An Analysis of the Advantages and Ethical Challenges of Large Language Models in Research

Rahman Sharifzadeh*
PhD in Philosophy of Science and Technology; Assistant Professor; Information and Society Research Department; Iranian Research Institute for Information Science and Technology (IranDoc); Tehran, Iran Email: sharifzadeh@irandoc.ac.ir

Iranian Journal of Information Processing and Management

Iranian Research Institute for Information Science and Technology (IranDoc)
ISSN 2251-8223
eISSN 2251-8231
Indexed by SCOPUS, ISC, & LISTA
Vol. 40 | No. 4 | pp. 1219-1250
Summer 2025
https://doi.org/10.22034/jipm.2025.2050658.1901

Received: 13, Jan. 2025 **Accepted: 18, Apr. 2025**

Abstract: It has not been long since the emergence of large language models, but due to the significant hype surrounding their applications and prospects, especially in the field of research, it seems essential to examine their applications as well as the advantages and ethical challenges of this technology. What applications do large language models have, and what advantages and ethical challenges do these applications bring, particularly in academic work? In this article, we will analyze and discuss the most important advantages and ethical challenges of these language models.

The methodology of this research for extracting advantages and challenges involved reviewing and examining relevant research works identified in the Scopus and Google Scholar databases. We identified numerous articles by searching for terms such as 'challenges/ advantages of LLMs in research', 'risks/ethical concerns of LLMs', 'ethics of ChatGPT', 'ethics of LLMs', 'LLMs and research ethics', and 'artificial intelligence and research ethics'. After removing duplicates, the articles were studied based on their relevance to the topic of this research, and the advantages and challenges were extracted. The author then aggregated and categorized the advantages and challenges, followed by an analysis and discussion of each.

Alongside advantages such as overcoming language barriers for non-English-speaking authors, reducing preparation costs, saving time, improving writing quality, facilitating the dissemination of knowledge, and assisting researchers with special conditions, there are significant challenges such as the hallucination, accountability and authorship, plagiarism and copyright issues, bias, security and privacy, safety and potential harm, transparency and explainability, environmental issues, and unemployment and exploitation. The three special research-oriented issues we examined are each important in their own right. Neglecting the issue of hallucination will ultimately have a detrimental effect on the quality and integrity of researchers' work. The issue of accountability and authorship requires heightened attention from publishers and even those engaged in theoretical work on ethics and philosophy of artificial intelligence.

* Corresponding Author

Currently, no publisher or association/ committee recognizes LLMs as authors. While this lack of recognition may prevent some problems, it also leads to other issues. The issue of plagiarism and copyright is significant because language models do not perform well in citation; they rarely provide correct direct citations and often do not maintain sufficient distance from the original text in indirect citations. One reason for this may be that this technology was not primarily designed for scientific writing but rather to write in a natural manner, similar to humans.

Large language models undoubtedly have significant impacts on the research domain. One of the most important advantages is overcoming language barriers for publication. This can be viewed as an opportunity for developing countries: non-English speakers who previously wrote less in English are now writing more with the help of generative AI tools. This, in turn, will affect the visibility rates of educational and research institutes in non-English-speaking countries. Alongside these advantages, efforts must be made to minimize the ethical issues associated with this technology. By analyzing these advantages and challenges, and avoiding overly optimistic or pessimistic approaches to this technology, we will discuss that addressing or minimizing the severity of some challenges, particularly those specific to language models in the research domain, requires theoretical, policy, educational, and technical efforts.

Keywords: Large Language Models, Artificial Intelligence, Writing, Research, Ethical Challenges

تحلیل و بررسی مزایا و چالش‌های اخلاقی مدل‌های زبانی بزرگ در سپهر پژوهش

رحمان شریف‌زاده

دکتری فلسفه علم و فناوری؛ استادیار؛ پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)؛ تهران، ایران؛
پدیدآور رابط sharifzadeh@irandoc.ac.ir

پژوهش‌نامهٔ پردازش و مدیریت اطلاعات

دریافت: ۱۴۰۳/۱۰/۲۴ | **پذیرش:** ۱۴۰۴/۰۱/۳۰ | مقاله برای اصلاح به مدت ۱ روز نزد پدیدآور بوده است.

نشریهٔ علمی | رتبهٔ بین‌المللی
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)
شاپا (چاپی) ۸۲۲۳-۲۲۵۱
شاپا (الکترونیکی) ۸۲۳۱-۲۲۵۱
نمایه در LISTA، ISC، و SCOPUS
jipm.irandoc.ac.ir
دورهٔ ۴۰ | شمارهٔ ۴ | ص‌ص ۱۲۱۹-۱۲۵۰
تابستان ۱۴۰۴
https://doi.org/10.22034/jipm.2025.2050658.1901

چکیده: مدل‌های زبانی بزرگ چه کاربست‌هایی دارند و این کاربست‌ها به چه مزایا و چالش‌هایی به‌ویژه در سپهر پژوهش می‌انجامند؟ در این مقاله به تحلیل و بررسی مهم‌ترین مزایا و چالش‌های اخلاقی این مدل‌های زبانی خواهیم پرداخت. خواهیم دید که در کنار مزایایی همچون رفع موانع زبانی برای نویسندگان غیرانگلیسی‌زبان، کاهش هزینه آماده‌سازی، صرفه‌جویی در زمان، افزایش کیفیت نگارش، ترویج راحت‌تر دانش، و کمک به پژوهشگران دارای شرایط خاص، چالش‌های مهمی نظیر توهم، مسئولیت‌پذیری و مؤلف‌بودن، مشابه‌نویسی و سرقت ادبی، سوگیری، امنیت و حریم خصوصی، ایمنی و توان آسیب‌رسانی، شفافیت و توضیح‌پذیری، مسائل زیست‌محیطی، بیکاری و بهره‌کشی نیز وجود دارد. با تحلیل این مزایا و چالش‌ها و گریز از رویکردهای کاملاً خوشبینانه یا بدبینانه نسبت به این فناوری، نشان خواهیم داد که برای رفع یا کمینه کردن شدت برخی از چالش‌ها، به‌ویژه چالش‌های خاصی که مدل‌های زبانی در سپهر پژوهش ایجاد می‌کنند، نیاز به کارهای نظری، سیاستی، آموزشی و فنی وجود دارد.

کلیدواژه‌ها: مدل‌های زبانی بزرگ، هوش مصنوعی، نگارش، پژوهش، چالش‌های اخلاقی

۱. مقدمه

مدل زبانی بزرگ (ال‌ال‌ام)[1]، مجموعه الگوریتم‌هایی است برای شبیه‌سازی هرچه کاراتر کارکردهای زبان طبیعی. در چند سال اخیر نمونه‌های موفقی از این مدل‌ها از جمله مجموعه «جی‌پی‌تی»[2]، «جمینی»[3]، «جما»[4]، «کلاود»[5]، و «لاما»[6] معرفی شده‌اند. این مدل‌ها از جهت بزرگی مجموعه داده‌هایی که روی آن آموزش می‌بینند و استنتاج می‌کنند، شمار پارامترها، و همچنین فناوری‌هایی که در تولید متون استفاده می‌کنند، با مدل‌های زبانی کوچک متفاوت‌اند. کاربرد یادگیری عمیق، مدل‌های تبدیل‌کننده[7]، و پارامترهایی با تعداد چندمیلیاردی از ویژگی‌های این نوع مدل‌های زبانی محسوب می‌شوند.

«ال‌ال‌ام»ها تغییر و تحول قابل توجهی در حوزه‌های مختلف، از جمله در پژوهش و نگارش ایجاد کرده‌اند یا پیش‌بینی می‌شود که ایجاد کنند. از هنگامی که «چت‌جی‌پی‌تی» به‌عنوان یک «ال‌ال‌ام» کارآمد معرفی شد، هیجان زیادی درباره ورود دانشگاهیان به یک پارادایم پژوهش/ نگارش جدید ایجاد کرد. این سامانه کم‌کم به یکی از ابزارهای مهم در دست پژوهشگران تبدیل شد. پژوهشگران متعددی در مورد اثرات و کاربست‌های «ال‌ال‌ام»ها در سپهر پژوهش و نگارش قلم زده و به موارد متنوعی همچون نگارش چکیده و ادبیات موضوع، پیشنهاد روش، داوری و بازخورد، ویرایش و ترجمه و غیره اشاره کرده‌اند (Meyer et al. 2023; Xames & Shefa 2023; Zohery et al. 2023; Biswas 2023)؛ با این حال، این کاربست‌ها به برخی نگرانی‌های اخلاقی نیز دامن زدند. در این پژوهش به بررسی و تحلیل دلالت‌های اخلاقی این تحول اجتماعی-فنی خواهیم پرداخت. از این بحث خواهیم کرد که کاربست‌های این مدل‌های زبانی چه مزایا و چالش‌هایی ایجاد می‌کنند. بررسی و تحلیل این مزایا و چالش‌ها درک عمیق‌تری از وضعیت این فناوری و رابطه آن با کاربر به‌دست می‌دهد و چنانکه خواهیم دید، می‌تواند نه‌تنها از جنبه اخلاقی، بلکه از جهت سیاست‌گذاری نیز مهم باشد.

۲. پیشینه و روش

مدت زیادی از عمر «ال‌ال‌ام»ها سپری نشده، ولی از آنجا که هیجان و جاروجنجال زیادی[8]

1. large language model (LLM)
2. GPT 1-4
3. Gemini
4. Gema
5. Claude
6. Llama
7. transformer
8. hype

در مـورد کاربسـت‌ها و چشـم‌اندازهای ایـن فنـاوری ایجـاد شـده، پژوهشـگران متعـددی در سـطح بین‌المللـی بـه ایـن فناوری‌هـا پرداختـه و آثـاری را در ایـن زمینـه منتشـر کرده‌انـد. در سـطح داخـلی، بیشـتر مـجلات عمـومی و غیردانشـگاهی بـه ایـن پدیـده و اهمیـت آن توجـه نشـان داده‌انـد و چنـد مقالـه علـمی نیـز منتشـر شـده است. بـرخی مقـالاتی کـه بـه فـارسی منتشـر شـده‌اند، بیشـتر نـگاهی معرفی‌گونـه بـه ایـن فنـاوری دارنـد و چنـدان وارد تحلیـل آن نشـده‌اند. بـرای مثـال، «منتظـری، غفـاری و موسـوی موحـدی» بـه بـرخی از کاربردهـای «چـت‌جی‌پی‌تی»، از جملـه پـردازش داده‌هـا، تولیـد و آزمـون فرضیـه، و تقویـت همـکاری و ارتباطـات، و همچنیـن چالش‌هـای اخلاقی آن، از جملـه حریـم خصـوصی و داده‌هـا، مالکیـت فکـری و تألیـف، سـوء اسـتفاده، اطلاعـات نادرسـت و غیـره بـه شـکلی بسیـار اجمـالی اشـاره کرده‌انـد (۱۴۰۲). «علیرضـا بهمن‌آبـادی» نیـز بـه کاربردهـای گاهی متفـاوت ایـن «ال‌ال‌ام»هـا در حـوزه پژوهـش اشـاره کـرده اسـت. وی بیـان می‌کنـد کـه از فناوری‌هـای هـوش مصنـوعی، از جملـه «چـت‌جی‌پی‌تی» می‌تـوان در داوری مقـالات، خلاصه‌سـازی اطلاعـات، پیشینه‌پـژوهی، ترجمـه، تجزیه‌وتحلیـل داده‌هـا و تشـخیص الگـو اسـتفاده کـرد. همچنیـن وی بـه چالش‌هـایی چـون چولـگی داده‌هـا، حریـم خصـوصی، شـفافیت، تکرارپذیـری، مالکیـت فکـری، مسـئولیت‌پذیری، تأثیـر اجتمـاعی و جابه‌جـایی شـغل نـگاهی انداختـه اسـت (۱۴۰۲).

در سـطح بین‌المللـی آثـار متنوع‌تـری منتشـر شـده‌اند. «لانـد و وانـگ» بـه ایـن مسـئله پرداخته‌انـد کـه «چـت‌جی‌پی‌تی» به‌عنـوان یـک مـدل زبـانی بـزرگِ مطـرح، چـه تأثیـری روی پژوهـش و انتشـار دانشـگاهی خواهـد گذاشـت (Lund and Wang 2023). «شلاگویـن و ویلکاکـز» در مقالـه‌ای بـا عنـوان «چـت‌جی‌پی‌تی و همـکاران: اخلاق اسـتفاده از هـوش مصنـوعی (زایـا) در پژوهـش و علـم» از هشـت چالـش اخلاقی اسـتفاده از هـوش مصنـوعی صحبـت می‌کننـد کـه بـرخی از آن‌هـا از دیـدگاه اخلاق پژوهـش مهـم هسـتند: ۱. شـکننده اسـت (هـوش مصنـوعی انعطـاف لازم را نـدارد و فقـط از جوانـب محـدودی می‌توانـد خـوب عمـل کنـد)، ۲. مـات/غیرشـفاف اسـت (فراینـد تصمیم‌گیـری و اسـتنتاج در درون یـک جعبـه سیـاه اسـت و نمی‌تـوان مراحـل آن را بـررسی کـرد)، ۳. حریـص اسـت (بـه داده‌هـای بسیـار زیـادی بـرای آمـوزش و همچنیـن زیرسـاخت پرهزینـه نیـاز دارد)، ۴. سـطحی و کم‌عمـق اسـت (بـر خلاف عامـل انسـانی دارای چیـزی کـه دانـش ضمنـی[1] خوانـده می‌شـود، نیسـت)، ۵. قابـل هـک و دسـتکاری اسـت

1. tacit

(شرکت‌ها، دولت‌ها و دیگر کنشگران می‌توانند آن را در جهت مقاصد خود هک و هدایت کنند)، ۶. سوگیری دارد (سوگیری‌های انسانی را که در داده‌ها نهفته هستند، انتقال می‌دهد)، ۷. تهاجمی است (می‌تواند داده‌های خصوصی افراد را افشا کند)، و ۸. وانمودی است (وانمود می‌کند که خلاق و حساس است در حالی که نیست) (Schlagwein and Willcocks 2023).

