

Transkribieren geht über Studieren

Das Abtippen von Interviews gehört zu den mühsameren Aufgaben von Journalistinnen und Journalisten. Moderne KI-Tools können helfen, aber bei manchen ist Vorsicht geboten.

VON RETO VOGT

Bis ein einstündiges Interview vom «Bändli» abgetippt und pfannenfertig im Redaktionssystem ist, dauert es gut und gerne einen halben Tag oder länger. Abspielen, stoppen, zurückspulen, vorspulen, abspielen ... Sie kennen das.

Doch jetzt gibt es KI-gestützte Transkriptionstools, die uns das mühselige Abtippen abnehmen und das Gesagte praktisch automatisch transkribieren. Wer sie nutzt, ist oft begeistert und um Faktor vier bis fünf schneller als mit der klassischen Methode.

Die Werkzeuge funktionieren alle ähnlich: Audiodatei hochladen, kurz warten und dann ist der Text schon da. Die Transkripte sind dabei interaktiv: Einzelne Passagen lassen sich anklicken, um direkt die zugehörige Passage abzuspielen. Mittlerweile sind die Resultate sogar bei schweizerdeutschen Gesprächen so präzise, dass Redaktionen dabei viele Arbeitsstunden einsparen (die sie theoretisch in den Feinschliff des Texts investieren können).

Doch mit den neuen Möglichkeiten kommen neue Fragen: Dürfen Gespräche überhaupt automatisiert verschriftlicht werden? Was passiert mit der Stimme der Gesprächspartnerinnen und Gesprächspartner, wenn sie von einer Software transkribiert wird? Und wie stellen Redaktionen sicher, dass heikle oder später nicht publizierte Informationen aus den Interviews nicht in falsche Hände geraten?

Was rechtlich gilt

Interviewte Personen haben das Recht an der eigenen Stimme. Sie müssen deshalb immer explizit zustimmen, dass erstens das Gespräch aufgezeichnet und es zweitens mit einem KI-Tool verarbeitet werden darf.

Dieser Konsens ist noch wichtiger, wenn das genutzte Tool zwischen den verschiedenen Sprecherinnen und Sprechern unterscheiden kann und ihnen das Gesagte richtig zuordnet. Das ist eine sehr praktische Funktion, weil sie die Nachbearbeitung erheblich vereinfacht. Aber die Unterscheidung von Stimmprofilen gilt datenschutzrechtlich als Verarbeitung biometrischer Daten, die laut Gesetz besonders schützenswert sind - unabhängig davon, ob ein dauerhaftes Stimmprofil angelegt wird oder nicht.

Die meisten der weitverbreiteten Werkzeuge (siehe Tabelle) bringen diese Funktion mit. Und solange die Zustimmung erfolgt, ist das rechtlich auch kein Problem. Typischerweise sprechen interviewte Personen ohnehin regelmässig in der Öffentlichkeit, und Sinn und Zweck des Gesagten ist es sowieso, publiziert zu werden. Aber es kann dennoch zu heiklen Situationen kommen. Zum Beispiel wenn der Interviewte Informationen während des Gesprächs als «off the record» deklariert. Oder bei Informantengesprächen, wenn die Auskunftspersonen anonym bleiben sollen: Dann kann die technische Möglichkeit der Stimmerkennung die Anonymität gefährden - selbst dann, wenn die Profile später gelöscht werden.

Anwendung in der Praxis

In der Praxis heisst das: Journalist:innen sollten Aufnahmegeräte stoppen, wenn Gesprächspartner etwas ausdrücklich als «off the record» kennzeichnen. Sie sollten zudem ganz auf Aufnahmen verzichten, wenn sie mit anonymen Informanten sprechen. In diesen Fällen überwiegt der Schutz der Quelle den Komfort der Technik.

Wie also gehen grosse Schweizer Medienhäuser mit diesen Herausforderungen um? Wie setzen sie Transkriptionstools konkret ein - und wo ziehen sie die Grenzen? Fakt ist: Die Tools sind im Alltag angekommen: «Unser Transkribierungstool wird selbst von Personen genutzt, die sonst keine grossen KI-Fans sind», sagt Thomas Benkö, AI Innovation Lead Newsroom beim «Blick». Der Verlag setzt auf das Open-Source-Modell Whisper - wie ChatGPT von OpenAI entwickelt - und betreibt es auf eigenen Servern in der Cloud. Das sei zwar aufwändiger, dafür aber auch sicherer: «Weil keine Informationen auf fremden Servern verarbeitet werden», so Benkö.

Auch bei Tamedia ist KI-gestützte Transkription längst fester Bestandteil im Redaktionsalltag. Seit 2019 setzt das Unternehmen auf den europäischen Anbieter Happyscribe, der unter anderem auch die Technologie von Whisper integriert hat. Im Unterschied zu Ringier betreibt Tamedia die Software nicht auf eigenen Servern, stattdessen «nutzen wir das Angebot direkt auf

Audiodatei hochladen, kurz warten und dann ist der Text schon da.



Weitverbreitete Tools in der Übersicht

Zahlreiche KI-Tools versprechen schnelle, präzise Transkriptionen - doch nicht alle eignen sich gleich gut für den journalistischen Einsatz. Die folgende Auswahl zeigt etablierte Angebote, die sich in Redaktionen bewährt haben. Gemeinsam ist ihnen: Sie kommen auch mit Schweizerdeutsch zurecht. Unterschiede gibt es bei Datenschutz, Hosting und Preismodellen - und genau hier lohnt sich ein genauer Blick.

Tool	Eigenschaften	Preis
Whisper	Open-Source-Modell von OpenAI, kann lokal oder auf eigenen Servern betrieben werden. Mehraufwand.	Nutzung gratis, es fallen allenfalls Hostingkosten an.
Happyscribe.com	Europäischer Anbieter, Datenverarbeitung in der EU. Datenschutzkonform.	Je nach Paket zwischen 9 und 49 Dollar monatlich.
Turboscribe.ai	US-Anbieter, Datenverarbeitung in den USA. Vorsicht bei Personendaten.	Drei Audiodateien (je 30 Minuten) pro Tag kostenlos, unlimitiert für 120 Dollar jährlich.
Töggli.ch	Schweizer Anbieter. Speicherung und Verarbeitung in der Schweiz. Spezialisiert auf Dialekte und CH-Vokabular.	Abrechnung nach Minuten: z. B. 108 Franken für 2 Stunden Schweizer Dialekt.

ihrer Plattform, die sich an strenge europäische Datenschutzregelungen hält», sagt Nadia Kohler, Head AI Lab bei Tamedia.

In beiden Redaktionen sind Interviews der Haupteinsatzzweck der Software. Sie können aber auch dazu genutzt werden, um fremdsprachige Videos aus dem Internet zu Recherchezwecken zu transkribieren und zu übersetzen oder um Untertitel für Videos zu erstellen.

Transkribieren bleibt Kopfsache

KI-Tools wie Whisper oder Happyscribe nehmen Journalistinnen und Journalisten viel Arbeit ab. Doch der Einsatz dieser Werkzeuge ist kein Selbstläufer. Erstens

bedingen sie Vorwissen in Sachen Datenschutz und den richtigen Umgang damit, zweitens existiert kein Tool, das ein pfannenfertiges Interview ausspuckt. Nach wie vor macht erst das journalistische Handwerk - Nachbearbeiten des Texts, Umsortieren von Antworten, sprachlicher Feinschliff - aus einem guten Transkript ein lesenswertes Gespräch.

Reto Vogt ist Studienleiter Digitale Medien und KI am MAZ und freier Journalist. mazz.ch