اندیشکده فرهنگ و توسعه

معرفی مقاله



فراتحلیلی از تأثیرات هوش مصنوعی مولد بر مهارت زبانی و پیامدهای عاطفی-شناختی در یادگیری زبان

نویسنده:میرکا سارلاساچینی گوناسکارا وپرابا کوماراگه | مجله کشف محاسبات | ۲۰۲۶

خلاصه

چکیده:

ابزارهای هوش مصنوعی مولد (GenAI) به طور فزاینده‌ای در آموزش زبان ادغام می‌شوند، با این حال اثربخشی کلی آنها و عوامل شکل‌دهنده تأثیر آنها هنوز مشخص نیست. این متاآنالیز، شواهدی از ۵۱ مطالعه که ۱۷۵ اندازه اثر مستقل را گزارش کرده‌اند، برای ارزیابی تأثیر ابزارهای GenAI بر نتایج یادگیری زبان دوم و خارجی، ترکیب می‌کند. نتایج نشان می‌دهد که ابزارهای GenAI تأثیرات بزرگ و از نظر آماری معناداری بر مهارت زبانی و نتایج عاطفی-شناختی دارند. تجزیه و تحلیل‌های تعدیل‌کننده نشان داد که تأثیرات بر اساس مدت زمان مداخله، محیط آموزشی، ویژگی‌های زبان‌آموز و زبان هدف متفاوت است. تأثیرات قوی‌تر عموماً در محیط‌های یادگیری غیررسمی، برای مهارت‌های تولیدی و در مطالعاتی که به زبان‌های کمتر آموزش داده شده می‌پردازند، مشاهده شد. این یافته‌ها از نظریه‌های اجتماعی-فرهنگی و تعاملی در مورد اکتساب زبان دوم پشتیبانی می‌کنند و نیاز به ادغام متناسب و حساس به زمینه ابزارهای GenAI را برجسته می‌کنند. این بررسی، راهنمایی‌های عملی برای مربیان، طراحان و سیاست‌گذاران ارائه می‌دهد و بر اهمیت تحقیقات مداوم در زمینه‌های متنوع و جمعیت‌های زبان‌آموز تأکید می‌کند.

کلمات کلیدی: فراتحلیل، آموزش زبان، هوش مصنوعی مولد، چت‌جی‌پی‌تی، یادگیری زبان دوم و زبان خارجی

خلاصه مبسوط مقاله:

ابزارهای هوش مصنوعی مولد مانند ChatGPT از اواخر ۲۰۲۲ به سرعت وارد آموزش زبان شده‌اند. دانشجویان و مدرسان در کلاس‌های رسمی و در محیط‌های یادگیری غیررسمی از آن‌ها برای کمک نوشتاری، تمرین گفتاری و بازخورد شخصی‌سازی‌شده استفاده می‌کنند. برخلاف سامانه‌های قاعده‌مبنا، این ابزارها تولید زبان طبیعی و گفت‌وگوی حساس به بافت را ممکن می‌کنند. با این حال اثربخشی کلی آن‌ها و عوامل شکل‌دهنده این اثر هنوز روشن نیست. فراتحلیل‌های پیشین بیشتر فناوری‌های پیش از هوش مصنوعی مولد را بررسی کرده‌اند. Lyu و همکاران برای چت‌بات‌ها در ۳۱ مطالعه اندازه اثر ۰/۶۰۸ و Xiu-Yi برای مداخلات هوش مصنوعی در ۴۹ مطالعه اندازه اثر استاندارد ۰/۹۸۲ گزارش کرده‌اند. تنها فراتحلیل مربوط به هوش مصنوعی مولد را Guan و همکاران انجام داده‌اند. آن پژوهش فقط ۱۵ مطالعه درباره یادگیری غیررسمی انگلیسی را بررسی کرده و اثر کوچک تا متوسط (۰/۴۴۷) یافته است. تاکنون هیچ فراتحلیلی اثر این ابزارها را بر یادگیری زبان دوم و خارجی در هر دو بافت رسمی و غیررسمی نسنجیده است. این شکاف دو پرسش مشخص را ایجاب می‌کند.