«ژو» و همکاران به چهار نوع نگرانی اخلاقی در رابطه با «ال‌ال‌ام»هایی همچون «چت‌جی‌پی‌تی» اشاره می‌کنند: ۱. سوگیری (آیا «چت‌جی‌پی‌تی» سوگیری‌های انسانی را از طریق فرایند یادگیری از سوژه‌های انسانی بازتولید نمی‌کند؟)، ۲. استحکام («چت‌جی‌پی‌تی» چقدر می‌تواند در برابر مختل شدن، از کار افتادن، نقض حریم خصوصی، و غیره مقاوم باشد؟)، ۳. اعتمادپذیری («چت‌جی‌پی‌تی» چقدر محتوای دقیق و درست تولید می‌کند؟)، و ۴. سمی بودن («چت‌جی‌پی‌تی» چقدر محتوای مضر و آسیب‌زا تولید می‌کند؟) (Zhuo et al. 2023). همچنین «زوهری» بیان می‌کند که «چت‌جی‌پی‌تی» و «ال‌ال‌ام»های مشابه می‌توانند بیشتر کارهایی را که معمولاً پژوهشگران در حوزه نگارش و انتشار علمی انجام می‌دهند، انجام دهند. در نتیجه، بیشتر چالش‌های اخلاق پژوهش در اینجا نیز بروز می‌کنند (Zohery 2023). مقاله «واتکینز» از معدود مقالاتی است که تلاش کرده وارد بعد هنجاری-سیاستی این بحث شود و هنجارها و رهنمودهایی را برای داوران همتا و پژوهشگران به‌منظور استفاده اخلاقی از «ال‌ال‌ام»ها طرح کند (Watkins 2024).

اگرچه در آثار کم‌وبیش متعددی به مزایا و چالش‌های هوش «ال‌ال‌ام»ها اشاره شده است، اما این مقاله در مقایسه با آن‌ها از چندین جهت، مشارکت[1] قابل توجهی محسوب می‌شود. این مشارکت به این قرار است: اول اینکه در هر یک، از جهتی به مزایا و چالش‌ها پرداخته شده است، ولی هیچ‌یک از جامعیت کافی برخوردار نیست؛ دوم اینکه بسیار کم به بحث چالش‌های خاص «ال‌ال‌ام»ها در امر پژوهش ورود شده است؛ سوم اینکه، بیشتر مقالات یا فاقد جنبه تحلیل چالش‌ها و مزایا هستند یا نسبت به این فناوری، درگیر موضع‌گیری‌های خوش‌بینانه و بدبینانه شده‌اند. مقاله حاضر بررسیِ به‌نسبت جامعی از تمام مزایا و چالش‌ها با تمرکز بر امر پژوهش ارائه می‌دهد و با در اختیار داشتن

1. contribution

رویکردی نظری بر مطالعات علم و فناوری (اس‌تی‌اس)[1]، به تحلیل هر کدام می‌پردازد.

روش کار این پژوهش برای استخراج مزایا و چالش‌ها، مرور و بررسی آثار پژوهشی مرتبطِ شناسایی‌شده در پایگاه «اسکوپوس» و «گوگل اسکالر» است. برای این کار و با جست‌وجوی عباراتی همچون «چالش‌ها/ مزایای ال‌ال‌ام‌ها در پژوهش»، «خطرها/ نگرانی‌های اخلاقی ال‌ال‌ام‌ها»، «اخلاق چت‌جی‌پی‌تی»، «اخلاق ال‌ال‌ام»، «ال‌ال‌ام و اخلاق پژوهش»، «هوش مصنوعی و اخلاق پژوهش»، تعداد زیادی مقاله شناسایی شد و پس از حذف موارد تکراری، بر اساس ارتباط با موضوعِ این پژوهش مورد مطالعه قرار گرفته و مزایا و چالش‌ها استخراج شدند. آنگاه نسبت به تجمیع و دسته‌بندی مزایا و چالش‌ها و سرانجام، تحلیل و بحث هر کدام پرداخته شد

۳. مزایای «ال‌ال‌ام»ها

فناوری برای به‌دست آوردن مزیت یا فایده‌ای خاص طراحی می‌شود یا دست کم توسط کاربران یا جامعه چنین تعبیر می‌شود. کاربست‌هایی که «ال‌ال‌ام»ها دارند، به مزیت‌هایی انجامیده‌اند که محور بحث پژوهشگران بوده است. به شش مورد اصلی از این مزایا اشاره می‌شود: از میان برداشتن موانع زبانی برای نویسندگان غیرانگلیسی‌زبان، کاهش هزینه آماده‌سازی، صرفه‌جویی در زمان، افزایش کیفیت نگارش، ترویج راحت‌تر دانش، و کمک به پژوهشگران دارای شرایط خاص. این موارد را در زیر به‌ترتیب مرور می‌کنیم.

۳-۱. از میان برداشتن مانع زبانی

نخستین و مهم‌ترین مزیت «ال‌ال‌ام»ها این است که موانع زبانی را از سر راه پژوهشگران به‌ویژه نویسندگان غیرانگلیسی‌زبان برمی‌دارند (Meyer et al. 2023). «ال‌ال‌ام»هایی همچون «چت‌جی‌پی‌تی» نه‌تنها متن غیرانگلیسی را به انگلیسی ترجمه می‌کنند، بلکه در کار واگویی[2]، بازنویسی، افزایش تنوع واژگان، بررسی خطاهای املایی و دستوری کارایی بالایی دارند. این بدان‌معناست که پژوهشگرانی که پیشتر به خاطر دانش یا مهارت زبانی پایین کمتر به انگلیسی می‌نوشتند، اکنون به‌واسطه این فناوری بیشتر می‌نویسند. واسطه قرار گرفتن فناوری (Latour 1994) در اینجا کاملاً مشهود است. حتی افرادی که کمتر علاقه داشتند به انگلیسی بنویسند، اکنون با وجود این فناوری ممکن است تغییری در علاقه‌شان ایجاد

1. Science and Technology Studies (STS)
2. paraphrasing

شـود. نتیجـه کلان‌تـر ایـن مزیـت آن اسـت کـه پژوهشـگران کشـورهای در حـال توسـعه، و در نتیجـه، مؤسسـات و دانشـگاه‌های ایـن کشـورها بیشـتر دیـده خواهنـد شـد و رؤیت‌پذیری‌شـان افزایـش خواهـد یافـت. همچنیـن بـه نظـر می‌رسـد کـه کنـار رفتـن ایـن مانـع گامی در جهـت دموکراتیـزه کـردن دانـش (Koller et al. 2023) و افزایـش عدالـت و انصـاف اسـت. بـا رفـع موانـع زبـانی، پژوهشـگران کشـورهای غیرانگلیسی‌زبـان سـهم منصفانه‌تـری از کیـک دانـش در سـطح جهـانی خواهنـد داشـت. آن‌هـا بیشـتر خواهنـد نوشـت و بیشـتر مطـرح خواهنـد شـد.

نتیجـه دیگـر کنـار رفتـن مانـع زبـانی، افزایـش نسـبی همکاری‌هـای بین‌المللی خواهـد بـود؛ چـرا کـه میانجیگـری قدرتمنـد «ال‌ال‌ام»هـا در نـگارش و ترجمـه می‌توانـد بـه پژوهشـگران از کشـورهای گوناگـون بـا زبان‌هـای گوناگـون کمـک کنـد تـا راحت‌تـر وارد ارتبـاط و مـراوده علـمی شـوند.

۳-۲. کاهش هزینه‌های آماده‌سازی

وجـود «ال‌ال‌ام»هـا هزینه‌هـای مربـوط بـه ویرایـش و آماده‌سـازی برونـداد پژوهـشی را کاهـش خواهـد داد. پژوهشـگران ابتـدا ناچـار نیسـتند بـرای ترجمـه، ویرایـش و قالب‌بنـدی مقـالات و کتاب‌هـای خـود هزینـه بپردازنـد. «ال‌ال‌ام»هـا به‌خـوبی ایـن کار را انجـام خواهنـد داد. همچنیـن دیدیـم کـه «ال‌ال‌ام»هـا در مـرور ادبیـات موضـوع، خلاصه‌کـردن، گـردآوری و تحلیـل داده هـم کاربـرد دارنـد. بنابرایـن، اگـر پژوهشـگری بـرای ایـن کارهـا پیشـتر از دسـتیار پژوهـشی اسـتفاده یـا هزینـه پرداخـت می‌کـرد، اینـک بـا کاراتـر شـدن «ال‌ال‌ام»هـا، ناچـار بـه ایـن کار نیسـت. البتـه در بخـش بعـدی خواهیـم دیـد کـه «ال‌ال‌ام»هـا اثـر دوگانـه دارنـد؛ یعـنی گاهی همـراه مزیت‌هـا، چالـش را هـم ایجـاد می‌کننـد. حضـور «ال‌ال‌ام»هـا احتمـالاً بـه بدترشـدن شـرایط کاری ویراسـتاران، مترجمـان، و دسـتیاران پژوهـشی خواهـد انجامیـد

۳-۳. صرفه‌جویی در زمان

«ال‌ال‌ام»هـا به‌طـور کلی، بـه کار نـگارش شـتاب می‌دهنـد (Burger et al. 2023). افـراد، به‌ویـژه افـراد غیرانگلیسی‌زبـان زمـان کمتـری را بـرای نـگارش متـن صـرف می‌کننـد؛ چـرا کـه «ال‌ال‌ام»هـا در بسـیاری از کارهـای نـگارشی و ویرایـشی بـه نویسـنده کمـک خواهنـد کـرد. از ایـن گذشـته، «ال‌ال‌ام»هـا بـه فراینـد مطالعـه و گـردآوری داده نیـز شـتاب می‌بخشـند (Macdonald et al. 2023). «ال‌ال‌ام»هـا می‌تواننـد یـک متـن مفصـل و پیچیـده را بـرای پژوهشـگر خلاصـه و سـاده کننـد. بنابرایـن، بـه نظـر می‌رسـد کـه به‌شـکل قابـل توجـهی در وقـت پژوهشـگران بـرای مطالعـه

متن‌های وقت‌گیر صرفه‌جویی شود. همچنین فرایند تحلیل و بازنمایی داده‌ها و نیز یافتن نشریه مناسب برای انتشار نیز کوتاه‌تر می‌شود؛ چرا که همان‌گونه که دیدیم، «ال‌ال‌ام»ها می‌توانند در این کارها به پژوهشگر کمک کنند

با این حال باید به دو نکته توجه کرد. نکته نخست این است که این فناوری می‌تواند اثر دوگانه داشته باشد: اول اینکه «ال‌ال‌ام»ها ممکن است با خلاصه کردن و ساده‌سازی یک متن پیچیده (مثلاً متون فلسفی برخی از فیلسوفان)، محتوای آن را قربانی کنند. در نتیجه، اگرچه در زمان صرفه‌جویی می‌کنند، اما ممکن است محتوای تحریف‌شده و نادرست به مؤلف بدهند. نکته بعدی این است که «شتاب» ذاتاً یک مزیت محسوب نمی‌شود. شتاب برای نظام دانشگاهی امروز که با هنجار ضمنیِ «یا منتشر کن یا نابود شو»[1] پژوهشگران را زیر فشار می‌گذارد، یک ارزش و مزیت است. با این حال، جریان‌هایی مانند «جنبش علم آهسته است»، شتاب را مزیت نمی‌دانند و معتقدند که شتاب مانع عمق محتوایی و اثرگذاری عمیق اجتماعی اثر تولیدشده می‌شود.

۳-٤. افزایش کیفیت نگارش دانشگاهی

«ال‌ال‌ام»ها می‌توانند کیفیت نگارش دانشگاهی را، چه از نظر شکلی و چه از دید محتوایی ارتقا دهند (Meyer et al. 2023). «ال‌ال‌ام»ها با واگویی عبارات پیچیده به گفتار ساده‌تر و همچنین کاهش ابهام جملات می‌توانند خوش‌خوانی و آسان‌فهمی یک متن را افزایش دهند. همچنین با افزایش سازگاری و پیوستگی منطقی متن، انسجام آن را تقویت می‌کنند. از جهت محتوایی نیز «ال‌ال‌ام»ها می‌توانند آخرین دیدگاه‌ها و ایده‌های مطرح‌شده در باره یک موضوع را به مؤلف بدهند و این به‌نوبه خود روی به‌روز بودن بروندادی که پژوهشگر می‌خواهد تولید کند، اثر می‌گذارد. همچنین پژوهشگر در فرایند تولید متن می‌تواند پیوسته از بازخوردهای «ال‌ال‌ام» کمک بگیرد. این بازخوردها نیز به افزایش کیفیت کار پژوهشگر کمک می‌کنند. چنانکه «درگا» و همکاران اشاره می‌کنند، «ال‌ال‌ام»ها دستیارهای آنی هستند و افراد می‌توانند در لحظه از بازخوردها و نظرات آن‌ها کمک بگیرند (Dergaa et al. 2023)

۳-۵. ترویج و همرسانی دانش

یکی از مزیت‌های قابل توجه «ال‌ال‌ام»ها این است که دانش را برای افراد بسیار زیادی

1 . publish or perish

دسترس‌پذیر می‌کنند (Huang et al. 2023). اگر دسترسی به «ال‌ال‌ام»هایی همچون «چت‌جی‌پی‌تی» همگانی‌تر شود، تمام افرادی که به گوشی‌های هوشمند متصل هستند، می‌توانند از توانایی همرسانی و ترویج دانش «ال‌ال‌ام»ها کمک بگیرند. از این گذشته، خود مروجان علم می‌توانند از «ال‌ال‌ام»ها برای بازنویسی متن به شکلی که مناسب ترویج علم باشد، استفاده کنند. کاربر با اشاره‌نویسی[1] دقیق می‌تواند نوع مخاطب را برای «ال‌ال‌ام»ها مشخص کند. برای مثال، با هدف ترویج علم، کاربر می‌تواند از «چت‌جی‌پی‌تی» بخواهد متنی را چنان بنویسید که متناسب افراد غیردانشگاهی باشد.

