Journal of
Professional Auditing Research
Fall 2026, V.6, No 24 pp 82-113

Classification and Auditing of Large Language Models (LLMs)

Morteza Asadi[1], Mina Farnoud Ahmadi[2]

Received: 2026/02/23
Approved: 2026/07/12

Review Paper

Highlights

Auditing LLMs ensures compliance with laws and regulations, thereby reducing legal risks and liabilities.

Auditing, by identifying and mitigating technical, operational, and social risks associated with LLMs, enhances system safety.

Auditing LLMs, through increased transparency of their performance, strengthens stakeholder trust and facilitates broader adoption of the technology.

Abstract

Large language models represent a major advance in artificial intelligence research. These models are based on complex neural networks trained on large datasets and can provide unique capabilities in various domains. The application of large language models in the audit field has also created challenges in performing repetitive tasks, evaluating internal controls, providing procedures, and planning audits. Studies have also defined AI auditing as a mechanism that ensures that AI systems are ethically, legally, and technically sound. However, existing auditing practices have failed to address the challenges posed by large language models that exhibit emerging capabilities. The present study, by reviewing the existing literature, while defining and classifying these models, has raised the opportunities and challenges of using large language models in the field of auditing and has presented a three-level framework for their auditing. The results of this study provide researchers with a structured basis for classifying large language models. The study also shows that audits, when conducted in a structured and coordinated manner at all three levels, can be a practical mechanism for managing some of the ethical and social risks posed by large language models.

Key Words: Large language models, Governance audit, Model audit, Application audit.

Extended Abstract

Introduction

The advent of Large Language Models (LLMs) marks a pivotal innovation in artificial intelligence, shaping the future of auditing and related professional fields. LLMs are increasingly

Journal of
Professional Auditing Research
Fall 2026, V.6, No 24 pp 82-113

used to automate repetitive tasks, assess internal controls, support audit procedure development, and streamline audit planning. Their widespread adoption introduces both significant advantages and complex risks, particularly related to data privacy, security, bias, and ethical reliability. Traditional auditing frameworks have proven inadequate in addressing the unique challenges posed by LLMs, necessitating novel approaches to AI audit and governance. This study aims to bridge this gap by classifying LLMs and proposing a comprehensive auditing framework that ensures responsible and effective integration of these technologies in the audit profession.

Materials and Methods

The research adopts a literature review and synthesis approach, drawing insights from academic articles, industry reports, and regulatory guidelines on LLMs, AI auditing, and ethical frameworks. Comparative analysis was conducted to distinguish feature sets, risks, and opportunities across internal, semi-internal, and external LLMs. The study presents a three-level audit framework—comprising Governance Audit, Model Audit, and Application Audit—designed to holistically address ethical, technical, and social concerns. Practical implications and adoption barriers were investigated through review of audit firm practices, regulatory compliance, and stakeholder perspectives.

Results and Discussion

Findings of this research highlight that LLMs have become highly prevalent, offering enhanced efficiency, effectiveness, and quality in audit processes, but their usage is hindered by privacy, security, and ethical challenges, exacerbated by insufficient oversight. Major audit firms are transitioning from reliance on external LLMs toward internal and semi-internal models to strengthen confidentiality and ethical standards, though adoption remains slow due to high costs and client preferences for external models. The proposed three-level audit framework delivers several practical benefits:

1. Operational Legitimacy and Compliance: Governance and application audits collaboratively ensure that the use of LLMs aligns with national and international regulations (e.g., GDPR, EU AI Act), minimizing legal risks.

2. Security Enhancement and Risk Reduction: Model audits document technical capabilities, biases, and risks, while application audits assess real-world impacts, together mitigating operational, reputational, and ethical risks.

3. Transparency and Trust Building: Systematic audit documentation enhances stakeholder trust and supports broader social acceptance of AI-driven audit systems.

4. Ethical and Responsible Development: Integrated audit procedures guide LLMs development towards fairness, privacy, and accountability.

5. User Empowerment: Transparent evaluation enables users to understand LLMs strengths and limitations, fostering digital literacy and responsible interaction.

6. Industrial Standardization: Aggregating audit outcomes enables the formation of best practices and standards for LLMs deployment in the auditing industry.

The framework demonstrates robust synergy across its levels, facilitating safe and sustainable adoption of LLM-based technologies. It also provides a structured pathway for regulatory bodies to update governance and ethical guidelines, aligning AI usage with evolving industry

Journal of
Professional Auditing Research
Fall 2026, V.6, No 24 pp 82-113

challenges.

Conclusion

The comprehensive implementation of the three-level audit framework (governance, model, and application audits) ensures that LLMs are utilized not only as advanced technical tools, but also responsibly, ethically, and in accordance with societal interests and user rights. This approach mitigates multifaceted risks, promotes transparency and trust, and drives the responsible development and deployment of LLMs within the audit sector. Therefore, the review of theoretical foundations and proposed frameworks shows that the management and evaluation of systems based on large language models (LLMs) requires a three-level (strategic, model, and application) and coherent approach, each level having its own specific role and mission in ensuring the development, deployment, and responsible exploitation of the aforementioned technologies. Overall, this three-level and complementary approach, each alongside the other, provides a systematic and efficient mechanism for ensuring the security, legitimacy, and accountability of intelligent language systems.

Aggregation of the results of multiple audits at different levels will gradually help to shape and evolve industry-wide standards, assessment frameworks, and best practices for the development, deployment, and management of Large Language Models (LLMs). This will provide a platform for comparing, evaluating, and assuring the quality of products and services on a broader scale. Ultimately, implementing a three-tier audit approach ensures that Large Language Models (LLMs) are not only technically powerful tools, but are also used responsibly, safely, ethically, and in the interests of society and users, which in turn will lead to the sustainable and trusted development of new technologies. Future research should focus on comparative studies of different LLM implementations, development of audit standards, and evaluation of LLM impact on auditor-client relationships and industry dynamics. Policymakers are encouraged to update AI governance regulations in alignment with the proposed auditing framework to address emerging risks and foster sustained trust in technological innovation.

Ethical considerations

In this study, the ethical principle of scientific integrity is observed and the intellectual property rights of the authors are respected. The authors declare that there is no conflict of interest.

10.22034/JPAR.2026.2086351.1515

1. Master of Accounting, Management Faculty, University of Tehran, Tehran, Iran. (Corresponding Author) mortezaasadi4361@gmail.com
2. Ph.D Student in Accounting, Islamic Azad University. Tabriz. Iran. minafarnoodahmadi@gmail.com
http://article.iacpa.ir

فصلنامه‌ی پژوهش‌های حسابرسی حرفه‌ای
سال ششم، شماره بیست و چهار، پاییز ۱۴۰۵، صص ۸۲-۱۱۳

طبقه‌بندی و حسابرسی مدل‌های زبانی بزرگ (LLMs)

مرتضی اسدی[۱]✉، مینا فرنود احمدی[۲]

مقاله‌ی ترویجی

تاریخ دریافت: ۱۴۰۴/۱۲/۰۴
تاریخ پذیرش: ۱۴۰۵/۰۴/۲۱

نکات برجسته

حسابرسی مدل‌های زبانی بزرگ با تضمین قابلیت انطباق با قوانین و مقررات، ریسک‌های حقوقی و مسئولیت‌های قانونی را کاهش می‌دهد.

حسابرسی با شناسایی و کاهش ریسک‌های فنی، عملیاتی و اجتماعی مرتبط با مدل‌های زبانی بزرگ، موجب افزایش ایمنی سیستم‌ها می‌شود.

حسابرسی مدل‌های زبانی بزرگ از طریق شفافیت عملکرد، ضمن افزایش اعتماد عمومی ذیفعان، به پذیرش گسترده فناوری کمک می‌کند.

