das sudanische Arabisch ist <a href= „<https://plc.sas.upenn.edu/sudanesearabic>“>ein eigener Dialekt</a>, der dem „gyptischen Dialekt“ ähnlich ist. In der Trainingsdatenauflistung kommt er nicht gesondert vor. Ebenso wie <a href= „<https://de.wikipedia.org/wiki/Jemenitisch-Arabisch>“>der jemenitische Dialekt</a>, 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.dwdcdn.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.dwdcdn.net immer

anzeigen</label></p></div></div></div><p>Wie geht das BAMF mit einer solchen Unsicherheit um, die die Ergebnisse mit sich bringen? Die Ergebnisse der Dialekterkennung seien lediglich Hinweise, h&#228;tten aber keinen Beweischarakter, so das Ministerium. Die Fragestellerin Clara B&#252;nger &#252;berzeugt das nicht: &#8222;Gerade bei hoher Arbeitsbelastung, unter Zeitdruck oder bei mangelnder Schulung besteht die Gefahr, dass BAMF-Mitarbeiter:innen die Ergebnisse der Software letztlich doch wie unumst&#246;liche Fakten behandeln, um schnell eine Entscheidung treffen zu k&#246;nnen&#8220;, so die Bundestagsabgeordnete gegen&#252;ber netzpolitik.org <a href=„<https://algorithmwatch.org/en/bamf-dialect-recognition-expanding/>“>und AlgorithmWatch</a>. Dann sei es f&#252;r Asylsuchende &#8222;u&#223;erst schwierig, daraus resultierende Fehleinsch&#228;tzungen anzufechten&#8220;.</p><p>B&#252;nger kritisiert auch die &#8222;Misstrauenskultur&#8220; im BAMF, die ihrer Auffassung nach der Ursprung f&#252;r die Einf&#252;hrung der technischen Assistenzsysteme ist. &#8222;Dabei wird unterstellt, dass Asylsuchende massenhaft falsche Angaben zu ihrer Identit&#228;t und Herkunft machen w&#252;rden, wof&#252;r es jedoch keinerlei Beweise gibt.&#8220; Statt &#8222;fehleranf&#228;lliger und teurer technischer L&#246;sungen&#8220; w&#252;nscht sie sich &#8222;eine angemessene Schulung der Mitarbeiter:innen im BAMF, um faire Asylverfahren zu erm&#246;glichen&#8220;.</p><p>Schon 2017 k&#252;ndigte das BAMF an, die Dialektanalyse im Jahr 2018 <a href=„<https://netzpolitik.org/2018/die-it-tools-des-bamf-fehler-vorprogrammiert/>“>wissenschaftlich begleiten lassen</a> zu wollen. Es sollten &#8222;technologische Fragestellungen&#8220; und &#8222;weiterf&#252;hrende sprachwissenschaftliche Aspekte&#8220; untersucht werden. Das k&#246;nnte zus&#228;tzlich Aufschluss dar&#252;ber geben, ob die computergenerierten Analysen &#252;berhaupt geeignet sein k&#246;nnen, belastbare Hinweise f&#252;r Asylverfahren zu geben. Eine wissenschaftliche Begleitung gab es noch nicht. Laut Innenministerium sei mittlerweile &#8222;die Einbindung einer deutschen Hochschule&#8220; in Vorbereitung. <a href=„<https://www.bamf.de/SharedDocs/Meldungen/DE/2017/20170726-am-vorstellung-modellprojekt-bamberg.html?nn=282388>“>Mehr als f&#252;nf Jahre</a>, nachdem die Dialekterkennung erstmals eingesetzt wurde, bei mittlerweile zehntausenden Menschen.</p><p><img src=„<https://vg03.met.vgwort.de/na/94c08d494327422ea26dba9acc4e0fb8>“ width=„1“ height=„1“ alt=„“ referrerpolicy=„no-referrer“ /></p> </html>

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

Permanent link:  
[https://schnipsl.qgelm.de/doku.php?id=wallabag:wb2asylverfahren\\_bamf-weitet-automatische-sprachanalyse-aus](https://schnipsl.qgelm.de/doku.php?id=wallabag:wb2asylverfahren_bamf-weitet-automatische-sprachanalyse-aus)

Last update: 2025/06/27 11:17