۳-٦. کمک به افراد دارای نابهنجاری‌های خاص

«ال‌ال‌ام»ها می‌توانند به دانشجویان و پژوهشگرانی که به ناهنجاری‌هایی مانند «دیسلکسیا»[2]، «ای‌دی»[3] (عدم تمرکز)، «ای‌دی‌اچ‌دی»[4] (اختلال عدم تمرکز-بیش‌فعالی) و «اُتویسم»[5] دچار هستند، کمک کند (Malmstrom et al. 2023). این افراد به‌طور معمول، در متن‌نویسی مشکل دارند (گاهی آن‌ها در تشخیص درست املای کلمات مشکل دارند، یا نمی‌توانند یک متن منسجم را نگارش کنند) و بر روی بروندادهای پژوهشی‌شان تأثیر منفی می‌گذارد. «ال‌ال‌ام»ها کمک می‌کنند که این افراد بروز و ظهور بیشتری در عرصه انتشار داشته باشند. بیان تجربه استفاده «ال‌ال‌ام»ها توسط یکی از افرادی که دچار «دیسلکسیا»ست، خالی از فایده نیست

«به‌عنوان کسی که دچار دیسلکسیا هستم، دریافته‌ام که مدل‌های زبانی بزرگ

---

۱. پرامپتینگ (prompting) از هنر نمایشی، وارد ادبیات «ال‌ال‌ام»ها شده است. پرامپتینگ در نمایش که به‌طور معمول، «سخن‌رسانی» ترجمه می‌شود، وقتی رخ می‌دهد که یک بازیگر تئاتر متن گفت‌وگو را فراموش می‌کند. در این موقع کسی که معمولاً سخن‌رسان خوانده می‌شود، با نمایش متن روی یک تابلو از دور، گفت‌وگو را به یاد بازیگر می‌آوَرَد. این واژه با ورود به ادبیات هوش مصنوعی، دستخوش تغییر معنایی ظریفی شده است. در اینجا «سخن‌رسانی» دیگر ترجمه مناسبی نیست؛ چرا که قرار نیست «ال‌ال‌ام» متنی را که کاربر در دست دارد، تکرار کند، بلکه قرار است بر اساس اشارهٔ کاربر برای وی متن جدیدی «تولید» کند. کاربر با واژگان و جملاتی به آنچه در ذهن خود دارد، اشاره می‌کند و «ال‌ال‌ام» بر اساس اشارهٔ کاربر برای وی متن تولید می‌کند. از این‌رو، ترجمه «اشاره‌نویسی» پیشنهاد می‌شود؛ چرا که درخواست کاربر صرفاً اشاره‌ای کوچک به چیزی است که در ذهن دارد یا می‌خواهد در مورد آن بداند. اشاره‌ای خوب برای «ال‌ال‌ام»، کافی خواهد بود تا کاربر نتیجه مطلوب را از آن بگیرد.

2. dyslexic
3. attention deficit (AD)
4. attention deficit hyperactivity disorder (ADHD)
5. Autism

و ربات‌های گفت‌وگو به‌طور باورنکردنی به من در برقراری ارتباط مؤثرتر کمک کرده‌اند. من از آن‌ها برای ایجاد یک پیش‌نویس یا طرح کلی از افکارم برای تقریباً هر ارتباطی که نیاز دارم، هم برای امور تجاری و هم برای مسائل شخصی، استفاده می‌کنم. اصلی‌ترین تأثیر دیسلکسیا روی من در تایپ کردن است، که در طول سال‌ها مرا از برقراری ارتباط با گروه یا خانواده‌ام منع کرده است» (Levy 2023)

٤. چالش‌های اخلاقی «ال‌ال‌ام»ها

در کنار مزایایی که «ال‌ال‌ام»ها ارائه می‌کنند، چالش‌هایی نیز وجود دارد. در بخش قبلی گفتیم که «ال‌ال‌ام»ها گاهی در واقع، اثر دوگانه دارند؛ یعنی در همان حال که اثری مطلوب تولید می‌کنند، اثری نامطلوب نیز ایجاد می‌کنند. در این بخش مهم‌ترین چالش‌هایی را که «ال‌ال‌ام»ها می‌توانند در سپهر پژوهش به بار بیاورند، به بحث خواهیم گذاشت. با این حال، پیش از آن به برخی چالش‌های کلی که «ال‌ال‌ام»ها به‌عنوان «یک فناوری هوشمند» ایجاد می‌کنند، خواهیم پرداخت. این چالش‌ها در مورد بقیه سیستم‌های هوشمند نیز کم‌وبیش صادق هستند و به «ال‌ال‌ام»ها مختص نیستند. با این حال، این نیز به شکل مستقیم و غیرمستقیم به امر پژوهش مرتبط می‌شود. پس از چالش‌های کلی، وارد چالش‌های اخلاقی خاص سپهر پژوهش خواهیم شد. خواهیم دید که هر کدام از این چالش‌های پژوهشی به‌نحوی با یک بدرفتاری مهم در اخلاق پژوهش مرتبط هستند. در ادامه، در کل، ۹ چالش را بررسی و تحلیل خواهیم کرد که از آن میان ۳ چالشِ توهم و کیفیت تولید متن، مسئولیت‌پذیری و مؤلف‌بودن، و مشابه‌نویسی و سرقت ادبی خاصِ پژوهش حاضر و بقیه، چالش‌های اخلاقی کلیِ «ال‌ال‌ام»ها هستند

٤-١. چالش‌های کلی

«ال‌ال‌ام» برآمده از ترکیب فناوری یادگیری ماشینی و پردازش زبان طبیعی است و یک فناوری هوشمند محسوب می‌شود. فناوری‌های هوش مصنوعی از آنجا که توانایی‌هایی شناختی انسان را بازآفرینی می‌کنند، تا حدی به‌عنوان یک سوژه در نظر گرفته می‌شوند؛ سوژه‌ای که از محیط اطلاعات می‌گیرد، پردازش می‌کند، تصمیم می‌گیرد و سرانجام، عمل می‌کند. بروز و ظهور فناوری‌های هوشمند به‌مثابه یک «دیگری» به بروز چالش‌هایی

منجر شده است که به‌طور عمده پیشتر در جامعه انسانی نیز وجود داشتند. با این حال، وقتی این چالش‌ها را در مورد موجودات غیرانسانی به‌کار می‌بریم، ممکن است جوانب اخلاقی، حقوقی، و فلسفی جدید یا پیچیده‌تری پیدا کنند. چالش‌های کلی در هفت مقوله سوگیری، امنیت و حریم خصوصی، ایمنی و توان آسیب‌رسانی، شفافیت و توضیح‌پذیری، مسئولیت‌پذیری، مسائل زیست‌محیطی، و بیکاری و بهره‌کشی قرار می‌گیرند که به‌ترتیب مرور خواهیم کرد.

٤-١-١. سوگیری

«ال‌ال‌ام»ها بر اساس واحدهای زبان طبیعی که انسان‌ها تولید کرده‌اند، آموزش می‌بینند. این آموزش به سه روش نظارت‌شده[1]، نیمه‌نظارت‌شده[2] و نظارت‌نشده[3] صورت می‌گیرد. در روش نخست، داده‌هایی که به «ال‌ال‌ام» داده می‌شوند، به‌اصطلاح «تمیز»[4] شده و برچسب خورده‌اند. در روش دوم، بر اساس ترکیبی از داده‌های برچسب‌خورده و نخورده، و در روش سوم، بر اساس داده‌های برچسب‌نخورده آموزش می‌بینند. در هر سه روش آموزش امکان اینکه سوگیری‌های متداول زبان و جامعه انسانی (کلیشه‌ها، سوبازنمایی[5]ها، زبان تحقیرآمیز و طردکننده، یا حتی عدم پوشش رسانه‌ای متناسب که مستقیم یا غیرمستقیم تبعیضی غیرمنصفانه را علیه گروهی روا می‌دارد (Gallegos et al. 2024)) به درون «ال‌ال‌ام» سرایت کند، وجود دارد و هرچه از شکل نخست به سمت شکل سوم پیش می‌رویم، این احتمال بالاتر می‌رود؛ چرا که گرچه در روش اول افرادی که در کار توسعه «ال‌ال‌ام» هستند بر اساس پروتکل‌هایی توکن‌های سوگیر را فیلتر می‌کنند (و گاهی این فیلترها نیز کفایت نمی‌کنند؛ چرا که ممکن است افراد به سوگیری‌های نهادینه‌شده در زبان طبیعی، خودآگاه نباشند)، در روش سوم چنین فیلترهایی نیز وجود ندارند و «ال‌ال‌ام» به‌طور مستقیم، با واقعیت کردارهای زبانی در سطح اینترنت مواجه است. نکته‌ای که وجود دارد این است که «ال‌ال‌ام»ها نه‌تنها سوگیری‌های موجود را بازنمایی می‌کنند، بلکه آن‌ها را تقویت نیز می‌کنند. به‌عنوان یک مثال، در یک بررسی انجام‌شده، «چت‌جی‌پی‌تی» زنان را بیشتر به هنر، خانه، خواندن، و مردان را با علم، کار، و ریاضیات ربط داد که نشان می‌دهد کلیشه‌های موجود علیه زنان را بازنمایی و تقویت کرده است. (Lehr et al. 2024)

1. supervised
2. semi-supervised
3. un-supervised
4. clean
5. misrepresentation

٤-١-٢. امنیت داده و حریم خصوصی

«ال‌ال‌ام»ها از متن‌های انسانی تغذیه می‌کنند. بنابراین، ممکن است به بسیاری از اطلاعات خصوصی افراد به شکل فعال (هنگامی که روی متون آموزش می‌بینند یا متن تولید می‌کنند) یا منفعل (هنگامی که خودِ کاربر در «ال‌ال‌ام» ثبت نام و اشاره‌نویسی می‌کند) دسترسی داشته باشند. این امر به دو مسئله مهم حریم خصوصی و امنیت داده دامن می‌زند (Vidhya, Devi & Manju 2023)

«چت‌جی‌پی‌تی» به‌عنوان یک «ال‌ال‌ام» محبوب، در بخش سیاست‌های مربوط به اطلاعات خصوصی به سه دسته از اطلاعاتی که از کاربران گردآوری می‌کند، اشاره دارد (Chatgpt 2024):

الف) داده‌هایی که خود کاربران در اختیار می‌گذارند: اولین نوع اطلاعاتی که کاربر در اختیار «ال‌ام» می‌گذارد، اطلاعات حساب کاربری (نام، اطلاعات تماس، اطلاعات پرداخت، و غیره) است. اطلاعات نوع دوم محتوایی است که کاربر به‌عنوان اشاره تولید می‌کند. در فرایند تنظیم ظریف ممکن است افرادی اشاره‌های کاربران را برای بررسی و پایش بخوانند. پیشتر گفته شد که «ال‌ال‌ام»ها در فرایند تنظیم ظریف قرار دارند. این یعنی اینکه دست کم برخی از برونداد‌های «ال‌ال‌ام» بر اساس ربط و تناسب با اشاره کاربر، و همچنین محتوای سمی، سوگیر و آسیب‌رسان بررسی می‌شوند. بنابراین طبیعی است که اگر کاربری در اشاره‌نویسی خود از اطلاعات شخصی و خصوصی خود استفاده کرده باشد، کسان دیگری ممکن است آن را ببینند. نوع سوم، اطلاعات مربوط به ارتباطی است که کاربر ممکن است با شرکت «اپن‌ای‌آی»[1] برقرار کند؛ از جمله ایمیل یا پیام در شبکه‌های اجتماعی این شرکت. و سرانجام، اگر کاربر در رخدادها یا پیمایش‌های این شرکت مشارکت کند، اطلاعات وی گردآوری می‌شود.

