



Examining the opportunities and challenges of using speech recognition (SR) technology in implementing oral history interviews from the perspective of oral history experts

Fatemeh Ahmadi

MA in Archival Studies, Al-Zahra University, Tehran, Iran

Abstract: The present study seeks to identify the capabilities and challenges of using speech recognition technology in implementing oral history interviews by studying and examining it, as well as using the experience and opinions of experts in the field of oral history. The method used to conduct the research is descriptive-survey. The tool of data collection is a questionnaire made by the researcher. Due to the difficulty of the researcher's access to the research community, the snowball sampling method was used to collect data. In this way, the 10 initial experts who cooperated in answering the questionnaire were asked to invite other experts they knew in this field to cooperate in answering the questionnaire. Ultimately, 30 oral history experts cooperated in this research by answering the questionnaire. The findings answer four questions regarding 1. Assessing the current status of implementing oral history interviews, 2. Assessing the applications of speech recognition technology, 3. The possibilities and conditions necessary for using speech recognition technology in transcribing oral history interviews, and 4. Identifying the opportunities and challenges of using this technology in transcribing oral history interviews. 67% of experts evaluate the current state of oral history interview transcription as undesirable. Among the evaluated criteria for technology applications, "saving time" with an average value of 3.7 is the most important according to experts. Among the criteria of facilities and conditions evaluated for the use of technology in oral history interview transcription, "security conditions and compliance with the principle of confidentiality" with an average value of 2.8 is considered one of the challenges of using this technology. "Inaccuracy in transcribing interviews with local dialects, etc." is another challenge of this technology.

Key Words: Oral History, Interview, Transcription, Speech Recognition, Artificial Intelligence.

بررسی فرصت‌ها و چالش‌های استفاده از فناوری تشخیص گفتار (SR) در پیاده‌سازی مصاحبه‌های تاریخ شفاهی از دیدگاه

صاحب‌نظران تاریخ شفاهی

فاطمه احمدی

دانش آموخته کارشناسی ارشد مطالعات آرشیوی، دانشگاه الزهرا (س)، تهران، ایران

تاریخ دریافت: ۱۴۰۴/۰۴/۳۰ تاریخ پذیرش: ۱۴۰۴/۰۲/۱۳

چکیده

پژوهش حاضر درصدد است تا با مطالعه و بررسی فناوری تشخیص گفتار و همچنین، استفاده از تجربه و نظرات صاحب‌نظران حوزه تاریخ شفاهی، قابلیت‌ها و چالش‌های به‌کارگیری آن در پیاده‌سازی مصاحبه‌های تاریخ شفاهی را شناسایی کند. روش استفاده‌شده برای انجام پژوهش توصیفی-پیمایشی است. ابزار جمع‌آوری داده‌ها پرسشنامه پژوهشگر ساخته است. به دلیل دشواری دسترسی پژوهشگر به جامعه پژوهش، از روش نمونه‌گیری گلوله‌برفی برای جمع‌آوری داده‌ها استفاده شد؛ به این صورت که از ۱۰ صاحب‌نظر اولیه که در پاسخ به پرسشنامه همکاری داشتند، درخواست شد تا سایر صاحب‌نظرانی که در این حوزه می‌شناسند را به همکاری در پاسخ‌گویی به پرسشنامه دعوت کنند. در نهایت، ۳۰ نفر از صاحب‌نظران تاریخ شفاهی با پاسخ‌گویی به پرسشنامه، در این پژوهش همکاری داشتند. یافته‌ها به ۴ پرسش پژوهش در خصوص: ۱. ارزیابی وضعیت فعلی پیاده‌سازی مصاحبه‌های تاریخ شفاهی، ۲. ارزیابی کاربردهای فناوری تشخیص گفتار، ۳. امکانات و شرایط لازم برای استفاده از فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی و ۴. شناسایی فرصت‌ها و چالش‌های استفاده از این فناوری در پیاده‌سازی مصاحبه‌های تاریخ شفاهی پاسخ می‌دهند. ۶۷ درصد از صاحب‌نظران وضعیت فعلی پیاده‌سازی مصاحبه‌های تاریخ شفاهی را نامطلوب ارزیابی می‌کنند. از میان معیارهای مورد ارزیابی برای کاربردهای فناوری، «صرفه‌جویی در زمان» با میانگین ارزش ۳/۷ بیشترین اهمیت را از نظر صاحب‌نظران دارد. در میان معیارهای امکانات و شرایط مورد ارزیابی برای استفاده از فناوری در پیاده‌سازی مصاحبه‌های تاریخ شفاهی، «شرایط امنیتی و رعایت اصل امانت‌داری» با میانگین ۲/۸ از مهم‌ترین چالش‌های استفاده از این فناوری به شمار می‌رود. «عدم دقت در پیاده‌سازی مصاحبه‌های با گویش‌های محلی و ...» از دیگر چالش‌های این فناوری است.

واژه‌های کلیدی: تاریخ شفاهی، مصاحبه، پیاده‌سازی، تشخیص گفتار، هوش مصنوعی.

* Corresponding Author: Fatemeh Ahmadi

E-mail address: farhad.rostami5215@gmail.com

Copyright2024@ Documents and National Library of the Islamic Republic of Iran. All rights reserved

مقدمه

یکی از مهم‌ترین مراحل تولید تاریخ شفاهی پیاده‌سازی مصاحبه‌هاست. پیاده‌سازی مصاحبه تاریخ شفاهی شاید کاری آسان به نظر برسد، اما روند طولانی و خسته‌کننده آن موجب دشواری کار پیاده‌سازی می‌شود. ممکن است مصاحبه‌های تولیدشده از یک طرح تاریخ شفاهی به بیش از صد ساعت برسد که پیاده‌سازی آن علاوه بر دشواری ناشی از طولانی‌بودن، صرف زمان زیادی را می‌طلبد که این مسئله خود موجب تأخیر در روند بهره‌وری از یک منبع تاریخ شفاهی می‌شود. این در حالی است که در کار تولید تاریخ شفاهی، غنیمت‌شمردن زمان نقش اساسی دارد؛ زیرا منبع اصلی تولید تاریخ شفاهی حافظه انسان‌هایی است که شاهدان رویدادها و وقایع هستند و بدیهی است که هرچه زمان بیشتری از یک رویداد بگذرد، قدرت حافظه‌ها در بیان روایت‌ها ضعیف‌تر خواهد شد و چه‌بسا دچار فراموشی شوند. از طرفی، رشد و توسعه فراگیر فناوری در هر حوزه مطالعاتی به گونه‌ای سرعت یافته است که نادیده‌گرفتن آن موجب از دست رفتن فرصت‌هایی جبران‌ناپذیر می‌شود. مصداق بارز آن، پیشرفت‌های روزافزون فناوری‌های مبتنی بر هوش مصنوعی است. یکی از این فناوری‌ها تبدیل گفتار به متنⁱ یا تشخیص گفتار (SR)ⁱⁱ است.

با وجود رشد روزافزون پژوهش‌های تاریخ شفاهی از نظر کمی، استفاده بهینه از آنها زمانی میسر می‌شود که تمام مراحل تولید یک منبع تاریخ شفاهی با برنامه‌ریزی صحیح و به‌درستی طی شوند. در روند تولید تاریخ شفاهی، اولین گام پس از ضبط و ثبت مصاحبه، پیاده‌سازی آن است.

