پرسش و پاسخ محاوره‌ای برای زبان‌های کم‌منبع: یک معماری جدیدِ تقویت‌شده با مدل‌های زبانی بزرگ

صفورا آقاداود جلفایی | دانشجوی دکتری مدیریت فناوری اطلاعات
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، تهران، ایران

آزاده محبی * | دکتری مهندسی طراحی سیستم‌ها
استادیار پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)، تهران،ایران

دریافت: ۱۴۰۴/۰۶/۳۱ | **پذیرش:** ۱۴۰۴/۰۹/۱۰ | مقاله برای اصلاح به مدت ۱۸ روز نزد پدیدآوران بوده است.

چکیده: سامانه‌های پرسش و پاسخ محاوره‌ای با ظهور مدل‌های زبانی بزرگ به طور قابل توجهی تکامل یافته‌اند ؛ با این حال، این پیشرفت‌ها عمدتاً به نفع زبان‌های پرمنبع بوده و زبان‌های کم‌منبع را نادیده گرفته‌اند. این مقاله چارچوب نوین تقویت‌شده با مدل‌های زبان بزرگ را معرفی می‌کند که به طور خاص برای پر کردن این شکاف زبانی طراحی شده است. معماری پیشنهادی دارای شش مؤلفه است: «پردازش ورودی» برای مدیریت ویژه‌ی زبان، «هسته مدل زبانی بزرگ تطبیق‌پذیر» ، «بهبود دانش» برای نگاشت بین زبانی، «مدیریت زمینه» برای راهبری کارآمد محاوره، «تولید پاسخ» همراه با انطباق فرهنگی، و «بازخورد انسانی» برای بهبود مستمر. برخلاف رویکردهای موجود، این چارچوب ملاحظات فرهنگی و زبانی را در سرتاسر فرایند پردازش لحاظ می‌کند. برای اعتبارسنجی این چارچوب، از یک ارزیابی کیفی از طریق گروه متمرکز متشکل از پنج متخصص پردازش زبان طبیعی استفاده شد. نتایج ارزیابی متخصصان، اثربخشی چارچوب پیشنهادی را در پرداختن به چالش‌های اساسی زبان‌های کم‌منبع، از جمله محدودیت داده، پیچیدگی‌های صرفی و ظرافت‌های فرهنگی، تأیید کرد. متخصصان به‌طور ویژه، رویکرد نوآورانه چارچوب در «پردازش فرهنگی یکپارچه» ، «کارایی منابع» از طریق مدیریت بهینه زمینه ، و «معماری ماژولار و مقیاس‌پذیر» را به عنوان دستاوردهای کلیدی برجسته ساختند.

کلیدواژه‌ها: پرسش و پاسخ محاوره‌ای ، درک ماشین، پردازش زبان طبیعی.

*پدیدآور رابط

نشریهٔ علمی (رتبه ین المللی)
پژوهشگاه علوم و فنّاوری اطلاعات ایران
شاپا(چاپی) ۸۲۲۳- ۲۲۵۱
شاپا(الکترونیکی) ۸۲۳۱- ۲۲۵۱
نمایه در LISTA، SCOPUS و ISC
http://jipm.irandoc.ac.ir
دوره xx | شمارهٔ x | ص ص xx-xx
۱۳xx x

نوع مقاله: مروری / پژوهشی

به این مقاله به شکل زیر استناد کنید:
درون متن:
(آقاداود جلفایی،محبی ، زودآیند)
در فهرست منابع:
آقاداودجلفایی، صفورا. محبی، آزاده. زودآیند. پرسش و پاسخ محاوره‌ای برای زبان‌های کم‌منبع ی‌ک معماری جدیدِ تقویت‌شده با مدل‌های زبانی بزرگ. *پژوهشنامه پردازش و مدیریت اطلاعات*.
http://Jipm.irandoc.ac.ir (دسترسی در روز/ماه/سال)

۱. مقدمه

در سیستم‌های پرسش و پاسخ[1]، کاربران قادر به دریافت پاسخ‌های دقیق به پرسش‌های زبان طبیعی هستند(J. Gao et al., 2019; Soares & Parreiras, 2020). این سیستم‌ها مؤلفه‌ای مهم از هوش مصنوعی محاوره‌ای[2] محسوب می‌شوند که دانش حوزه‌های پردازش زبان طبیعی[3]، بازیابی اطلاعات و یادگیری ماشین را در بر دارند.(Gupta et al., 2020) . هوش مصنوعی محاوره‌ای به ماشین‌ها امکان درک، پردازش و پاسخگویی به انسان‌ها به زبان طبیعی را می‌دهد(Kusal et al., 2022). طراحی سیستمی که بتواند همتراز یا فراتر از توانایی انسان در یک گفتگوی تعاملی مشارکت داشته باشد، و به پرسش‌هایی در موضوعات گوناگون پاسخ دهد یکی از دستاوردهای هوش مصنوعی است(J. Gao et al., 2019).

یک سیستم پرسش و پاسخ محاوره‌ای[4] باید قابلیت تفسیر صحیح پرسش به همراه ارائه پاسخ‌های هوشمندانه را داشته باشد. برای تحقق این امر، معمولا دسترسی به اطلاعات مرتبط در گفتگوهای پیشین نیز ضروری است (Vakulenko et al., 2021; Zaib et al., 2022). سامانه‌های محاوره‌ای باید پاسخ‌های دقیق ارائه دهند (پرسش و پاسخ)، درخواست کاربر را برآورده سازند (تکمیل وظیفه) و قادر به گفتگویی پیوسته و مناسب با کاربر باشند (گفتگوی اجتماعی) (J. Gao et al., 2019).

مدل‌های زبانی بزرگ[5] موجب بهبود چشمگیر سامانه‌های محاوره‌ای در پاسخگویی به پرسش‌های کاربران در طیف گسترده‌ای از وظایف زبانی شده‌اند. سامانه‌های پاسخگویی محاوره‌ای پیش از ظهور مدل‌های زبانی بزرگ، با محدودیت‌هایی از جمله درک ضعیف زمینه، حوزه‌های دانش محدود و ناتوانی در پردازش پرسش‌های پیچیده مواجه بودند. این سامانه‌ها غالباً پاسخ‌های نادرست یا نامرتبط ارائه می‌دادند که به تعاملات غیرطبیعی

[1] Question answering
[2] Conversational artificial intelligence
[3] Natural language processing
[4] Conversational question answering
[5] Large language models

می‌انجامید. همچنین، در پردازش اطلاعاتِ گسترده و به‌هنگام در موضوعات متنوع با چالش‌هایی روبرو بودند. معرفی مدل‌های زبانی بزرگ بسیاری از این چالش‌ها را مرتفع ساخته و قابلیت‌ها و تجربه کاربری هوش مصنوعی محاوره‌ای را به‌طور معناداری ارتقاء داده است (Dam et al., 2024).

با این حال، سامانه‌های مبتنی بر مدل‌های زبانی بزرگ اغلب فاقد آگاهی از اهداف[1] هستند؛ یعنی توانایی درک، حفظ و پیگیری فعالانه اهداف محاوره‌ای مشخص در حین تعامل با کاربران را ندارند. این ضعف عمدتاً از آنجا ناشی می‌شود که مدل‌های زبانی بزرگ معمولا برای پیروی از دستورالعمل‌های کاربران آموزش دیده‌اند، نه برای پیگیری اهداف محاوره‌ای خاص (Deng et al., 2023). هرچند مدل‌های زبانی بزرگ قابلیت تولید پاسخ را بهبود بخشیده‌اند، اما توسعه سیستم‌های هوش مصنوعی محاوره‌ای که قادر به حفظ رفتارهای مبتنی بر هدف باشند، به‌ویژه برای زبان‌های کم‌منبع[2] که محدودیت‌های بیشتری هم دارند، همچنان چالش‌برانگیز است.

توسعه هوش مصنوعی محاوره‌ای برای زبان‌های کم‌منبع به دلایل متعددی چالش‌برانگیز است؛ از جمله داده‌های آموزشی محدود، منابع محاسباتی ناکافی، چالش‌های ارزیابی، مشکلات فنی مرتبط با ساختارهای ریخت‌شناسی و نیاز به سازگاری فرهنگی (Andersland, 2024; Yong et al., 2024).

با توجه به این شکاف عمیق میان پیشرفت‌ها در زبان‌های پرمنبع و نیازهای اساسی زبان‌های کم‌منبع، این پژوهش با هدف پاسخگویی به سوالات کلیدی زیر شکل گرفته است :۱ .مؤلفه‌های ضروری یک چارچوب جامع برای سیستم‌های پرسش و پاسخ محاوره‌ای که به طور خاص برای زبان‌های کم‌منبع طراحی شده باشد، کدامند؟ ۲ .این چارچوب چگونه می‌تواند به چالش‌های منحصربه‌فرد این زبان‌ها، از جمله کمبود داده، پیچیدگی‌های ساختاری و ظرافت‌های فرهنگی، به طور مؤثری پاسخ دهد؟

در راستای پاسخ به این سوالات، این پژوهش اهداف اصلی مشخصی را دنبال می‌کند:

[1] Goal awareness
[2] Low-resource languages

نخســـت، طراحی یک چارچوب جامع و ماژولار که قابلیت انطباق با ویژگی‌های زبان‌های مختلف کم‌منبع را داشته باشد؛ دوم، یکپارچه‌سازی پردازش‌های زبانی و فرهنگی در تمامی مراحل برای تولید پاسـخ‌هایی دقیق و متناسـب با زمینه فرهنگی؛ و در نهایت، ارائه راه‌حلی کارآمد از نظر محاسباتی که در محیط‌های با منابع محدود نیز قابل پیاده‌سازی باشد.

برای دسـتیابی به این اهداف، در این مقاله یک چارچوب نوین شـش‌مؤلفه‌ای برای پاسـخگویی محاوره‌ای به پرسـش‌ها معرفی و تشـریح می‌شـود. این چارچوب برخلاف معماری‌های موجود که عمدتاً برای زبان انگلیسـی طراحی شـده‌اند، چالش‌های زبان‌های کم‌منبع مانند فارسـی را به عنوان محور اصـلی طراحی خود در نظر گرفته و راه‌حل‌هایی یکپارچه برای غلبه بر آن‌ها ارائه می‌دهد. در ادامه، پس از بررسی کارهای مرتبط، به تفصیل مؤلفه‌های این چارچوب و نحوه پاسـخگویی آن‌ها به چالش‌های مطرح‌شـده، ارائه خواهد شد .

۲. کارهای مرتبط

پیش از ظهور مدل‌های زبانی بزرگ، سـیسـتم‌های پرسـش و پاسـخ محاوره‌ای از طریق عامل‌های محاوره‌ای پیاده‌سـازی می‌شـدند که در دو دسـته قرار می‌گرفتند: عامل‌های مبتنی بر پایپ لاین[1] که با پردازش ترتیبی برای حوزه‌های با داده محدود مناسـب بودند (Harms et al., 2019)، و عامل‌های سرتاسری[2] که مستقیماً از داده‌های بزرگ یاد می‌گرفتند تا تعاملات طبیعی‌تری ایجاد کنند اما با چالش‌های امنیتی همراه بودند (Chen et al., 2017; Dinan et al., 2021). در این سیستم‌ها مؤلفه‌های کلیدی مانند مدل‌سازی تاریخچه گفتگو (Qu et al., 2019; Qian et al., 2022)، درک پرسش (Burgin et al., 2022; V. Kumar & Callan, 2020)، بازیابی دانش (Christmann et al., 2019; Yang et al., 2019) و تولیدپاسخ (Burgin et al., 2022; Campos, Otegi, et al., 2020) وجود داشـتند. پژوهش‌ها در این حوزه روی حوزه‌های نظیر بهبود فرآیند مکالمه از طریق اسـتفاده از بازخورد (Campos, Cho, Otegi, et al., 2020)، حل مشـکل فراموشـی در گفتگوهای چند مرحله‌ای (Qian et al., 2022)، مدیریت داده‌های سـاختاریافته و بدون سـاختار (Burgin et al., 2022) و بهبود اسـتدلال با یادگیری چندوظیفه‌ای (Kongyoung et al., 2022) متمرکز بودند. نتایج

[1] Pipeline
[2] end-to-end

این پژوهش‌ها، امکان مدیریت انواع پرسش‌ها و تلفیق منابع دانش گوناگون را فراهم کردند. جدول ۱ نمونه ای از چارچوب ها و معماری های این سیستم‌ها را قبل از ظهور مدل های زبانی نشان می‌دهد.