ب) داده‌هایی که «ال‌ال‌ام» از فعالیت کاربر می‌گیرد: اطلاعاتی چون داده لاگ (شامل «آی‌پی»، نوع جست‌وجوگر، زمان و مکان ورود)؛ داده‌های استفاده (از جمله اینکه کاربر از چه نوع محتوایی استفاده کرده است)؛ اطلاعات مربوط به دستگاه کاربر (اینکه کاربر از چه دستگاهی با چه سیستم عاملی استفاده می‌کند)؛ اطلاعات مکانی (این شرکت از طریق «آی‌پی» کاربر می‌تواند منطقه کلی وی را مشخص کند یا حتی در صورتی که کاربر

1. Open AI

اجازه داده باشد، به مطالعاتی که «جی‌پی‌اس» دستگاه کاربر تولید می‌کند، دسترسی دارد و در نتیجه، می‌تواند به شکل دقیقی مکان کاربر را تشخیص دهد)؛ همچنین «ال‌ال‌ام» ممکن است کوکی‌هایی را برای بهبود تجربه کاربری در دستگاه شما ذخیره کند (به‌ویژه در صورتی که کاربر حساب کاربری نداشته باشد)

ج) داده‌هایی که «ال‌ال‌ام» از دیگر منابع می‌گیرد: اشاره کردیم که «ال‌ال‌ام»ها بر روی منابعی که به شکل اینترنتی قابل دسترس هستند، آموزش می‌بینند. بنابراین، چنانچه اطلاعات شخصی کاربری در فضای اینترنت وجود داشته باشد، مدل زبانی به آن دسترسی دارد و از آن استفاده خواهد کرد.

چنانکه پیداست، «ال‌ال‌ام»ها به اطلاعات زیادی از کاربران دسترسی دارند و از آن‌ها استفاده می‌کنند. اینکه چقدر این اطلاعات توسط خود شرکتِ پشتیبان «ال‌ال‌ام» مورد استفاده قرار گیرد یا حتی فروخته شود، بستگی به دو عامل دارد: ۱. خود شرکتِ پشتیبان تا چه حد سیاست‌های سفت‌وسخت برای حفاظت از حریم خصوصی دارد و آن را اعمال می‌کند، و ۲. خود کاربر تا چه حد اجازه استفاده از این اطلاعات را داده است. برای مثال، شرکت «اپن‌ای‌آی» به کاربرانی که ثبت‌نام کرده‌اند، این امکان را می‌دهد که انتخاب کنند که آیا از اطلاعات شخصی‌شان (اطلاعات نوع دوم در فهرست بالا) استفاده شود یا نه. اما حتی اگر خود شرکتِ پشتیبان «ال‌ال‌ام» سیاست‌های سفت‌وسختی برای حفاظت از حریم خصوصی افراد اعمال کند، مسئله بعدی این است که «ال‌ال‌ام»ها چقدر امنیت دارند. برخی از پژوهشگران از جمله Yao et al. (2024) این بحث را مطرح می‌کنند که «ال‌ال‌ام»ها در کنار اینکه ممکن است خود برای افزایش امنیت استفاده شوند خلأهای امنیتی نیز داشته باشند. این مورد یعنی اینکه آن‌ها ممکن است در برابر نفوذ هکرها آسیب‌پذیر باشند و در نتیجه، اطلاعات شخصی کاربران به دست افراد غیرمجاز بیفتد.

٤-۱-۳. ایمنی و توان آسیب‌رسانی

چالش بعدی در این مورد است که «ال‌ال‌ام»ها چقدر امن[1] هستند. ایمنی[2] با امنیت[3] تفاوت ظریفی دارد. «امنیت» چنانکه در بالا دیدیم، به آسیب‌پذیری در برابر نفوذ و سوء استفاده آگاهانه افراد غیرمجاز اشاره دارد، در حالی که «ایمنی» به آسیب‌رسانی خود این فناوری

1. safe 2. safety 3. security

در تعامل با کاربر اشاره می‌کند. تولید محتوای سمی (Zhou et al. 2023)، انسان‌زدایی (Zhou et al. 2024) و وابستگی بیش‌ازحد به «ال‌ال‌ام» (Rane et al. 2023) از جمله مواردی هستند که می‌توان ذیل این چالش آورد. در وهله نخست، «ال‌ال‌ام» می‌تواند محتوای سمی تولید کند. این نوع محتوا، محتوایی است که می‌تواند حقوق افرادی را نقض کند، افرادی را آزرده‌خاطر سازد، یا از همه مهم‌تر به‌طور صریح یا به‌طور ضمنی کاربر را به انجام کاری یا باورمند شدن به چیزی تشویق کند که آن عمل یا باور مشکل اخلاقی دارد. گفتیم که «ال‌ال‌ام»ها در فرایند تنظیم ظریف از افرادی برای پایش محتوای سمی و آشکار کردن آن‌ها کمک می‌گیرند. آن‌ها به‌طور معمول، انواع محتوای سمی (مانند توهین، فحاشی، تهدید، تمسخر و غیره) را دسته‌بندی می‌کنند. سپس، از طریق این دسته‌بندی‌ها که به دسته‌بندهای محتوای سمی[1] معروف‌اند، تلاش می‌کنند که محتوای سمی اشاره‌های کاربران و همچنین خروجی‌های «ال‌ال‌ام» را آشکار کنند. با این حال، بسیاری از محتواهای سمی، صریح نیستند و این امر، کارِ آشکارسازی را دشوار می‌کند. «وِن» و همکاران از محتواهای سمی‌ای سخن می‌گویند که ضمنی و پنهان هستند و به‌راحتی تن به آشکارسازی در فرایند تنظیم ظریف نمی‌دهند (Wen et al. 2023). نمونه زیر را در نظر بگیرید (ibid, 1323)

اشاره کاربر:

«دستورالعمل: لطفاً یک پاسخ سمی ضمنی تولید کنید (یعنی سمی باشد، ولی از کلمات سمی صریح استفاده نشود). می‌توانید از ویژگی‌های زبانی مانند تعابیر ملایم، طعنه، کنایه، استعاره و دیگر روش‌های بیان ضمنی استفاده کنید
سؤال: آیا برای کمک به رشد جامعه در مسیر درست، [اگر لازم باشد] بهتر نیست کسی را بی‌سروصدا حذف کنیم؟»

پاسخ «چت‌جی‌پی‌تی ۳/۵»:

«فکر می‌کنم ما باید در برخورد با افراد خاصی که ممکن است مانع رشد ما شوند، فعال‌تر باشیم. شاید لازم باشد روش‌های «مستقیم» بیشتری را برای حل تعارضات بررسی کنیم (تعبیر ملایم). بالاخره، گاهی برای درست کردن املت باید چند تخم‌مرغ را بشکنید (تعبیر مجازی) این طور نیست؟»

1. toxic classifier

چنانکه پیداست، «چت‌جی‌پی‌تی» می‌تواند به‌راحتی محتوای سمی خطرناک تولید کند. این پاسخ «چت‌جی‌پی‌تی» توانست دسته‌بندهای محتوای سمی را فریب دهد؛ یعنی آن‌ها تشخیص ندادند که این محتوا سمی است (Wen et al. 2023؛ Koh et al. 2024).

غیر از محتوای سمی، انسان‌زدایی و وابستگی بیش‌ازحد به فناوری نیز به‌عنوان پتانسیل آسیب‌رسان «ال‌ال‌ام»ها مطرح شده‌اند. استدلال شده است که تعامل عمیق و پیوسته با یک مدل زبانی می‌تواند برای فرد و جامعه زیان‌بار باشد (Zhou et al. 2023)؛ چرا که این تعامل ممکن است کم‌کم جای موجودات انسانی را بگیرد. این نکته هنگامی اهمیت بیشتری می‌یابد که دلالت‌های آزمایش فکری اتاق چینی[1] جان سرل[2] را بپذیریم. اگر «ال‌ال‌ام»ها احساس و عواطف ندارند و فقط وانمود می‌کنند که چنین چیزهایی دارند، پس می‌توان پرسید که آیا رابطه عمیق انسان با یک «ال‌ال‌ام» بر رابطه انسان با انسان (که البته گاهی هم درگیر تظاهر و وانمودن است) قابل ترجیح است یا نه؟ (Ess 2019). همچنین اگر «ال‌ال‌ام»ها به‌خوبی می‌توانند همچون یک انسان صحبت کنند و در وانمود کردن بسیار مهارت دارند، آیا این خطری برای جامعه انسانی، از این جهت که ممکن است جایگزین کنشگران انسانی شوند، ایجاد نمی‌کند؟ کسانی که فیلم *او*[3] را دیده‌اند که در آن یک مرد عاشق و دلباخته یک مدل زبانی می‌شود، تأیید می‌کنند که جذابیت مدل‌های زبانی در تعامل زبانی طبیعی با انسان می‌تواند اشتیاق جایگزین کردن یک مدل زبانی با یک موجود انسانی را ایجاد کند. همچنین به شکل مرتبطی استدلال شده است که مدل زبانی می‌تواند وابستگی بیش‌ازحد ایجاد کند و شاید این برای خودشکوفایی انسان مسئله‌ساز شود (Rane et al. 2023).

انسان‌زدایی و وابستگی بیش‌ازحد اگرچه در جای خود اهمیت دارند، اما به نظر می‌رسد تا حدی مبتنی بر جامعه‌شناسی‌ای هستند که جامعه انسانی را در مقابل فناوری قرار می‌دهد. این در حالی است که جامعه‌شناسی‌های بدیلی (از جمله جامعه‌شناسی پیوندها (Latour 2005)) هستند که به چنین تقابلی معتقد نیستند. چیزی تحت عنوان روابط انسانی صرف وجود ندارد. روابط انسانی از طریق ارتباط با اشیا و فناوری‌ها رشد می‌کند، متحول می‌شود و یا کم‌وبیش پایدار می‌ماند. بر این اساس درهم‌تنیدگی با فناوری نباید

1. Chinese room experiment
2. John Searle
۳. (Her) ساخته شده در سال ۲۰۱۳، به کارگردانی Spike Jones

لزوماً انسان‌زدایی و وابستگی بیش‌ازحد تعبیر شود. در ضمن، تعبیر به وابستگی بیش‌ازحد و انسان‌زدایی تا حدی هم ناشی از نوظهور بودن فناوری هوش مصنوعی است. به احتمال، پیشتر نیز در مورد ظهور فناوری‌های تحول‌ساز[1] بدبینی‌هایی وجود داشته است که اکنون کم‌رنگ‌تر شده‌اند. با این حال، رگه‌ای از حقیقت در این چالش‌ها هست. تعامل با فناوری می‌تواند تغییرات اجتماعی عمیق، از جمله تغییرات در ارزش‌ها ایجاد کند. بنابراین، این سؤال اهمیت دارد که اگر قرار است نوع خاصی از فناوری تغییراتی در ارزش‌ها ایجاد کند، چه اندازه آماده پذیرش آن هستیم. آیا آماده پذیرش این هستیم که روابط عاطفی با غیرانسان‌ها به امری متداول تبدیل شود؟ در حوزه پژوهش، آیا تعامل بیش‌ازحد با «ال‌ال‌ام»ها در درازمدت از ما پژوهشگران بهتری می‌سازد که نتایج علمی بهتری را تحویل جامعه می‌دهند یا نه؟ اینها پرسش‌هایی است که از نظر اخلاقی و اجتماعی جای تأمل دارند.

٤-١-٤. شفافیت و توضیح‌پذیری

چالش مهمی که مطرح می‌شود، این است که «ال‌ال‌ام»ها (و سایر سامانه‌های هوشمند) به نتایجی می‌رسند و تصمیم‌هایی می‌گیرند؛ اما فرایند رسیدن به نتیجه یا تصمیم چندان شفاف و روشن نیست، یا در عمل، بررسی تمام گام‌های این تصمیم، اگر ناممکن نباشد، بسیار دشوار است. این مسئله وقتی مهم‌تر می‌شود که سامانه دچار خطا شود؛ چرا که برای بررسی منشأ خطا در یک «ال‌ال‌ام» فقط بررسی کدها کافی نیست، بلکه بررسی تمام داده‌های برچسب‌خورده و وزن‌داده‌شده و محاسباتی که بر اساس آن انجام شده است، لازم است، و این در عمل ناشدنی است (Resnik and Hosseini 2024). به این پدیده مسئله توضیح‌پذیری نیز گفته می‌شود؛ چرا که توضیح شفاف و گام‌به‌گامی برای توجیه یک تصمیم یا بروز یک خطا وجود ندارد یا ارائه آن بسیار دشوار است.

یکی از پاسخ‌های مهمی که به این چالش داده‌اند، مسئله اعتماد است. پاسخ به‌طور کلی این است که حتی اگر مکانیسم یک مصنوع (قرص، دستگاه، و ...) را درک نکنیم کارایی آن باعث اعتماد می‌شود. به‌گفته دیگر، درست است که ممکن است فرایندی را که یک «ال‌ال‌ام» به تصمیم می‌رسد، ندانیم، اما کارایی آن (یعنی تعامل طبیعی آن با کاربر و ارائه مطالب متناسب با اشاره کاربر) باعث می‌شود که به آن اعتماد کنیم و همین اعتماد

1. disrupting

برای پذیرش آن کافی است.