پیاده‌سازی مصاحبه که از آن به شیوه‌های گوناگون مانند پیاده‌کردن، متن‌نویسی، ترانویسی، گفتارنویسی و خارج نویسی یاد شده است، نیازمند شنیدن مصاحبه و در عین حال، نوشتن سخنان رد و بدل شده در جلسه مصاحبه است (حسن آبادی، ۱۳۸۶، به نقل از عزیزی، ۱۴۰۰، ص. ۲۵). پیاده‌سازی اسناد شنیداری به طری

چشمگیر میزان دسترسی به آنها را ارتقا می‌بخشد. نسخه متنی مصاحبه را می‌توان به راحتی ارزیابی کرد و استفاده از آن نیز به تجهیزات خاصی نیاز ندارد. اگر نسخه متنی طی مدت زمان اندکی پس از تهیه مصاحبه تهیه شود، ابهامات موجود درباره جوانب شاهد و مدرک‌بودن آن از بین خواهند رفت (عزیزی، ۱۴۰۰، ص. ۲۵).

پیاده‌سازی یا رونویسی نقشی محوری در تحلیل مصاحبه‌ها ایفا می‌کند. اگرچه تجزیه و تحلیل نباید محدود به رونوشت باشد، بلکه باید شامل نحوه صحبت‌کردن، حالات چهره و حرکات نیز باشد، رونوشت ابزار ضروری برای تحلیل مصاحبه‌های تاریخ شفاهی و آرشیوکردن آنهاست. به عبارت دیگر، عدم رونویسی استفاده از مصاحبه را به شدت محدود خواهد کرد. یکی از مشکلاتی که در این زمینه وجود دارد این است که نه فقط بخش جالب توجهی از مصاحبه‌ها بدون رونویسی آرشیو می‌شوند، بلکه امروزه مصاحبه‌های بیشتری تولید می‌شوند که برای رونویسی آنها برنامه‌ریزی نشده است. این مصاحبه‌ها کاربردی محدود دارند و در بسیاری از موارد فراموش خواهند شد. در چنین شرایطی، بدیهی است که تاریخ شفاهی در حال حاضر و در آینده نیاز شدیدی به استفاده از فناوری تشخیص خودکار گفتارⁱⁱⁱ و سایر فناوری‌های استخراج صدا دارد (Draxler et al., 2024, p. 38). در ایران، برخلاف بسیاری از دیگر مراکز تاریخ شفاهی در دنیا، اساس کار بر چاپ و انتشار و نه آرشیو مصاحبه‌های انجام‌شده در این زمینه گذاشته شده است (حسن آبادی، ۱۳۹۶، ص. ۹). این امر یک نقطه قوت در تاریخ شفاهی ایران محسوب می‌شود؛ با وجود این، دشواری و زمان‌بر بودن پیاده‌سازی نباید مورد غفلت قرار گیرد. اگرچه استفاده از فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی با چالش‌هایی روبه‌رو است، شناخت قابلیت‌ها و محدودیت‌های به‌کارگیری این فناوری موجب می‌شود تا به دنبال راهکارهایی برای برطرف‌کردن محدودیت‌ها و استفاده بهینه از امکانات آن باشیم. این پژوهش سعی دارد تا با

مرور متون وارد کردند. در یافته‌ها، اصلی‌ترین مزیت بهبود گردش کار بود و اساسی‌ترین موانع کاهش دقت و تأثیر اختلال‌های محیطی شناخته شدند و استفاده از ابزارهای مناسب مهم‌ترین تسهیلگر استفاده از فناوری بود (کمالی و شیخ طاهری، ۱۳۹۷، ص. ۷۰).

در خارج از ایران، پژوهش‌هایی جالب توجه در حوزه علوم کامپیوتر و هوش مصنوعی در خصوص کاربرد فناوری تشخیص گفتار در تاریخ شفاهی انجام شده‌اند. در ادامه، چند مورد از جدیدترین آنها که با پژوهش حاضر مرتبط هستند، بررسی می‌شوند:

دراکسلر^{i۷} و همکاران (۲۰۲۴) در مقاله‌ای با عنوان «خدمات فناوری گفتار برای پژوهش‌های تاریخ شفاهی»، درگاه رونویسی و خدمات وب مرتبط با پردازش گفتار در BAS^v، راهکارهای گفتاری توسعه‌یافته در LINDAT^{vi}، نحوه پردازش گفتار با Whisper^{vi} و سایر چالش‌ها و تحولات آینده در این زمینه را بررسی می‌کنند. در این مقاله، تعدادی از نوآوری‌ها و رویکردهای ابزاری در راستای بهبود تشخیص خودکار گفتار برای مصاحبه‌های تاریخ شفاهی، در یک سپر داده ایمن و محافظت‌شده که نشت اطلاعات را به حداقل برساند، ارائه می‌شود (Draxler et al., 2024, p. 38).

گرف^{vi i} (۲۰۲۲) در پایان‌نامه دکتری خود با عنوان «تشخیص قوی گفتار از طریق انطباق برای مصاحبه‌های تاریخ شفاهی آلمان»، ترکیبی از روش‌های مختلف انطباق دامنه را برای غلبه بر عدم وجود داده‌های یادگیری نماینده و مقابله با غیرقابل پیش‌بینی بودن مصاحبه‌های تاریخ شفاهی مطالعه می‌کند. این پژوهش سعی دارد با استفاده از تطبیق بین زبانی در یک ساختار چندمرحله‌ای، قابلیت دسترسی گسترده به ساختارهای گفتاری انگلیسی را افزایش دهد. هدف اصلی این پژوهش توسعه و بهبود عملکرد سیستم‌های تشخیص خودکار گفتار برای مصاحبه‌های تاریخ شفاهی آلمانی بود. برای رسیدن به این هدف، با وجود کمبود داده‌ها، سازگاری داده‌ها و مدل‌های ارائه‌شده بررسی می‌شود و به طور ویژه، بر عملکرد دنیای

بررسی پژوهش‌های انجام‌شده در خصوص فناوری تشخیص گفتار و همچنین، از طریق بررسی دیدگاه‌ها و تجربیات صاحب‌نظران تاریخ شفاهی در زمینه پیاده‌سازی مصاحبه و کاربرد فناوری تشخیص گفتار در آن، قابلیت‌ها و محدودیت‌های استفاده از این فناوری در تاریخ شفاهی را شناسایی کند.

پیشینه پژوهش

طبق بررسی‌های انجام‌شده، در ایران، در زمینه کاربرد فناوری تشخیص گفتار در علوم انسانی، پژوهش‌هایی محدود به نگارش درآمده‌اند. به برخی از آنها که با پژوهش حاضر ارتباطی نزدیک دارند، به ترتیب روزآمدترین تاریخ انتشار، اشاره می‌شود:

ربیعی و میرزاییان (۱۴۰۱) در مقاله‌ای با عنوان «کاربردهای پردازش زبان طبیعی در علم اطلاعات و دانش‌شناسی با تأکید بر کتابخانه‌های دیجیتال» با استفاده از روش کتابخانه‌ای و تحلیل اسنادی، کاربرد پردازش زبان طبیعی در کتابخانه‌های دیجیتال و عرصه علم اطلاعات را بررسی می‌کنند. نتایج این مطالعه نشان می‌دهد پردازش زبان طبیعی در مواردی مانند نظام‌های پرسش و پاسخ و به‌کارگیری فناوری خطایاب املائی، ابهام‌زدایی از عبارت‌های پرسش کاربران و پیش‌بینی واژه‌های مدنظر آنها، یاری‌رساندن به کاربران دارای معلولیت‌های جسمی، نظرکاوی و تحلیل احساس واژگان استفاده‌شده توسط کاربران و ... قابل ردیابی است (ربیعی و میرزاییان، ۱۴۰۱، ص. ۱۹۸).