جدول۱ : چارچوب های پرسش و پاسخ محاوره ای پیش از مدل‌های زبانی بزرگ

چارچوب	ویژگی‌های کلیدی	مؤلفه‌های اصلی
(Qu et al., 2019)	تمرکز بر مدل‌سازی و انتخاب تاریخچه	انتخاب تاریخچه، رمزگذار، ماژول استدلال، پیش‌بینی‌کننده خروجی
(Su et al., 2019)	تمایز بین انواع پرسش	استخراج منطق، درک پرسش، پرسش و پاسخ راستی‌آزمایی، پرسش و پاسخ جستجوی دانش
(Burgin et al., 2022)	رویکرد ترکیبی برای داده‌های ساختاریافته و غیرساختاریافته	طبقه‌بندی قصد، بررسی‌کننده
(Christmann et al., 2019)	مدیریت منابع ناهمگن (پایگاه دانش، متن، جداول)	بازنمایی پرسش، ضبط شواهد، تولید پاسخ
(Christmann et al., 2023)	ترکیب چندین منبع با توضیحات	ایجاد گراف ناهمگن، شبکه عصبی گرافی با توجه سطح پرسش

علیرغم ایجاد معماری بنیادی برای هوش مصنوعی محاوره‌ای، این چارچوب‌های پیش از LLM با محدودیت‌های قابل توجهی در رابطه با مقیاس‌پذیری و درک عمیق متنی مواجه بودند. این سیستم‌ها به شدت به مجموعه داده‌های حاشیه‌نویسی شده گسترده و اجزای خط لوله‌ای سفت و سخت متکی بودند، که آنها را به ویژه برای زبان‌های کم منبع که در آنها داده‌های آموزشی کمیاب و پیچیدگی مورفولوژیکی بالا است، نامناسب می‌کرد. این محدودیت‌ها، تغییر الگو به سمت معماری‌های انعطاف‌پذیرتر و بهبود یافته با مدل های زبانی بزرگ را ضروری ساخت.

در چشم‌انداز کنونی هوش مصنوعی محاوره‌ای، مدل‌های زبانی بزرگ با فراهم کردن تعاملات طبیعی‌تر، قابلیت‌های سیستم‌های پرسش و پاسخ را به طور چشمگیری بهبود بخشیده‌اند (Dam et al., 2024). این مدل‌ها، از طریق فرآیندی سه‌مرحله‌ای شامل تنظیم دقیق[1]، یادگیری تقویتی مبتنی بر بازخورد انسانی[2] و ارزیابی انسانی آموزش می‌بینند تا با

[1] Fine tuning
[2] Reinforcement learning from human feedback

هدف کاربر همسو شوند (Zhao et al., 2024; Dam et al., 2024). برای غلبه بر چالش‌های کلیدی مانند نمایش زمینه و کسب دانش، رویکردها از بازنویسی پرسش (Z. Wu et al., 2021) به سمت مدل‌سازی مستقیم زمینه مکالمه (Dinan et al., 2021; Harms et al., 2019) با مدل‌های زبانی بزرگ و استفاده از روش‌های تولید تقویت‌شده با بازیابی[1] برای تکمیل ورودی‌ها با پایگاه دانش بیرونی (Asai et al., 2023; Qin et al., 2023) تکامل یافته‌اند.

مطالعات اخیر رویکردهای نوآورانه‌ای را معرفی کرده‌اند، از جمله رویکردهای ترکیبی که مدل های زبانی بزرگ را با عامل‌های سنتی ادغام می‌کنند (Foosherian et al., 2023). چارچوب ConvRAG بازیابی اطلاعات را در سطح مکالمه بهینه می‌کند (Ye et al., 2024). چارچوب RAISE (N. Liu et al., 2024) با الهام از ReAct (Yao, Zhao, et al., 2023) و با استفاده از حافظه دوگانه، کنترل‌پذیری را افزایش می‌دهد. مدلی دیگر تنظیم دقیق، معماری معلم-شاگرد[2] و یادگیری تقویتی[3] را برای پیمایش گراف دانش[4] ترکیب می‌کند و چارچوبی ارائه می دهد که با برنامه‌ریزی پیش از مکالمه و ایجاد یک رویه عملیاتی استاندارد[5]، گفتگوهای هدفمند و انطباق‌پذیر را با حداقل ورودی انسانی ممکن می‌سازد (Z. Li, Peng, et al., 2024).

پیشرفته‌ترین مدل‌های زبانی بزرگ مانند GPT و LLaMA با وجود توانایی‌های استثنایی، در پردازش زبان‌های کم‌منبع به دلیل کمبود داده با چالش مواجه هستند (Z. Li, Shi, et al., 2024; Robinson et al., 2023). اگرچه روش‌هایی مانند تنظیم دقیق۶ و یادگیری مستمر[7] بررسی شده‌اند، اما اغلب به مشکلاتی نظیر فراموشی فاجعه‌آمیز[8] منجر

[1] Retrieval-augmented generation (RAG)
[2] teacher-student model
[3] reinforcement learning
[4] Knowledge graph
[5] Standard Operating Procedure (SOP)
[6] Language specific fine tuning
[7] Continuous learning
[8] Catastrophic forgetting

می‌شوند (Adelani et al., 2021; A. Kumar et al., 2022).

در این میان، یادگیری درون‌زمینه‌ای[1] به عنوان یک راه‌حل کلیدی مطرح شـــده که به مدل‌ها اجازه می‌دهد با چند نمونه، دانش خود را به زبان‌های جدید منتقل کرده و نیاز به داده را کاهش دهند(Alam et al., 2024; Cahyawijaya et al., 2024). با این حال، نوع میان زبانی[2] آن همچنان در رقابت با روش‌های مبتنی بر ترجمه برای زبان‌های بسیار کم‌منبع با چالش روبروست (Cahyawijaya et al., 2024; Tanwar et al., 2023). برای رفع این محدودیت‌ها، رویکردهای دیگری نیز توسعه یافته‌اند؛ از جمله ساخت مدل‌های چندوجهی برای زبان‌هایی مانندAmharic[3] با استفاده از افزایش داده[4] از طریق ترجمه (Andersland, 2024)، و چارچوب میان زبانی به کمک ترجمه[5] که با به‌کارگیری ترجمه در فرآیند زنجیره-تفکر[6] و یادگیری برنامه‌ریزی‌شـــده، هزینه‌های افزودن زبان‌های جدید را کاهش می‌دهد (Upadhayay & Behzadan, 2024). در نهایت، مطالعات نشـــان می‌دهند که اگرچه مدل های زبانی بزرگ در وظایف ســـاده برای زبان‌های کم‌منبع عملکرد خوبی دارند، اما برای وظایف زبانی پیچیده، تخصص انسانی همچنان برتر است و نیازمند برچسب گذاری انســـانی اســـت (Nasution & Onan, 2024). جدول ۲ اجزای کلیدی از نمونه چارچوب‌ها و معماری‌های این ســـیســـتم ها را بعد از ظهور مدل‌های زبانی بزرگ نشـــان می‌دهد.

[1] In-context learning (ICL)
[2] Cross-lingual in-context learning - X-ICL
[3] زبان رسمی کشور اتیوپی
[4] Data augmentation
[5] Translation-Assisted Cross-Linguality
[6] chain-of-thought

جدول۲ : چارچوب‌های پرسش و پاسخ محاوره ای پس از بروز مدل‌های زبانی بزرگ

چارچوب	ویژگی‌های کلیدی	اجزای اصلی
(Foosherian et al., 2023)	بهبود عامل‌های مبتنی بر خط لوله سنتی، ایجاد تعادل بین قابلیت‌ها و حریم خصوصی	تولید داده طبقه‌بندی قصد و منظور از طرح پرسش فرمول‌بندی پاسخ
(Ye et al., 2024)	بازیابی در سطح گفتگو، بهبود وضوح پرسش	ابزار روشن‌سازی پرسش بازیابی اطلاعات خاص بررسی‌کننده مفید بودن اطلاعات
(N. Liu et al., 2024)	تقلید از حافظه انسانی مدیریت گفتگوهای پیچیده	دفترچه یادداشت (حافظه کوتاه‌مدت) ماژول بازیابی (حافظه بلندمدت) رویکرد چندمرحله‌ای فاز آموزش مدل زبانی بزرگ
(L. Liu et al., 2024)	ترکیب تنظیم دقیق و یادگیری تقویتی تمرکز بر بازفرمول‌بندی پرسش	تنظیم دقیق مدل زبانی بزرگ معماری معلم-شاگرد مکان‌یابی پاسخ مبتنی بر یادگیری تقویتی
(Z. Li, Peng, et al., 2024)	ترکیب برنامه‌ریزی پیش از گفتگو با مدیریت گفتگوی هدف‌محور	ایجاد رویه عملیاتی استاندارد آفلاین برنامه‌ریزی اقدام آنلاین تولید پاسخ هدایت‌شده با رویه عملیاتی استاندارد تولید داده

اگرچه تغییر به معماری‌های مبتنی بر مدل های زبانی بزرگ از طریق استدلال پیشرفته، تولید تقویت‌شده با بازیابی و برنامه‌ریزی تقویت‌شده با حافظه، هوش مصنوعی محاوره‌ای را متحول کرده است، اما این چارچوب‌ها عمدتاً انگلیسی‌محور و منابع‌محور باقی مانده‌اند. معماری‌های پیشرفته فعلی اغلب در پرداختن به محدودیت‌های منحصر به فرد زبان‌های کم‌منبع، مانند پیچیدگی مورفولوژیکی و کمبود داده‌های آموزشی با کیفیت بالا، شکست می‌خورند. علاوه بر این، راه‌حل‌های موجود معمولاً سازگاری فرهنگی را به عنوان یک مرحله پس‌پردازش ثانویه به جای یک جزء جدایی‌ناپذیر از خط لوله استدلال در نظر می‌گیرند که منجر به تعاملات از نظر فنی صحیح اما از نظر فرهنگی ناموزون می‌شود.

۳. روش پژوهش

پژوهش حاضر با رویکردی کیفی و توسعه‌ای در دو مرحله اصلی انجام شده است: ۱) طراحی چارچوب نظری و ۲) اعتبارسنجی چارچوب از طریق ارزیابی متخصصان.

مرحله اول(طراحی چارچوب): برای رسیدن به چارچوب پیشنهادی، از روش تحلیل شکاف عملکردی و مرور ساختاریافته ادبیات پژوهش استفاده شد .در این مرحله، ابتدا چارچوب‌ها و معماری‌های موجود برای پرسش و پاسخ محاوره‌ای مورد مطالعه قرار گرفتند .سپس، چالش‌های کلیدی پردازش زبان‌های کم‌منبع که در پژوهش‌های اخیر شناسایی شده‌اند (مانند کمبود داده، پیچیدگی‌های صرفی، و نیاز به انطباق فرهنگی) استخراج گردید. با تحلیل شکاف میان قابلیت‌های چارچوب‌های موجود و نیازمندی‌های زبان‌های کم‌منبع، مؤلفه‌های ضروری برای یک چارچوب جامع شناسایی و طراحی شدند. این فرآیند بر پایه مطالعه و تحلیل عمیق بیش از ۶۰ مقاله کلیدی در این حوزه شکل گرفت.

مرحله دوم (اعتبارسنجی چارچوب): پس از طراحی اولیه چارچوب، برای اعتبارسنجی ساختار و کارایی مفهومی آن، از روش کیفی گروه متمرکز[1] استفاده شد .در این مطالعه، پنج پژوهشگر متخصص در حوزه پردازش زبان طبیعی مشارکت داشتند. این متخصصان بر اساس معیارهایی نظیر داشتن مدرک دکتری تخصصی در هوش مصنوعی، سابقه انتشار مقالات علمی-پژوهشی در کنفرانس‌ها و مجلات معتبر حوزه پردازش زبان طبیعی، و تجربه عملی در پروژه‌های مرتبط با زبان‌های کم‌منبع (به‌ویژه فارسی) انتخاب شدند .این گروه متشکل از اعضای هیئت علمی دانشگاه و پژوهشگران ارشد بودند. فرآیند ارزیابی شامل مصاحبه‌های نیمه‌ساختاریافته بر اساس پرسشنامه‌ای دقیق (ارائه‌شده در پیوست ۱) و بحث‌های عمیق پیرامون مؤلفه‌های چارچوب بود تا اثربخشی آن‌ها در مواجهه با چالش‌های زبان‌های کم‌منبع سنجیده شود .نتایج این ارزیابی در بخش «ارزیابی و مقایسه» ارائه شده است.