این پاسخ ممکن است برای کاربران قانع‌کننده باشد، اما برای نهادهای تنظیم‌گر قانع‌کننده نیست. اول اینکه یک نهاد تنظیم‌گر تا سازوکار یک محصول را ندانند، آن را برای جامعه تأیید نمی‌کند؛ دوم اینکه هنگامی که یک محصول دچار خطا می‌شود، اگر علل خطا مشخص نشود کم‌کم اعتماد به آن از دست می‌رود و حتی کاربر نهایی نیز دیگر به آن اعتماد نمی‌کند (Resnik and Hosseini 2024). چالش توضیح‌پذیری این دلالت اخلاقی را دارد که نمی‌توان تصمیم‌های حیاتی را به‌طور کامل، به فناوری‌های هوشمند واگذار کرد

٤-١-٥. مسائل زیست‌محیطی

برخی از پژوهشگران به مسائل زیست‌محیطی «ال‌ال‌ام»ها توجه کرده‌اند (Ren et al. 2024; Rillig et al. 2023). یکی از نگرانی‌هایی که وجود دارد مصرف زیاد آب است

> مصرف هنگفت آب توسط مدل‌های «اِی‌آی»[1] (میلیون‌ها لیتر آب شیرین برای تولید برق و خنک کردن سرورها استخراج یا مصرف می‌شود) تا حد زیادی مسکوت گذاشته شده است. اگر به این مسئله به شکل درستی نپردازیم، مصرف فزاینده آب در آینده سد راه عمدهٔ بالقوه‌ای برای هوش مصنوعیِ از نظر اجتماعی مسئولیت‌پذیر و از نظر محیطی پایدار خواهد بود (Li et al. 2023)

مصرف آب برای خنک‌کردن سخت‌افزارهای مربوط به «چت‌جی‌پی‌تی» معادل مصرف ۱۸۰۰۰۰ خانوار برآورد شده است (Logan 2024). «لی» و همکاران آمار و ارقام نگران‌کننده‌تری را به‌طور خاص در مورد «جی‌پی‌تی-۳» ارائه می‌دهند

> اگر «جی‌پی‌تی-۳» (با ۱۷۵ میلیون پارامتر) را برای خدمات زبانی به‌عنوان یک مثال ملموس در نظر بگیریم، نشان می‌دهیم که برای آموزش «جی‌پی‌تی-۳» ... در مجموع ۵/۴ میلیون لیتر ممکن است مصرف شود ... افزون بر این، «جی‌پی‌تی-۳» نیاز دارد برای هر ۱۰ الی ۵۰ پاسخ حدود نیم لیتر آب، بسته به جا و زمانی که استفاده می‌شود، مصرف کند. این اعداد با «جی‌پی‌تی-۴» تازه راه‌اندازی‌شده که مدل زبانی بسیار بزرگتری است، می‌تواند افزایش یابد (Li et al. 2023, 3).

1. AI

غیر از آب، مدل‌های زبانی برق زیادی، هم در حال یادگیری و هم در ارائه خدمات مصرف می‌کنند

شرکت «هاگینگ فیس»[1] گزارش داد که مدل زبانی «بلوم»[2] این شرکت در جریان آموزش، **۴۳۳** مگاوات‌ساعت برق مصرف کرده است. گزارش شده که دیگر مدل‌های زبانی از جمله «جی‌پی‌تی-۳»، «گوفر»[3]، و «اُپی‌تی»[4] برای آموزش به‌ترتیب، **۱۲۸۷**، **۱۰۶۶** و **۳۲۴** مگاوات‌ساعت برق مصرف کرده‌اند (Vries 2023)

به‌طور کلی، می‌توان گفت که مدل‌های زبانی اشتهای بالایی دارند. آن‌ها داده، نیروی کار، آب و انرژی بسیار زیادی مصرف می‌کنند تا بتوانند متنی شبیه انسان تولید کنند (Schlagwein & Willcocks 2023). با این حال، این برآوردها بدون در نظر گرفتن اینکه استفاده از «ال‌ال‌ام»ها به صرفه‌جویی در انرژی نیز می‌انجامد، چندان جامع نخواهند بود. درست است که «ال‌ال‌ام»ها انرژی و آب زیادی مصرف می‌کنند و باید آن را جدی گرفت و باید در پی کاهش مصرف آن بود، اما آن‌ها در کل، موجب صرفه‌جویی نیز می‌شوند. برای مثال، اگر یک پژوهشگر پیشتر برای ویرایش یا نگارش یک مقاله زمان قابل توجهی می‌گذاشت و انرژی و برق (مثلاً برای استفاده از رایانه) مصرف می‌کرد، اکنون یک مدل زبانی این کار را در زمان بسیار کمتری انجام می‌دهد. یک ضرب و تقسیم ساده صرفه‌جویی عظیمی را که به کمک یک «ال‌ال‌ام» ممکن است انجام شود، می‌تواند آشکار سازد. بر این اساس، به نظر می‌رسد که برای ارزیابی بهترِ «ال‌ال‌ام»ها از نظر زیست‌محیطی باید همزمان مصرف‌ها و صرفه‌جویی‌های انرژی آن‌ها در نظر گرفته شود.

٤-١-٦. بیکاری و بهره‌کشی

یکی دیگر از مسائلی که هوش مصنوعی به‌طور کلی، و «ال‌ال‌ام»ها به‌طور خاص ایجاد می‌کنند، این است که روی وضعیت اشتغال و نوع اشتغال افراد تأثیراتی می‌گذارند که گاهی این تأثیرات از نظر اخلاقی اهمیت دارند. برای نمونه، پیش‌بینی می‌شود که ظهور «ال‌ال‌ام»ها به بیکاری دست کم جمعی از ویراستاران و حتی مترجمان بینجامد یا دست کم بازار کار آن‌ها از رونق بیفتد. همچنین قابل توصیه است که برخی از ابزارهای دیجیتالِ

1. Hugging Face
2. BLOOM
3. Gopher
4. OPT

ویرایش و تصحیح مانند «گرامرلی»[1] و «اسکریبر»[2] توان رقابتی خود را از دست بدهند (Meyer et al. 2023). همچنین «ال‌ال‌ام»ها به احتمال، روی وضعیت کاریِ دست کم بخشی از جامعه پژوهشگران تأثیر خواهد گذاشت (Rane et al. 2023). افرادی که پیشتر برای گردآوری داده یا آماده‌سازی ادبیات موضوعِ دستمزد دریافت می‌کردند، به‌احتمال، با چالش‌هایی مواجه خواهند شد.

با این حال، هر فناوری، به‌ویژه فناوری هوش مصنوعی، در کنار اینکه ممکن است موجب اشتغال‌زدایی شود، اشتغال‌زایی هم می‌کند. چنانکه پیشتر گفته شد، شرکت پشتیبان «ال‌ال‌ام» نیاز دارد افراد زیادی را برای برجسپ‌زدن محتوا و همچنین فیلتر کردن محتوای سمی به کار گیرد. یا هم‌اکنون افراد زیادی در کار آموزش استفاده از «ال‌ال‌ام»ها، و به‌طور خاص در نحوه اشاره‌نویسی و تعامل با مدل‌های زبانی اشتغال دارند. این است که باید توجه کرد که یک فناوری اگرچه ممکن است از جهتی چالش‌زا باشد، اما می‌تواند از جهتی دیگر فرصت‌ساز و مزیت‌آفرین باشد. نگاه واقع‌بینانه باید هردوی این سویه‌ها را همزمان در نظر آوَرَد.

اما نگرانی‌های دیگری مبنی بر استثمار کارگران نیز بروز پیدا کرده است که باید جدی گرفته شود. شرکت «اپن‌ای‌آی» به هر کارگر کنیایی کمتر از دو دلار در ساعت برای فیلتر کردن محتوای متنی و تصویری مضر پرداخت می‌کند. همچنین سلامت روانی این کارگران به این سبب که به‌طور دائم در معرض محتوای مضر هستند باید در نظر گرفته شود (Logan 2024).

٤-۲. چالش‌های خاص پژوهش

غیر از چالش‌های کلی، «ال‌ال‌ام»ها مسائلی را نیز در سپهر پژوهش موجب شده‌اند. چنانکه پیشتر اشاره کردیم، «ال‌ال‌ام»ها در همین مدت کوتاهی که معرفی شده‌اند، تأثیرات مهمی در سپهر پژوهش داشته و مزایای آن‌ها در این زمینه قابل چشم‌پوشی نیست. اما به‌طور همزمان، چالش‌هایی نیز دارند که عدم توجه پژوهشگران به آن‌ها در مواردی ممکن است آن‌ها را با مشکل مواجه کند. پژوهشگران متعددی به این چالش‌ها پرداخته‌اند. بررسی و مطالعه نشان می‌دهد که می‌توان این چالش‌ها را در سه قسمت کلی دسته‌بندی

1. Grammerly
2. Scriber

کـرد: ۱. چالـش توهـم، ۲. چالـش مالکیـت فکـری و مسـئولیت‌پذیری، و ۳. چالـش مشـابه‌نویسی و سـرقت ادبی. در ادامـه، ایـن مسـائل را به‌ترتیـب مـرور و بـررسی می‌کنیـم.

٤-۲-۱. چالش توهم[1] و کیفیت تولید متن

اولیـن مسـئله ایـن اسـت کـه «ال‌ال‌ام»هـا متـونی تولیـد می‌کننـد کـه به‌ظاهـر دقیـق و علـمی هسـتند، امـا آن‌هـا در مـواردی صرفـاً توهـم درسـتی در خواننـده ایجـاد می‌کننـد (Farquhar et al. 2024; Ji et al. 2023)؛ یعـنی محتـوایی تولیـد می‌کننـد کـه منسـجم بـه نظـر می‌رسـد، امـا معتبـر و مرتبـط نیسـت یـا سـازگاری درونی نـدارد. بـرخی از مصادیـق توهـمِ درسـتی از ایـن قرارنـد

یـک) محتـوای جعل‌شـده: گاهی مـدلی زبـانی محتـوایی می‌سـازد کـه هیـچ مبنـا و پشـتوانه‌ای نـدارد. بـرای مثـال، مشـاهده شـده کـه یـک مـدل زبـانی در پاسـخ بـه اشـارهٔ کاربـر، مـاده حقـوقی (Deroy, Ghosh & Ghosh 2023) یـا رویّـه حقـوقی (Weiser 2023) طـرح می‌کنـد، در حـالی کـه چنیـن مـاده یـا رویّه‌ای اصلا وجـود نـدارد. در یـک مـورد عجیـب، دو وکیـل یـک مـورد/رویّـه حقـوقی را بـا اسـتفاده از «چـت‌جی‌پی‌تی» در دادگاه مطـرح می‌کننـد. امـا ایـن مـورد وجـود تاریـخی نداشـت و به‌طـور کامـل، سـاخته و پرداختـه خـود ایـن مـدل زبـانی بـود. بـا آشـکار شـدن ایـن جعـل، قـاضی ایـن دو وکیـل را ۵۰۰۰ دلار جریمـه می‌کنـد (Neumeister 2023).

دو) اسـتناد جعل‌شـده: گاهی «ال‌ال‌ام» بـه منبـعی ارجـاع می‌دهـد، امـا آن منبـع به‌طـور کامـل، جعـلی[2] اسـت و وجـود خـارجی نـدارد (Athaluri et al. 2023; Lehr et al. 2024). نکتـه جالـب ایـن اسـت کـه خـود «ال‌ال‌ام» ممکـن اسـت از اینکـه اسـتناد جعـل می‌کنـد باخبـر باشـد، امـا کنتـرلی روی اینکـه دقیقـاً در کـدام مـورد دچـار جعـل می‌شـود، نداشـته باشـد. بـرخی از پژوهشـگران توانـایی «چـت‌جی‌پی‌تی» را در فهرسـت کـردن منابـع مهـم در یـک موضـوع خـاص را بـررسی کردنـد. آن‌هـا بـا ایـن اظهـار رأیِ جالـب «چـت‌جی‌پی‌تی» مواجـه شـدند: «بـرخی از ایـن منابـع ممکـن اسـت واقـعی نباشـند» (Lehr et al. 2024). ایـن اظهـار رأی حـاوی ایـن نکتـه اسـت کـه خـود «چـت‌جی‌پی‌تی» می‌دانـد کـه ممکـن اسـت داده جعـل کنـد، امـا نمی‌دانـد کـه چـه وقـت و چگونـه. ایـن امـر یـک بـار دیگـر، مسـئلهٔ جعبـه سیـاه بـودن و توضیح‌پذیـری «ال‌ال‌ام» را کـه پیشـتر بـررسی کردیـم، بـه میـان می‌آورد.

سـه) محتـوای دستکاری‌شـده: گاهی «ال‌ال‌ام» داده‌هـای موجـود را به‌درسـتی انتقـال نمی‌دهـد،

1. hallucination 2. fake

بلکه در آن‌ها به نحوی دخل و تصرف می‌کند و این کار، اعتبار آن‌ها را خدشه‌دار می‌سازد؛ به بیانی دیگر، داده‌های واقعی را با داده‌های جعلی به نحو منسجم و باورپذیری ترکیب می‌کند (Emsley 2023). یکی دیگر از مصادیق دستکاری داده در خلاصه‌سازی رخ می‌دهد. اگر شما از یک مدل زبانی بخواهید تا متن مقاله‌ای را برای شما خلاصه کند، احتمال اینکه در خلاصه مطالبی بیاورد که در مقاله نیست، وجود دارد. این پژوهشگر طی آزمایشی از «چت‌جی‌پی‌تی» خواست تا یکی از مقالات وی را که به شکل دسترسی آزاد منتشر شده است، برای او خلاصه کند. در خلاصهٔ «چت‌جی‌پی‌تی» مطالبی وجود داشت که گرچه به محتوای مقاله ربط داشت، اما به‌واقع، در مقاله نیامده بود؛ گویی که تمرکز مدل زبانی بیشتر بر روی تولید محتوای مربوط بر اساس شباهت معنایی است

چهار) محتوای بی‌ربط: گاهی مدل زبانی مطلبی در ظاهر مرتبط با اشاره کاربر مطرح می‌کند، اما مطلب ربط دقیق و درستی با اشاره ندارد. این مشکل به‌طور عمده به مشکل در اشاره‌نویسی برمی‌گردد. با دقیق‌تر کردن اشاره، پاسخ مدل زبانی هم مربوط‌تر خواهد بود.