چکیده

مدل‌های زبانی بزرگ، نشان‌دهنده پیشرفتی بزرگ در حوزه تحقیقات هوش‌مصنوعی است. این مدل‌ها، مبتنی بر یادگیری ماشینی است که بر روی مجموعه‌ای بزرگ از داده‌ها آموزش دیده‌اند و می‌توانند در حوزه‌های مختلف، قابلیت‌های منحصربفردی ارائه کنند. کاربرد مدل‌های زبانی بزرگ در حوزه حسابرسی، ضمن انجام وظایف تکراری، ارزیابی کنترل‌های داخلی، ارائه رویه‌ها، و برنامه‌ریزی حسابرسی، چالش‌هایی را نیز ایجاد کرده است. همچنین مطالعات، حسابرسی هوش‌مصنوعی را به عنوان سازوکاری تعریف کرده‌اند که اطمینان می‌دهد سیستم‌های هوش‌مصنوعی از نظر اخلاقی، قانونی و فنی، قابل‌اتکا و مستحکم هستند. با این حال، رویه‌های حسابرسی موجود در مواجهه با چالش‌های ناشی از مدل‌های زبانی بزرگ که قابلیت‌های نوظهوری را نشان می‌دهند، ناتوان بوده‌اند. مطالعه حاضر با مروری بر ادبیات موجود، ضمن تعریف و طبقه‌بندی این مدل‌ها، فرصت‌ها و چالش‌های استفاده از مدل‌های زبانی بزرگ را در حوزه حسابرسی مطرح کرده و چارچوبی سه‌سطحی برای حسابرسی آن‌ها ارائه کرده است. نتایج این مطالعه مبنایی ساختاریافته برای طبقه‌بندی مدل‌های زبانی بزرگ در اختیار محققان قرار می‌دهد. همچنین نتایج مطالعه نشان داده است که هنگامی که حسابرسی‌ها، به شیوه‌ای ساختاریافته و هماهنگ در هر سه سطح انجام شوند، می‌توانند سازوکاری عملی برای مدیریت برخی از ریسک‌های اخلاقی و اجتماعی ناشی از مدل‌های زبانی بزرگ باشند.

واژه‌های کلیدی: مدل‌های زبانی بزرگ، حسابرسی راهبری، حسابرسی مدل، حسابرسی کاربردی.

10.22034/JPAR.2026.2086351.1515

۱. کارشناسی ارشد، دانشکده مدیریت، گروه حسابداری، دانشگاه تهران، تهران، ایران. (نویسنده مسئول) mortezaasadi4361@gmail.com
۲. دانشجوی دکتری حسابداری، دانشگاه آزاد اسلامی تبریز، تبریز، ایران. minafarnoodahmadi@gmail.com

http://article.iacpa.ir

۱-مقدمه

فناوری در دهه‌های اخیر پیشرفت شگرفی داشته است (نقدی و همکاران، ۱۴۰۳) و ظهور مدل‌های زبانی بزرگ (LLMs[۱]) تحولات قابل توجهی را در تمامی حوزه‌ها از جمله حسابرسی ایجاد کرده است (گو و همکاران[۲]، ۲۰۲۴)؛ بطوریکه عدم استفاده از این فناوری‌ها در حوزه حسابرسی، چالش‌های متعددی را برای فعالان حرفه به همراه خواهد داشت (علوی و همکاران، ۱۴۰۴). حسابرسان و حسابداران از جمله متخصصانی هستند که بیشتر از همه با قابلیت‌های به سرعت در حال توسعه سیستم‌های هوش مصنوعی مانند مدل‌های زبان بزرگ، مواجه‌اند (اولریچ و همکاران[۳]، ۲۰۲۴). در حالیکه تحقیقات قبلی نشان می‌دهد که مدل‌های زبانی بزرگ اثربخشی، کارایی و کیفیت کلی رویه‌های حسابرسی را افزایش می‌دهند (امت و همکاران[۴]، ۲۰۲۴)، برخی از محدودیت‌های مدل‌های زبانی بزرگ، چالش‌های اخلاقی قابل توجهی را در مورد حریم خصوصی داده‌ها، امنیت، محرمانگی، و نگرانی‌های مربوط به قابلیت اتکا را ایجاد می‌کند (راس و ژانگ[۵]، ۲۰۲۵). این نگرانی‌ها به دلیل احتمال طرح دعاوی قضایی و آسیب به اعتبار مؤسسه‌های حسابرسی، باعث افزایش ریسک حسابرسی شده است (امت و همکاران، ۲۰۲۴). تحقیقات موجود بیشتر بر اهمیت مدیریت محدودیت‌های مدل‌های زبانی بزرگ تأکید کرده‌اند (وود و ایلوریچ[۶]، ۲۰۲۵). از انجایی که چالش‌های ذکر شده بیشتر در مدل‌های زبانی بزرگ خارجی مشهودتر است، چهار موسسه بزرگ حسابرسی به طور فزاینده‌ای در مدل‌های زبانی بزرگ داخلی و نیمه داخلی سرمایه‌گذاری می‌کنند (دیلویت[۷]، ۲۰۲۴). استفاده از مدل‌های زبانی بزرگ داخلی و نیمه داخلی، با حفظ اطلاعات محرمانه صاحبکاران، موجب رعایت اصول اخلاق حرفه‌ای شده است (انجمن حسابداران رسمی آمریکا[۸]، ۱۹۹۳). رعایت چنین استانداردهای اخلاقی یکی از ابزارهای مهمی تلقی می‌شود که از طریق آن حسابرسان می‌توانند اعتماد عمومی را جلب کرده و وضعیت اخلاقی حرفه را بهبود بخشند (پاپاکریستو و بکیاریس[۹]، ۲۰۲۰). با این حال، ادبیات حسابرسی هنوز به طور واضح مدل‌های زبانی بزرگ را قالب داخلی، نیمه داخلی یا خارجی تعریف و طبقه‌بندی نکرده است، و ویژگی‌ها، مزایا و چالش‌های آنها بطور منسجم مطرح نشده است.

چالش‌های مطرح‌شده توسط مدل‌های زبانی بزرگ لزوماً از چالش‌های مرتبط با پردازش زبان طبیعی (NLP[۱۰]) یا سایر سیستم‌های مبتنی بر یادگیری ماشینی (ML[۱۱]) متمایز نیستند. با این حال، استفاده گسترده و عمومیت مدل‌های زبانی بزرگ، این چالش‌ها را درخور توجه کرده است. همه این دلایل، تحلیل مدل‌های زبانی بزرگ از دیدگاه‌های اخلاقی را نیازمند نوآوری در ابزارهای ارزیابی ریسک، معیارها و چارچوب‌ها کرده است (تامکین و همکاران[۱۲]، ۲۰۲۱). چندین مکانیسم نظارتی برای اطمینان از قانونی، اخلاقی و ایمن بودن مدل‌های زبانی بزرگ، طراحی، پیشنهاد یا آزمایش شده‌اند (آوین و همکاران[۱۳]، ۲۰۲۱). برخی از آنها از نظر فنی جهت‌گیری شده‌اند، از جمله پیش‌پردازش داده‌های آموزشی، تنظیم دقیق مدل‌های زبانی بزرگ بر روی داده‌هایی با ویژگی‌های مطلوب و رویه‌هایی برای آزمایش مدل قبل از استقرار (تامکین و همکاران، ۲۰۲۱). برخی دیگر به دنبال پرداختن به ریسک‌های اخلاقی و اجتماعی مرتبط با مدل‌های زبانی بزرگ از طریق استراتژی‌های کاهش اثرات منفی اجتماعی-فنی هستند، به عنوان مثال: پروتکل‌های انسانی در حلقه (وانگ و همکاران[۱۴]، ۲۰۲۱)،

و ابزارهای ارزیابی کیفی مبتنی بر روش‌های قوم‌نگاری (ماردا و نارایان[۱۵]، ۲۰۲۱). برخی دیگر به دنبال تضمین شفافیت در فرآیندهای توسعه هوش مصنوعی هستند مانند: استفاده ساختاریافته از کارت‌های مدل (میشـــل و همکاران[۱۶]، ۲۰۱۹)، برگه‌های داده (گبرو و همکاران[۱۷]، ۲۰۲۱)، کارت‌های سیســـتم (گرین و همکاران[۱۸]، ۲۰۲۲) و واترمارک کردن خروجی‌های سیستم (کرچنبایر و همکاران[۱۹]، ۲۰۲۳). با این حال، اثربخشی و امکان‌سنجی این مکانیسم‌های نظارتی هنوز توسط تحقیقات تجربی اثبات نشده است (انگلر[۲۰]، ۲۰۲۳)، اما آنچه واضح است این است که دیگر کاربرد رویه‌های حسابرسی سنتی برای حسابرسی فناوری‌های هوش‌مصنوعی و از جمله مدل‌های زبانی بزرگ کافی نیست. در مطالعه حاضر با مروری بر ادبیات مرتبط، طبقه‌بندی‌های مختلف مدل‌های زبانی بزرگ، و ویژگی‌های مربوط به هر یک از انواع آن تشریح شده و فرصت‌ها و چالش‌های کاربرد مدل‌های زبانی بزرگ در حوزه حسابرسی ارائه شده است. همچنین برای اثربخشی مکانیسم‌های نظارتی در حوزه مدل‌های زبانی بزرگ، چارچوبی سه لایه در راستای حسابرسی مدل‌های زبانی بزرگ ارائه شده است که شامل: حسابرسی راهبری، حسابرسی مدل، و حسابرسی کاربردی است. نتایج حاصل از مطالعه به مؤسسه‌های حسابرسی کمک می‌کند تا در خصوص ادغام مدل‌های زبانی بزرگ در حسابرسی‌ها،آگاهانه و مسئولانه تصمیم‌گیری کنند. همچنین نتایج مطالعه نشان می‌دهد هنگامی که حسابرسی‌ها، به شیوه‌ای ساختاریافته و هماهنگ در هر سه سطح چارچوب حسابرسی مذکور انجام شـــوند، می‌توانند مکانیسمی عملی و مؤثر برای شناسایی و مدیریت برخی از ریسک های اخلاقی و اجتماعی ناشی از مدل‌های زبانی بزرگ فراهم آورند.