[1] Focus Group

۴. چارچوب پیشنهادی

چارچوب پیشنهادی، همان‌طور که در شکل ۱ نشان داده شده، از شش مؤلفه اصلی ۱)پردازش ورودی، ۲)هسته مدل زبانی بزرگ، ۳)تقویت دانش، ۴)مدیریت زمینه، ۵)تولید پاسخ و ۶)بازخورد انسانی تشکیل شده است.

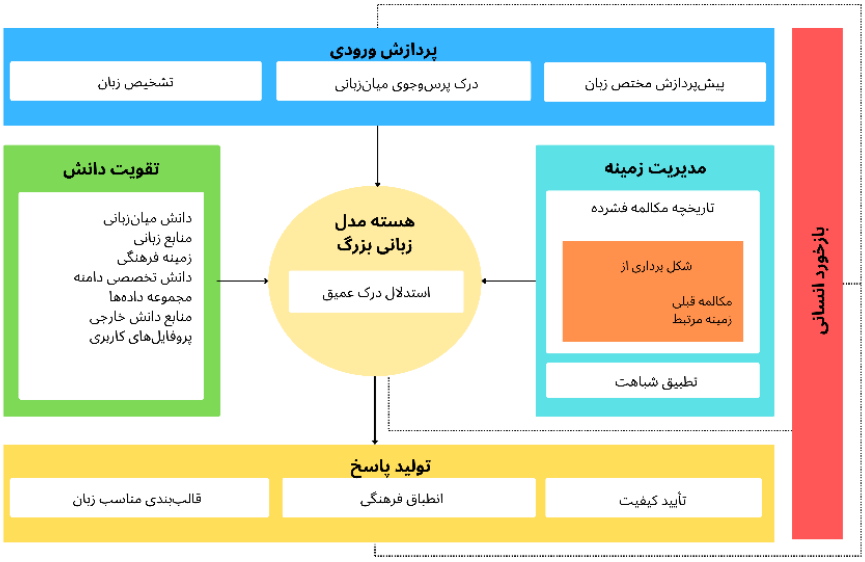
این ساختار با استفاده ازتحلیل شکاف‌ها در چارچوب‌های موجود (Campos, Cho, Otegi, Soroa, Agirre, et al., 2020; Kongyoung et al., 2022; Su et al., 2019) و رسیدگی به چالش‌های خاص پردازش زبان‌های کم‌منبع که در ادبیات پژوهش شناسایی شده‌اند (Kargaran et al., 2023; Yong et al., 2024)، پیشنهاد شده است. مطالعات پیشین بر لزومِ مدیریت تخصصی ورودی (Nooralahzadeh et al., 2020)، درک زمینه‌ای (Mao & Yu, 2024; Merx et al., 2024; Robinson et al., 2023) و انطباق فرهنگی (Azime et al., 2025) برای پردازش مؤثر زبان‌های کم‌منبع تأکید کرده‌اند.

اگرچه چارچوب پیشنهادی عمدتاً بر زبان فارسی متمرکز است، اما طراحی این چارچوب به چالش‌های مشترک بسیاری از زبان‌های کم‌منبع رسیدگی می‌کند. به عنوان مثال، مؤلفه پردازش ورودی می‌تواند خطوط هجایی[1] زبان Amharic (Nigusie & Tegegne, 2023)، ماهیت آهنگین[2] زبان Yoruba (Adelani et al., 2021) و صرف ترکیبی[3] زبان کردی (Hassani, 2018) را مدیریت کند. به طور مشابه، ماژول انطباق فرهنگی در بخش تولید پاسخ، سطوح مختلف رسمی بودن را که در زبان‌هایی مانند Swahili (Githiora, 2018) و Amharic ضروری است، در نظر می‌گیرد؛ زبان‌هایی که در آن‌ها سلسله‌مراتب اجتماعی به طور قابل توجهی بر الگوهای ارتباطی تأثیر می‌گذارد.

چارچوب شش‌مؤلفه‌ای پیشنهادی به طور نظام‌مند به این چالش‌ها رسیدگی کرده و در عین حال، کارایی محاسباتی و سازگاری[4] را تضمین می‌کند. هر مؤلفه دارای زیرمؤلفه‌ها یا

[1] syllabic scripts
[2] tonal
[3] agglutinative morphology
[4] adaptability

مراحلی است که در بخش‌های بعدی به تفصیل شرح داده شده است.



شکل ۱ : چارچوب پرسش و پاسخ محاوره‌ای برای زبان های با منابع کم

پردازش ورودی

پردازش ورودی اولین مؤلفه است که تعامل اولیه با پرس‌وجوهای کاربر به هر زبانی را مدیریت می‌نماید. این مؤلفه برای زبان‌های کم‌منبع به ویژه ضروری است، و سه ماژول کلیدی را در بر می‌گیرد (جدول ۳):

تشخیص زبان[1]: این ماژول زبان را شناسایی کرده و سطح دسترسی به منابع آن را نیز ارزیابی می‌کند تا مبنای تصمیم‌گیری در مراحل پردازش بعدی قرار گیرد برای مثال، تشخیص اینکه یک پرس‌وجو به زبان فارسی است و آگاهی از منابع پردازش زبان طبیعی برای آن (Kargaran et al., 2023).

درک پرس‌وجوی بین‌زبانی[2]: این ماژول، قصد و نیت پرس‌وجو را حتی در صورت کمیاب بودن منابع پردازشی مستقیم در زبان هدف، مشخص می‌کند. این کار با بهره‌گیری

[1] Language Detection
[2] Cross-lingual Query Understanding

از یادگیری انتقالی[1] از زبان‌های پرمنبع و بردارهای نهفته بین‌زبانی[2] انجام می‌شود (Miao et al., 2024; Nooralahzadeh et al., 2020)، به عنوان مثال پرس‌وجوی پزشکی به زبان بنگالی از طریق نگاشت آن به الگوهای مشابه در زبان انگلیسی.

پیش‌پردازش مختص زبان[3]: این ماژول پردازش‌های مربوط به ویژگی‌های منحصربه‌فرد در زبان کم‌منبع را انجام می‌دهد، مانند: نرمال‌سازی نوشتار[4]، تحلیل صرفی[5] و توکن‌سازی تخصصی[6]. برای مثال، مدیریت صرف پیچیده در زبان ترکی یا تنوع رسم‌الخط در گویش‌های عربی. تأثیر این مؤلفه بر پردازش زبان‌های کم‌منبع بسیار قابل توجه است (Howcroft & Gkatzia, 2022; Sun et al., 2021).

اگر پردازش ورودی نتواند به درستی یک پرس‌وجوی کم‌منبع را مدیریت کند، کل زنجیره پردازش در مراحل بعدی با اختلال مواجه خواهد شد. به عنوان مثال، هنگام پردازش یک پرس‌وجو به زبان فارسی، این مؤلفه نه تنها باید زبان را تشخیص دهد، بلکه باید رسم‌الخط را نرمال‌سازی کند، صرف پیچیده را مدیریت نماید و قصد پرس‌وجو را به مفاهیمی که برای مدل زبانی بزرگ قابل پردازش است، نگاشت دهد؛ در حالی که تمام این کارها را با ابزارها و منابع زبانی محدودتری در مقایسه با زبان‌های پرمنبعی مانند انگلیسی انجام می‌دهد.

---

[1] transfer learning
[2] Cross-lingual embeddings
[3] Language-specific Preprocessing
[4] Script normalization
[5] morphological analysis
[6] Specialized tokenization

جدول۳ : ماژول‌های پردازش ورودی، شرح، تکنیک‌ها و نمونه پیاده‌سازی‌ها

ماژول	تشریح	نمونه فرایند (حوزه پزشکی)	تکنیک ها و ارجاعات
تشخیص زبان	شناسایی زبان و ارزیابی میزان منابع در دسترس	ورودی : سردرد شدید دارم خروجی: زبان: فارسی ، خط: فارسی–عربی ، سطح منابع: متوسط	Language ID with Resource Assessment (Kargaran et al., 2023) Script and Dialect Detection (Alam et al., 2024; Joshi et al., 2025) Resource Level Classification
درک پرسش میان‌زبانی	درک هدف پرسش در زبان‌های مختلف	ورودی : سردرد شدید دارم خروجی: حوزه: پزشکی، هدف: توصیف علائم، مفهوم اصلی: نگاشت به "having severe headache" نوع پرسش: مشاوره سلامت	Cross-lingual Transfer Learning (Nooralahzadeh et al., 2020) Multilingual Embeddings (Miao et al., 2024)
پیش‌پردازش ویژه زبان	مدیریت ویژگی‌های زبان‌شناختی منحصربه‌فرد زبان هدف	ورودی : سردرد شدید دارم خروجی: نرمال‌سازی‌شده: کاراکترهای استانداردشده، واژه‌بندی‌شده: ["سردرد"، "شدید"، "دارم"]، تحلیل صرفی: زمان حال، اول شخص آماده برای پردازش مدل زبان بزرگ	Script Normalization (Howcroft & Gkatzia, 2022) Morphological Analysis Specialized Tokenization

هسته مدل زبانی بزرگ

هســـته مدل زبانی بزرگ به عنوان مؤلفه مرکزی برای درک عمیق و تولید زبان عمل می‌کند. این بخش به مثابه مرکز شناختی سیستم است که به جای درک سطحی پرس‌وجو، بر درک عمیق معنایی و استدلال تمرکز دارد. هنگامی که هسته، پردازش یک پرس‌وجوی ورودی را آغاز می‌کند، به جای نگاشت آن به یک «نیت» ساده (مانند سیستم‌های سنتی)، بر درک عمیق معنایی و اســـتدلال تمرکز می‌نمایـد. برای مثال، هنگـام پردازش یـک

پرس‌وجوی پزشکی درباره علائم کووید به زبان فارسی، هسته صرفاً متوجه نمی‌شود که این یک سؤال مرتبط با سلامت است (که لایه ورودی قبلاً آن را مشخص کرده)، بلکه وارد استدلال عمیق‌تری می‌شود: تعیین می‌کند که چه دانش پزشکی خاصی مورد نیاز است، چگونه باید اطلاعات را برای مخاطب عام قالب‌بندی کرد، چه عوامل فرهنگی را باید در زمینه فارسی‌زبان در نظر گرفت و چگونه یک پاسخ جامع را ساختاربندی نمود. سپس هسته تصمیم می‌گیرد که چه اطلاعاتی را در پاسخ بگنجاند و چگونه آن را ساختاربندی کند، پیش از آنکه این تصمیمات را برای قالب‌بندی نهایی زبانی به لایه تولید پاسخ منتقل نماید. این ویژگی، هسته را هم از پردازش ورودی (که مسئول درک اولیه است) و هم از تولید پاسخ (که مسئول تولید نهایی زبانی است) متمایز می‌سازد. این بخش به عنوان موتور استدلال عمل می‌کند که از طریق پردازش عمیق معنایی و تصمیم‌گیری، پلی میان درک و تولید ایجاد می‌نماید.

به طور خاص برای چالش‌های زبان‌های کم‌منبع، چارچوب پیشنهادی روش‌های انطباق با پارامتر بهینه[1] را در هسته پیاده‌سازی می‌کند. این رویکرد به طور قابل توجهی نیازمندی‌های محاسباتی را کاهش داده و در عین حال، عملکرد را برای زبان‌های کم‌منبع حفظ کرده و اغلب بهبود می‌بخشد (Gurgurov et al., 2025). روش‌های انطباق سبک‌وزن[2]، انتقال بین‌زبانی مؤثر را با حداقل پارامترهای اضافی امکان‌پذیر می‌سازند، که این امر پیاده‌سازی سیستم را از نظر فنی امکان‌پذیر می‌کند و از نظر اقتصادی برای استقرار در بسترهای زبانی متنوع، نیز مقرون‌به‌صرفه است. جدول ۴ این روش‌ها را نشان می دهد.