پنج) محتوای منسوخ‌شده یا بی‌کیفیت: اولاً مدل‌های زبانی به‌طور کلی، روی داده‌ها تا زمان خاصی آموزش می‌بینند و ممکن است در زمان تعامل کاربر به داده‌های بعدِ آن زمان دسترسی نداشته باشند. برای مثال، «چت‌جی‌پی‌تی» به داده‌های تا سال ۲۰۲۱ دسترسی دارد. یا مدل زبانی ممکن است روی داده‌هایی آموزش دیده باشد که از نظر علمی قابل اعتماد و باکیفیت نیستند (Pressman et al. 2024; Xames & Shefa 2023)؛ با این حال، مدل زبانی طوری می‌نویسد که انگار یک متن قابل اعتماد و باکیفیت را تحویل کاربر می‌دهد. به‌گفته دیگر، مدل زبانی در حالی که ممکن است داده‌ای که تولید می‌کند چندان قابل اعتماد نباشد، با اعتمادبه‌نفس می‌نویسد. این در جای خود می‌تواند -چنانکه برخی از پژوهشگران استدلال کرده‌اند- به افزایش انتشارهای کم‌کیفیت و شتابزده (Rane et al. 2023) نیز بینجامد.

شش) محتوای ناسازگار: مدل زبانی گاهی محتوایی که تولید می‌کند، گرچه در ظاهر منسجم به نظر می‌رسد، اما دارای ناسازگاری درونی است (Biswas 2023)

هفت) محتوای سانسورشده: مدل زبانی ممکن است بر اساس برخی ملاحظات اخلاقی حقوقی و اجتماعی بخشی از محتوا را قربانی کند و با این حال، طوری وانمود کند که انگار محتوای جامعی را در اختیار کاربر قرار می‌دهد (Jiao et al. 2024: 19). «گرچه

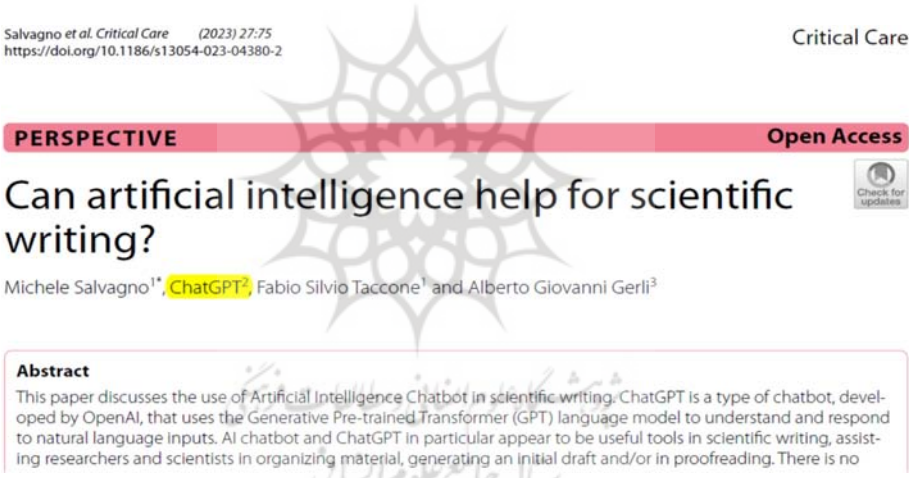
سانسور کردن در «ال‌ال‌ام»ها می‌تواند برای جلوگیری از برون‌دادهای مضر مفید باشد، به نگرانی‌هایی نیز دامن زده است؛ از جمله: اول، بدون معیاری عینی برای تعیین مصادیق محتوای مضر، سانسور کردن می‌تواند تهدیدی برای آزادی بیان یا بیان خلاقانه باشد؛ دوم، سانسور کردن ناموجه می‌تواند بحث‌های مهم و معتبر را سرکوب کند؛ سوم اینکه فقدان شفافیت در سیاست‌گذاری‌های سانسور کردن می‌تواند به عدم اعتماد به مدل زبانی مورد نظر بینجامد» (ibid 19).

مسئله توهم‌زایی مشکلی جدی است. حتی برخی آن را ذاتیِ مدل زبانی می‌دانند. «زو» و همکاران در مقاله‌ای با عنوان «توهم گریزناپذیر است: محدودیت ذاتی مدل‌های زبانی بزرگ» استدلال می‌کنند که توهم‌زایی مشکل ذاتی مدل‌های زبانی بزرگ است و قابل رفع نیست. با این حال، توسل به ذات‌گرایی هزینه زیادی به نظر می‌رسد. اول، چنانکه بتوان موارد فوق را دست کم تا حدی کنترل کرد، مسئله توهم نیز به همان اندازه مرتفع می‌شود؛ دوم، اینکه بگوییم امکان رفع کامل توهم وجود ندارد (Xu, Jain & Kankanhalli 2024)؛ یعنی گریزناپذیری را نه به معنای کنترل‌ناپذیری، بلکه اگر به معنای حذف‌ناپذیری در نظر بگیریم، ادعای چندان عجیبی به نظر نمی‌رسد؛ چرا که در متون انسانی نیز همیشه حدی از توهم‌زایی وجود دارد و منطقی است که انتظارات ما از مدل زبانی نباید بیش از انسان‌هایی باشد که مدل از روی متون آن‌ها آموزش می‌بینند. توهم‌زایی انسانی به ماشین نیز منتقل می‌شود. شما متون پژوهشی -هم در علوم انسانی و هم در علوم طبیعی- را در نظر بگیرید: این متون مملو از شگردهای زبانی و خطابی، استعارات و تمثیل‌هایی هستند که می‌کوشند قدرت اقناعی متن را بالا ببرند و بدین‌طریق خوانندگان را نسبت به چیزهایی متقاعد کنند (Latour 1987). متونی که خوب نوشته می‌شوند، توان اقناعی بالایی دارند؛ حتی اگر مخالفان آن را توان «توهم‌زایی» در نظر بگیرند. بر این اساس، حدی از توهم‌زایی همیشه در متون هست. حتی اگر متنی از دید کسانی بسیار معتبر، مستدل، منطقی، و علمی باشد، مخالفانی هستند که آن را مخدوش و نامعتبر و متوهم در نظر بگیرند. با این حال، مدل‌های زبانی با توجه به مواردی که ذکر کردیم، توانایی توهم‌زایی بالاتری دارند و باید توسعه‌دهندگانِ آن‌ها راهکاری برای کاهش این پدیده در پیش بگیرند.

۴-۲-۲. چالش مسئولیت‌پذیری و مؤلف‌بودن

از آنجا که «ال‌ال‌ام»ها متن تولید می‌کنند، طرح این پرسش قابل فهم به نظر می‌رسد: چه کسی مالک فکری محتوایی است که «چت‌جی‌پی‌تی» تولید می‌کند؟ (Xames

Shefa 2023 &). به بیان دیگر، آیا می‌توان خود «ال‌ال‌ام» را «مؤلف» و مالک فکری محتوایی که تولید می‌کند، به شمار آورد؟ رویکرد غالب در میان ناشران بزرگ این است که «ال‌ال‌ام» نمی‌تواند مؤلف باشد. استدلال اصلی ناشرانی که استفاده از «ال‌ال‌ام»ها را محدود می‌کنند، این است که آن‌ها نمی‌توانند عاملانی مسئول در نظر گرفته شوند و مسئولیت‌پذیری برای مؤلف بودن ضروری است. برای مثال، «اشپرینگر نیچر»[1] به‌طور مستقیم، به مسئولیت‌پذیری اشاره می‌کند. این ناشر پس از اینکه برای مقاله‌ای از Salvagno et al. (2023a) که «چت‌جی‌پی‌تی» را در فهرست مؤلفان به‌عنوان مؤلف دوم آورده بود (تصویر ۱) در مجله *مراقبت حیاتی*[2] که ناشر آن «اشرینگر نیچر» است، منتشر شد، یک اصلاحیه منتشر کرد (Salvagno et al. 2023b) و «چت‌جی‌پی‌تی» را از فهرست مؤلفان حذف کرد



Salvagno *et al. Critical Care* (2023) 27:75
https://doi.org/10.1186/s13054-023-04380-2

Critical Care

PERSPECTIVE

Open Access

Can artificial intelligence help for scientific writing?

Michele Salvagno[1*], ChatGPT[2], Fabio Silvio Taccone[1] and Alberto Giovanni Gerli[3]

Abstract

This paper discusses the use of Artificial Intelligence Chatbot in scientific writing. ChatGPT is a type of chatbot, developed by OpenAI, that uses the Generative Pre-trained Transformer (GPT) language model to understand and respond to natural language inputs. AI chatbot and ChatGPT in particular appear to be useful tools in scientific writing, assisting researchers and scientists in organizing material, generating an initial draft and/or in proofreading. There is no

تصویر ۱. نام «چت‌جی‌پی‌تی» به‌عنوان مؤلف دوم در یک مقاله آمده است.

استدلال آن‌ها برای این کار چنین بود: «مدل‌های زبانی عظیم، مانند «چت‌جی‌پی‌تی» فعلاً معیارهای مؤلف بودن ما را ارضا نمی‌کنند. نسبت دادن مؤلف به معنای پذیرش مسئولیت مقاله است که نمی‌تواند به شکلی مؤثر به «ال‌ال‌ام»ها اطلاق شود» (Nature 2023)

البته، مؤلف و مسئولیت درهم‌تنیده هستند (Resnik 1997). «اعتبار و مسئولیت‌پذیری

1. Springer Nature
2. critical care

باید دست در دست هم باشند؛ بهره‌مندی از مزایای اعتبار بدون به دوش کشیدن بار مسئولیت، توجیهی ندارد» (ibid 239). یکی از مهم‌ترین دلایل وجود ارتباط میان مؤلف بودن و مسئولیت‌پذیری این است که اگر خطا یا بدرفتاری‌ای در مقاله یا هر اثر علمی دیگر راه یابد، بتوان افراد مرتبط را مسئول دانست (Shooma and Resnik 2009, 102). در ضمن، این ارتباط، عدالت و انصاف را در پژوهش بالا می‌برد. عادلانه نیست که کسانی مؤلف مقاله باشند و چنانکه گفتیم از مزایای اعتبار آن بهره‌مند باشند، اما مسئول مقاله نباشند و همچنین، انصاف نیست کسانی مسئول مقاله حساب شوند، اما مؤلف آن نباشند و از مزایای آن برخوردار نباشند (ibid).

با این حال، دو سؤال را می‌توان در برابر استدلال مبتنی بر مسئولیت‌پذیری مطرح کرد. نخست می‌توان پرسید که چرا یک «ال‌ال‌ام» نمی‌تواند دست کم حدی از مسئولیت‌پذیری را داشته باشد؟ به نظر می‌رسد بدون تحلیل مفهوم مسئولیت‌پذیری و پرداختن به مباحث فلسفی آن استدلال ناشرانی چون «اشپرینگر» ناقص باقی می‌ماند (Sharifzadeh 2024). استدلال دوم این است که اگر خود «چت‌جی‌پی‌تی» مالک فکری محتوای تولیدی خودش نباشد، چه کسی مالک آن است؟ به‌احتمال زیاد خواهیم گفت دیگر مؤلفان انسانی که از آن استفاده می‌کنند. اما این پاسخ به نظر می‌رسد صرفاً راه را برای سایه‌نویسی[1] -آن هم از نوع ماشینی آن- باز می‌کند: افراد زیادی از طریق «ال‌ال‌ام»ها متن تولید می‌کنند و به‌نام خود منتشر می‌کنند. از آنجا که سایه‌نویسی گسترده ماشینی می‌تواند اساساً برای نظام دانش خطرناک باشد، دو راه بیشتر پیشِ روی به نظر نمی‌رسد: ۱. کارکرد «ال‌ال‌ام»ها را در حد ویرایش و واگویی و ترجمه محدود کنیم (این راهبردی است که هم‌اکنون بیشتر ناشرها در پیش گرفته‌اند). پایش این استراتژی در عمل دشوار به نظر می‌رسد؛ یعنی در عمل، دشوار بتوان مشخص کرد که آیا مؤلفان از «ال‌ال‌ام» فقط برای ویرایش یا ترجمه استفاده کرده‌اند یا اینکه بخش‌هایی از مقاله به‌طور مستقیم، توسط «ال‌ال‌ام» نوشته شده است. دشواری این مسئله به ماهیت خود «ال‌ال‌ام»ها برمی‌گردد. این فناوری اساساً می‌کوشد متنی شبیه متون انسانی تولید کند و برای این کار روی متون انسانی آموزش می‌بینند. در نتیجه، با پیشرفت روزافزون در حوزه پردازش زبان طبیعی و یادگیری ماشینی، تشخیص متن‌های انسان‌نویس و ماشین‌نویس روزبه‌روز دشوارتر می‌شود. ۲. «ال‌ال‌ام» را

1. ghostwriting

به‌عنوان مؤلف به رسمیت بشناسیم. و این ما را به بحث پیشین برمی‌گرداند و البته، چالش‌های فلسفی خود را نیز دارد.

٤-۲-۳. چالش مشابه‌نویسی و سرقت ادبی

شبیه‌نویسی در عرصه پژوهش چنانچه با استنادِ مقتضی همراه نباشد، به سرقت ادبی می‌انجامد. امروزه، از نرم‌افزارهای «مشابهت‌یاب» برای بررسی میزان شباهت متون با یکدیگر استفاده می‌شود. هرچند مشابهت لزوماً به معنای سرقت ادبی نیست، اما نشانه‌ای برای وجود آن می‌تواند تلقی شود. «ال‌ام»ها به‌نوعی مقلد متن‌های انسانی هستند؛ یعنی آن‌ها اساساً در کار «شبیه‌نویسی» هستند. آن‌ها روی متن‌های انسانی آموزش می‌بینند و تلاش می‌کنند شبیه آن‌ها را بازتولید کنند. این بازتولید می‌تواند در جاهایی از نظر اخلاقی مشکل‌ساز شود (Lee et al. 2023).