۲- ادبیات و مبانی نظری

مدل‌های زبانی بزرگ (LLMs)

مدل‌های زبانی بزرگ (LLMs)، نوعی از یادگیری ماشین است که برای تحلیل مجموعه داده‌های گسترده، انجام وظایف، پاسخ به پرسش‌ها و ارائه پاسخ‌هایی دقیق و شبیه انسان، آموزش دیده و طراحی شده است(دانگ و همکاران[۲۱]، ۲۰۲۴). فرآیند آموزش مدل‌های زبانی بزرگ معمولاً شامل دو مرحله است: ۱- پیش‌آموزش ، ۲-بهینه‌سازی مدل(تنظیم دقیق) (فوهر و همکاران[۲۲]، ۲۰۲۳). پیش‌آموزش برای عملکرد مؤثر مدل‌های زبانی بزرگ بسیار مهم است، زیرا مدل‌های زبانی بزرگ بر اساس داده‌های گسترده آموزش می‌بینند تا دانشی جامع در مورد وظایف مختلف - چه وظایف عمومی و چه وظایف خاص - با حداقل مداخله انسانی و از طریق یادگیری خودنظارتی کسب کنند (گو و همکاران، ۲۰۲۴).

مدل‌هـــای زبانی بزرگ رایجی کـــه از انتقال‌دهنده‌ها (مبدل‌ها) اســـتفاده می‌کنند شـــامل: ارائه رمزگذاری‌شده دو طرفه از انتقال‌دهنده‌ها (BERT[۲۳]) و انتقال‌دهنده‌های پیش‌آموخته مولد (GPT[۲۴]) است. ارائه رمزگذاری‌شده دو طرفه از انتقال‌دهنده‌ها (BERT) با در نظر گرفتن راست‌چین و چپ‌چین بودن متن، نمایش‌های متنی را یاد می‌گیرد، در حالی که انتقال‌دهنده‌های پیش‌آموخته مولد (GPT) یک مدل خودهمبســـتگی یک طرفه اســـت که در آن هر کلمه جدید تحت تأثیر کلمات قبلی قرار می‌گیرد (دانگ و همکاران، ۲۰۲۴). ویژگی‌های دو طرفه ارائه رمزگذاری‌شده دو طرفه از انتقال‌دهنده‌ها (BERT)، توانایی آن را در درک زمینه‌ها، از جمله تشخیص موجودیت‌های نام‌گذاری‌شده (NER[۲۵]) و

پاسخ به سؤالات، افزایش می‌دهد. علی‌رغم این قابلیت‌ها، ارائه رمزگذاری‌شده دو طرفه از انتقال‌دهنده‌ها (BERT) اغلب نیاز به بهینه‌سازی (تنظیم دقیق) قابل‌توجهی برای درک بهتر واژگان خاص و ویژگی‌های منحصربفـــرد یک دامنه خاص دارد. در مقابل، انتقال‌دهنده‌های پیش‌آموخته مولد (GPT) عموماً در کارهایی مانند نوشـــتن خلاق، تولید و تکمیل متن، و تولید رمز عملکرد خوبی دارد. این الگوریتم در انجام کارهایی که از چندین نمونه یاد می‌گیرد (یادگیری چند مرحله‌ای) و کارهایی که صرفا برای انجام آنها آموزش ندیده است (یادگیری صفر مرحله‌ای) عملکرد چشمگیری دارد (دانگ و همکاران، ۲۰۲۴).

تعریف و دسته‌بندی مدل‌های زبانی بزرگ داخلی، نیمه داخلی و خارجی

مدل‌های زبانی بزرگ (LLMs) خارجی، مدل‌های عمومی طراحی شده برای استفاده عمومی است، و بطور خاص برای رفع نیازهای یک شرکت حسابرسی خاص یا صنعت حسابرسی طراحی نشده است. این مدل‌ها متعلق به ارائه دهنده خدمات است؛ به این معنی که ارائه دهنده خدمات دارای حقوق قانونی و مالکیت بر آنها است. یکی از ویژگی‌های ذاتی مدل‌های زبانی بزرگ خارجی این است که توسط ارائه دهنده خدمات میزبانی می‌شوند و به طور گسترده بر روی مجموعه داده‌های عمومی حجیمی آموزش دیده‌اند، که گاهی اوقات ممکن است دارای سوگیری و فاقد دانش خاص در دامنه معینی باشد. علاوه بر این، هر داده‌ای که از اســـتفاده آنها تولید می‌شـــود، متعلق به ارائه‌دهنده خدمات است و توسط آن کنترل می‌شود و کاربران، هیچ کنترلی بر داده‌ها ندارند. چنین مدل‌هایی معمولاً محدودیت‌هایی برای اســـتفاده عمومی دارند، زیرا عموم مردم نمی‌توانند به ویژگی‌های تعبیه شده در آنها دسترسی داشته باشـــند. نمونه‌هایی از مدل‌های زبانی بزرگ خارجی شامل ChatGPT توسط OpenAI، PaLM2، Bard، T5 توسط Google و LlaMA1 و LlaMA2 توسط Meta است (ایلاد و موگویرا[۲۶]، ۲۰۲۵).

مدل‌های زبانی بزرگ داخلی به عنوان مدل‌هایی تعریف می‌شوند که منحصراً برای یک شرکت و برای رفع نیازهای خاص آن طراحی شـــده اســـت؛ که به صورت رایگان برای اســـتفاده داخلی شرکت حسابرسی در دسترس است و حقوق اختصاصی آن در اختیار شرکت است و برای استفاده عمومی در دســـترس نیست. برخلاف مدل‌های زبانی بزرگ خارجی، مدل‌های زبانی بزرگ داخلی به صورت داخلی میزبانی می‌شوند و از زیرساخت‌های شرکت حسابرسی، مانند سرورهای داخلی، سیســـتم‌های ابری یا سایر سیستم‌های مرتبط، اســـتفاده می‌کنند. این تنظیمات تضمین می‌کند که داده‌های حســـاس، محرمانه و اختصاصی صاحبکاران، برخلاف مدل‌های زبانی بزرگ خارجی، تحت کنترل شـــرکت باقی می‌ماند. بنابراین، در استفاده از مدل‌های زبانی بزرگ داخلی، داده‌های تولید شـــده متعلق به شرکت حسابرسی اســـت و این مدل‌ها ممکن است در انجام حسابرسی‌ها مورد اســـتفاده قرار گیرند. علاوه بر این، آموزش داده‌ها بر اســـاس داده‌های خاص، مرتبط و قابل اتکا متناســـب با نیازهای شرکت حسابرسی انجام می‌شـــود. نمونه‌هایی از مدل‌های زبانی بزرگ داخلی شامل فناوری‌های در حال توسعه مؤسسه دیلویت مانند Re-، DocAI، ChatAI، PairD، ScribeAI، RiskAI، viewAI، فناوری‌های کاربردی مؤسســـه ارنست اندیانگ[۲۷] مانند EY.ai، EYQ، مدل‌های کاربردی مؤسسه پرایس واترهاوس کوپرز[۲۸] مانند ChatPwC، و فناوری مورد استفاده مؤسسه کی پی ام جی[۲۹] با نام KymChat است (ایلاد و موگویرا، ۲۰۲۵).