[1] parameter-efficient adaptation
[2] lightweight

جدول٤: روش‌های یکپارچه هسته مدل زبانی بزرگ، شرح، تکنیک‌ها و نمونه پیاده‌سازی‌ها

روش‌های یکپارچه	تشریح	نمونه (حوزه علمی)	تکنیک‌ها و ارجاعات
انطباق با پارامتر بهینه	امکان انطباق دقیق با زبان‌های کم‌منبع را بدون نیاز به منابع محاسباتی گسترده یا مجموعه داده‌های بزرگ فراهم می‌کند	تبدیل یک مدل زبانی عمومی برای پردازش پرسش‌ها به زبان فارسی یا کردی با حداقل داده‌های آموزشی اضافی	LoRA (Hu et al., 2021) Sequential Bottleneck Adapters (Houlsby et al., 2019) Sequential Bottleneck with Invertible Layers (Seq_bn_inv)(Pfeiffer, Vulić, et al., 2020) AdapterFusion (Pfeiffer, Kamath, et al., 2020)
پردازش میان‌زبانی[1]	امکان انتقال دانش تخصصی از زبان‌های پرمنبع به زبان‌های کم‌منبع را با حفظ دقت حوزه‌ای فراهم می‌کند	ورودی: مقالات جدید در مورد یادگیری عمیق چیست؟ فرایند: نگاشت یادگیری عمیق به "deep learning" شناسایی نوع پرسش علمی تشخیص بررسی مقالات اخیر	mT5 (Xue et al., 2021) XLM-R (Conneau et al., 2020) BLOOM(Workshop et al., 2023) XGLM(X. V. Lin et al., 2022) RemBERT(Deng et al., 2023)
استراتژی‌های پرامپت نویسی	راهنمایی مدل زبان بزرگ برای پردازش محتوا	ورودی : پرس و جوی علمی فرایند: استفاده از نمونه‌های مقالات آکادمیک، ساختاردهی پاسخ علمی، حفظ دقت علمی نمونه پرامپت : با توجه به مقالات اخیر درباره X یافته‌های کلیدی را در قالب Y خلاصه کنید.	Few-shot Learning(Nguyen et al., 2024; Song et al., 2023) Chain of Thought(Upadhayay & Behzadan, 2024; Wei et al., 2023) Tree of thought (Yao, Yu, et al., 2023)

[1] Cross-lingual Processing

تقویت دانش

مؤلفه تقویت دانش به عنوان مخزن جامعی عمل می‌کند که از هسته مدل زبانی بزرگ در درک و پردازش زبان‌های کم‌منبع پشتیبانی می‌نماید. این بخش هفت نوع دانش ضروری را نگهداری می‌کند:

۱. دانش بین‌زبانی[1] برای نگاشت مفاهیم بین زبان‌ها،
۲. منابع زبانی[2] حاوی قواعد دستوری و زبان‌شناختی،
۳. زمینه فرهنگی[3] که هنجارهای اجتماعی را در بر می‌گیرد،
۴. دانش تخصصی[4] در حوزه‌های مختلف،
۵. مجموعه داده‌ها[5] که اطلاعات ساختاریافته را فراهم می‌کند،
۶. منابع دانش خارجی[6] که دانش مرجع تکمیلی ارائه می‌دهند،
۷. پروفایل‌های کاربری[7] حاوی داده‌های تعاملات پیشین.

مجموعه داده‌ها، که بخش حیاتی این مؤلفه را تشکیل می‌دهند، معمولاً توسط انسان برچسب‌گذاری شده و حاوی حاشیه‌نویسی‌هایی شامل نیت، موجودیت، نوع، رابطه و سه‌تایی هستند (Perez-Beltrachini et al., 2023). این سه‌تایی‌ها که برای نمایش دانش ساختاریافته به کار می‌روند، معمولاً از ساختار (موضوع، گزاره، مفعول) یا (موجودیت ۱، رابطه، موجودیت ۲) پیروی می‌کنند؛ برای مثال، رابطه پایتختی بین «تهران» و «ایران» با سه‌تایی (تهران، پایتخت است، ایران) نمایش داده می‌شود. چالش اساسی در زبان‌های کم‌منبع، کمیابی شدید یا نبود کامل مجموعه داده‌های باکیفیت است. اتکا به برچسب‌گذاری انسانی برای ایجاد این داده‌ها، اگرچه دقیق است، اما به دلیل هزینه‌های بسیار بالا و زمان‌بر بودن، اغلب در عمل امکان‌پذیر نیست (Kusal et al., 2022). در مقابل مدل‌های زبانی بزرگ پتانسیل آن را دارند که هزینه‌های مالی و محاسباتی را در

[1] cross-lingual knowledge
[2] Language resources
[3] Cultural context
[4] Domain-specific knowledge
[5] Datasets
[6] External knowledge sources
[7] User profiles

حاشـیه‌نویسـی داده‌ها به طور قابل توجهی کاهش دهند، اما همزمان ممکن اسـت منجر به ایجاد توهم[1] در نتایج بشوند (Kholodna et al., 2024). اگرچه پذیرش سطحی از نویز و توهم در نتایج در این فرآیند اجتناب‌ناپذیر است، اما این رویکرد می‌تواند به ایجاد مجموعه داده‌های اولیه در محیط‌های کم‌منبع کمک شـایانی کند. علیرغم این محدودیت‌ها، مؤلفه «تقویت دانش» با فراهم کردن زمینه و نگاشـت‌های ضـروری (شـامل دانش بین‌زبانی و فرهنگی)، به طور فعال به هسته در درک و پردازش پرس‌وجوها، به‌ویژه در غیاب یا نویزی بودن داده‌های سـاختاریافته، کمک می‌کند. در جدول ۵، تمامی انواع دانشـی که توسـط مؤلفه تقویت دانش نگهداری می‌شـوند، به همراه تشـریحات و تکنیک‌های پیشـرفته مرتبط که توسط پژوهش‌های اخیر پشتیبانی می‌شوند، ارائه شده است.

مدیریت زمینه

مؤلفه مدیریت زمینه با نگهداری یک شـکل فشـرده یا برداری۲ از تاریخچه مکالمه مرتبط و استفاده از تطبیق شباهت برای شناسایی زمینه مرتبط، به هسته کمک می‌کند. به عنوان مثال، زمانی که کاربر یک سـؤال تکمیلی می‌پرسـد، این مؤلفه می‌تواند به سـرعت مرتبط‌ترین بخش‌های مکالمه قبلی را بازیابی کرده و در اختیار هسـته قرار دهد، به جای اینکه مجبور باشـد هر بار کل تاریخچه مکالمه را پردازش کند. این امر به ویژه برای زبان‌های کم‌منبع که در آن‌ها کارایی پردازش اهمیت دارد، مفید است. یک چالش اساسی در زبان‌های کم‌منبع، آن اسـت که داده‌های آموزشـی ناکافی منجر به تولید بردارهای برداری تک‌زبانه ضعیف می‌شود که در درک ظرافت‌های معنایی و تطبیق شباهت، کارایی لازم را ندارند. برای غلبه بر این چالش، چارچوب پیشـنهادی به جای اتکا به بردارهای بومی، از «بردارهای نهفته میان‌زبانی که در مؤلفه‌های «پردازش ورودی» و «تقویت دانش» نیز به کار گرفته می‌شوند، بهره می‌برد .این بردارها که از طریق هم‌راستاسازی با زبان‌های پرمنبع آموزش دیده‌اند، امکان تطبیق شـباهت معنایی را حتی با وجود منابع محدود فراهم

[1] Hallucination
[2] embedding

می‌سازند. به عنوان مثال، زمانی که کاربر یک سؤال تکمیلی می‌پرسد، این مؤلفه می‌تواند به سرعت مرتبط‌ترین بخش‌های مکالمه قبلی را بازیابی کرده و در اختیار هسته قرار دهد، به جای اینکه مجبور باشد هر بار کل تاریخچه مکالمه را پردازش کند.

جدول۵ : انواع منابع دانشی

انواع دانش	تشریح	نمونه‌ها (حوزه پزشکی)	تکنیک‌ها
دانش میان‌زبانی	نگاشت مفاهیم بین زبان‌های مختلف	English "headache" → Persian "سردرد" English "take twice daily" → Persian "روزی دو بار میل کنید"	Cross-lingual Embeddings (Conneau et al., 2020) Neural Machine Translation (Tan et al., 2019)
منابع زبانی	قواعد دستوری و زبان‌شناختی	صیغه‌های فعل پزشکی : بخور، میل کنید	Morphological Learning (Fam & Lepage, 2024) Grammar Induction (Kai et al., 2024)
زمینه فرهنگی	هنجارها و باورهای اجتماعی و فرهنگی	مشارکت خانواده در تصمیمات پزشکی، مفاهیم طب سنتی	Cultural Information Extraction (Ringel et al., 2019)
دانش تخصصی	اطلاعات تخصصی حوزه	نام‌های محلی داروها، ساختار نظام سلامت	Knowledge Graph Construction (J. Li et al., 2024)
مجموعه داده‌ها	مجموعه‌های داده ساختاریافته	جفت‌های پرسش و پاسخ پزشکی به زبان فارسی، پیکره‌های موازی (متون پزشکی انگلیسی–فارسی)، گفتگوهای حاشیه‌نویسی‌شده	Data Augmentation (Soudani et al., 2023; Q. Wu et al., 2021) Few-shot Learning (Brown, 2020) LLMs (Kholodna et al., 2024)
منابع دانشی خارجی	مواد مرجع اضافی	وب‌سایت‌ها/اسناد پزشکی، راهنماهای مراقبت سلام، مقالات علمی، پروتکل‌های پزشکی محلی	RAG (Y. Gao et al., 2024; Lewis et al., 2020)
پروفایل های کاربری	داده‌های تاریخچه تعاملات	پرسش‌های پزشکی گذشته، ترجیحات زبانی، علایق حوزه‌ای، الگوهای پرسش	User Modeling Pattern Mining

مدیریت زمینه مانند یک سیستم حافظه هوشمند عمل می‌کند و به هسته کمک می‌کند تا به جای پردازش مجدد و مداوم زمینه گذشته، بر وظیفه فعلی تمرکز نماید. تنها هدف این

مؤلفه، کمک به کارآمدتر شدن هسته از طریق مدیریت تاریخچه مکالمه به شیوه‌ای است که بار پردازشی را کاهش دهد؛ امری که کارایی محاسباتی را ارتقا داده و به‌ویژه در سناریوهایی که منابع پردازشی نیز (مانند خود داده‌ها) محدود هستند، ضروری است .

تولید پاسخ

مؤلفه «تولید پاسخ»، درک پردازش‌شده را از هسته دریافت کرده و با استفاده از مدل‌های زبانی، یک فرآیند سه‌مرحله‌ای هدایت‌شده را برای ایجاد پاسخ‌های مناسب به زبان کم‌منبع هدف دنبال می‌کند. این رویکرد به جای اتکای به یک تولیدِ تک‌مرحله‌ای و کنترل‌نشده، بر افزایش کنترل‌پذیری و دقت در زبان‌های کم‌منبع تمرکز دارد.

این فرآیند با تولید مختص زبان [1] آغاز می‌شود، که در آن به مدل زبانی دستور داده می‌شود تا خروجی هسته را براساس قواعد دستوری و نحوی زبان هدف شکل دهد؛ به عنوان مثال، هنگام تولید پاسخ به زبان فارسی، این مرحله از صرف صحیح افعال و ترتیب درست کلمات اطمینان حاصل می‌کند. سپس فرآیند به مرحله انطباق فرهنگی [2] می‌رود، که در آن پاسخ صحیح از نظر زبانی، مجدداً توسط مدل زبانی برای تطابق با زمینه‌های فرهنگی پالایش می‌شود ؛ مانند تنظیم سطوح گفتار رسمی/غیررسمی یا جایگزینی اصطلاحات با عبارات متناسب با فرهنگ مقصد. در نهایت، فرآیند با راستی‌آزمایی کیفیت [3] به پایان می‌رسد، که به عنوان یک بررسی نهایی (که می‌تواند توسط مدل زبانی به عنوان یک ارزیاب یا با رویکرد Self-Refine انجام شود) برای تضمین روانی، انسجام و دقت پاسخ عمل می‌کند.

این مراحل در یک پایپ‌لاین مبتنی بر مدل زبانی با یکدیگر کار می‌کنند: خروجی تولید مختص زبان به «انطباق فرهنگی» وارد می‌شود، که آن نیز به نوبه خود ورودی راستی‌آزمایی کیفیت را فراهم می‌کند. این ساختار تضمین می‌کند که کیفیت هر پاسخ متناسب با زبان و به تناسب فرهنگی ارتقا یابد. جدول ۶ مراحل تولید پاسخ را نشان می‌دهد.