کمالی و شیخ طاهری (۱۳۹۷) در مقاله خود با عنوان «مستندسازی گزارش‌های پرستاری با استفاده از فناوری تشخیص گفتار (مزایا، موانع و چالش‌ها و تسهیلگرها)»، با استفاده از روش مروری، مطالعات مختلف استفاده از فناوری تشخیص گفتار در حوزه پرستاری در سال‌های ۱۹۹۰ تا ۲۰۱۷ میلادی در پایگاه‌های اطلاعاتی PubMed, Science Direct, Google Scholar, Magira و SID را بررسی و در نهایت، ۱۰ مطالعه در

- بررسی وضعیت فعلی پیاده‌سازی مصاحبه‌های تاریخ شفاهی؛

- بررسی دیدگاه صاحب‌نظران تاریخ شفاهی در خصوص کاربرد فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی؛

- بررسی دیدگاه صاحب‌نظران تاریخ شفاهی در خصوص امکان استفاده از فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی؛

- بررسی دیدگاه صاحب‌نظران تاریخ شفاهی در خصوص فرصت‌ها و چالش‌های استفاده از فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی.

پرسش‌های پژوهش

با توجه به اهداف دنبال‌شده در پژوهش، پرسش‌های زیر مطرح می‌شوند:

۱. وضعیت فعلی پیاده‌سازی مصاحبه‌های تاریخ شفاهی چگونه است؟

۲. دیدگاه صاحب‌نظران تاریخ شفاهی در خصوص کاربرد فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی چیست؟

۳. دیدگاه صاحب‌نظران تاریخ شفاهی درباره امکانات و شرایط لازم به منظور استفاده از فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی چیست؟

۴. از دیدگاه صاحب‌نظران تاریخ شفاهی، استفاده از فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی چه فرصت‌ها و چالش‌هایی را در بر می‌گیرد؟

فناوری تشخیص گفتار (SR)

امروزه، فناوری تشخیص خودکار گفتار یکی از شناخته‌شده‌ترین و موفق‌ترین رشته‌های هوش مصنوعی است. تشخیص خودکار گفتار به بخشی ضروری از بسیاری از برنامه‌های روزمره تبدیل شده است. ترجمه دقیق گفتار به متن با واژگان نسبتاً نامحدود کاربرد

واقعی سیستم تمرکز و اضافه‌شدن احتمالی به دامنه هدف برای تمام سیستم‌های پیشنهادی در نظر گرفته می‌شود (Gref, 2022, p. 190).

گابریل^x و همکاران (۲۰۲۰) در پژوهشی با عنوان «فناوری تبدیل گفتار به متن به عنوان ابزار مستندسازی برای مترجمان: رویکردی جدید برای گردآوری یک مجموعه داده ویژه و استخراج واژگان از سخنرانی‌های ضبط‌شده ویدیویی»، سه هدف اصلی را دنبال می‌کنند: ۱. معرفی مناسب‌ترین برنامه کاربردی ASR (تشخیص خودکار گفتار) برای ایجاد مجموعه‌های داده، با مقایسه چندین ابزار ASR و ارزیابی عملکرد آنها، ۲. استفاده از ASR به منظور استخراج واژگان از رونویسی‌های به‌دست‌آمده از سخنرانی‌های ضبط‌شده ویدیویی و ۳. ترویج پذیرش ASR به عنوان یک ابزار مستندسازی جدید در میان مترجمان. به اظهار پژوهشگران، این پژوهش یکی از اولین مطالعاتی است که امکان فناوری پردازش خودکار گفتار را برای برآوردن نیازهای مقدماتی مترجمان از نظر واژگان و دانش زمینه/دامنه بررسی می‌کند (Gaber et al., 2020, p. 263).

له^x و همکاران (۲۰۱۹) در مقاله‌ای با عنوان «استخراج صدا؛ تجزیه و تحلیل گفتار پیشرفته برای تاریخ شفاهی»، امکانات و چالش‌های فناوری استخراج صدا برای رونویسی خودکار مصاحبه‌های تاریخ شفاهی را بررسی می‌کنند. این پژوهش مروری بر فناوری‌های استفاده‌شده برای تجزیه و تحلیل گفتار، عملکرد سیستم‌های استخراج صوت در «Fraunhofer IAIS» و آرشیو تاریخ شفاهی (حافظه آلمانی) در دانشگاه هاگن دارد و از نتایج به‌دست‌آمده و روش‌های پژوهشی جاری، پیشنهادهایی را برای بهبود یکپارچه‌سازی سیستم‌ها ارائه می‌دهد (Leh et al., 2019, pp. 1-9).

اهداف پژوهش

با توجه به آنچه گفته شد، پژوهش حاضر درصدد دستیابی به اهداف زیر است:

گسترده آن را به عنوان دستیار گفتار در تلفن‌های هوشمند، بلندگوهای هوشمند و سایر دستگاه‌ها امکان‌پذیر کرده است. در بسیاری از کاربردها، سیستم‌های تشخیص خودکار گفتار به دقت تشخیص نزدیک به انسان دست می‌یابند. با وجود این، تشخیص خودکار گفتار یک فناوری بدون مشکل نیست. هنوز هیچ سیستمی وجود ندارد که دقت رونویسی زیادی را در همه حوزه‌ها، برنامه‌ها و موقعیت‌های گفتاری ارائه دهد. این مشکل به ویژه برای زبان‌های غیرانگلیسی وجود دارد (1). (Gref, 2022, p. 1).

عملکرد چشمگیر مدل‌های گفتاری مبتنی بر هوش مصنوعی، به ویژه در خصوص استحکام آنها در برابر اختلال ^{xi}، گستره وسیع زبان‌های پشتیبانی‌شده و امکان تطبیق آنها با دیگر زبان‌ها با کمک یادگیری عمیق، تولید رونوشت‌های مناسب در حوزه‌های پژوهش‌هایی مانند تاریخ شفاهی که در آن زبان گفتاری منبع اصلی اطلاعات است را تا حد زیادی تسهیل می‌کند. با این حال، بسته به نیازهای پژوهشگر، چالش‌های دیگری مانند تخصیص حس مناسب گوینده، خروجی رویدادهای گفتاری دقیق‌تر مانند صداهای همراه با تردید، لکنت، کوتاه‌کردن کلمات و غیره وجود دارند (Draxler et al., 2024, p. 38). البته چارچوب تفسیرها متفاوت است. برای مثال، یک مورخ شفاهی ممکن است به یک مصاحبه ضبط‌شده به عنوان روایتی میان‌ذهنی ^{xi i} از یک تجربه گذشته نگاه کند، در حالی که مورخ دیگری ممکن است همان منبع را فقط به دلیل اطلاعات واقعی که منتقل می‌کند، در نظر بگیرد. یک دانشمند علوم اجتماعی احتمالاً سعی می‌کند مضامین مشترک و شباهت‌ها و تفاوت‌ها را در مجموعه‌ای کامل از مصاحبه‌ها کشف کند، در حالی که یک زبان‌شناس، برای اهداف محاسباتی، بر شمارش بسامدها ^{xi ii}، شناسایی هم‌نشینی‌ها و رویدادهای هم‌زمان تکیه می‌کند (Scagliola et al., 2020, p. 127).