مدل‌های زبانی بزرگ نیمه‌داخلی، ویژگی‌های مشترک زیادی با مدل‌های زبانی بزرگ داخلی دارند، به جز اینکه حقوق مالکیت ممکن است توسط ارائه‌دهنده خدمات یا به طور مشترک با مؤسسه حسابرسی باشد و میزبانی آن‌ها ممکن است به صورت داخلی یا توسط ارائه‌دهنده خدمات مدیریت شود. با این حال، مؤسسه حسابرسی معمولاً مالکیت داده‌های تولید شده توسط این مدل‌ها را حفظ می‌کند، که بر اساس توافق قراردادی بین ارائه‌دهنده خدمات و مؤسسه حسابرسی انجام می‌شود. آموزش داده‌ها در مدل‌های زبانی بزرگ نیمه‌داخلی احتمالاً بر روی مجموعه داده‌های بزرگ و عمومی انجام می‌شود، که با در نظر گرفتن الگوریتم‌هایی برای فیلتر کردن و اولویت‌بندی محتوای خاص صنعت حسابرسی، یا گنجاندن مجموعه داده‌های خاص که محرمانگی داده‌ها را حفظ کند، تنظیم شده‌اند. این رویکرد مدل‌ها را قادر می‌سازد تا زبان فنی و دانش منحصر به فرد صنعت حسابرسی را یاد بگیرند و در نتیجه عملکرد خود را بهبود بخشند. یک نمونه معمول از مدل‌های زبانی بزرگ نیمه‌داخلی، همکاری ارنست اندیانگ، پرایس واترهاوس کوپرز، کی پی ام جی با مایکروسافت برای استفاده از خدمات Azure OpenAI در حسابرسی‌ها، یا همکاری کی پی ام جی با مایکروسافت برای استفاده از Microsoft 365 Co-Pilot برای انجام حسابرسی‌ها است (ایلاد و موگویرا، ۲۰۲۵). جدول (۱) تمایز بین مدل‌های زبانی بزرگ خارجی، نیمه‌داخلی و داخلی را نشان داده است.

جدول(۱): تمایز بین مدل‌های زبانی بزرگ خارجی، نیمه داخلی و داخلی

ویژگی‌ها	مدل‌های زبانی بزرگ خارجی	مدل‌های زبانی بزرگ نیمه داخلی	مدل‌های زبانی بزرگ داخلی
مالکیت مدل‌های زبانی بزرگ	متعلق به ارائه‌دهنده خدمات	متعلق به ارائه‌دهنده خدمات، یا هم ارائه‌دهنده خدمات و هم مؤسسه حسابرسی	متعلق به مؤسسه حسابرسی
نهاد میزبان	ارائه‌دهنده خدمات خارجی	شرکت میزبان/ارائه‌دهنده خدمات خارجی	شرکت میزبان
مالکیت داده‌ها	ارائه‌دهنده خدمات	مؤسسه حسابرسی (احتمالاً بر اساس قرارداد توافق)	مؤسسه حسابرسی
آموزش داده	به طور گسترده آموزش دیده روی مجموعه داده‌های عمومی عظیم	آموزش دیده روی مجموعه داده‌های بزرگ و عمومی با الگوریتم‌هایی که برای فیلتر کردن محتوای خاص حسابرسی یا گنجاندن مجموعه داده‌های خاص دامنه که محدود و محرمانه هستند، تنظیم شده‌اند.	آموزش دیده در مورد داده‌های خاص، مرتبط و قابل اتکا متناسب با نیاز موسسه حسابرسی
ماهیت استفاده	قابل استفاده برای عموم	قابل استفاده توسط موسسات حسابرسی در انجام حسابرسی	قابل استفاده توسط موسسات حسابرسی در انجام حسابرسی

محدودیت استفاده	محدودیت استفاده برای برخی از کاربردهای عمومی	محدودیت برای استفاده عمومی اما بدون محدودیت استفاده برای مؤسسه حسابرسی	محدودیت برای استفاده عمومی اما بدون محدودیت استفاده برای مؤسسه حسابرسی

فرصت‌ها و چالش‌های استفاده از مدل‌های زبانی بزرگ(LLMs) در حوزه حسابرسی

فرصت‌های استفاده از مدل‌های زبانی بزرگ(LLMs) در حسابرسی

مدل‌های زبانی بزرگ به دلیل توانایی‌شـــان در تولید پاسخ‌های به موقع و شبیه به انسان، و در عین حال فیلتر کردن اطلاعات نامربوط در خروجی تولید شده، مورد توجه قرار می‌گیرند (فتوح و موگویرا[۳۰]، ۲۰۲۵). علاوه بر این، مدل‌های زبانی بزرگ می‌توانند حجم اسنبوهی از داده‌های متنی را تحلیل کرده و نظارت مداوم و جمع‌آوری شواهد حسابرسی از داده‌های بدون ساختار را در زمان واقعی تسهیل کنند (امت و همکاران، ۲۰۲۴). مدل‌های زبانی بزرگ همچنین می‌توانند با تسهیل ارزیابی ریســک از طریق شناسایی پرچم‌های قرمز بالقوه، برنامه‌ریزی حسابرسی را بهبود بخشند و حسابرسان را قادر سازند تا ریسک‌های حسابرسی نوظهور، فعالیت‌های مشکوک و حوزه‌هایی را که حسابرسان در آنها از نظر دانش و تخصص کمبود دارند، شناسایی کنند (اولریچ و وود، ۲۰۲۵؛ فتــوح و موگویــرا، ۲۰۲۵). در حالی که برخی مطالعات بر پتانســیل مدل‌های زبانی بزرگ برای تسهیل استخراج تقلب‌های اساسی تأکید می‌کنند (لی و واسارهلی[۳۱]، ۲۰۲۴)، شواهد تجربی اخیر نتایج امیدوارکننده‌ای را نشان می‌دهد که عنوان می‌کند مدل‌های زبانی بزرگ می‌توانند تشخیص جعل و تقلب در صورت‌های مالی را بهبود و تســریع بخشند (بوسکو و همکاران[۳۲]، ۲۰۲۴). علاوه بر این، مدل‌های زبانی بزرگ می‌توانند رویه‌های حسابرســی مناســبی را نسبت به میزان ریسک توصیه کنند، بنابراین حسابرســان را قادر می‌ســازند تا دامنه حسابرسی‌ها را تعریف کرده و کار میدانی حسابرسی را ، به ویژه در تسریع تهیه مصاحبه‌ها و چک لیست حوزه‌های موجود در دامنه حسابرسی (امت و همکاران، ۲۰۲۴)، و رویه‌های پیگیری تسهیل کنند (اولریچ و وود). همچنین، استفاده از مدل‌های زبانی بزرگ موجب تسهیل ارزیابی کنترل‌های داخلی، و شناسایی نقاط ضعف کنترل‌های داخلی شده و زمینه‌های بهبود را پیشنهاد می‌دهند (ژائو و وانگ[۳۳]، ۲۰۲۴).

از ســویی دیگر، این مدل‌ها می‌توانند وظایف روزمره، تکراری، وقت‌گیر و مســتعد خطا، مانند تهیــه و اصلاح گزارش‌های حسابرســی و گزارش‌هــای کاری (اوترو و آگــو[۳۴]، ۲۰۲۵)، مطابق با انتظارات را انجام دهند. زمان صرفه‌جویی شــده از این فعالیت‌هــای روزمره می‌تواند یک مزیت رقابتی برای مؤسسه‌های حسابرســی فراهم کند (فتوح و موگویرا، ۲۰۲۵) و به حسابرسان اجازه دهــد تا بــر کار میدانی، تعامل با صاحبکاران (امت و همــکاران، ۲۰۲۴) و وظایف پیچیده‌ای که مستلزم قضاوت حسابرسان است، تمرکز کنند (فتوح و موگویرا، ۲۰۲۵). مطالعات نشان داده است که طرحواره‌های شــناختی مدل‌های زبانی بزرگ، مشابه طرحواره‌های حسابرسان مالی است، که نشانگر این است که مدل‌های زبانی بزرگ می‌توانند وظایف حسابرسی خاصی را به طور مؤثر انجام

دهند (وی و همکاران[۳۵]، ۲۰۲۳). مدل‌های زبانی بزرگ به لطف توانایی‌های پردازش زبان طبیعی خـــود که تحلیل داده‌های متنی کارآمد و مؤثر در زمان واقعی را ممکن می‌ســـازد، قادر به انجام وظایفی هســـتند که معمولاً توسط حسابرسان انجام می‌شود (اولیری[۳۶]، ۲۰۲۴). چنین مدل‌های زبانی بزرگی می‌توانند با ارائه رهنمودهایی در مورد اســـتانداردهای حسابرسی مربوطه که مربوط به کار حسابرســـی است، تجربه یادگیری حسابرســـان جوان را افزایش دهند (دانگ و همکاران، ۲۰۲۳؛ فتـــوح و موگویرا، ۲۰۲۵). یکی دیگر از پتانســـیل‌های مهم مدل‌های زبانی بزرگ، توانایی آنها در پیش‌ارزیابی و تقویت رویه‌های حسابرســـی برنامه‌ریزی‌شده، قبل از اجرا است که منجر به درک بهتر رویه‌های حسابرسی احتمالی، افزایش اعتماد حسابرسان به برنامه‌های حسابرسی شده و در نهایت منجر به بهبود برنامه‌ها و نتایج حسابرســـی می‌شود (واسارهلی و همکاران[۳۷]، ۲۰۲۳). جدول(۲) شـــامل خلاصه‌ای جامع از توانمندی‌ها و فرصت‌های ایجاد شده توسط مدل‌های زبانی بزرگ در حسابرسی را ارائه کرده است.