---
[1] Language-Specific Generation
[2] Cultural Adaptation
[3] Quality Verification

جدول٦ : مراحل فرایند در تولید پاسخ و تشریحات، تکنیک‌ها و نمونه‌ها

مرحله فرایند	تشریح	نمونه (پرسش پزشکی)	تکنیک‌ها
تولید مختص زبان	شکل‌دهی خروجی بر اساس قواعد دستوری و نحوی زبان هدف توسط مدل زبانی	ورودی: complete the task by tomorrow خروجی: لطفا کار را تا فردا تمام کنید(اعمال ترتیب صحیح کلمات فارسی، صرف فعل)	Multilingual Generation (J. Li et al., 2024) Grammar-Aware Generation (Gemechu & Kanagachidambaresan, 2023) Syntax-Controlled NLG (Jia et al., 2025)
انطباق فرهنگی	اصلاح پاسخ‌ها برای تناسب فرهنگی توسط مدل زبانی	ورودی: لطفا کار را تا فردا تمام کنید خروجی: خواهشمند است کار محوله را تا فردا به اتمام برسانید افزودن زمینه کسب‌وکار رسمی، ادب مناسب	Cultural Style Transfer Politeness Adaptation Context-Aware Generation
راستی‌آزمایی کیفیت	اطمینان از دقت زبان‌شناختی، تناسب فرهنگی و قابلیت اعتماد محتوا	بررسی‌ها:دستور/نحو ، تناسب فرهنگی، وضوح پیام ، دقت محتوا	Quality Metrics Coherence Checking Accuracy Validation

بازخورد انسانی

مؤلفه بازخورد انسانی به عنوان یک سازوکار حیاتی برای ارزیابی و بهبود کل سیستم پردازش زبان کم‌منبع عمل می‌کند. این بخش از طریق ارزیابی مستقیم انسانی و جمع‌آوری بازخورد ساختاریافته، به بهبود کیفیت پاسخ، دقت زبانی و تناسب فرهنگی کمک می‌نماید. روش‌های اخیر، نحوه جمع‌آوری و بهره‌برداری از بازخورد انسانی را با استفاده از بهینه‌سازی مستقیم ترجیحات[1] (Rafailov et al., 2024) به طور قابل توجهی بهبود بخشیده‌اند. این روش‌ها، با بهینه‌سازی مستقیم خروجی‌های مدل زبانی بر اساس ترجیحات

[1] Direct Preference Optimization (DPO)

انسانی و بدون نیاز به مدل‌سازی پاداش پیچیده، یک جایگزین کارآمدتر برای روش سنتی یادگیری تقویتی از بازخورد انسانی[1] به حساب می‌آیند. هوش مصنوعی مبتنی بر اصول و قواعد با کمک انسان[2] (Y. Bai et al., 2022) کمک می‌کند تا هم‌راستایی پاسخ‌ها با دستورالعمل‌های رفتاری مشخص، همراه با حفظ کیفیت زبان، تضمین شود. پالایش تکرارشونده از طریق بازخورد انسانی[3] (Madaan et al., 2023) امکان بهبود تدریجی پاسخ‌های مدل را از طریق چندین دور بازخورد فراهم می‌سازد. علاوه بر این، روش‌هایی مانندیادگیری فعال برای جمع‌آوری بازخورد[4] (Ji et al., 2024) به اولویت‌بندی نمونه‌هایی که به فوریت بیشتری به بازخورد انسانی نیاز دارند، کمک می‌کنند؛ این امر به ویژه در زبان‌های کم‌منبع که بازخورد متخصصان ممکن است محدود باشد، ارزشمند است.

۵. ارزیابی و مقایسه

این چارچوب که ساختار اولیه آن از طریق تحلیل جامع معماری‌های موجود شکل گرفت، برای اعتبارسنجی در اختیار متخصصان قرار گرفت. با استفاده از روش گروه متمرکز (Huyler & McGill, 2019) با مشارکت پنج پژوهشگر پردازش زبان طبیعی ارزیابی انجام شده است. فرآیند ارزیابی شامل مصاحبه‌های ساختاریافته با استفاده از یک پرسشنامه دقیق (ارائه شده در پیوست ۱) و به دنبال آن، بحث‌های عمیق در مورد اجزای چارچوب و اثربخشی آن‌ها در پرداختن به چالش‌های زبان‌های کم‌منبع بود. هر جلسه تخصصی شامل بررسی دقیقی از معماری چارچوب بود و به طور خاص بر این موضوع تمرکز داشت که چگونه هر یک از اجزا به چالش‌های مشخصی در پردازش زبان‌های کم‌منبع رسیدگی می‌کنند .جدول ۷ چالش‌های کلیدی شناسایی‌شده و نحوه پاسخگویی اجزای چارچوب پیشنهادی به آن‌ها را خلاصه می‌کند. ستون راه حل در این جدول، بر اساس نتایج مستند بیان شده‌اند.

[1] Reinforcement learning from human feedback

[2] Constitutional AI with human
[3] Iterative Refinement through Human Feedback
[4] Active Learning for Feedback Collection

متخصصان به طور ویژه بر رویکرد نوآورانه چارچوب در مدیریت عناصر فرهنگی از طریق مؤلفه بهبود دانش و فرآیند تولید پاسخ تأکید کردند و اشاره نمودند که این چارچوب چگونه جنبه‌های مختلف فرهنگی مانند سطوح رسمیت، سلسله‌مراتب اجتماعی و الگوهای ارتباطی غیرمستقیم را که در بسیاری از زبان‌های کم‌منبع رایج است، به طور مؤثر مدیریت می‌کند. مؤلفه بازخورد انسانی نقشی حیاتی در این فرآیند انطباق فرهنگی ایفا می‌کند، زیرا یادگیری مستمر و پالایش درک فرهنگی را امکان‌پذیر می‌سازد. از طریق بازخورد نظام‌مند از گویشوران بومی و کارشناسان فرهنگی، سیستم می‌تواند ظرافت‌های فرهنگی را بیاموزد و خود را با آن‌ها تطبیق دهد، مدیریت خود را بر سطوح رسمیت متناسب با بافت بهبود بخشد و الگوهای ارتباطی محلی را بهتر درک کند. این حلقه بازخورد به ویژه برای زبان‌های کم‌منبع که در آن‌ها مستندات فرهنگی از پیش موجود ممکن است محدود باشد، ارزشمند است و به سیستم اجازه می‌دهد تا پایگاه دانش فرهنگی خود را از طریق تعاملات واقعی و راهنمایی متخصصان بسازد و اصلاح کند.

جدول۷ : چالش‌های زبان‌های کم‌منبع در پرسش و پاسخ محاوره‌ای و راه‌حل‌های مبتنی بر چارچوب

چالش	توصیف	راه‌حل چارچوب و مراجع مربوطه	تأثیر
محدودیت داده‌های آموزشی	پیکره‌های متنی و مجموعه داده‌های حاشیه‌نویسی‌شده ناکافی برای آموزش مدل‌های مؤثر	سازگاری کارآمد پارامتر در هسته مدل زبانی بزرگ (Gurgurov et al., 2025) افزایش داده در تقویت دانش (Soudani et al., 2023; Q. Wu et al., 2021) تکنیک‌های یادگیری کم‌نمونه (Brown, 2020) استفاده از مدل‌های زبانی بزرگ (Jolfaei & Mohebi, 2025 , Kholodna et al., 2024)	امکان آموزش مؤثر حجم کمتر داده زبانی به جای نیاز به مجموعه داده‌های عظیم
پیچیدگی صرفی	ساختارهای واژگانی پیچیده و قواعد دستوری که مدل‌سازی آن‌ها دشوار است	پیش‌پردازش اختصاصی زبان در پردازش ورودی(Howcroft & Gkatzia, 2022)	مدیریت صحیح ویژگی‌های زبانی منحصربه‌فرد مانند الصاق، صرف پیچیده فعل، و خط‌های غیرلاتین

چالش	توصیف	راه‌حل چارچوب و مراجع مربوطه	تأثیر
نکات ظریف فرهنگی	عبارات فرهنگی‌خاص، سطوح رسمیت و الگوهای ارتباطی	سازگاری فرهنگی در تولید پاسخ(Tao, Y et al., 2024) زمینه فرهنگی در تقویت دانش(Li. H et al., 2024) بازخورد انسانی برای تناسب فرهنگی(Zhang et al., 2025)	تولید پاسخ‌های مناسب زمینه ای با رعایت سلسله‌مراتب اجتماعی، سطوح رسمیت و ارجاعات فرهنگی
محدودیت‌های انتقال میان‌زبانی	دشواری انتقال دانش از زبان‌های پرمنبع	تعبیه های میان‌زبانی(Miao et al., 2024) تکنیک‌های یادگیری انتقالی (Nooralahzadeh et al., 2020)	نگاشت مؤثر مفاهیم بین زبان‌ها و امکان انتقال دانش از زمینه‌های پرمنبع به کم‌منبع
کارآیی محاسباتی	منابع محاسباتی محدود در زبان های کم‌منبع	مدیریت زمینه برای استفاده بهینه از حافظه سازگاری کارآمد پارامتر(LoRA ، Seq_bn، Seq_bn_inv) (Hu et al., 2021 , Houlsby et al., 2019 , Pfeiffer, Vulić, et al., 2020)	کاهش نیازبه منابع محاسباتی نسبت به تنظیم دقیق کامل مدل
تنوع خط و گویش	خط‌ها یا تنوعات گویشی متعدد در یک زبان	نرمال‌سازی خط در پردازش ورودی، تشخیص خط و گویش، توکن‌سازی تخصصی (Howcroft & Gkatzia, 2022)	استانداردسازی قالب ورودی و مدیریت تنوعات در سیستم‌های نوشتاری
چالش‌های ارزیابی	فقدان معیارهای ارزیابی استاندارد و محک‌های سنجش	مؤلفه بازخورد انسانی (Ji et al., 2024) بهینه‌سازی ترجیح مستقیم (Rafailov et al., 2024) فرایند اصلاح تکراری (Madaan et al., 2023)	ایجاد حلقه بهبود مستمر با روش‌شناسی ارزیابی ساختاریافته
شکاف‌های دانش حوزه‌خاص	فقدان واژگان تخصصی و دانش حوزه‌ای	دانش حوزه‌ای در تقویت دانش (Neha et al., 2025) یکپارچه‌سازی منابع دانش خارجی و تکنیک‌های تولید تقویت‌شده با بازیابی(Arslan et al., 2024)	تقویت مدل زبانی با تخصص حوزه‌ای در حوزه‌های پزشکی، حقوقی و فنی

چالش	توصیف	راه‌حل چارچوب و مراجع مربوطه	تأثیر
حفظ زمینه	دشواری حفظ زمینه مکالمه در چندین نوبت	مؤلفه مدیریت زمینه ، بازیابی انتخابی زمینه ، بهینه‌سازی تاریخچه مکالمه(Christmann et al., 2023)	افزایش کیفیت مکالمه چندنوبتی با حفظ زمینه مرتبط
ریسک توهم	ریسک بالاتر تولید اطلاعات نادرست به دلیل کمبود داده	اعتبارسنجی منابع دانش ، تأیید کیفیت در تولید پاسخ(Chen et al., 2024)	کاهش نرخ توهم از طریق اتکای پاسخ‌ها به منابع معتبر و قابل راستی‌آزمایی
کمبود منابع لازم برای پردازش زبان طبیعی	فقدان ابزارهای پایه پردازش زبان طبیعی (توکن‌ساز، تجزیه‌گر و غیره)	پیش‌پردازش اختصاصی زبان اشتراک‌گذاری منابع میان‌زبانی سازگاری ابزارهای موجود	ایجاد خط لوله پردازش زبان طبیعی کارآمد حتی برای زبان ها یی که منابع پردازشی زبان در حداقل است.

۶. بحث و بررسی

چارچوب پیشنهادی چندین دستاورد کلیدی دارد که از طریق ارزیابی متخصصان اعتبارسنجی شده است چندین دستاورد کلیدی را در مقایسه با رویکردهای موجود برجسته می‌سازد:

۱. پردازش فرهنگی یکپارچه: برخلاف چارچوب‌های موجود که انطباق فرهنگی را به عنوان یک مرحله پس‌پردازش در نظر می‌گیرند، چارچوب ما درک فرهنگی را در سرتاسر فرآیند خود ادغام می‌کند،که منجر به پاسخ‌هایی با هم‌راستایی عمیق‌تر می‌شود.