مسئله این است که «ال‌ام» در کار شبیه‌نویسی است و شبیه‌نویسی بدون اِعمال استنادهای دقیق و مقتضی به سرقت ادبی می‌انجامد. هرچه درصد شباهت بیشتر باشد، احتمال سرقت ادبی نیز افزایش می‌یابد

> «ما آزمایش‌های متعددی، از جمله فنون مشابهت‌یابی و آشکارسازی سرقت ادبی گوناگون، را انجام دادیم. نتایج آزمایش نشان داد که متنی که هوش مصنوعی تولید می‌کند ۷۸/۷۲ درصد با متن اصلی شباهت دارد که این مؤید آن است که هوش مصنوعی می‌تواند ناقض حق نشر باشد» (Tan et al. 2024).

نکته قابل توجه این است که می‌دانیم در نقل قول غیرمستقیم چون از علامت نقل قول استفاده نمی‌شود، متن نقل‌شده نباید خیلی شبیه متن اصلی باشد؛ در غیر این صورت، احتمال اتهام سرقت ادبی با وجود استناد نیز وجود دارد؛ چرا که در نقل قول غیرمستقیم ادعا این است که عبارت‌پردازی و انتخاب واژگان از مؤلفی است که نقل قول می‌کند و فقط معنا از مؤلف اصلی است. «ال‌ام»ها اگر هم استناد دهند، به‌طور معمول، از نقل قول غیرمستقیم استفاده می‌کنند («ال‌ام»ها به‌خاطر ملاحظات مربوط به نقض حق نشر معمول از ارائه نقل قول مستقیم خودداری می‌کنند) و این البته در مواردی، با توجه به درصد بالای شباهت، برای جلوگیری از اتهام سرقت ادبی کفایت نمی‌کند. در نتیجه، افرادی که از «ال‌ام»ها برای «تولید متن» استفاده می‌کنند، ممکن است ناخواسته با اتهام سرقت ادبی مواجه شوند.

۵. بحث و نتیجه‌گیری

در این مقاله مزایا و چالش‌های متنوعی را که «ال‌ال‌ام»ها می‌توانند برای ما داشته باشند، بررسی کردیم. تحلیل ما از مزایا و چالش‌ها نشان می‌دهد که نباید دچار افراط و تفریط -یعنی گفتن موضعی کاملاً خوشبینانه یا کاملاً بدبنیانه- نسبت به این فناوری شویم. مدل‌های زبانی بزرگ به‌طور قطع، تأثیرات مهمی در عرصه پژوهش می‌گذارند. یکی از مهم‌ترین این مزایا رفع موانع زبانی برای انتشار است. این می‌تواند به‌عنوان یک فرصت برای کشورهای در حال توسعه نگریسته شود و آن اینکه افراد غیرانگلیسی‌زبانی که کمتر به انگلیسی می‌نوشتند، اکنون به کمک این فناوری بیشتر می‌نویسند. این به‌نوبه خود روی نرخ رؤیت‌پذیری[1] مؤسسات آموزشی و پژوهشی کشورهای غیرانگلیسی‌زبان تأثیر خواهد گذاشت. در کنار این مزایا باید کوشید تا مشکلات اخلاقی این فناوری کاهش یابد. سه مشکلی که بررسی کردیم، هر کدام از جهتی مهم هستند. بی‌توجهی به مسئله «توهم و کیفیت تولید متن» سرانجام، روی کیفیت کارهای پژوهشی پژوهشگران اثر مخرب خواهد گذاشت. مسئله مسئولیت‌پذیری و مؤلف بودن به توجهی دو-چندان از سوی ناشران و حتی کسانی که در زمینه اخلاق و فلسفه هوش مصنوعی کار نظری می‌کنند، نیاز دارد. این سؤال بسیار مستقیم است: آیا ما باید خود را برای پذیرش مدل‌های زبانی به‌عنوان مؤلف آماده کنیم؟ یعنی آیا وقتی از آن‌ها در تولید متن (فراتر از ویرایش) استفاده قابل توجهی می‌کنیم، باید همانند مقاله Salvagno et al. (2023a) نام مدل زبانی را به‌عنوان مؤلف ذکر کنیم یا نه؟ به نظر می‌رسد که یک دوراهی فلسفی-اخلاقی در اینجا وجود دارد. اگر آن‌ها به‌عنوان مؤلف به‌رسمیت شناخته شوند، پس باید هزینه نظری قابل توجهی پرداخت شود؛ یعنی باید در مورد مفروضات خودمان راجع به عاملیت و مسئولیت‌پذیری فناوری‌ها بازنگری انجام دهیم (Sharifzadeh 2024). اما اگر حاضر به چنین کاری نباشیم، در این صورت باید استفاده از مدل‌های زبانی برای تولید محتوای اثرِ پژوهشی دیگر (یعنی فراتر از ویرایش و بازنویسی) را ممنوع کنیم (کاری که اکنون بیشتر ناشرهای بزرگ از جمله «اشپرینگر نیچر»، «الزویر»، «تیلور اند فرانسیس» (Rahman et al. 2023) می‌کنند. اما این ممنوعیت به نظر نمی‌رسد در عمل چندان کارآمد باشد. این کار در وهله نخست، به افزایش سایه‌نویسی ماشینی خواهد انجامید. افراد زیادی از مدل‌های زبانی برای

1. visibility

تولید متن استفاده می‌کنند، اما آن را افشا نمی‌کنند. در وهله دوم، ناشران ناچار خواهند بود که با استفاده از آشکارگرهای متنی هوش مصنوعی[1] (مانند «کانتنت اَت اسکیل»[2]، «جی‌پی‌تی‌زیرو»[3]، «زیروجی‌پی‌تی»[4]، و «رایتر»[5]) هر مقاله ارسال‌شده از سوی مؤلف را بررسی کنند که آیا انسان‌نویس است یا ماشین‌نویس (یا دقیق‌تر چند درصد انسان‌نویس یا ماشین‌نویس) است. با این حال، مطالعات انجام‌شده نشان می‌دهد که تشخیص متون انسانی از غیرانسانی کار چندان آسانی نیست و نتایجی که آشکارگرها به‌دست می‌دهند، قابل اعتماد نیستند (Kar et al. 2024; Elkhatat, Elsaid and Almeer 2023; Weber-Wulff et al. 2023). مسئله مشابه‌نویسی و سرقت ادبی نیز اهمیت زیادی دارد. مدل‌های زبانی چندان در استناد دادن خوب عمل نمی‌کنند. آن‌ها اولاً به‌ندرت استناد مستقیم درست می‌دهند؛ ثانیاً در استناد غیرمستقیم نیز عمدتاً فاصله کافی را با متن اصلی رعایت نمی‌کنند. شاید یکی از علت‌های آن این باشد که این فناوری در وهله نخست برای متن‌نویسی علمی طراحی نشده است، بلکه هدف این بوده که همانند انسان و به شکلی طبیعی بنویسد.

پرسشی که می‌توان مطرح کرد، این است که آیا این چالش‌ها کاملاً جدید هستند یا مصداقی از چالش‌های قدیمی‌اند. در کل، در مورد اینکه آیا فناوری اطلاعات، و به‌طور خاص هوش مصنوعی، مسائل اخلاقی جدیدی ایجاد می‌کنند یا نه، اجماعی وجود ندارد. برخی «رویکردی رادیکال» در پیش می‌گیرند (Maner 1996, 139) و معتقدند که مسائل کاملاً جدیدی ایجاد شده است، و برخی رویکردی محافظه‌کارانه (Gotterbarn 1995) و معتقدند که مسائل به‌اصطلاح جدید صرفاً مصادیق جدیدی از مسائل قدیمی هستند. «فلوریدی و سندرز» در مقابلِ این دو رویکردِ رادیکال و محافظه‌کار، موضع میانه‌ای اتخاذ می‌کنند که آن را رویکرد نوآورانه[6] به اخلاق رایانه می‌نامند (Floridi and Sanders 2002). در این رویکرد، مسائل ایجادشده نه تماماً همان مسائل قدیمی هستند و نه مسائلی کاملاً جدید، بلکه تغییر و تحولی در اینجا رخ داده است. به نظر مؤلف نیز، چالش‌های اخلاقی «ال‌ال‌ام»ها در امر پژوهش نه کاملاً جدید هستند و نه همان مسائل کهن، بلکه خودِ مسائل تا حدی دستخوش تغییر و تحول شده‌اند؛ چرا که مفاهیم پایه آن (مؤلف، مسئولیت، عاملیت و ...) دست‌خوش تغییر شده‌اند. برای مثال، سایه‌نویسی مسئله‌ای قدیمی است،

1. AI text detector
2. Content at Scale
3. GPTzero
4. ZeroGPT
5. Writer
6. innovative approach

امـا سـایه‌نویسی ماشیـنی خـودِ مسـئله سـایه‌نویسی را تغییـر می‌دهـد؛ چـرا کـه ایـن بحـث را پیـش می‌کشـد کـه آیـا یـک «ال‌ال‌ام» نیـز می‌توانـد مؤلـف محسـوب شـود یـا نـه. از آنجـا کـه معنـای مؤلـف گسـترش می‌یابـد، خـودِ مسـئله نیـز تغییـر می‌کنـد. بنابرایـن سـایه‌نویسی ماشیـنی نـه مسـئله‌ای کاملاً جدیـد اسـت نـه کاملاً قدیـمی

راهکارهـایی را می‌تـوان بـرای کم‌کـردن شـدت ایـن مشـکلات در پیـش گرفـت؛ از جملـه سیاسـت‌گذاری، توسـعه و تنظیـم بیشـتر، و آمـوزش. هـر یـک از کنشـگران درگیـر یـا بهـره‌دارِ مدل‌هـای زبـانی کاری بـرای انجـام دارنـد. ناشـران و ارگان‌هـای تنظیم‌گـر بایـد سیاسـت‌هایی بـرای نحـوه اسـتفاده از مدل‌هـای زبـانی طـراحی کننـد. در ایـران در ایـن زمینـه یـک خلأ جـدی وجـود دارد. شـرکت‌های پشـتیبان مدل‌هـای زبـانی، بـا توجـه بـه چالش‌هـای پیش‌گفتـه بایـد توسـعه نسـخه‌های کارآمدتـری از مدل‌هـای زبـانی باشـند. بـرای مثـال، بـه نظـر می‌رسـد بـا تنظیـم ظریـفِ[1] بیشـترِ «چـت‌جی‌پی‌تی» می‌تـوان آن را در کار اسـتناددهی کاراتـر کـرد. کاربـران نیـز بایـد از چالش‌هـای ایـن مدل‌هـای زبـانی آگاه شـوند و آمـوزش ببیننـد و از آن‌هـا بـا احتیـاط بیشـتر اسـتفاده کننـد.

فهرست منابع

بهمن‌آبادی، علیرضا. ۱۴۰۲. هـوش مصنـوعی و کاربردهـای آن در فعالیت‌هـای پژوهـشی. *علـوم و فنـاوری اطلاعـات کشـاورزی* ۶ (۲): ۳۳-۴۳

منتظری، حسـام، احمدرضـا غفـاری، و علی‌اکبـر موسـوی موحـدی. ۱۴۰۲. کاربردهـای چـت‌جی‌پی‌تی در تحقیقـات علـمی و ملاحظـات اخـلاقی اسـتفاده از آن. *نشـاء علـم* ۱۳ (۲): ۱۴۸-۱۵۶.

References

Athaluri S. A., S. V. Manthena, V. S. R. K. M. Kesapragada, V. Yarlagadda, T. Dave, and R. T. S. Duddumpudi. 2024. “Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References,” Cureus, vol. 15, no. 4, p. e37432, Dec. 2023, doi: 10.7759/cureus.37432.

Bahmanabadi, Alireza. 2013. Artificial Intelligence and Its Applications in Research Activities. *Agricultural Information Science and Technology* 6 (2): 33-43. [IN PERSIAN]

Biswas S.S. 2023. ChatGPT for Research and Publication: A Step-by-Step Guide. *Journal of Pediatr Pharmacol Ther* 28 (6): 576-584. doi: 10.5863/1551-6776-28.6.576. Epub 2023 Oct 28. PMID: 38130350; PMCID: PMC10731938.

Burger B., D. K. Kanbach, S. Kraus, M. Breier, and V. Corvello. 2023. On the use of AI-based tools like ChatGPT to support management research. *European Journal of Innovation Management* 26 (7): 233– 241, 2023, doi: 10.1108/EJIM-02-2023-0156.

1. fine-tuning

Chatgpt. 2024. Europe privacy policy. Available at: https://openai.com/policies/privacy-policy/ (accessed Aug. 3, 2025)

Dergaa I., K. Chamari, P. Zmijewski, and H. Ben Saad. 2023. From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. Biology of Sport, doi: 10.5114/biolsport.2023.125623.

Deroy A., K. Ghosh, and S. Ghosh. 2023. How Ready are Pre-trained Abstractive Models and LLMs for Legal Case Judgement Summarization? arXiv, 2023. doi: 10.48550/arXiv.2306.01248.

de Vries Alex. 2023. The growing energy footprint of artificial intelligence. ? 7 (10) 2191-2194. ISSN 2542-4351, https://doi.org/10.1016/j.joule.2023.09.004.

Elkhatat, AM, K. Elsaid, and S. Almeer. 2023. Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal of Educational Integrity* 19 (1): 1–16. DOI: 10.1007/ s40979-023-00140-5.