جدول(۲): فرصت‌های استفاده از مدل‌های زبانی بزرگ (LLMs) در حسابرسی

فرصت‌های ایجاد شده توسط مدل‌های زبانی بزرگ (LLMs)	نکات اصلی	محققان
تولید گزارش‌های مرتبط، به موقع و شبیه به انسان	مدل‌های زبانی بزرگ گزارش‌های به موقع و انسانی تولید می‌کنند و اطلاعات نامربوط را فیلتر می‌کنند.	Dong et al., 2024; Fotoh & Mugwira,2025
تحلیل داده‌های متنی، نظارت مستمر، و جمع‌آوری شواهد در لحظه	مدل‌های زبانی بزرگ حجم زیادی از داده‌های متنی را تحلیل می‌کنند و شواهد حسابرسی کلیدی را از داده‌های بدون ساختار به صورت بلادرنگ جمع‌آوری کرده و نظارت می‌کنند.	Gu et al., 2024; Li & Vasarhelyi, 2024; Oteri & Agu, 2025
بهبود برنامه‌ریزی حسابرسی و تشخیص تقلب	مدل‌های زبانی بزرگ ارزیابی و کاهش ریسک را تسهیل می‌کنند و به حسابرسان در تشخیص ریسک‌های بالقوه و کاستی‌های دانش کمک می‌کنند. آنها همچنین حسابرسان را قادر می‌سازند تا تقلب در صورت‌های مالی را تشخیص دهند.	Eulerich & Wood, 2025; Zhao & Wang, 2023
توصیه رویه‌های حسابرسی مبتنی بر ریسک و رویه‌های پیگیری	مدل‌های زبانی بزرگ رویه‌های حسابرسی را بر اساس میزان ریسک توصیه می‌کنند، تعریف دامنه حسابرسی را تسهیل کرده، رویه‌های کار میدانی حسابرسی و پیگیری را انجام می‌دهند، و پیشرفت رویه‌های حسابرسی قبلی را پیگیری می‌کنند.	Eulerich & Wood, 2025; Zhao & Wang, 2023
ارزیابی کنترل‌های داخلی	مدل‌های زبانی بزرگ ارزیابی کنترل‌های داخلی را تسهیل می‌کنند، نقاط ضعف را مشخص کرده و بهبودهایی را پیشنهاد می‌دهند.	Eulerich & Wood, 2025; Fotoh & Mugwira, 2025

انجام وظایف حسابرسی پیش پا افتاده	مدل‌های زبانی بزرگ وظایف تکراری ساده، وقت‌گیر و مستعد خطا مانند تهیه و اصلاح گزارش‌های حسابرسی و گزارش‌های کاری را مدیریت می‌کنند.	Emett et al., 2024; Eulerich & Wood, 2025; Fotoh & Mugwira, 2025
صرفه‌جویی در زمان فرآیندهای حسابرسی	زمان صرفه‌جویی‌شده از فعالیت‌های تکراری ساده می‌تواند برای شرکت‌های حسابرسی یک مزیت رقابتی ایجاد کند.	Fotoh & Mugwira, 2025
افزایش تمرکز بر وظایف اصلی حسابرسی	صرفه‌جویی در زمان انجام کارهای تکراری ساده باعث آزاد شدن زمان حسابرسان شده و به آنها اجازه می‌دهد تا بر کارهای میدانی حسابرسی، تعامل با صاحبکاران و وظایف پیچیده‌ای که نیاز به قضاوت دارند، تمرکز کنند.	Emett et al., 2024; Fotoh & Mugwira, 2025
شباهت طرحواره‌های شناختی مدل‌های زبانی بزرگ با طرحواره‌های حسابرسان باتجربه	مدل‌های زبانی بزرگ طرحواره‌های شناختی مشابه حسابرسان باتجربه دارند که آنها را قادر می‌سازد تا به دلیل آموزش بر روی داده‌های ساختاریافته، وظایف حسابرسی خاصی را به طور کارآمد انجام دهند.	Wei et al., 2023
ارتقای یادگیری برای حسابرسان تازه کار	مدل‌های زبانی بزرگ با ارائه راهنمایی در مورد استانداردهای حسابرسی مرتبط با کارهای حسابرسی، تجربه یادگیری حسابرسان تازه‌کار را افزایش می‌دهند.	Dong et al., 2024; Fotoh & Mugwira, 2025
پتانسیل همکاری مدل‌های زبانی بزرگ با حسابرسان	مدل‌های زبانی بزرگ می‌توانند به طور قابل توجهی با حسابرسان همکاری کنند و وظایفی مانند تحلیل نسبت‌های مالی و تسهیل استخراج متن از داده‌ها، استخراج اطلاعات مرتبط و تشخیص ناهنجاری را انجام دهند.	Gu et al., 2024; Li & Vasarhelyi, 2024
تجربه یادگیری شخصی‌سازی‌شده برای حسابرسان و آموزش‌های خاص صنعت برای مدل‌های زبانی بزرگ	مدل‌های زبانی بزرگ با درک ترجیحات و رویکردهای نوشتاری حسابرسان، یک تجربه یادگیری شخصی و زمینه‌ای ارائه می‌دهند و دقت و ارتباط پاسخ‌ها را بهبود می‌بخشند. علاوه بر این، مدل‌های زبانی بزرگ می‌توانند به طور خاص برای تامین نیازهای صنایع و شرکت‌های خاص آموزش ببینند.	Gu et al., 2024
پیش ارزیابی رویه‌های حسابرسی	مدل‌های زبانی بزرگ می‌توانند رویه‌های حسابرسی برنامه‌ریزی‌شده را از قبل ارزیابی کرده، و شناخت اقدامات احتمالی حسابرسی را افزایش دهد، و اعتماد به طرح‌های حسابرسی، برنامه‌ها و نتایج حسابرسی را افزایش دهند.	Vasarhelyi et al., 2023
افزایش کارایی رویه‌های حسابرسی	مدل‌های زبانی بزرگ اجرای رویه‌های حسابرسی کارآمد، قابل اعتماد و مؤثر را تسریع می‌کنند.	Gu et al., 2024

چالش‌های استفاده از مدل‌های زبانی بزرگ (LLMs) در حسابرسی

علیرغم مزایای بالقوه ادغام مدل‌های زبانی بزرگ(LLMs) با فناوری‌های نوظهور (لی و همکاران[۳۸]، ۲۰۲۴)، تحقیقات معاصر چالش‌های متعددی را در ارتباط با پذیرش مدل‌های زبانی بزرگ در حسابرسی، صرف نظر از طبقه‌بندی آنها در قالب مدل‌های زبانی بزرگ به عنوان داخلی، نیمه‌داخلی، یا خارجی، ارائه کرده است.