Saarela، Gunasekara و Kumarage برای پاسخ به این شکاف دو چارچوب نظری را برگزیده‌اند. در نظریه فرهنگی اجتماعی رشد زبان از راه ابزارها و تعامل اجتماعی میانجی‌گری می‌شود. بر این پایه ابزارهای مولد می‌توانند به‌عنوان میانجی دیجیتال، عملکرد یادگیرنده را در منطقه تقریبی رشد او داربست‌بندی کنند. در رویکرد تعامل‌گرا نیز درون‌داد قابل‌فهم، مذاکره معنا و بازخورد اصلاحی نقش اصلی دارند و گفت‌وگوی تطبیقی و بی‌درنگ هوش مصنوعی مولد این امکان‌ها را فراهم می‌کند. بر اساس این چارچوب‌ها دو پرسش مطرح شده است. پرسش نخست اثر این ابزارها را بر پیامدهای یادگیری زبان دوم و خارجی در مقایسه با رویکردهای آموزشی دیگر می‌سنجد. پرسش دوم می‌پرسد کدام عوامل مربوط به یادگیرنده، مداخله و بافت این اثر را تعدیل می‌کنند. پیامدها به دو دسته پیشرفت زبانی و پیامدهای عاطفی و شناختی، مانند انگیزه، اضطراب، درگیری و خودتنظیمی، تقسیم شده‌اند. پاسخ به این پرسش‌ها نیازمند گردآوری نظام‌مند شواهد تجربی بود.

گردآوری شواهد از رهنمود پریزما ۲۰۲۰ پیروی کرده است. جست‌وجو در چهار پایگاه Scopus (۷۵ مقاله)، Web of Science (۷۰)، ACM Digital Library (۶) و IEEE Xplore (۴) برای مقاله‌های داوری‌شده منتشرشده از ۲۰۲۲ به بعد انجام شد و دو مقاله دیگر از راه پی‌گیری ارجاعات افزوده شد. معیارهای ورود چهار شرط داشت: بافت یادگیری زبان دوم یا خارجی، گروه آزمایش با ابزار هوش مصنوعی مولد، گروه مقایسه، و پیامد کمی همراه با اطلاعات آماری کافی برای محاسبه اندازه اثر. مطالعه‌های بدون گروه کنترل (۵ مورد)، بدون اطلاعات آماری کافی (۱۲ مورد) و با پیامدهای نامرتبط با یادگیری زبان (۳ مورد) کنار گذاشته شدند. ۳۸ مطالعه نیز به دلیل ناهمخوانی روش‌شناختی، نبود ابزار مشخص هوش مصنوعی مولد، نبود داوری همتا یا تمرکز بر دیدگاه معلمان حذف شد. نمونه نهایی ۵۱ مطالعه منتشرشده میان ۲۰۲۳ و ۲۰۲۵ با ۱۷۵ اندازه اثر بود. ۴۹/۰ درصد مطالعه‌ها طرح آزمایشی با گروه کنترل، ۳۹/۲ درصد شبه‌آزمایشی و ۱۱/۸ درصد پیش‌آزمون و پس‌آزمون درون‌آزمودنی داشتند. ۳۴/۹ درصد از چین بودند و ChatGPT در ۵۵/۶ درصد مطالعه‌ها بررسی شده بود. این ترکیب نمونه شیوه تحلیل داده‌ها را تعیین کرد.

شیوه تحلیل برای تجمیع این مطالعه‌های ناهمگن طراحی شده است. نخست d کوهن از میانگین‌ها و انحراف معیارهای گروه‌ها محاسبه شد و برای اصلاح سوگیری نمونه کوچک به g هجز تبدیل شد. اندازه‌ها با مدل اثر تصادفی تجمیع شدند، زیرا ناهمگنی