۲. استفاده بهینه از منابع: این چارچوب از طریق مؤلفه‌های مدیریت زمینه و تقویت دانش، کارایی منابع را بهبود می‌بخشد. این معماری، به‌ویژه با بازیابی گزینشی زمینه و ذخیره‌سازی دانش ساختاریافته، چارچوب را برای سناریوهای کم‌منبع که در آن‌ها کارایی محاسباتی حیاتی است، مناسب‌تر می‌سازد. این دو مؤلفه مستقیماً به

چالش‌های محاسباتی مطرح‌شده در پیشینه پاسخ می‌دهند. این معماری، به‌ویژه با بازیابی گزینشی زمینه و ذخیره‌سازی دانش ساختاریافته (RAG)، چارچوب را برای سناریوهای کم‌منبع که در آن‌ها کارایی محاسباتی حیاتی است، مناسب‌تر می‌سازد.

۳. معماری پیمانه‌ای: طراحی پیمانه‌ای (ماژولار) امکان ادغام آسان زبان‌ها و بافت‌های فرهنگی جدید را فراهم می‌کند؛ موضوعی که متخصصان آن را برای پیاده‌سازی عملی، حیاتی تشخیص دادند.

۴. سیستم یادگیری تطبیق‌پذیر: ترکیب مؤلفه‌های تقویت دانش و بازخورد انسانی سیستمی را ایجاد می‌کند که به طور مداوم در حال بهبود است و متخصصان آن را برای توسعه بلندمدت قابلیت‌های زبان‌های کم‌منبع بسیار مؤثر ارزیابی کردند. این یافته، اهمیت سیستم‌های یادگیری را برای زبان‌های کم‌منبع که در آن‌ها داده‌ها به صورت تدریجی در دسترس قرار می‌گیرند ، تأیید می‌کند.

۵. یادگیری انتقال بین‌زبانی: توانایی بهبودیافته در بهره‌گیری از دانش زبان‌های پرمنبع برای پردازش زبان‌های کم‌منبع از طریق تکنیک‌های پیشرفته یادگیری انتقال.

۶. ادغام جامع بازخورد انسانی: تلفیق نظام‌مند بازخورد انسانی که بهبود مستمر جنبه‌های زبانی و فرهنگی را امکان‌پذیر می‌سازد و پاسخی به نیاز مبرم برای مکانیسم‌های هم‌راستایی پویا در پیشینه پژوهش است.

اگرچه چارچوب ما قابلیت‌های امیدوارکننده‌ای برای پردازش زبان‌های کم‌منبع و انطباق فرهنگی نشان می‌دهد، ارزیابی متخصصان چندین محدودیت و زمینه را برای کارهای آتی نیز آشکار ساخت. این دیدگاه‌ها مسیر ارزشمندی را برای تحقیق و توسعه بیشتر فراهم می‌کنند و چالش‌های فنی و عملی را که باید برای پیاده‌سازی استوارتر مورد توجه قرار گیرند، برجسته می‌سازند. این مسائل شامل موارد زیر هستند:

۱. محدودیت‌های فنی:

- نیاز به آزمون‌های گسترده‌تر با زبان‌های بسیار کم‌منبع.
- نیازمندی‌های محاسباتی برای راه‌اندازی اولیه سیستم ممکن است در

محیط‌های با منابع محدود، چالش‌برانگیز باشد.

- وابستگی به کیفیت منابع بین‌زبانی موجود.

۲. ارزیابی فرهنگی:

- نیاز به توسعه معیارهای استاندارد برای سنجش تناسب فرهنگی.
- چالش در کمّی‌سازی دقت فرهنگی.
- نیاز به ارزیابی نظام‌مندترِ انطباق فرهنگی.

۳. چالش‌های پیاده‌سازی:

- نیاز به ورودی اولیه قابل توجه از سوی متخصصان برای زبان‌های جدید.
- پیچیدگی یکپارچه‌سازی با سیستم‌های موجود.
- نیاز به تخصص ویژه در مرحله استقرار[1]

۴. محدودیت‌های منابع:

- در دسترس بودنِ محدودِ کارشناسان فرهنگی برای برخی زبان‌ها.
- چالش در به‌روز نگه‌داشتن دانش فرهنگی.
- نیازمندی‌های منابع برای پردازش مستمر بازخورد.

۵. معیارهای ارزیابی:

- فقدان روش‌های ارزیابی استاندارد برای جنبه‌های فرهنگی.
- نیاز به معیارهای کمّی بهتر برای سنجش عملکرد سیستم.
- چالش در مقایسه نتایج میان زبان‌های مختلف.

۶. مسائل مقیاس‌پذیری:

- وجود گلوگاه‌های[2] بالقوه در مدیریت همزمان چندین زبان.

---

[1] deployment
[2] bottlenecks

- نیازمندی‌های منابع ممکن است با افزودن زبان‌های جدید به صورت غیرخطی افزایش یابد.
- نیاز به اشتراک‌گذاری کارآمدتر دانش بین‌زبانی.

۷. نتیجه‌گیری

در این پژوهش، چارچوبی نوین برای سیستم پرسش و پاسخ محاوره‌ای در زبان‌های کم‌منبع ارائه شد که بر پردازش کارآمد، انطباق فرهنگی، و بهبود مستمر از طریق بازخورد انسانی تمرکز دارد. معماری این چارچوب چندین مؤلفه کلیدی را یکپارچه می‌کند: «پردازش ورودی» برای درک اولیه زبان، یک «هسته مرکزی مدل زبانی بزرگ» برای پردازش عمیق معنایی، «تقویت دانش» برای مدیریت کارآمد اطلاعات، «مدیریت زمینه» برای مدیریت بهینه مکالمه، و «تولید پاسخ» برای ارائه خروجی متناسب با فرهنگ. یکی از ویژگی‌های متمایز چارچوب پیشنهادی، رویکرد جامع آن به انطباق فرهنگی است که توسط یک سازوکار نظام‌مند بازخورد انسانی پشتیبانی می‌شود و یادگیری مستمر و پالایش درک فرهنگی را امکان‌پذیر می‌سازد.

ارزیابی تخصصی با مشارکت پنج پژوهشگر پردازش زبان طبیعی، اثربخشی این چارچوب را در پرداختن به چالش‌های حیاتی پردازش زبان‌های کم‌منبع تأیید کرد. این ارزیابی چندین دستاورد مهم را برجسته ساخت، از جمله پردازش فرهنگی یکپارچه در سرتاسر فرآیند، بهره‌وری کارآمد از منابع از طریق مدیریت بهینه بافت، و یک سیستم یادگیری تطبیق‌پذیر که بهبود دانش را با بازخورد انسانی ترکیب می‌کند. معماری ماژولار و مقیاس‌پذیر چارچوب به‌طور ویژه به دلیل پتانسیل پیاده‌سازی عملی آن در زبان‌های کم‌منبع مختلف، مورد توجه قرار گرفت.

با این حال، این ارزیابی محدودیت‌ها و زمینه‌هایی را برای کارهای آتی نیز آشکار ساخت؛ از جمله نیاز به آزمون‌های گسترده‌تر با زبان‌های بسیار کم‌منبع، توسعه معیارهای استاندارد برای سنجش تناسب فرهنگی، و چالش‌ها در پیاده‌سازی و مدیریت منابع. این محدودیت‌ها مسیرهای روشنی را برای تحقیقات آینده مشخص می‌کنند، به‌ویژه در زمینه

توسعه معیارهای ارزیابی استوارتر و بهبود کارایی در اشتراک‌گذاری دانش بین‌زبانی. علی‌رغم این چالش‌ها، چارچوب پیشنهادی گام مهمی رو به جلو در دسترس‌پذیرتر کردن و متناسب‌سازی فرهنگی هوش مصنوعی محاوره‌ای برای جوامع با زبان‌های کم‌منبع محسوب می‌شود و در عین حال، بنیادی را برای توسعه‌های آتی در این حوزه حیاتی ایجاد می‌کند.

فهرست منابع

Adelani, D. I., Abbott, J., Neubig, G., D'souza, D., Kreutzer, J., Lignos, C., Palen-Michel, C., Buzaaba, H., Rijhwani, S., Ruder, S., Mayhew, S., Azime, I. A., Muhammad, S. H., Emezue, C. C., Nakatumba-Nabende, J., Ogayo, P., Anuoluwapo, A., Gitau, C., Mbaye, D., … Osei, S. 2021. MasakhaNER: Named Entity Recognition for African Languages. *Transactions of the Association for Computational Linguistics, 9, 1116–1131. https://doi.org/10.1162/tacl_a_00416*

Adelani, D. I., Ruiter, D., Jesujoba, O., Adebonojo, D,. Ayeni, A., Adeyemi, M., Awokoya, A. E., and España-Bonet,. C. 2021. The Effect of Domain and Diacritics in Yoruba–English Neural Machine Translation. *In Proceedings of Machine Translation Summit XVIII: Research Track, pages 61–75, Virtual. Association for Machine Translation in the Americas.*

Alam, F., Chowdhury, S. A., Boughorbel, S., & Hasanain, M. 2024. Llms for low resource languages in multilingual, multimodal and dialectal settings. *In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: Tutorial Abstracts (pp. 27-33).*

Andersland, M. 2024. Amharic llama and llava: Multimodal llms for low resource languages. *arXiv preprint arXiv:2403.06354*.

Arslan, M., Ghanem, H., Munawar, S., & Cruz, C. 2024. A Survey on RAG with LLMs. Procedia computer science, 246, 3781-3790.

Asai, A., Wu, Z., Wang, Y., Sil, A., & Hajishirzi, H. 2023. *Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection* (arXiv:2310.11511). arXiv. http://arxiv.org/abs/2310.11511

Azime, I. A., Tonja, A. L., Belay, T. D., Chanie, Y., Balcha, B. F., Abadi, N. H., Ademtew, H. B., Nerea, M. A., Yadeta, D. D., Geremew, D. D., tesfau, A. A., Slusallek, P., Solorio, T., & Klakow, D. 2025. *ProverbEval: Exploring LLM Evaluation Challenges for Low-resource Language Understanding* (arXiv:2411.05049). arXiv. https://doi.org/10.48550/arXiv.2411.05049

Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., Chen, C., Olsson, C., Olah, C., Hernandez, D., Drain, D., Ganguli, D., Li, D., Tran-Johnson, E., Perez, E., … Kaplan, J. 2022. *Constitutional AI: Harmlessness from AI Feedback* (arXiv:2212.08073). arXiv. http://arxiv.org/abs/2212.08073

Brown, T. B. 2020. Language models are few-shot learners. *arXiv Preprint arXiv:2005.14165*.

Burgin, E., Dutta, S., Assem, H., & Patel, R. N. 2022. Cage: A Hybrid Framework for Closed-Domain Conversational Agents. 636–640.

Cahyawijaya, S., Lovenia, H., & Fung, P. 2024 . LLMs Are Few-Shot In-Context Low-Resource Language Learners *(arXiv:2403.16512). arXiv. http://arxiv.org/abs/2403.16512*

Campos, J. A., Cho, K., Otegi, A., Soroa, A., Agirre, E., & Azkune, G. 2020. Improving Conversational Question Answering Systems after Deployment using Feedback-Weighted Learning. *Proceedings of the 28th International Conference on Computational Linguistics*, 2561–2571. https://doi.org/10.18653/v1/2020.coling-main.230

Chen, X., Wang, L., Wu, W., Tang, Q., & Liu, Y. 2024. Honest AI: Fine-Tuning" Small" Language Models to Say" I Don't Know", and Reducing Hallucination in RAG. arXiv preprint arXiv:2410.09699.

Chen, H., Liu, X., Yin, D., & Tang, J. 2017. A Survey on Dialogue Systems: Recent Advances and New Frontiers. *SIGKDD Explor. Newsl.*, *19*(2), 25–35. https://doi.org/10.1145/3166054.3166058

Christmann, P., Saha Roy, R., Abujabal, A., Singh, J., & Weikum, G. 2019. Look before you hop: Conversational question answering over knowledge graphs using judicious context expansion. 729–738.