اولاً، مسائل مربوط به حریم خصوصی داده‌ها، محرمانگی و امنیت داده‌ها بسیار مهم هستند و پیامدهای اخلاقی قابل توجهی دارند (استریت و همکاران[۳۹]، ۲۰۲۳؛ امت و همکاران، ۲۰۲۴). برای اینکه مدل‌های زبانی بزرگ دقیق و قابل‌اتکا باشند، نیاز به آموزش جامع در مورد داده‌های اختصاصی صاحبکاران دارند (ژائو و وانگ، ۲۰۲۳). برای مثال، پتانسیل مدل‌های زبانی بزرگ برای افشای ناخواسته اطلاعات محرمانه، حساس یا اختصاصی صاحبکاران، نگرانی‌های اخلاقی قابل توجهی را ایجاد کرده است (اوترو و آگو، ۲۰۲۵). دسترسی غیرمجاز به داده‌های صاحبکاران، نقض داده‌ها یا هک کردن منجر به طرح دعوی قضایی، آسیب به اعتبار و از بین رفتن اعتماد به حرفه حسابرسی می‌شود (گو و همکاران، ۲۰۲۴). علاوه بر این، داده‌های محرمانه چنین صاحبکارانی می‌تواند برای آموزش مدل‌های زبانی بزرگ استفاده شود و هنجارهای اخلاقی را نقض کند که حسابرسان را ملزم به رعایت محرمانگی اطلاعات صاحبکاران و عدم انتشار چنین اطلاعاتی بدون تأیید صاحبکاران می‌کند (انجمن حسابداران رسمی آمریکا، ۱۹۹۳؛ انجمن استانداردهای بین‌المللی اخلاقی برای حسابداران[۴۰]، ۲۰۲۳). در نتیجه، ضروری است که اطمینان حاصل شود که اقدامات لازم در خصوص حفظ حریم خصوصی انجام شده است و داده‌های حساس صاحبکاران در طول پردازش توسط مدل‌های زبانی بزرگ ایمن باقی می‌مانند (استریت و همکاران). علاوه بر این، تفسیر نادرست داده‌ها توسط مدل‌های زبانی بزرگ ممکن است باعث ایجاد توجیهات یا گاهی اوقات استدلال‌های کاملاً اشتباه شوند؛ زیرا این داده‌ها فاقد توانایی استدلال اخلاقی هستند و بنابراین نمی‌توانند تصمیمات حسابرسی را که نیاز به استدلال دارد، اتخاذ کنند (فتوح و موگویرا، ۲۰۲۵). توانایی مدل‌های زبانی بزرگ در تولید پاسخ‌هایی شبیه به انسان، آن tr ها را مستعد انتشار اطلاعات جعلی می‌کند (اولیری، ۲۰۲۳) و در مقایسه با حسابرسان انسانی، تصور نادرستی از کارایی آنها ایجاد می‌کند (استریت و ویلک[۴۱]، ۲۰۲۳). همچنین از آنجایی که کیفیت پاسخ‌های تولید شده توسط مدل‌های زبانی بزرگ تا حد زیادی به کیفیت دستورالعمل‌های ارائه شده بستگی دارد (گو و همکاران، ۲۰۲۴)، دستورالعمل‌های تولید شده ضعیف می‌توانند منجر به پاسخ‌های ضعیف نادرست شوند (فتوح و موگویرا، ۲۰۲۵). از سویی دیگر، نگرانی‌هایی در مورد حقوق مالکیت معنوی خروجی‌های تولید شده توسط مدل‌های زبانی بزرگ وجود دارد، زیرا این مدل‌ها بر اساس حجم زیادی از داده‌های دارای حق چاپ آموزش دیده‌اند و هیچگونه شفافیت یا پاسخگویی در مورد مالکیت این داده‌ها وجود ندارد. در نتیجه، خروجی‌های تولید شده توسط برخی از مدل‌های زبانی بزرگ می‌تواند بطور بالقوه حق چاپ و سایر حقوق مالکیت معنوی را

نقض کند (استریت و ویلک، ۲۰۲۳).

باید در نظر داشت که مدل‌های زبانی بزرگ در برابر حملات سایبری آسیب پذیر هستند (اولریچ و وود، ۲۰۲۵)، که این موضوع نیاز به اقدامات قوی امنیت سایبری برای مقابله با چنین تهدیدهایی را برجسته می‌کند. همچنین چالش‌هایی در رابطه با پاسخگویی و شفافیت این مدل‌ها وجود دارد (فتوح و موگویرا، ۲۰۲۵). داده‌های آموزشی آنقدر گسترده هستند که حتی توسعه‌دهندگان اغلب منابع آنها را نمی‌دانند و ردیابی منشأ آنها را دشوار می‌کند (اولیری، ۲۰۲۳). در نتیجه، این ابهام در مورد منابع آنها، منطق پیرامون تولید خروجی و فرآیند آموزش داده‌ها، سوالات مربوط به شفافیت و قابلیت‌اتکای آنها را تشدید می‌کند (استریت و ویلک، ۲۰۲۳؛ فتوح و موگویرا، ۲۰۲۵). همچنین چالش اثر اتاق پژواک (echo chamber effect)[۴۲] وجود دارد که از بازتولید خطاها و سوگیری‌های مدل‌های زبانی بزرگ از منابع داده ناشی می‌شود. مدل‌های زبانی بزرگ به دلیل آموزش داده‌ها بر اساس داده‌های مغرضانه و غیرمتنوع، ممکن است خروجی‌های مغرضانه‌ای تولید کنند (اولریچ و وود، ۲۰۲۵). به عنوان مثال، عدم دسترسی به اطلاعات حسابداری خاص ممکن است این موضوع را تشدید کند (فتوح و موگویرا، ۲۰۲۵).

از سویی دیگر، خروجی‌های تولید شده توسط برخی از مدل‌های زبانی بزرگ ممکن است در اثر عدم انجام بهینه‌سازی مدل، قدیمی و ناقص باشند (استریت و ویلک، ۲۰۲۳). به عنوان مثال، زمان بین یک اطلاعیه حسابرسی جدید و زمانی که مدل‌های زبانی بزرگ از این اطلاعیه مطلع می‌شوند، می‌تواند بر خروجی تولید شده توسط مدل‌های زبانی بزرگ تأثیر بگذارد. این چالش نگرانی‌های بیشتری را در مورد دقت و قابلیت‌اتکای چنین مدل‌های زبانی بزرگ در انجام وظایف حسابرسی خاص (فتوح و موگویرا، ۲۰۲۵) و عدم توانایی آنها در ارائه رهنمودهای آینده‌نگر ایجاد می‌کند (استریت و همکاران، ۲۰۲۳). علاوه بر این، احتمال انسان‌انگاری وجود دارد، جایی که حسابرسان به اشتباه ویژگی‌ها، صفات و احساسات انسانی را به مدل‌های زبانی بزرگ نسبت می‌دهند (واسارهلی و همکاران، ۲۰۲۳). برخلاف انسان‌ها، مدل‌های زبانی بزرگ و سایر فناوری‌ها فاقد ویژگی‌هایی مانند خودشناسی، کنترل، آگاهی و انگیزه هستند (موناکو و همکاران[۴۳]، ۲۰۲۳) و بنابراین نمی‌توانند به عنوان متخصصان حسابرسی قابل اعتماد، عمل کنند؛ زیرا نمی‌توانند مانند حسابرسان قضاوت حرفه‌ای انجام دهند (استریت و همکاران، ۲۰۲۳) (Street et al., 2023). بنابراین، حسابرسان باید آگاه باشند که فرآیندهایی که مدل‌های زبانی بزرگ برای نتیجه‌گیری اتخاذ می‌کنند، بسیار غیرانسانی است (واسارهلی و همکاران، ۲۰۲۳)، و مستلزم تردید حرفه‌ای و نظارت دقیق حسابرسان است. اتکای بیش از حد به مدل‌های زبانی بزرگ ممکن است بر استقلال، تردیدحرفه‌ای و قضاوت حسابرسان که برای عملکرد مؤثر حسابرسی حیاتی هستند، تأثیر منفی بگذارد (فتوح و موگویرا، ۲۰۲۵). در نتیجه، مدل‌های زبانی بزرگ به عنوان دستیارهای آزمایشی (گو و همکاران، ۲۰۲۴)، نه برای جایگزینی حسابرسان، بلکه برای افزایش کارایی و اثربخشی وظایف حسابرسی طراحی شده‌اند (استریت و همکاران، ۲۰۲۳). جدول (۳) خلاصه‌ای جامع از چالش‌های فعلی مدل‌های زبانی بزرگ در حسابرسی را نشان داده است.