چشمگیر میان مطالعه‌ها پیش‌بینی می‌شد، و ناهمگنی با آماره Q و شاخص I^2 سنجیده شد. محاسبات با پایتون نسخه ۳/۱۱/۱۳ و بسته statsmodels انجام شد. پیامدها به پیشرفت زبانی و پیامدهای عاطفی و شناختی تقسیم شدند. پیامدهایی که کاهش آن‌ها مطلوب است، مانند اضطراب، بی‌حوصلگی، ناامیدی و بی‌انگیزگی، وارونه‌کدگذاری شدند تا همه اندازه‌های مثبت پیامد مطلوب را نشان دهند. این وارونه‌کدگذاری به‌صورت یک تعدیل‌گر دوتایی جداگانه ثبت شد. برای دو مطالعه که انحراف معیار گزارش نکرده بودند، انحراف معیار تجمیعی سایر مطالعه‌ها جایگزین شد. هر مطالعه را یک پژوهشگر کدگذاری کرد و دیگری مستقل بازبینی کرد و اختلاف‌ها با اجماع حل شد. شاخص پایایی بین کدگذاران مانند کاپای کوهن محاسبه نشد. اعتبار نتایج نیز با چند آزمون بررسی شد.

این آزمون‌ها شامل سوگیری انتشار، کیفیت روش‌شناختی و حساسیت بودند. بر اساس نتایج، پیشرفت زبانی (۷۱ اندازه اثر) به اندازه اثر تجمیعی ۱/۱۱۱ با فاصله اطمینان ۹۵ درصد از ۰/۵۶۶ تا ۱/۶۵۶ رسید و پیامدهای عاطفی و شناختی (۱۰۴ اندازه اثر) اندازه ۱/۰۲۹ با فاصله از ۰/۴۹۰ تا ۱/۵۶۷ داشتند. ناهمگنی بسیار بالا بود و I^2 برای دو دسته به ترتیب ۹۲/۷ و ۹۷/۲ درصد شد. تحلیل سوگیری انتشار دو نتیجه داد. تعداد ایمن در برابر شکست روزنتال ۵۳٬۱۰۹ شد و از آستانه ۸۸۵ بسیار فراتر رفت. آزمون Egger اما اثرهای مطالعه کوچک را معنادار نشان داد (عرض از مبدأ ۳/۰۶ و p کمتر از ۰/۰۰۱) و تحلیل برش و پر کردن هیچ مطالعه گمشده‌ای نیافت. در همان بخش اندازه اثر تجمیعی ۰/۸۱ (۰/۶۱ تا ۱/۰۲) گزارش شده است. ارزیابی خطای سوگیری با RoB 2 نشان داد بیشتر مطالعه‌ها خطر پایین یا نگرانی‌هایی دارند و شمار کمی پرخطر هستند. حذف مطالعه‌های دارای نمونه کمتر از ۵۰ (۱۵۰ اندازه اثر) و اندازه‌های فراتر از سه انحراف معیار (۱۴۷ اندازه اثر) نتیجه کلی را تغییر نداد. نتایج کلی در تحلیل تعدیل‌گرها به چند الگوی مشخص رسید.

تحلیل تعدیل‌گرها نشان داد اثر به‌طور منظم با ویژگی‌های مطالعه تغییر می‌کند و همه آزمون‌های Q میان گروهی p کمتر از ۰/۰۰۱ داشتند. مدت مداخله قوی‌ترین تعدیل‌گر بود. مداخله‌های بلندمدت (بیش از ۱۲ هفته) اثر ۱/۶۶۸، میان‌مدت (۶ تا ۱۲ هفته) ۰/۶۴۹ و کوتاه‌مدت ۰/۱۶۵ داشتند. اثر در بافت غیررسمی (۰/۷۴۱) بزرگ‌تر از بافت رسمی (۰/۴۳۶) بود. ابزارهای تولید بازخورد (۱/۷۰۸) و دستیاران بهره‌وری (۱/۴۵۴) بیشترین اثر را داشتند، هرچند این دو گروه تنها ۹ و ۴ اندازه اثر دارند. یادگیرندگان پیشرفته (۰/۷۱۵) بیش از مبتدی (۰/۴۶۶) و متوسط (۰/۳۶۵) بهره بردند و بزرگسالان (۰/۸۳۹) بیش از دانش‌آموزان دبیرستان (۰/۳۰۳) سود بردند. در مهارت‌ها، شنیداری (۱/۰۵۰ با ۲ اندازه اثر)، گفتاری (۰/۶۸۶) و نوشتاری (۰/۵۴۵) بالاترین اثر را داشتند. اثر در آفریقا (۱/۷۸۴ با ۳ اندازه اثر) و برای زبان کره‌ای (۰/۷۰۴) چشمگیر بود. پیامدهای وارونه‌کدگذاری‌شده، مانند کاهش اضطراب، اثر بیشتری (۰/۶۰۱) از پیامدهای مثبت (۰/۴۳۲) داشتند. این الگوها پیامدهای نظری و آموزشی روشنی دارند.