Christmann, P., Saha Roy, R., & Weikum, G. 2023. Explainable Conversational Question Answering over Heterogeneous Sources via Iterative Graph Neural Networks. 643–653.

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. 2020. *Unsupervised Cross-lingual*

Dam, S. K., Hong, C. S., Qiao, Y., & Zhang, C. 2024. A Complete Survey on LLM-based AI Chatbots *(arXiv:2406.16937). arXiv. http://arxiv.org/abs/2406.16937*

Deng, Y., Lei, W., Huang, M., & Chua, T.-S. 2023. Rethinking Conversational Agents in the Era of LLMs: Proactivity, Non-collaborativity, and Beyond. *Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region*, 298–301. https://doi.org/10.1145/3624918.3629548

Dinan, E., Abercrombie, G., Bergman, A. S., Spruit, S., Hovy, D., Boureau, Y.-L., & Rieser, V. 2021. *Anticipating Safety Issues in E2E Conversational AI: Framework and Tooling (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2107.03451*

Fam, R., & Lepage, Y. 2024. A study of universal morphological analysis using morpheme-based, holistic, and neural approaches under various data size conditions. *Annals of Mathematics and Artificial Intelligence*. https://doi.org/10.1007/s10472-024-09944-8

Foosherian, M., Purwins, H., Rathnayake, P., Alam, T., Teimao, R., & Thoben, K. D. 2023. Enhancing pipeline-based conversational agents with large language models. *arXiv preprint arXiv:2309.03748*.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. 2024. Retrieval-Augmented Generation for Large Language Models: A Survey (arXiv:2312.10997). arXiv. http://arxiv.org/abs/2312.10997

Gao, J., Galley, M., & Li, L. 2019. Neural Approaches to Conversational AI. *Foundations and Trends® in Information Retrieval*, *13*(2–3), 127–298. https://doi.org/10.1561/1500000074

Gemechu, E. A., & Kanagachidambaresan, G. R. 2023. Grammar-aware phrase dataset generated using a novel python package. Data in Brief, 48, 109237. https://doi.org/10.1016/j.dib.2023.109237

Gupta, S., Rawat, B. P. S., & Yu, H. 2020. Conversational machine comprehension: A literature review. *arXiv Preprint arXiv:2006.00671*.

Gurgurov, D., Vykopal, I., Genabith, J. van, & Ostermann, S. 2025. Small Models, Big Impact: Efficient Corpus and Graph-Based Adaptation of Small Multilingual Language Models for Low-Resource Languages *(arXiv:2502.10140). arXiv. https://doi.org/10.48550/arXiv.2502.10140*

Harms, J.-G., Kucherbaev, P., Bozzon, A., & Houben, G.-J. 2019. Approaches for Dialog Management in Conversational Agents. *IEEE Internet Computing*, *23*(2), 13–22. https://doi.org/10.1109/MIC.2018.2881519

Hassani, H. 2018. BLARK for multi-dialect languages: towards the Kurdish BLARK. *Language Resources and Evaluation, 52(2), 625-644.*

Howcroft, D. M., & Gkatzia, D. 2022. Most NLG is Low-Resource: Here's what we can do about it. *Proceedings of the 2nd Workshop on Natural Language Generation, Evaluation, and Metrics (GEM)*, 336–350. https://doi.org/10.18653/v1/2022.gem-1.29

Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., de Laroussilhe, Q., Gesmundo, A., Attariyan, M., & Gelly, S. 2019. Parameter-Efficient Transfer Learning for NLP (Version 2). arXiv. https://doi.org/10.48550/ARXIV.1902.00751

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., & Chen, W. 2021. LoRA: Low-Rank Adaptation of Large Language Models. *CoRR, abs/2106.09685*. https://arxiv.org/abs/2106.09685

Huyler, D., & McGill, C. M. 2019. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, by John Creswell and J. David Creswell. Thousand Oaks, CA: Sage Publication,. *New Horizons in Adult Education and Human Resource Development*, *31*(3), 75–77.

https://doi.org/10.1002/nha3.20258

Jia, X., Mao, Z., Zhang, Z., Lv, Q., Wang, X., & Wu, G. 2025. Syntax-controlled paraphrases generation with VAE and multi-task learning. Computer Speech & Language, 89, 101705. https://doi.org/10.1016/j.csl.2024.101705

Ji, K., He, J., & Gu, Q. 2024. Reinforcement Learning from Human Feedback with Active Queries *(arXiv:2402.09401). arXiv. http://arxiv.org/abs/2402.09401*

Joshi, A., Dabre, R., Kanojia, D., Li, Z., Zhan, H., Haffari, G., & Dippold, D. 2025. Natural language processing for dialects of a language: A survey. *ACM Computing Surveys, 57(6), 1-37.*

Kargaran, A. H., Imani, A., Yvon, F., & Schütze, H. 2023. GlotLID: Language Identification for Low-Resource Languages. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 6155–6218. *https://doi.org/10.18653/v1/2023.findings-emnlp.410*

Kai, J., Hou, S., Huang, Y., & Lin, Z. 2024. Leveraging Grammar Induction for Language Understanding and Generation (arXiv:2410.04878). arXiv. http://arxiv.org/abs/2410.04878

Kholodna, N., Julka, S., Khodadadi, M., Gumus, M. N., & Granitzer, M. 2024. LLMs in the Loop: Leveraging Large Language Model Annotations for Active Learning in Low-Resource Languages. In A. Bifet, T. Krilavičius, I. Miliou, & S. Nowaczyk (Eds.), *Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track* (Vol. 14950, pp. 397–412). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-70381-2_25

Kongyoung, S., Macdonald, C., & Ounis, I. 2022. monoQA: Multi-Task Learning of Reranking and Answer Extraction for Open-Retrieval Conversational Question Answering. 7207–7218.

Kumar, A., Shrotriya, H., Sahu, P., Mishra, A., Dabre, R., Puduppully, R., Kunchukuttan, A., Khapra, M. M., & Kumar, P. 2022. IndicNLG Benchmark: Multilingual Datasets for Diverse NLG Tasks in Indic Languages. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 5363–5394. https://doi.org/10.18653/v1/2022.emnlp-main.360

Kumar, V., & Callan, J. 2020. Making information seeking easier: An improved pipeline for conversational search. 3971–3980.

Kusal, S., Patil, S., Choudrie, J., Kotecha, K., Mishra, S., & Abraham, A. 2022. AI-based conversational agents: A scoping review from technologies to future directions. *IEEE Access*.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., & Rocktäschel, T. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in Neural Information Processing Systems, 33, 9459–9474.

Li, J., Song, S., Yan, S., Hu, G., Lai, C., & Zhou, Y. 2024. DANTE: Dialog graph enhanced prompt learning for conversational question answering over KGs. Knowledge-Based Systems, 301, 112294. https://doi.org/10.1016/j.knosys.2024.112294

Li, H., Jiang, L., Hwang, J. D., Kim, H., Santy, S., Sorensen, T., ... & Choi, Y. 2024. Culture-gen: Revealing global cultural perception in language models through natural language prompting. arXiv preprint arXiv:2404.10199.

Li, Z., Peng, J., Wang, Y., Shen, T., Zhang, M., Su, L., Wu, S., Wu, Y., Wang, Y., Wang, Y., Hu, W., Li, J., Wang, S., Xiao, J., & Xiong, D. 2024. Planning with Large Language Models for Conversational Agents *(arXiv:2407.03884). arXiv. http://arxiv.org/abs/2407.03884*

Li, Z., Shi, Y., Liu, Z., Yang, F., Payani, A., Liu, N., & Du, M. 2024. Language Ranker: A Metric for Quantifying LLM Performance Across High and Low-Resource Languages *(arXiv:2404.11553). arXiv. https://doi.org/10.48550/arXiv.2404.11553*

Lin, X. V., Mihaylov, T., Artetxe, M., Wang, T., Chen, S., Simig, D., Ott, M., Goyal, N., Bhosale, S., Du, J., Pasunuru, R., Shleifer, S., Koura, P. S., Chaudhary, V., O'Horo, B., Wang, J., Zettlemoyer, L., Kozareva, Z., Diab, M., … Li, X. 2022. Few-shot Learning with Multilingual Language Models (arXiv:2112.10668). arXiv. http://arxiv.org/abs/2112.10668

Liu, L., Hill, B., Du, B., Wang, F., & Tong, H. 2024. Conversational Question Answering with Language Models Generated Reformulations over Knowledge Graph. *Findings of the Association for Computational Linguistics ACL 2024*, 839–850. https://doi.org/10.18653/v1/2024.findings-acl.48

Liu, N., Chen, L., Tian, X., Zou, W., Chen, K., & Cui, M. 2024. From LLM to Conversational Agent: A

Memory Enhanced Architecture with Fine-Tuning of Large Language Models (arXiv:2401.02777). arXiv. http://arxiv.org/abs/2401.02777

Madaan, A., Tandon, N., Gupta, P., Hallinan, S., Gao, L., Wiegreffe, S., Alon, U., Dziri, N., Prabhumoye, S., Yang, Y., Gupta, S., Majumder, B. P., Hermann, K., Welleck, S., Yazdanbakhsh, A., & Clark, P. 2023. Self-Refine: Iterative Refinement with Self-Feedback (arXiv:2303.17651). arXiv. http://arxiv.org/abs/2303.17651

Mao, Z., & Yu, Y. 2024. Tuning LLMs with Contrastive Alignment Instructions for Machine Translation in Unseen, Low-resource Languages (arXiv:2401.05811). arXiv. http://arxiv.org/abs/2401.05811

Merx, R., Mahmudi, A., Langford, K., Araujo, L. A. de, & Vylomova, E. 2024. Low-Resource Machine Translation through Retrieval-Augmented LLM Prompting: A Study on the Mambai Language (arXiv:2404.04809). arXiv. https://doi.org/10.48550/arXiv.2404.04809

Miao, Z., Wu, Q., Zhao, K., Wu, Z., & Tsuruoka, Y. 2024. Enhancing Cross-lingual Sentence Embedding for Low-resource Languages with Word Alignment. *Findings of the Association for Computational Linguistics: NAACL 2024*, 3225–3236. https://doi.org/10.18653/v1/2024.findings-naacl.204

Neha, F., Bhati, D., & Shukla, D. K. 2025. Retrieval-Augmented Generation (RAG) in Healthcare: A Comprehensive Review. AI, 6(9), 226.

Nasution, A. H., & Onan, A. 2024. ChatGPT Label: Comparing the Quality of Human-Generated and LLM-Generated Annotations in Low-Resource Language NLP Tasks. *IEEE Access*, *12*, 71876–71900. https://doi.org/10.1109/ACCESS.2024.3402809

Nguyen, X.-P., Aljunied, S. M., Joty, S., & Bing, L. 2024. *Democratizing LLMs for Low-Resource Languages by Leveraging their English Dominant Abilities with Linguistically-Diverse Prompts* (arXiv:2306.11372). arXiv. https://doi.org/10.48550/arXiv.2306.11372

Nigusie, G., & Tegegne, T. 2022. Amharic Text Complexity Classification Using Supervised Machine Learning. *In International Conference on Advances of Science and Technology (pp. 1-12). Cham: Springer Nature Switzerland.*

Nooralahzadeh, F., Bekoulis, G., Bjerva, J., & Augenstein, I. (2020). *Zero-Shot Cross-Lingual Transfer with Meta Learning* (arXiv:2003.02739). arXiv. http://arxiv.org/abs/2003.02739

Pfeiffer, J., Kamath, A., Rücklé, A., Cho, K., & Gurevych, I. 2020. AdapterFusion: Non-Destructive Task Composition for Transfer Learning. *https://doi.org/10.48550/ARXIV.2005.00247*

Pfeiffer, J., Vulić, I., Gurevych, I., & Ruder, S. 2020. MAD-X: An Adapter-Based Framework for Multi-Task Cross-Lingual Transfer. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 7654–7673. https://doi.org/10.18653/v1/2020.emnlp-main.617

Perez-Beltrachini, L., Jain, P., Monti, E., & Lapata, M. 2023. Semantic parsing for conversational question answering over knowledge graphs. *arXiv Preprint arXiv:2301.12217*.

Qian, J., Zou, B., Dong, M., Li, X., Aw, A., & Hong, Y. 2022. Capturing conversational interaction for question answering via global history reasoning. 2071–2078.