جدول(۳): چالش‌های استفاده از مدل‌های زبانی بزرگ (LLMs) در حسابرسی

چالش‌های ایجاد شده توسط مدل‌های زبانی بزرگ (LLMs)	نکات اصلی	محققان
حریم خصوصی داده‌ها، محرمانگی، نگرانی‌های امنیتی، و نقض‌های اصول اخلاقی بالقوه و آسیب‌های اعتباری	آموزش مدل‌های زبانی بزرگ با داده‌های محرمانه صاحبکاران بدون تأیید آنها، نگرانی‌های اخلاقی قابل توجهی را در مورد حریم خصوصی، محرمانگی و امنیت داده‌ها، به ویژه خطر افشای سهوی داده‌های محرمانه، حساس یا اختصاصی صاحبکاران، که به طور بالقوه منجر به طرح دعوی قضایی، آسیب به اعتبار و از دست دادن اعتماد به حرفه حسابرسی می‌شود، ایجاد می‌کند.	Emett et al., 2024; Eulerich & Wood, 2025
مدل‌های زبانی بزرگ مستعد انتشار اطلاعات نادرست هستند.	استفاده از مدل‌های زبانی بزرگ می‌تواند منجر به تولید اظهارات نادرست، تفسیر نادرست داده‌ها و توجیه نادرست خطاها شود و در نتیجه منجر به انتشار اطلاعات جعلی و ایجاد تصور نادرست از کارایی می‌شود.	Fotoh & Mugwira, 2025 Vasarhelyi et al., 2023
فقدان استدلال اخلاقی در استفاده از مدل‌های زبانی بزرگ	مدل‌های زبانی بزرگ توانایی استدلال اخلاقی ندارند و این امر توانایی آنها را در تصمیم‌گیری‌های حسابرسی که نیاز به استدلال انسانی دارد، محدود می‌کند.	Fotoh & Mugwira, 2025
وابستگی به کیفیت سریع	کیفیت پاسخ‌های مدل‌های زبانی بزرگ به شدت به کیفیت سوالات ارائه شده بستگی دارد؛ سوالات ضعیف منجر به پاسخ‌های ضعیف می‌شود.	Gu et al., 2024; Fotoh & Mugwira, 2025
نگرانی‌های مربوط به مالکیت معنوی	خروجی‌های تولید شده توسط مدل‌های زبانی بزرگ به دلیل استفاده از حجم زیادی از داده‌های دارای حق چاپ بدون پاسخگویی به مالکیت، می‌توانند به طور بالقوه حق چاپ و سایر حقوق مالکیت معنوی را نقض کنند.	Fotoh & Mugwira, 2025 Street & Wilck, 2023
خطر انسان‌پنداری و فرآیندهای غیرانسانی دوره‌های آموزشی مدل‌های زبانی بزرگ	این خطر وجود دارد که حسابرسان ممکن است به اشتباه ویژگی‌ها، صفات و احساسات انسانی را به مدل‌های زبانی بزرگ نسبت دهند، در عین حال فرآیندهایی که مدل‌های زبانی بزرگ برای نتیجه‌گیری استفاده می‌کنند به طور قابل توجهی با استدلال انسانی متفاوت است و نیاز به تردید و نظارت دقیق حسابرسان دارد.	Vasarhelyi et al., 2023
تأثیر اتکای بیش از حد به مدل‌های زبانی بزرگ	اتکای بیش از حد به مدل‌های زبانی بزرگ ممکن است بر استقلال، تردیدحرفه‌ای و قضاوتحرفه‌ای حسابرسان که برای عملکرد مؤثر حسابرسی بسیار مهم هستند، تأثیر منفی بگذارد.	Fotoh & Mugwira, 2025

فقدان دانش تخصصی در صنعت	مدل‌های زبانی بزرگ اغلب فاقد دانش تخصصی در صنعت، از جمله درک دقیق از مقررات خاص حسابداری و حسابرسی مانند IFRS[۴۴]، ISA [۴۵] و GAAP[۴۶] است.	Fotoh & Mugwira, 2025; Oteri & Agu, 2025
آسیب‌پذیری در برابر حملات سایبری	مدل‌های زبانی بزرگ در برابر حملات سایبری آسیب‌پذیرند، که نشان دهنده نیاز به اقدامات قوی امنیت سایبری است.	Eulerich and Wood, 2025
نگرانی‌های مربوط به پاسخگویی، شفافیت و قابلیت اتکا	منشأ نامشخص داده‌های آموزشی برای مدل‌های زبانی بزرگ، همراه با منطق مرتبط با خروجی‌های تولید شده، قابلیت ردیابی را پیچیده کرده و نگرانی‌هایی را در مورد منشأ، شفافیت و قابلیت اتکای داده‌ها ایجاد می‌کند.	O'Leary, 2023; Street & Wilck, 2023 Oteri & Agu, 2025
اثر اتاق پژواک و سوگیری در خروجی‌های مدل‌های زبانی بزرگ	مدل‌های زبانی بزرگ می‌توانند خطاها و سوگیری‌ها را از منابع داده‌ای غیرمتنوع بازتولید کنند، و مدل‌های آتی ممکن است این سوگیری‌ها را به طور نامحسوس درونی کرده و تشدید کنند. در حوزه حسابداری و حسابرسی به دلیل دسترسی محدود به انواع خاصی از اطلاعات این امر تشدید می‌شود.	Eulerich and Wood, 2025; Fotoh and Mugwira, 2025
خروجی‌های قدیمی و نگرانی‌های مربوط به دقت در مدل‌های زبانی بزرگ	برخی از خروجی‌های مدل‌های زبانی بزرگ ممکن است به دلیل شکاف در به‌روزرسانی داده‌های آموزشی با اطلاعات جدید، قدیمی بوده و به‌روز نباشند.این امر بر دقت خروجی تولید شده تأثیر می‌گذارد و عملکرد مدل‌های زبانی بزرگ را در انجام وظایف حسابرسی خاص تحت تأثیر قرار می‌دهد.	Fotoh & Mugwira, 2025; Street & Wilck, 2023

حسابرسی هوش‌مصنوعی

حسابرسی را می‌توان به عنوان یک سازوکار نظارتی در نظر گرفت زیرا می‌توان از آن برای نظارت بر رفتار و عملکرد استفاده کرد (فلینت[۴۷]، ۱۹۸۸) و سابقه طولانی در ارتقاء دقت و شفافیت رویه‌ها در زمینه‌هایی مانند حسابداری مالی دارد (لابری و استنکی[۴۸]، ۲۰۱۹). بنابراین ایده پشت حسابرسی هوش‌مصنوعی ساده است: درست همانطور که تراکنش‌های مالی را می‌توان از نظر صحت، کامل بودن و قانونی بودن حسابرسی کرد، طراحی و استفاده از سیستم‌های هوش‌مصنوعی را نیز می‌توان از نظر استحکام فنی، انطباق با قوانین یا پایبندی به اصول اخلاقی از پیش تعریف شده حسابرسی کرد. حسابرسی هوش‌مصنوعی یک حوزه مطالعاتی نسبتاً جدید است که در سال ۲۰۱۴ توسط مقاله سندویچ و همکارانش با عنوان «الگوریتم‌های نظارتی» مطرح شد (سندویگ[۴۹]، ۲۰۱۴). حسابرسی تقریباً بر تمام ابعاد هوش‌مصنوعی، از مستندسازی رویه‌های طراحی گرفته تا آزمون و تأیید مدل (استادت و ریچ[۵۰]، ۲۰۲۱)، نظارت دارد. بنابراین، حسابرسی هوش‌مصنوعی هم یک عمل چندوجهی و هم یک حوزه تحقیقاتی چندرشته‌ای است که از علوم کامپیوتر (آدلر و همکاران[۵۱]، ۲۰۱۸)، حقوق (لوکس و همکاران[۵۲]، ۲۰۲۱)، مطالعات رسانه‌ای و

ارتباطات (باندی[۵۳]، ۲۰۲۱) و مطالعات سازمانی بهره می‌برد.

محققان مختلف، حسابرسی هوش‌مصنوعی را به روش‌های مختلفی تعریف کرده‌اند. بطوریکه می‌توان بین مفاهیم محدود و گسترده حسابرسی هوش‌مصنوعی تمایز قائل شد. گروهی از تعاریف، رویکردی تأثیرمحور از حسابرسی هوش‌مصنوعی ارائه کرده است که بر بررسی و ارزیابی خروجی‌های سیستم‌های هوش‌مصنوعی برای داده‌های ورودی مختلف تمرکز دارد (متاکسا و همکاران[۵۴]، ۲۰۲۱). در حالیکه گروهی دیگر از تعاریف، رویکردی فرآیندمحور دارد و بر ارزیابی کفایت فرآیندهای توسعه نرم‌افزار و سیستم‌های مدیریت کیفیت ارائه‌دهندگان فناوری تمرکز دارد (برگوت و همکاران[۵۵]، ۲۰۲۳). بطور کلی، حسابرسی هوش‌مصنوعی را می‌توان به عنوان یک فرآیند سیستماتیک و مستقل، برای جمع‌آوری و ارزیابی شواهد مربوط به اقدامات یا ویژگی‌های یک واحد تجاری و انتقال نتایج آن ارزیابی، به ذینفعان مربوطه تعریف کرد. باید توجه داشت که واحد تجاری مورد نظر، می‌تواند یک سیستم هوش‌مصنوعی، یک سازمان، یک فرآیند یا هر ترکیبی از آنها باشد (موکاندر و آکسنتی[۵۶]، ۲۰۲۳).

رویکردی سه سطحی برای حسابرسی مدل‌های زبانی بزرگ (LLMs)

رویکرد سه سطحی برای حسابرسی مدل‌های زبانی بزرگ (LLMs)، بر سه سطح تمرکز دارد. سطح اول، ارائه دهندگان فناوری را شامل می‌شود که مدل‌های زبانی بزرگ را توسعه می‌دهند و باید تحت حسابرسی‌های راهبری قرار گیرند که رویه‌های سازمانی، ساختارهای پاسخگویی و سیستم‌های مدیریت کیفیت آنها را ارزیابی کند. سطح دوم، مدل‌های زبانی بزرگ باید پس از آموزش اولیه، اما قبل از تطبیق و استقرار در کاربردهای خاص، تحت حسابرسی‌های مدل قرار گیرند و قابلیت‌ها و محدودیت‌های آنها ارزیابی شود. سطح سوم، برنامه‌های کاربردی پایین‌دستی که از مدل‌های زبانی بزرگ استفاده می‌کنند باید تحت حسابرسی‌های مداوم قرار گیرند که همسویی اخلاقی و انطباق قانونی عملکردآنها و تأثیر این برنامه‌های کاربردی را در طول زمان ارزیابی کنند (موکاندر و همکاران[۵۷]، ۲۰۲۴). شکل(۱)، رویکرد سه‌سطحی حسابرسی مدل‌های زبانی بزرگ را به تصویر کشیده است.