بر اساس این الگوها، نویسندگان نتیجه می‌گیرند هوش مصنوعی مولد را باید داربستی تعاملی دانست که عاملیت و تولید زبانی یادگیرنده را تقویت می‌کند. این نتیجه با دیدگاه‌های فرهنگی اجتماعی و تعامل‌گرا سازگار است، زیرا بازخورد و تمرین گفت‌وگومحور بیشترین اثر را داشتند. به لحاظ آموزشی، این ابزار راه‌حل یکسان برای همه نیست و باید با سطح مهارت، اهداف آموزشی و بافت نهادی هماهنگ شود. ادغام بلندمدت، استفاده در بافت غیررسمی یا ترکیبی و به‌کارگیری ابزار برای بازخورد و کمک نوشتاری اثر را افزایش می‌دهد و نقش معلم در هدایت و کنترل اعتبار محتوا حفظ می‌شود. محدودیت‌ها نیز روشن است. ناهمگنی بسیار بالا است و بخشی از آن از تجمیع پیامدهای متنوع در دو دسته کلی می‌آید. برخی زیرگروه‌ها مطالعه‌های اندک دارند و بیشتر مطالعه‌ها بر

انگلیسی، یادگیرندگان بزرگسال، سنجش‌های خودگزارشی و کوتاه‌مدت تکیه دارند. توزیع جغرافیایی آسیامحور است و هیچ مطالعه‌ای از ایالات متحده یا بریتانیا نیست. پژوهش‌های آینده باید مدت طولانی‌تر، سنجش‌های عینی و زبان‌ها و مناطق متنوع‌تر را بررسی کنند.

نکات کلیدی

- **مطالعه موردی:** فراتحلیل ۵۱ مطالعه تجربی و شبه‌تجربی منتشرشده میان ۲۰۲۳ و ۲۰۲۵ با ۱۷۵ اندازه اثر، درباره کاربرد ابزارهای هوش مصنوعی مولد در یادگیری زبان دوم و خارجی در بافت‌های رسمی و غیررسمی
- **مفهوم مرکزی:** هوش مصنوعی مولد به‌مثابه داربست تعاملی که با بازخورد بی‌درنگ و گفت‌وگوی تطبیقی، عاملیت و تولید زبانی یادگیرنده را تقویت می‌کند
- **چارچوب تحلیلی:** نظریه فرهنگی اجتماعی (منطقه تقریبی رشد و داربست‌بندی) و نظریه تعامل‌گرا (درون‌داد قابل‌فهم، مذاکره معنا و بازخورد اصلاحی)، با روش فراتحلیل اثر تصادفی، g هجز و تحلیل تعدیل‌گرها بر پایه رهنمود پریزما ۲۰۲۰
- **یافته کلیدی:** اندازه اثر پیشرفت زبانی ۱٫۱۱۱ و پیامدهای عاطفی و شناختی ۱٫۰۲۹ گزارش شده است. اثر در مداخله‌های بلندمدت (۱٫۶۶۸)، بافت غیررسمی (۰٫۷۴۱)، ابزارهای تولید بازخورد (۱٫۷۰۸) و یادگیرندگان پیشرفته و بزرگسال بیشتر بود.
- **استدلال اصلی:** اثر این ابزارها به مدت مداخله، بافت، نوع ابزار و ویژگی‌های یادگیرنده وابسته است. بنابراین آن را باید متناسب با بافت به کار برد و راه‌حل یکسان برای همه شمرد، به‌ویژه که ناهمگنی بسیار بالا، اثرهای مطالعه کوچک و تمرکز نمونه بر آسیا و انگلیسی احتیاط در تعمیم را لازم می‌کند.



اندیشکده فرهنگ و توسعه

باشگاه مقالات فرتو