توکل ^{xi v} نیز معتقد است اگر تاریخ شفاهی موضوع محور باشد، یعنی پژوهشگر تاریخ شفاهی در نظر

داشته باشد از منظر موضوع، خاطرات و زندگینامه اشخاصی را ثبت و ضبط کند، نرم‌افزارهای تشخیص گفتار منجر به خلاصه‌سازی و از دست رفتن بخشی از اطلاعات می‌شوند، ولی اگر هدف پژوهشگر، تاریخ شفاهی شیء محور باشد، برای مثال قصد داشته باشد تاریخ شفاهی یک سازمان را تدوین کند، در این صورت، استفاده از نرم‌افزارهای تشخیص گفتار می‌تواند بسیار کارآمد باشد (ارتباط شخصی، ۱۳ دی ۱۴۰۳).

تشخیص خودکار گفتار که به اختصار تشخیص گفتار نامیده می‌شود، یکی از هوشمندترین شاخه‌های پردازش گفتار است که به دستگاه امکان می‌دهد گفتار کاربر را بفهمد و آن را از طریق یک برنامه رایانه‌ای به یک مجموعه کلمات تبدیل کند و در نتیجه، نوعی ارتباط طبیعی بین انسان و ماشین ایجاد کند. این فناوری می‌تواند به صورت نمایش گرافیکی بسامدهای منتشرشده به عنوان تابعی از زمان تعریف شود. تمام روش‌های پردازش گفتار (تجزیه و تحلیل گفتار، شناسایی سخنگو، تأیید سخنگو) ایجاد رابط‌های صوتی (رابط ماشین-انسانی) یا انجام تعامل صوتی را ممکن می‌کنند (Benkerzaz et al., 2019, p. 78).

در ادامه، برای آشنایی بیشتر با فناوری تشخیص گفتار، ساختار و عملکرد آن به طور مختصر بررسی می‌شود:

معماری سیستم‌های تشخیص خودکار گفتار ^{xv}

هدف اصلی سیستم تشخیص گفتار این است که دستگاه قادر به گوش دادن و درک اطلاعات گفتاری یا صوتی برای تصمیم‌گیری درست باشد. اولین مرحله این سیستم تجزیه و تحلیل علائم ^{xvi} گفتار است تا آخرین نتیجه مجموعه‌ای از کلمات گفتاری باشد. بین این دو مرحله، سیستم چندین مرحله را دنبال می‌کند که معمولاً بر اساس رویکرد آماری هستند. به طور کلی، سیستم تشخیص گفتار از پنج واحد تشکیل شده است که عبارت‌اند از:

می‌شوند که در طول مشاهده داده‌های زبانی تخمین زده می‌شوند.

۳. مدل تلفظ: ساخت مدل زبان؛ نحوه نوشتن یک کلمه.

۴. مدل صوتی: این مدل امکان پیش‌بینی محتمل‌ترین واج‌ها را در صدای ورودی فراهم می‌کند.

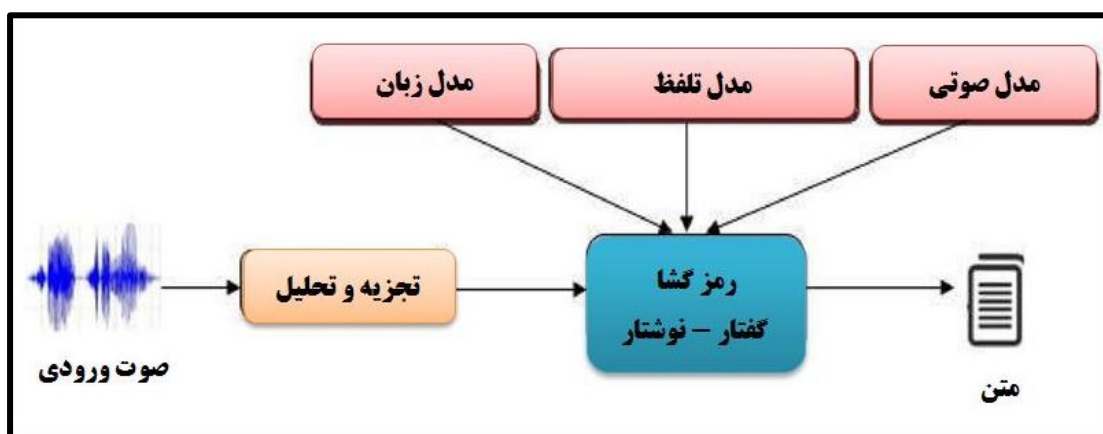
۵. رمزگشای گفتار به متن: ترکیبی از مدل‌های قبلی است که محتمل‌ترین رونویسی متن را برای یک عبارت گفتاری معین فراهم می‌کند. شکل (۱) معماری کلی سیستم‌های تشخیص گفتار را نشان می‌دهد.

۱. تجزیه و تحلیل: اولین مرحله پردازش گفتار تجزیه و تحلیل است که امکان انتخاب اندازه قالب برای تقسیم‌بندی سیگنال ورودی را فراهم می‌کند تا تحلیل دیگری را در بخش گفتار انجام دهد.

۲. مدل زبان: که به دو گروه تقسیم می‌شود:

• قطعی (دستوری): توسط متخصصان زبان طراحی شده است.

• تصادفی (آماري): مدل‌های زبانی آماری نتیجه برآورد مدل‌های زبانی بدون نظارت بر یادگیری هستند. در بیشتر موارد، با مجموعه‌ای از پارامترهای خالی شروع



شکل ۱: معماری کلی سیستم‌های تشخیص خودکار گفتار

تشخیص گفتار باید شبیه یادگیری در انسان باشد و وظیفه تشخیص گفتار تبدیل زبان گفتار به متن قابل خواندن توسط رایانه است (بهشتی، ۱۴۰۱، ص. ۵).

روش‌شناسی

این پژوهش از نوع کاربردی است و با روش پیمایشی-توصیفی انجام شده است. ابزار گردآوری داده‌ها پرسشنامه پژوهشگر ساخته است که با توجه به چارچوب نظری پژوهش طراحی شد و پس از بررسی روایی آن توسط ۲ نفر از متخصصان تاریخ شفاهی و مطالعات آرشیوی و دو نفر از اساتید علم اطلاعات و دانش‌شناسی، پرسشنامه ویرایش و بین جامعه پژوهش منتشر شد.