Qin, Y., Cai, Z., Jin, D., Yan, L., Liang, S., Zhu, K., Lin, Y., Han, X., Ding, N., Wang, H., Xie, R., Qi, F., Liu, Z., Sun, M., & Zhou, J. (2023). WebCPM: Interactive Web Search for Chinese Long-form Question Answering. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 8968–8988. https://doi.org/10.18653/v1/2023.acl-long.499

Qu, C., Yang, L., Qiu, M., Croft, W. B., Zhang, Y., & Iyyer, M. (2019). *BERT with history answer embedding for conversational question answering*. 1133–1136.

Rafailov, R., Sharma, A., Mitchell, E., Ermon, S., Manning, C. D., & Finn, C. 2024. *Direct Preference Optimization: Your Language Model is Secretly a Reward Model* (arXiv:2305.18290). arXiv. http://arxiv.org/abs/2305.18290

Ringel, D., Lavee, G., Guy, I., & Radinsky, K. 2019. Cross-Cultural Transfer Learning for Text Classification. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 3871–3881. https://doi.org/10.18653/v1/D19-1400*

Robinson, N. R., Ogayo, P., Mortensen, D. R., & Neubig, G. (2023). *ChatGPT MT: Competitive for High- (but not Low-) Resource Languages* (arXiv:2309.07423). arXiv.

https://doi.org/10.48550/arXiv.2309.07423

Soares, M. A. C., & Parreiras, F. S. (2020). A literature review on question answering techniques, paradigms and systems. *Journal of King Saud University-Computer and Information Sciences*, *32*(6), 635–646.

Song, Y., Wang, T., Cai, P., Mondal, S. K., & Sahoo, J. P. 2023. A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities. ACM Computing Surveys, 55(13s), 1–40. https://doi.org/10.1145/3582688

Soudani, H., Kanoulas, E., & Hasibi, F. (2023). Data augmentation for conversational ai. 5220–5223.

Su, L., Guo, J., Fan, Y., Lan, Y., Zhang, R., & Cheng, X. (2019). An Adaptive Framework for Conversational Question Answering. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(01), 10041–10042. https://doi.org/10.1609/aaai.v33i01.330110041

Sun, K., Li, Z., & Zhao, H. (2021). Multilingual pre-training with universal dependency learning. *Advances in Neural Information Processing Systems*, *34*, 8444-8456.

Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. 2024. Cultural bias and cultural alignment of large language models. PNAS nexus, 3(9), pgae346.

Tan, X., Ren, Y., He, D., Qin, T., Zhao, Z., & Liu, T.-Y. 2019. Multilingual Neural Machine Translation with Knowledge Distillation (arXiv:1902.10461). arXiv. http://arxiv.org/abs/1902.10461

Tanwar, E., Dutta, S., Borthakur, M., & Chakraborty, T. (2023). Multilingual LLMs are Better Cross-lingual In-context Learners with Alignment. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 6292–6307. https://doi.org/10.18653/v1/2023.acl-long.346

Upadhayay, B., & Behzadan, V. (2024). *TaCo: Enhancing Cross-Lingual Transfer for Low-Resource Languages in LLMs through Translation-Assisted Chain-of-Thought Processes* (arXiv:2311.10797). arXiv. http://arxiv.org/abs/2311.10797

Vakulenko, S., Longpre, S., Tu, Z., & Anantha, R. (2021). *Question rewriting for conversational question answering*. 355–363.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. 2023. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models (arXiv:2201.11903). arXiv. http://arxiv.org/abs/2201.11903

Workshop, B., Scao, T. L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A. S., Yvon, F., Gallé, M., Tow, J., Rush, A. M., Biderman, S., Webson, A., Ammanamanchi, P. S., Wang, T., Sagot, B., Muennighoff, N., … Wolf, T. 2023. *BLOOM: A 176B-Parameter Open-Access Multilingual Language Model* (arXiv:2211.05100). arXiv. http://arxiv.org/abs/2211.05100

Wu, Q., Feng, S., Chen, D., Joshi, S., Lastras, L. A., & Yu, Z. (2021). DG2: Data augmentation through document grounded dialogue generation. arXiv Preprint arXiv:2112.08342.

Wu, Z., Luan, Y., Rashkin, H., Reitter, D., Hajishirzi, H., Ostendorf, M., & Tomar, G. S. (2021). Conqrr: Conversational query rewriting for retrieval with reinforcement learning. *arXiv preprint arXiv:2112.08558*.

Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., & Raffel, C. 2021. mT5: A massively multilingual pre-trained text-to-text transformer *(arXiv:2010.11934). arXiv. http://arxiv.org/abs/2010.11934*

Yang, L., Hu, J., Qiu, M., Qu, C., Gao, J., Croft, W. B., Liu, X., Shen, Y., & Liu, J. (2019). *A hybrid retrieval-generation neural conversation model*. 1341–1350.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). *ReAct: Synergizing Reasoning and Acting in Language Models* (arXiv:2210.03629). arXiv. http://arxiv.org/abs/2210.03629

Ye, L., Lei, Z., Yin, J., Chen, Q., Zhou, J., & He, L. (2024). Boosting Conversational Question Answering with Fine-Grained Retrieval-Augmentation and Self-Check. *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2301–2305. https://doi.org/10.1145/3626772.3657980

Yong, Z.-X., Menghini, C., & Bach, S. H. (2024). *Low-Resource Languages Jailbreak GPT-4*

(arXiv:2310.02446). arXiv. https://doi.org/10.48550/arXiv.2310.02446

Zaib, M., Zhang, W. E., Sheng, Q. Z., Mahmood, A., & Zhang, Y. (2022). Conversational question answering: A survey. *Knowledge and Information Systems*, *64*(12), 3151–3195.

Zhang, H., Chen, K., Bai, X., Xiang, Y., & Zhang, M. 2025. Evaluating and Improving Cultural Awareness of Reward Models for LLM Alignment. arXiv preprint arXiv:2509.21798.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., … Wen, J.-R. (2024). *A Survey of Large Language Models* (arXiv:2303.18223). arXiv. http://arxiv.org/abs/2303.18223

پیوست ۱

پرسش‌ها

پردازش ورودی

الف. ماژول تشخیص و ارزیابی زبان، در شناسایی زبان‌های کم‌منبع و تعیین روش‌های پردازش مناسب، تا چه حد مؤثر عمل می‌کند؟

ب. ماژول درک پرس‌وجوی فرازبانی، توانایی سیستم در فهم پرس‌وجوها به زبان‌های کم‌منبع را تا چه میزان افزایش می‌دهد؟

ج. برای پردازش زبان‌های کم‌منبع که از نظر صَرفی پیچیده هستند، چه تکنیک‌های پردازشیِ زبان-محورِ دیگری را پیشنهاد می‌کنید؟

هسته مدل زبانی بزرگ

الف. اثربخشی تکنیک‌های انطباق با پارامتر بهینه مانند LoRA را برای پردازش زبان‌های کم‌منبع چگونه ارزیابی می‌کنید؟

ب. محدودیت‌های هسته مدل زبانی بزرگ در پردازش زبان‌های بسیار کم‌منبع چیست و این محدودیت‌ها چگونه باید برطرف شوند؟

تقویت دانش

الف. مؤلفه تقویت دانش در این چارچوب (شامل دانش فرازبانی، منابع زبانی، زمینه فرهنگی و غیره) چگونه امکان درک مطلب در محیط‌های زبان کم‌منبع را فراهم می‌آورد؟

ب. توصیه‌های شما برای ساخت و نگهداری مؤثر منابع دانش برای زبان‌های بسیار کم‌منبع چیست؟

ج. یکپارچه‌سازی زمینه فرهنگی تا چه حد چالش‌های انطباق فرهنگی در پردازش زبان‌های کم‌منبع را برطرف می‌سازد؟

مدیریت زمینه

الف. رویکرد تاریخچه فشرده مکالمه تا چه اندازه چالش‌های حفظ درک متنی را در سناریوهای کم‌منبع برطرف می‌کند؟

ب. برای فرآیند تطبیق شباهت به منظور بازیابی زمینه مرتبط در گفتگوی چندنوبتی، چه تغییراتی را پیشنهاد می‌کنید؟

تولید پاسخ

الف. ماژول انطباق فرهنگی تا چه حد مسئله تولید پاسخ‌های متناسب با فرهنگ را در زبان‌های کم‌منبع برطرف می‌کند؟

ب. چه روش‌های راستی‌آزمایی کیفیت دیگری را برای تضمین دقت زبانی و تناسب فرهنگی در پاسخ‌های تولیدشده پیشنهاد می‌کنید؟

بازخورد انسانی

الف. در هنگام جمع‌آوری و بهره‌برداری از بازخورد انسانی برای زبان‌های کم‌منبع، چارچوب باید چه چالش‌ها یا نگرانی‌های خاصی را مدیریت کند؟

ب. نقش بازخورد انسانی در بهبود سیستم و عملکرد کلی برای زبان‌های کم‌منبع چقدر اهمیت دارد؟

ارزیابی کلی چارچوب

الف. چارچوب توصیف‌شده، مسائل مختص زبان‌های کم‌منبع را در بستر پاسخگویی به پرسش محاوره‌ای تا چه حد به‌خوبی پوشش می‌دهد؟

ب. بر اساس تجربه خود، آیا ویژگی‌ها یا بهبودهای دیگری را برای شخصی‌سازی بیشتر چارچوب به منظور دستیابی به عملکرد بهتر برای زبان‌های کم‌منبع پیشنهاد می‌کنید؟

ج. چارچوب پیشنهادی در مقایسه با روش‌های موجود، با توجه به ویژگی‌های خاص فرهنگی و زبانیِ زبان‌های کم‌منبع، چگونه عمل می‌کند؟

Conversational Question Answering for Low-Resource Languages: A Novel Architecture Enhanced by Large Language Models

Safoura Aghadavoud Jolfaei, PhD Candidate in Iranian Research Institute for Information Science and Technology (IranDoc), Tehran, Iran.
aghadavood@students.irandoc.ac.ir

Azadeh Mohebi*, PhD in Systems Design Engineering
Assistant Professor in Iranian Research Institute for Information Science and Technology (IranDoc),Tehran, Iran.
mohebi@irandoc.ac.ir

Abstract

Conversational Question Answering (CQA) systems have evolved significantly with the advent of Large Language Models (LLMs); however, these advancements have predominantly favored high-resource languages, often overlooking low-resource ones. This paper introduces a novel LLM-enhanced framework specifically designed to bridge this linguistic gap. The proposed architecture comprises six components: "Input Processing" for language-specific handling, an "Adaptive LLM Core," "Knowledge Enhancement" for cross-lingual mapping, "Context Management" for efficient conversation navigation, "Response Generation" incorporating cultural adaptation, and "Human Feedback" for continuous improvement. Unlike existing approaches, this framework integrates cultural and linguistic considerations throughout the entire processing pipeline. To validate the framework, a qualitative evaluation was conducted using a focus group consisting of five Natural Language Processing (NLP) experts. Expert evaluation results confirmed the proposed framework's effectiveness in addressing fundamental challenges of low-resource languages, including data scarcity, morphological complexities, and cultural nuances. Experts particularly highlighted the framework's innovative approach to "integrated cultural processing," "resource efficiency" via optimized context management, and its "modular and scalable architecture" as key achievements. This research demonstrates that integrating human feedback and cultural adaptation within an efficient architecture offers a practical solution for developing Conversational Question Answering systems in low-resource languages.

Keywords: Conversational Question Answering, Interactive Question Answering, Machine Comprehension, Natural Language Processing.

*Corresponding Author

پژوهش‌نامهٔ پردازش و مدیریت اطلاعات

X xx ۱۳ | دورهٔ xx | شمارهٔ x

صفورا آقاداود جلفایی

دانشجوی دکتری رشته مدیریت فناوری اطلاعات، گرایش کسب‌وکار هوشمند است. مدل‌های زبانی بزرگ، سیستم‌های پرسش و پاسخ محاوره‌ای و پردازش زبان‌های کم‌منبع از جمله علایق پژوهشی وی است.

آزاده محبی

دارای مدرک دکتری در رشتهٔ مهندسی طراحی سیستم‌ها از دانشگاه واترلو کاناداست. ایشان هم‌اکنون استادیار پژوهشکدهٔ فناوری اطلاعات پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) است. تعامل انسان و کامپیوتر، داده‌کاوی، سیستم‌های هوشمند، بازشناسی الگو، متن‌کاوی و بازیابی اطلاعات از جمله علایق پژوهشی وی است.