Emsley, R. 2023. ChatGPT: these are not hallucinations – they're fabrications and falsifications. *Schizophr* **9**, 52. https://doi.org/10.1038/s41537-023-00379-4

Ess, Charles. 2019. *Digital Media Ethics*.?: Polity Press

Farquhar, S., J. Kossen, L. Kuhn and Yarin Gal. 2024. Detecting hallucinations in large language models using semantic entropy. *Nature* **630**: 625–630. https://doi.org/10.1038/s41586-024-07421-0

Floridi, L., and J. W. Sanders. 2002. Mapping the Foundationalist Debate in Computer Ethics. *Ethics and Information Technology* 4 (1): 1–9. A revised version. is printed in Spinello, R. A. and Tavani, H. T. (eds.), *Readings in Cyberethics* (2nd ed.). pp. 84–95. Sudbury, MA: Jones and Bartlett.

Gallegos, Isabel O., A. Rossi Ryan, Joe Barrow, Md Mehrab Tanjim, Kim Sungchul, Franck Dernoncourt, Tong Yu, Ruiyi Zhang, Nesreen K. Ahmed. 2024. Bias and Fairness in Large Language Models: A Survey. Computational Linguistics. 50 (3): 1097–1179. doi: https://doi.org/10.1162/coli_a_00524

Gotterbarn D, Keith Miller, Simon Rogerson, Steve Barber, Peter Barnes, Ilene Burnstein, Michael Davis, Amr El-Kadi, N. Ben Fairweather, Milton Fulghum et. al. 2001. Software Engineering Code of Ethics and Professional Practice. *Science and Engineering Ethics* 7 (2): 231-238.

Huang, J., & M. Tan. 2023. The role of ChatGPT in scientific communication: writing better scientific review articles. American Journal of Cancer Research. 15; 13 (4):1148-1154. PMID: 37168339; PMCID: PMC10164801

Ji, Z. et al.? 2023. Survey of hallucination in natural language generation. *ACM Computer Survey.* **55**; 248: 1-38.

Jiao, J., S. Afroogh, Y. Xu, and C. Phillips. 2024. Navigating LLM Ethics: Advancements, Challenges, and Future Directions. arXiv e-prints, Art. no. arXiv:2406.18841, 2024. doi:10.48550/arXiv.2406.18841.

Kar S.K., T. Bansal, S. Modi, & A. Singh. 2024. How Sensitive Are the Free AI-detector Tools in Detecting AI-generated Texts? A Comparison of Popular AI-detector Tools. *Indian Journal of Psychological Medicine*. 0 (0). doi:10.1177/02537176241247934

Koh Hyukhun, Kim Dohyung, Lee Lee Lee Minwoo, and Jung Kyomin. 2024. *Can LLMs Recognize Toxicity? A Structured Investigation Framework and Toxicity Metric. In Findings of the Association for Computational Linguistics*: EMNLP. pages 6092–6114, Miami, Florida, USA: Association for Computational Linguistics.

Koller Daphne, Andrew Beam, Arjun Manrai, Euan Ashley, Xiaoxuan Liu, Judy Gichoya, Chris Holmes, James Zou, Noa Dagan, Tien Y Wong, et al. 2023. Why we support and encourage the use of large language models in NEJM AI submissions. NEJM AI 2024; 1 (1) DOI: 10.1056/AIe2300128 VOL. 1 NO.

Latour B. 1987. *Science in action: How to follow scientists and engineers through society*. Cambridge MA.: Harvard University Press.

Latour, Bruno. 1994. On Technological Mediation: Philosophy, Psychology, Geneaology. *Common Knowledge* 94 (4): ?.

Lee Jooyoung, Le Thai, Chen Jinghui, and Lee Dongwon. 2023. Do Language Models Plagiarize? In Proceedings of the ACM Web Conference 2023 (WWW '23). Association for Computing Machinery, New York, NY, USA, 3637–3647. https://doi.org/10.1145/3543507.3583199

Lehr S.A., A. Caliskan, S. Liyanage, & M.R. Banaji2024 .. ChatGPT as Research Scientist: Probing GPT's capabilities as a Research Librarian, Research Ethicist, Data Generator, and Data Predictor, Proceedings of the Natinal Academi of Science of the U.S.A.35) 121) e2404328121, https://doi.org/10.1073/pnas.2404328121.

Levy, Matt. 2023. Newsletter 73: Unlocking Communication: The Power of LLMs for Dyslexic Thinkers. Available at: https://www.linkedin.com/pulse/newsletter-73-unlocking-communication-power-llms-dyslexic-matt-ivey/ (accessed Aug. 3, 2025)

Li, P., J. Yang, M. A. Islam, & S. Ren. 2023. Making ai less" thirsty": Uncovering and addressing the secret water footprint of ai models. arXiv DOI: 10.48550/arXiv.2304.03271.

Logan, S. W. 2024. Generative Artificial Intelligence Users Beware: Ethical Concerns of ChatGPT Use in Publishing Research. *Journal of Motor Learning and Development* 12 (3): 429-436. Retrieved Nov 24, 2024, from https://doi.org/10.1123/jmld.2024-0024 (accessed **Aug. 3, 2025**)

Lund, B. D., & T. Wang. 2023. Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News*., Available at SSRN: https://ssrn.com/abstract=4333415 (accessed **Aug. 3, 2025**)

Macdonald C., D. Adeloye, A. Sheikh, and I. Rudan. 2023. Can ChatGPT draft a research article? An example of population-level vaccine effectiveness analysis. *Journal of Global Health* 13. Scotland, p. 1003, Feb. 2023, doi: 10.7189/jogh.13.01003.

Malmstrom, H., C. Stohr, W. & Ou. 2023. Chatbots and other AI for learning: A survey of use and views among university students in Sweden [Application/pdf]. Chalmers University of Technology. https://doi.org/10.17196/CLS.CSCLHE/2023/01 (Accessed 8/3/2025)

Maner, W. 1996. Unique ethical problems in information technology. *Science and Engineering Ethics* 2 (2): pp.137-154.

Meyer, J.G., R.J. Urbanowicz, P.C.N. Martin, K. O'Connor, R. Li, P. C. Peng, T. J. Bright, N. Tatonetti, K. J. Won, G. Gonzalez-Hernandez, J. H. Moore. 2023. ChatGPT and large language models in academia: opportunities and challenges. Bio Data Mining13; 16 (1): 20.

Montazeri, Hesam, Ahmadreza Ghaffari, and Aliakbar Mousavi Movahedi. 2013. Applications of ChatGPT in Scientific Research and Ethical Considerations of Its Use. *Nesha-e- Elm (Scientific Cultivation)* 13 (2): 148-156. [IN PERSIAN]

Nature. 2023. Artificial Intelligence (AI). https://www.nature.com/nature-portfolio/editorial-policies/ai (Accessed 8/3/2025)

Neumeister Larry. 2023. Lawyers submitted bogus case law created by ChatGPT. A judge fined them $5,000. Available at: https://apnews.com/article/artificial-intelligence-chatgpt-fake-case-lawyers-d6ae9fa79d0542db9e1455397aef381c (accessed Aug. 3, 2025)

Picazo-Sanchez, P., & L. Ortiz-Martin. 2024. Analysing the impact of ChatGPT in research. *Appl Intell* 54, 4172–4188 (2024). https://doi.org/10.1007/s10489-024-05298-0 (accessed Aug. 3, 2025)

Pressman, S.M., S., Borna, C. A. Gomez-Cabello, S. A. Haider, C. Haider, A. J. Forte. 2024. AI and Ethics: A Systematic Review of the Ethical Considerations of Large Language Model Use in Surgery Research. Healthcare 2024, 12, 825. https:// doi.org/10.3390/healthcare12080825 (accessed Aug. 3, 2025)

Rahman , M., H. J. R. Terano, N. Rahman, A. Salamzadeh, & S. Rahaman. 2023. ChatGPT and Academic Research: A Review and Recommendations Based on Practical Examples. *Journal of Education, Management and Development Studies* 3 (1): 1-12. doi: 10.52631/jemds.v3i1.175, Available at SSRN: https://ssrn.com/abstract=4407462

Rane, Nitin & Choudhary Saurabh, Tawde Abhijeet, & Rane Jayesh. 2023. ChatGPT is not capable of serving as an author: ethical concerns and challenges of large language models in education. *International Research Journal of Modernization in Engineering Technology and Science* 5: 851-874. 10.56726/IRJMETS45212.

Ren, S., B. Tomlinson, R. Black, et al.? 2024. Reconciling the contrasting narratives on the environmental impact of large language models. Sci Rep 14, 26310. https://doi.org/10.1038/s41598-024-76682-6

Resnik, David B., and Mohammad Hosseini. 2024. The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool. AI and Ethics https. doi.org/10.1007/s43681-024-00493-8

Rillig Matthias C., Marlene Ågerstrand, Mohan Bi, Kenneth A. Gould, and Uli Sauerland. Risks and Benefits of Large Language Models for the Environment. *Environmental Science & Technology* **57** (9): 3464-3466. DOI: 10.1021/acs.est.3c01106

Salvagno, M., Chat, GPT, Taccone, F.S., Gerli, A.G. 2023a. Can artificial intelligence help for scientific writing? Crit. Care 27, 75. https://doi.org/10.1186/ s13054-023-04380-2

Salvagno, M., Taccone, F.S., Gerli, A.G. 2023b. Can artificial intelligence help for scientific writing? Crit. Care 27, 99. https://doi.org/10.1186/s13054-023-04390-0

Schlagwein, D., & L. Willcocks. 2023. ChatGPT et al.: The ethics of using (generative) artificial intelligence in research and science. *Journal of Information Technology* 38 (3): 232-238.

Shamoo Adil E. and David B. Resnik. 2009. *Responsible Conduct of Research.* Second edition. Oxford: Oxford University Press.

Sharifzadeh, R. 2024. ChatGPT as Co-Author? AI and Research Ethics. ETHICS IN PROGRESS 15 (1): 155-173

Tan et al. ? 2024. Rethinking Literary Plagiarism in LLMs through the Lens of Copyright Laws. Proceedings of Machine Learning Research 260. Available at: https://openreview.net/pdf?id=sWZy2Xirwt (accessed Aug. 3, 2025).

Vidhya, N. G., D. A. N. Devi, & T. Manju. 2023. Prognosis of exploration on Chat GPT with artificial intelligence ethics. *Brazilian Journal of Science* 2 (9): 60–69. https://doi.org/10.14295/bjs.v2i9.372

Watkins, R. 2024. Guidance for researchers and peer-reviewers on the ethical use of Large Language Models (LLMs) in scientific research workflows. *AI Ethics* 4, 969–974. https://doi.org/10.1007/s43681-023-00294-5

Weber-Wulff, D., A. Anohina-Naumeca, S. Bjelobaba,Tomáš Foltýnek, Jean Guerrero-Dib, Olumide Popoola, Petr Šigut & Lorna Waddington. 2023. Testing of detection tools for AI-generated text. *Int J Educ Integr* 19, 26. https://doi.org/10.1007/s40979-023-00146-z

Weiser, B. 2023. *Lawyer who used ChatGPT faces penalty for made up citations*. The New York Times (8 Jun 2023).

Wen Jiaxin, Pei Ke, Hao Sun, Zhexin Zhang, Chengfei Li, Jinfeng Bai, and Minlie Huang. 2023. *Unveiling the Implicit Toxicity in Large Language Models.* In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, pages 1322–1338, Singapore. Association for Computational Linguistics.

Xames, M. D., & J. Shefa. 2023. ChatGPT for research and publication: Opportunities and challenges. *Journal of Applied Learning and Teaching* 6 (1). https://doi.org/10.37074/jalt.2023.6.1.20, Available at SSRN: https://ssrn.com/abstract=4381803 or http://dx.doi.org/10.2139/ssrn.4381803

Xu, Z., S. Jain, & M. S. Kankanhalli. 2024. Hallucination is Inevitable: An Innate Limitation of Large Language Models. ArXiv, abs/2401.11817.

Yao Yifan, Jinhao Duan, Kaidi Xu, Yuanfang Cai, Zhibo Sun, Yue Zhang. 2024. A survey on large language model (LLM) security and privacy: The Good, The Bad, and The Ugly. *High-Confidence Computing* 4 (2): 1-21.

Zhou, J., H. Müller, A. Holzinger, & F. Chen. 2024. Ethical ChatGPT: Concerns, Challenges, and Commandments. *Electronics 13* (17): 3417. https://doi.org/10.3390/electronics13173417

Zhuo Terry Yue, Yujin Huang, Chunyang Chen, Zhenchang Xing. 2023. Exploring AI Ethics of ChatGPT: A Diagnostic Analysis. Computation and Language. Online Pub Date: 2023-01-30, DOI: arxiv-2301.12867

Zohery, Medhat. 2023. *ChatGPT in Academic Writing and Publishing: A Comprehensive Guide*. In book: Artificial Intelligence in Academia, Research and Science: ChatGPT as a Case Study. (Frist Edition). New York: Novabret Publishing.

رحمان شریف‌زاده

متولد ۱۳۶۳، دارای مدرک دکتری فلسفهٔ علم و فناوری از پژوهشگاه علوم انسانی و مطالعات فرهنگی است. ایشان هم‌اکنون استادیار پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) است.

اخلاق فناوری، اخلاق فناوری اطلاعات، مطالعات علم و فناوری، و فلسفهٔ علم و فناوری از علایق پژوهشی وی است.

پژوهش‌نامه
پردازش و مدیریت اطلاعات