یادگیری در سیستم‌های تشخیص گفتار

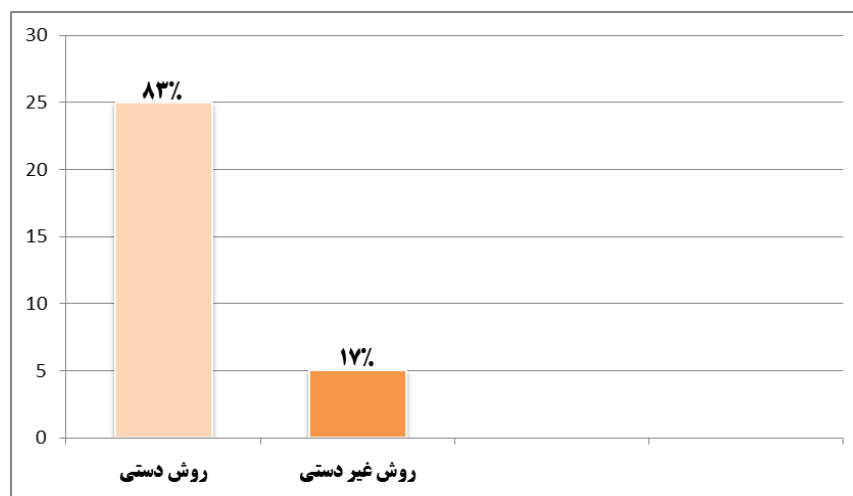
در تشخیص گفتار، یادگیری به فرایندی گفته می‌شود که ورودی آن داده و خروجی آن یک مدل تشخیص گفتار است. سیستم‌های تشخیص گفتار یکی از مدل‌های هوش مصنوعی هستند که ساختار مغز انسان را شبیه‌سازی می‌کنند. از این رو، یادگیری‌ای که در سیستم‌های تشخیص گفتار فقط با استفاده از نمونه داده‌های برچسب‌دار اتفاق بیفتد، شبیه مدل یادگیری زبان اکتسابی در انسان نیست. در واقع، همان‌طور که نوزادان با گوش دادن به بزرگسالان در اطراف خود زبان را یاد می‌گیرند، یادگیری در تشخیص گفتار نیز باید به این صورت باشد که به مرور و با به دست آوردن دانش و تجربه به دست آید. بنابراین، می‌توان گفت یادگیری در

یافته‌ها

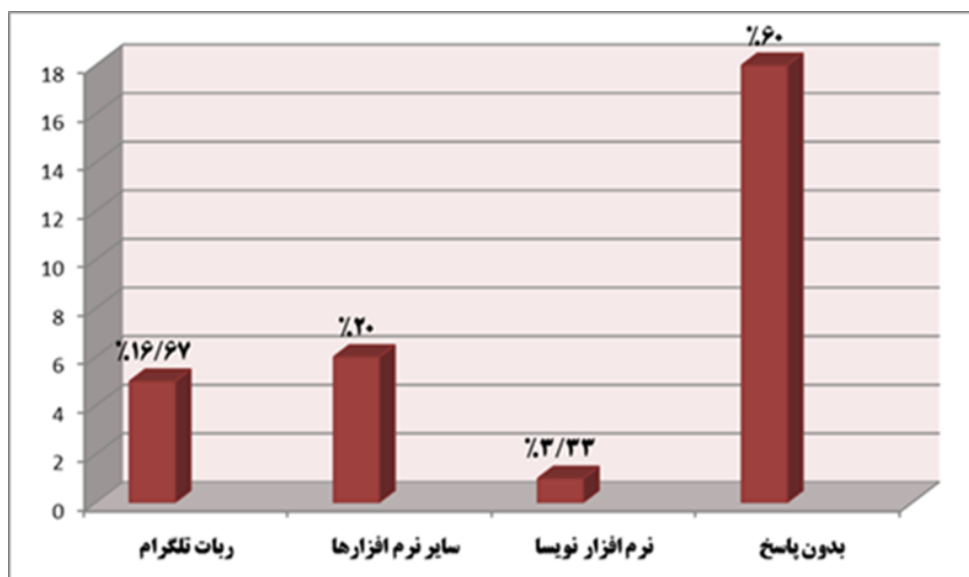
پرسش ۱. وضعیت فعلی پیاده‌سازی مصاحبه‌های تاریخ شفاهی چگونه است؟

برای پاسخ به این پرسش، ۷ پرسش در پرسشنامه مطرح شدند. در این بخش، با طرح ۶ پرسش، با گزینه‌های اسمی (بله و خیر) سعی شده است تا وضعیت و شرایط پیاده‌سازی مصاحبه‌های تاریخ شفاهی بررسی شود. در پرسش اول، روش استفاده‌شده برای پیاده‌سازی مصاحبه‌ها بررسی و مشخص شد ۸۳ درصد به روش سنتی، فایل مصاحبه را می‌شنوند و پیاده‌سازی می‌کنند و پس از آن، فایل رونوشت را با استفاده از واژه‌پرداز وارد رایانه می‌کنند. در پرسش بعدی، در ادامه پرسش قبل، از پاسخ‌دهندگان خواسته شد تا در صورتی که پاسخ آنها خیر است یا اینکه تا کنون از روش غیردستی استفاده کرده‌اند، پیاده‌سازی مورد استفاده خود را بنویسند. نتایج نشان می‌دهد ۶۰ درصد از پاسخ‌دهندگان روش غیردستی را تا کنون نیازموده‌اند. همچنین، ۲۰ درصد از پاسخ‌دهندگان نرم‌افزار مورد استفاده خود را بیان نکرده‌اند. نمودار (۱) توزیع فراوانی روش استفاده‌شده برای پیاده‌سازی مصاحبه‌ها و نمودار (۲) توزیع فراوانی روش‌های پیاده‌سازی‌های غیردستی به کار گرفته‌شده را نشان می‌دهد.

جامعه آماری پژوهش را اساتید تاریخ، پژوهشگران و کارشناسانی که در حوزه تاریخ شفاهی فعالیت دارند، تشکیل می‌دهند. روش نمونه‌گیری استفاده‌شده نمونه‌گیری گلوله‌برفی است. نمونه‌گیری گلوله‌برفی یک روش نمونه‌گیری هدفمند برای جمع‌آوری داده‌ها در پژوهش‌های کیفی است. این روش زمانی استفاده می‌شود که دسترسی به افرادی با ویژگی‌های هدف دشوار باشد. در این روش، افراد مورد مطالعه موجود آزمودنی‌های دیگر را از میان آشنایان خود پیشنهاد می‌دهند (Naderifar et al., 2017, p. 2). پرسشنامه در ابتدا از طریق پست الکترونیکی و پیام‌رسان‌های داخلی و خارجی برای ۲۰ نفر از صاحب‌نظران تاریخ شفاهی ارسال شد که از میان ایشان، ۱۰ نفر در پاسخ‌گویی به پرسشنامه همکاری کردند. در مرحله بعد، از پاسخ‌دهندگان درخواست شد تا صاحب‌نظرانی که در این زمینه می‌شناسند را برای همکاری در پاسخ به پرسشنامه پیشنهاد دهند. در نهایت، ۲۰ نفر از صاحب‌نظران تاریخ شفاهی در پاسخ‌گویی به پرسشنامه، با پژوهش حاضر همکاری داشتند. پس از جمع‌آوری پرسشنامه‌ها، داده‌های موجود با استفاده از روش‌های آماری توصیفی و تحلیلی و با استفاده از نرم‌افزار صفحه گسترده Excel تجزیه و تحلیل شدند.



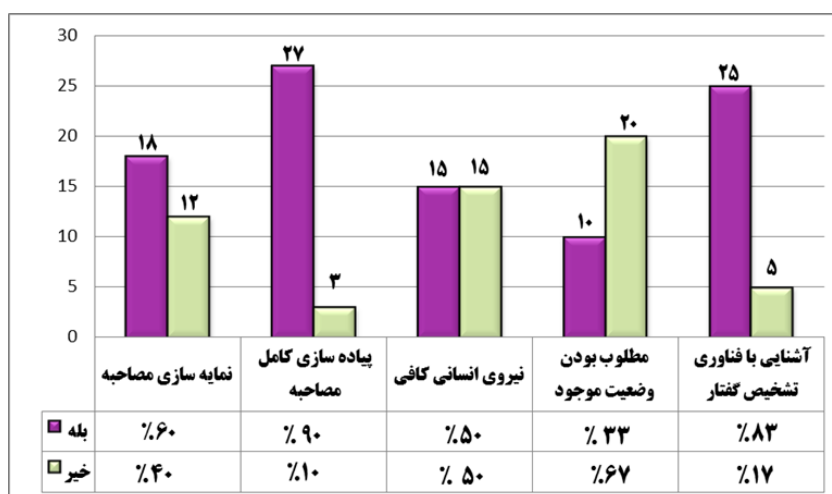
نمودار ۱: روش استفاده‌شده برای پیاده‌سازی مصاحبه‌ها



نمودار ۲: روش‌های پیاده‌سازی غیر دستی مصاحبه‌ها

پرسش ششم نظر صاحب‌نظران تاریخ شفاهی را در رابطه با مطلوب بودن وضعیت فعلی پیاده‌سازی مصاحبه‌ها جویا می‌شود که ۶۷ درصد پاسخ‌دهندگان شرایط فعلی را نامطلوب می‌دانند. در پرسش هفتم، درباره آشنایی با فناوری تشخیص گفتار (تبدیل گفتار به متن) از صاحب‌نظران پرسیده شد که ۸۳ درصد اظهار آشنایی با این فناوری را داشتند. نمودار (۳) توزیع فراوانی معیارهای مورد ارزیابی در پرسش‌های سوم تا هفتم را به نمایش می‌گذارد

در پرسش سوم، در خصوص نمایه‌سازی مصاحبه‌ها پرسیده شد که ۶۰ درصد از پاسخ‌دهندگان اظهار داشتند مصاحبه‌ها نمایه‌سازی می‌شوند. در پرسش چهارم، از پاسخ‌دهندگان در خصوص نوع پیاده‌سازی (کامل/ خلاصه) پرسیده می‌شود. ۹۰ درصد از پاسخ‌دهندگان مصاحبه‌ها را به طور کامل پیاده‌سازی می‌کنند. پرسش پنجم از پاسخ‌دهندگان در خصوص وجود نیروی انسانی کافی برای پیاده‌سازی مصاحبه‌ها می‌پرسد که ۵۰ درصد نیروی انسانی موجود را کافی و ۵۰ درصد ناکافی می‌دانند.



نمودار ۲: توزیع فراوانی معیارهای مورد ارزیابی در پرسش‌های سوم تا هفتم پرسشنامه

فناوری تشخیص گفتار برای کمک به پیاده‌سازی مصاحبه‌ها را توجیه می‌کند.

پرسش ۲. دیدگاه صاحب‌نظران تاریخ شفاهی در خصوص کاربرد فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی چیست؟

برای پاسخ به پرسش دوم، ۷ پرسش به صورت معیار برای ارزیابی کاربردهای فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی از دیدگاه صاحب‌نظران در نظر گرفته شدند. جدول (۱) معیارهای ارزیابی کاربردهای فناوری تشخیص گفتار از دیدگاه صاحب‌نظران را نشان می‌دهد.

باتوجه به آنچه از یافته‌های بخش اول پرسشنامه به دست آمد، به نظر می‌رسد بیشتر پژوهشگران تاریخ شفاهی با نرم‌افزارهای تشخیص گفتار که به صورت رسمی و توسط شرکت‌های معتبر و متخصص در زمینه هوش مصنوعی ارائه می‌شوند، آشنایی کافی ندارند. استفاده از نرم‌افزارهای منبع باز، اگرچه رایگان است، کارآیی لازم را برای کمک به پیاده‌سازی مصاحبه‌های تاریخ شفاهی ندارد. داده‌های نمودار (۳) که نشان می‌دهد ۶۰ درصد از پاسخ‌دهندگان به کار نمایه‌سازی اهمیت می‌دهند، ۹۰ درصد مصاحبه را به طور کامل پیاده‌سازی می‌کنند، ۵۰ درصد نیروی انسانی برای پیاده‌سازی مصاحبه‌ها را ناکافی می‌دانند و ۶۷ درصد وضعیت فعلی پیاده‌سازی مصاحبه‌ها را نامطلوب می‌دانند، توجه به

جدول ۱: معیارهای ارزیابی کاربرد فناوری تشخیص گفتار از دیدگاه صاحب‌نظران تاریخ شفاهی

ردیف	معیار	فراوانی					درصد فراوانی					میانگین ارزش	تغییرات معیار
		خیلی زیاد	زیاد	متوسط	کم	خیلی کم	خیلی زیاد	زیاد	متوسط	کم	خیلی کم		
۱	میزان صرفه‌جویی در زمان	۱۱	۷	۶	۴	۲	۳۷٪	۲۳٪/۳	۲۰٪	۱۳٪/۳	۶٪/۶	۳/۷	۱/۲۶
۲	افزایش کیفیت کار پیاده‌سازی مصاحبه	۱	۴	۸	۱۴	۳	۳٪/۳	۱۳٪/۳	۲۶٪/۶	۴۶٪/۶	۱۰٪	۲/۵۳	۰/۹۵
۳	افزایش کمیت کار پیاده‌سازی مصاحبه	۱	۱۱	۱۱	۴	۳	۳٪/۳	۳۶٪/۶	۳۶٪/۶	۱۳٪/۳	۱۰٪	۳/۱	۱
۴	افزایش بهره‌وری (بهره) بیشتر از تولیدات تاریخ شفاهی	۲	۶	۱۳	۷	۲	۶٪/۶	۲۰٪	۴۳٪/۳	۲۳٪/۳	۶٪/۶	۲/۹۶	۰/۹۸
۵	کمک به سهولت نمایه‌سازی مصاحبه	۴	۲	۱۹	۴	۱	۵٪/۳	۶٪/۶	۶۳٪/۳	۱۳٪/۳	۳٪/۳	۳/۱	۰/۹۲
۶	کمک به مدیریت بهتر منابع تاریخ شفاهی	۳	۶	۸	۱۲	۱	۱۰٪	۲۰٪	۲۶٪/۶	۴۰٪	۳٪/۳	۲/۹۳	۱
۷	میزان اثربخشی استفاده از فناوری تشخیص گفتار	۳	۷	۷	۱۲	۱	۱۰٪	۲۳٪/۳	۲۳٪/۳	۴۰٪	۳٪/۳	۲/۹۶	۱

پرسش ۳. دیدگاه صاحب‌نظران تاریخ شفاهی درباره امکانات و شرایط لازم برای استفاده از فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی چیست؟

برای پاسخ به پرسش سوم، ۵ پرسش به صورت معیار ارزیابی امکانات و شرایط مورد نیاز برای استفاده از فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی از دیدگاه صاحب‌نظران در نظر گرفته شدند. جدول (۲) معیارهای ارزیابی امکانات و شرایط موجود برای استفاده از فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی از دیدگاه صاحب‌نظران را نشان می‌دهد.

جدول ۲: معیارهای ارزیابی امکانات و شرایط لازم برای استفاده از فناوری تشخیص گفتار در تاریخ شفاهی

ردیف	معیار	فراوانی					درصد فراوانی					میانگین	انحراف معیار
		فراوانی	نسبت	درصد	فراوانی	نسبت	درصد	فراوانی	نسبت	درصد	فراوانی		
۱	امکان استفاده از اینترنت پرسرعت	۵	۵	۱۰۰٪	۳	۳۳٪	۱۶٪	۳	۳۳٪	۱۶٪	۰	۳/۵	۰/۹۵
۲	امکان استفاده از سیستم‌عامل‌ها و برنامه‌های رایانه‌ای روزآمد	۶	۸	۸۰٪	۶	۲۶٪	۲۶٪	۶	۲۶٪	۲۶٪	۰	۳/۴	۱
۳	نیروی انسانی آشنا به مهارت‌های رایانه‌ای	۴	۱۰	۹۰٪	۳	۳۳٪	۳۰٪	۳	۳۳٪	۲۳٪	۰	۳/۳۶	۰/۹۸
۴	امکان اختصاص اعتبار مالی برای استفاده از فناوری تشخیص گفتار	۶	۵	۸۰٪	۶	۲۶٪	۲۶٪	۶	۲۶٪	۳۶٪	۰	۳/۲	۱/۱۵
۵	شرایط امنیتی و رعایت اصل امانت‌داری	۱	۷	۱۳٪	۳	۳۳٪	۴۰٪	۳	۳۳٪	۲۳٪	۱۰٪	۲/۸	۰/۹۹

انسانی آشنا به مهارت‌های رایانه‌ای» با میانگین ۳/۳۶ و انحراف معیار ۰/۹۸ از نظر صاحب‌نظران قابل دسترس هستند. معیار «شرایط امنیتی و رعایت اصل امانت‌داری» با کمترین میانگین ارزش ۲/۸ و انحراف معیار ۰/۹۹ نشان می‌دهد از نظر صاحب‌نظران، این معیار می‌تواند یکی از چالش‌های به‌کارگیری فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی به شمار رود.

یافته‌های این بخش نشان می‌دهد معیار «صرفه‌جویی در زمان» با میانگین ارزش ۳/۷ و انحراف معیار ۱/۲۷ بیشتر از دیگر معیارها مورد توجه صاحب‌نظران قرار گرفته است. پس از آن، معیارهای «کمک به سهولت نمایه‌سازی» با میانگین ۳/۱ و انحراف معیار ۰/۹۳ و «افزایش کمیت» با میانگین ۳/۱ و انحراف معیار ۱ از نظر صاحب‌نظران اهمیت داشتند. معیار «افزایش کیفیت کار پیاده‌سازی» با میانگین ارزش ۲/۵۳ و انحراف معیار ۰/۹۵ که کمترین میزان پراکندگی بین پاسخ‌ها را داشت، نشان می‌دهد از نظر صاحب‌نظران، این معیار می‌تواند یکی از چالش‌های استفاده از فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی به شمار رود.

یافته‌های این بخش نشان می‌دهد از نظر صاحب‌نظران، معیار «امکان استفاده از اینترنت پرسرعت» با میانگین ارزش ۳/۵ و داشتن کمترین میزان پراکندگی در پاسخ، با انحراف معیار ۰/۹۵، بیشتر از سایر معیارها برای به‌کارگیری فناوری تشخیص گفتار فراهم است. پس از آن، معیارهای «سیستم‌عامل‌ها و برنامه‌های رایانه‌ای روزآمد» با میانگین ۳/۴ و انحراف معیار ۱ و «نیروی

پرسش ۴. از دیدگاه صاحب‌نظران تاریخ شفاهی، استفاده از فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی چه فرصت‌ها و چالش‌هایی را در بر می‌گیرد؟

برای پاسخ به این پرسش، طی یک پرسش از نوع پرسش باز، از صاحب‌نظران درخواست شد تا نظر خود را درباره فرصت‌ها و چالش‌هایی که در استفاده از فناوری تشخیص گفتار می‌بینند، ارائه دهند. یافته‌های این پرسش به روش تحلیل محتوا مطالعه شد. با توجه به اظهارات صاحب‌نظران، فناوری تشخیص گفتار از نظر «افزایش سرعت در کار پیاده‌سازی و صرفه‌جویی در زمان» می‌تواند اثربخشی زیادی در پیاده‌سازی مصاحبه‌های تاریخ شفاهی داشته باشد، اما به دلیل «خطاهای ناشی از عدم تشخیص لهجه‌ها، گویش‌ها، زیروبم صدای برخی از اشخاص به خصوص افراد سالمند با مشکل تکلم»، «استفاده از فناوری تشخیص گفتار از طریق نرم‌افزارهای منبع باز»، منجر به دوباره‌کاری می‌شود. بیشتر صاحب‌نظران به ضرورت استفاده از فناوری تشخیص گفتار به متن اذعان داشتند، اما استفاده از نرم‌افزارهای منبع باز موجب نارضایتی ایشان از این فناوری شده است. در جدول (۳) یافته‌های این بخش به تفکیک فرصت و چالش بیان شده است.

جدول ۳: فرصت‌ها و چالش‌های استفاده از فناوری تشخیص

گفتار در تاریخ شفاهی

فرصت‌ها	چالش‌ها
افزایش سرعت	عدم دقت لازم برای پیاده‌سازی مصاحبه‌های با گویش‌ها و لهجه‌های مختلف
صرفه‌جویی در زمان	عدم امنیت کافی
سهولت در پیاده‌سازی	موجب دوباره‌کاری و لزوم بازبینی متن می‌شود
سهولت دسترسی کاربران تاریخ شفاهی به مصاحبه‌ها	عدم قابلیت انجام ماشینی ویراستاری زبانی در حین پیاده‌سازی
مناسب برای تبدیل مصاحبه به کتاب	
مناسب برای مصاحبه‌های شیء محور	

نتیجه

با توجه به یافته‌های پژوهش، در بخش اول در خصوص ارزیابی وضعیت فعلی پیاده‌سازی مصاحبه‌های تاریخ شفاهی، ۹۰ درصد از صاحب‌نظران معتقد هستند مصاحبه‌ها باید به طور کامل پیاده‌سازی شوند و ۶۰ درصد نمایه‌سازی مصاحبه‌ها را مهم می‌دانند و ۵۰ درصد اذعان می‌دارند نیروی انسانی کافی برای انجام این امور وجود ندارد. از این رو، وضعیت فعلی پیاده‌سازی مصاحبه‌های تاریخ شفاهی با دشواری‌هایی همراه است که اگر فناوری تشخیص گفتار در آن به کار گرفته شود، می‌تواند منجر به بهبود وضعیت آن شود.

در بخش دوم، با ارزیابی معیارهای در نظر گرفته‌شده برای شناسایی کاربردهای فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی، معیار «صرفه‌جویی در زمان» با زیادترین میانگین، بیشترین اهمیت را از دیدگاه صاحب‌نظران دارد. پس از آن، معیارهای «کمک به سهولت نمایه‌سازی» و «افزایش کمیت» از نظر صاحب‌نظران کاربردی و اثربخش به نظر می‌رسند. اما معیار «افزایش کیفیت کار پیاده‌سازی» با کمترین میانگین ارزش نشان می‌دهد از نظر صاحب‌نظران، این معیار می‌تواند یکی از چالش‌های استفاده از فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی باشد.

در بخش سوم، از نظر صاحب‌نظران، معیار «امکان استفاده از اینترنت پرسرعت» با زیادترین میانگین ارزش بیشتر از سایر معیارها برای به‌کارگیری فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی فراهم است. پس از آن، معیارهای «سیستم‌عامل‌ها و برنامه‌های رایانه‌ای روزآمد» و «نیروی انسانی آشنا به مهارت‌های رایانه‌ای» از نظر صاحب‌نظران قابل تأمین هستند. معیار «شرایط امنیتی و رعایت اصل امانت‌داری» از نظر صاحب‌نظران یکی از چالش‌های به‌کارگیری فناوری تشخیص گفتار در پیاده‌سازی مصاحبه‌های تاریخ شفاهی به شمار می‌رود.

یکی از راهکارهای مورد استفاده دانشمندان حوزه هوش مصنوعی برای ارائه خدمات بهتر و جدیدتر باشد.

منابع فارسی

بهشتی، معصومه (۱۴۰۱). ارائه یک مدل برای تبدیل صوت به متن فارسی با استفاده از یادگیری عمیق [پایان‌نامه کارشناسی ارشد]. گروه کامپیوتر دانشکده مهندسی دانشگاه زنجان.

حسن آبادی، ابوالفضل (۱۳۹۶). درآمدی بر روش‌شناسی تدوین در تاریخ شفاهی انقلاب اسلامی ایران. *تاریخ‌نامه انقلاب*، ۱ (دفتر اول و دوم بهار و تابستان)، ۹-۲۹. https://tarikhname-ri-khomeini.ac.ir/article_50874.html

ربیعی، محبوبه، و میرزاییان، وحیدرضا (۱۴۰۱). کاربردهای پردازش زبان طبیعی در علم اطلاعات و دانش‌شناسی با تأکید بر کتابخانه‌های دیجیتال. *فصلنامه علمی بازیابی دانش و نظام‌های بازیابی*، ۹، ۱۹۷-۲۶۲. <https://doi.org/10.22054/jks.2022.64237.1478>
عزیزی، غلامرضا (۱۴۰۰). علم آرشیو و منابع شفاهی. *دو فصلنامه پژوهش‌های تاریخ شفاهی*، ۱۳ (۲)، ۱۵-۳۰. https://jorah.nlai.ir/article_2733.html
کمالی مهدیه، و شیخ طاهری، عباس (۱۳۹۷). مستندسازی گزارش‌های پرستاری با استفاده از فناوری تشخیص گفتار (مزایا، موانع و چالش‌ها و تسهیلگرها). *مجله انفورماتیک سلامت و زیست‌پزشکی*، ۵ (۱)، ۸۲-۷۰. <https://jhbmi.ir/article-1-268-fa.html>

منابع لاتین

Benkerzaz, S., Elmir, Y., & Dennai, A. (2019). A study on automatic speech recognition. *Journal of Information Technology Review*, 8(10), 77-85. <https://www.researchgate.net/publication/337155654>
Draxler, C., Heuvel, H. V. D., van Hessen, A., Ircing, P., & Lehečka, J. (2024). *Speech technology services for oral history research*. <https://arxiv.org/abs/2405.02333>

استفاده از فناوری تشخیص گفتار، مانند هر فناوری دیگری، همان‌قدر که فرصت‌ها و امکاناتی را فراهم می‌آورد، چالش‌هایی را نیز به همراه دارد. اما چالش‌ها نباید مانع شوند تا از کاربردهای این فناوری چشم‌پوشی کنیم. این فناوری مبتنی بر هوش مصنوعی است که به طرز شگفت‌آور و روزافزون در حال پیشرفت و گسترش در علوم و حوزه‌های مختلف است. با پیشرفت روزافزون فناوری، ما ملزم هستیم فناوری را در خدمت دانش درآوریم؛ چاره کار، کنار گذاشتن آن نیست.

این پژوهش برای بررسی فرصت‌ها و چالش‌های استفاده از فناوری تبدیل گفتار به متن در پیاده‌سازی مصاحبه‌های تاریخ شفاهی انجام شد تا با شناخت فرصت‌ها، امکان توجه به این فناوری و برنامه‌ریزی برای به‌کارگیری آن در پیاده‌سازی مصاحبه‌های تاریخ شفاهی فراهم شود. امروزه در کشور شرکت‌های دانش‌بنیان بسیاری در حال خدمت‌رسانی در زمینه هوش مصنوعی هستند که یکی از محصولات آنها خدمات پردازش گفتار است؛ شرکت‌هایی مانند مرکز تحقیقات هوش مصنوعی پارت که سرویس گفتار به متن آن با نام آوانگار در حال خدمت‌رسانی در این زمینه است. همچنین، فارس‌آوا نام خدمات پردازش گفتار شرکت دانش‌بنیان عامراندیش هوشمند است. پیشنهاد می‌شود سازمان‌های فعال در زمینه تولید تاریخ شفاهی و همچنین، سایر مراکز اسنادی و آرشیوی، چنانچه در نظر داشته باشند از این فناوری استفاده کنند، از شرکت‌های معتبر در این زمینه مشاوره دریافت کنند.

همچنین، با شناخت چالش‌ها و اطلاع‌رسانی در خصوص آنها، این امکان برای متخصصان حوزه هوش مصنوعی که ارائه‌دهنده خدمات فناوری تشخیص گفتار هستند، فراهم می‌شود تا با بررسی و مطالعه این چالش‌ها، درصدد یافتن راهکارهایی برای حل آنها باشند. فناوری‌های هوش مصنوعی مبتنی بر یادگیری هستند. استفاده از مدل‌های یادگیری عمیق و ایجاد مجموعه داده‌های خاص از گویش‌ها و لهجه‌های مختلف می‌تواند

- Naderifar, M., Goli, H., & Ghaljaei, F. (2017). Snowball sampling: A purposeful method of sampling in qualitative research. *Strides in Development of Medical Education*, In Press.
<http://dx.doi.org/10.5812/sdme.67670>
- Scagliola, S., Corti, L., Calamai, S., Karrouche, N., Beeken, J., van Hessen, A., ..., & Truong, K. P. (2020). Cross disciplinary overtures with interview data: Integrating digital practices and tools in the scholarly workflow. In K. Simov, & M. Eskevich (Eds.), *Selected papers from the CLARIN Annual Conference 2019* (pp. 126-136). (Linköping Electronic Conference Proceedings; Vol. 172). Linköping University Electronic Press.
<https://doi.org/10.3384/ecp2020172015>
- Gaber, M., Corpas Pastor, G., & Omer, A. (2020). Speech-to-Text technology as a documentation tool for interpreters: A new approach to compiling an ad hoc corpus and extracting terminology from video-recorded speeches. *TRANS: Revista de Traductología*, 24, 263-281. Retrieved from:
<https://revistas.uma.es/index.php/trans/article/view/7876>
- Gref, M. (2022). *Robust speech recognition via adaptation for German oral history interviews*. – Bonn [PhD thesis]. Bonn University.
<https://nbn-resolving.org/urn:nbn:de:hbz:5-68271>
- Leh, A., Gref, M., & Köhler, J. (2019). Audio mining: Advanced speech analytics for oral history. *Words and Silences, Memory and Narration, Journal of the International Oral History Association*, 1-9.
<https://ioha.org/journal/articles/audio-mining/>

¹ Speech to text

² Speech Recognition

ⁱⁱⁱ Automatic Speech Recognition

^{iv} Draxler

^v <https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface>

^{vi} <https://lindat.mff.cuni.cz/services/translation/>

^{vii} <https://openai.com/index/whisper/>

^{viii} Gref

^{ix} Gaber

^x Leh

^{xi} Noise

^{xii} Intersubjective

^{xiii} Frequency

^{xiv} فائزه توکلی، کارشناس پژوهشی و عضو گروه تاریخ پژوهشگاه علوم انسانی و مطالعات فرهنگی.

^{xv} ترجمه‌شده از مقاله «A study on Automatic Speech Recognition» (Benkerzaz et al., 2019, pp. 79-80).

^{xvi} Signal