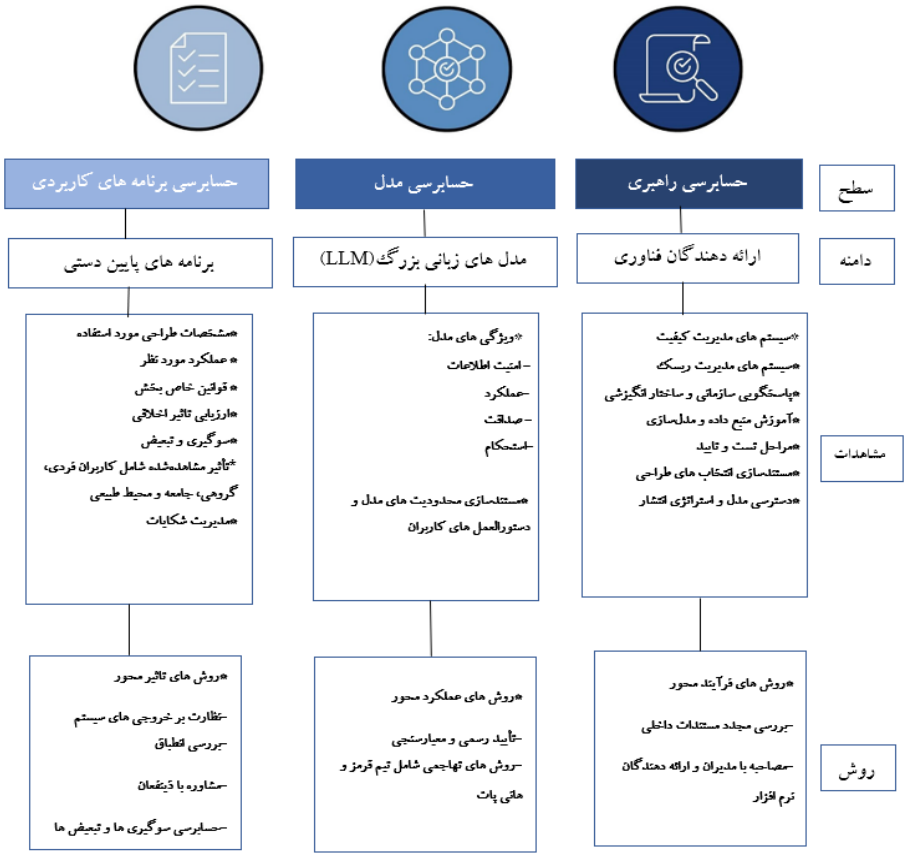
حسابرسی راهبری

ارائه‌دهندگان فناوری که روی مدل‌های زبانی بزرگ کار می‌کنند باید تحت حسابرسی راهبری قرار گیرند که رویه‌های سازمانی، ساختارهای انگیزشی و سیستم‌های مدیریتی آنها را ارزیابی می‌کند. شواهد نشان داده است که چنین ویژگی‌هایی بر طراحی و استقرار فناوری‌ها تأثیر می‌گذارند (برونداج و همکاران[۵۸]، ۲۰۲۰). علاوه بر این، تحقیقات نشان داده است که استراتژی‌های کاهش ریسک زمانی بهترین عملکرد را دارند که به طور شفاف، مداوم و با حمایت سطح اجرایی اتخاذ شوند. ارائه‌دهندگان فناوری مسئول تشخیص ریسک‌های مرتبط با مدل‌های زبانی بزرگ خود هستند و به طور منحصر بفردی در موقعیت مناسبی برای مدیریت برخی از این ریسک‌ها قرار دارند. بنابراین، بسیار مهم است که رویه‌های سازمانی و ساختارهای راهبری آنها جامع و کافی باشد (هادج[۵۹]، ۲۰۱۵).

حسابرسی‌های راهبری سابقه طولانی در حوزه‌هایی مانند مدیریت فناوری اطلاعات (ایلیسن[۶۰]، ۲۰۱۰) و مهندسی سیستم‌ها و امنیت داده‌ها دارد (فالکو و همکاران[۶۱]، ۲۰۲۱). وظایف آن شامل ارزیابی ساختارهای راهبری داخلی، فرآیندهای توسعه خدمات و سیستم‌های مدیریت کیفیت، برای ارتقای شفافیت، و اطمینان از وجود سیستم‌های مدیریت ریسک مناسب، و ایجاد مشورت در مورد پیامدهای اخلاقی و اجتماعی در طول چرخه عمر توسعه نرم‌افزارها است. همچنین، حسابرسی‌های راهبری می‌توانند پاسخگویی را بهبود بخشند. در کل، حسابرسی‌های راهبری، عناصری از حسابرسی‌های انطباق، در مورد کامل بودن و شفافیت اسناد، و حسابرسی‌های ریسک، در مورد کفایت سیستم مدیریت ریسک را در بر می‌گیرد (اینگلر[۶۲]، ۲۰۲۱). بنابراین، حسابرسی‌های راهبری ارائه دهندگان مدل‌های زبانی بزرگ، باید بر سه وظیفه کلی متمرکز شوند:

(۱) بررسی کفایت ساختارهای راهبری برای اطمینان از اینکه فرآیندهای توسعه مدل از بهترین شیوه‌ها پیروی می‌کنند و سیستم‌های مدیریت کیفیت می‌توانند ریسک‌های خاص مدل‌های زبانی بزرگ را در بر گیرند. در حالی که ارائه دهندگان فناوری دارای متخصصان مدیریت کیفیت داخلی هستند، سوگیری ممکن است مانع از تشخیص نقص‌های بحرانی توسط آنها شود؛ مشارکت حسابرسان خارجی به رفع این مشکل کمک می‌کند (بائور[۶۳]، ۲۰۱۶). باید در نظر داشت که حسابرسی‌های راهبری زمانی بیشترین تأثیر را دارند که حسابرسان و ارائه دهندگان فناوری برای تشخیص ریسک‌ها همکاری کنند (چوپرا و سینگ[۶۴]، ۲۰۱۸).

(۲) ایجاد یک مسیر حسابرسی از فرآیند توسعه مدل‌های زبانی بزرگ برای ارائه شواهد مستند از توسعه قابلیت‌های این مدل‌ها، شامل اطلاعاتی در مورد هدف مورد نظر، مشخصات و گزینه‌های طراحی، و همچنین نحوه آموزش و آزمایش آن از طریق تولید کارت‌های مدل (میشل و همکاران، ۲۰۱۹) و کارت‌های سیستم (گرین و همکاران، ۲۰۲۲). این شامل استفاده ساختاریافته از برگه‌های داده برای مستندسازی نحوه منبع‌یابی، برچسب‌گذاری و گردآوری مجموعه داده‌های مورد استفاده برای آموزش و اعتبارسنجی مدل‌های زبانی بزرگ می‌شود. ایجاد چنین مسیرهای حسابرسی، اهداف مرتبطی را دنبال می‌کند.



شکل(۱) : رویکرد سه لایه برای حسابرسی مدل‌های زبانی بزرگ (LLMs) (منبع: Mökander et al., 2024)

حسابرسی راهبری

تعیین مشـــخصات طراحی از قبل، بررسی پایبندی سیســـتم به الزامات قضایی پایین‌دستی را تســـهیل می‌کند (فالکو و همکاران، ۲۰۲۱). علاوه بر این، اطلاعات مربوط به موارد استفاده مورد نظر باید در توافق‌نامه‌های صدور مجوز با توســـعه‌دهندگان پایین‌دستی گنجانده شود و در نتیجه احتمال آســـیب از طریق استفاده مخرب را محدود کند. در نهایت، الزام ارائه‌دهندگان خدمات به مستندســـازی و توجیه انتخاب‌های طراحی خود، با روشـــن کردن بده‌بستان‌ها، باعث ایجاد تأمل اخلاقی می‌شود (کونتراکتور و همکاران[۶۵]، ۲۰۲۲).

(۳) ترســـیم نقش‌ها و مسئولیت‌ها در ســـازمان‌هایی که مدل‌های زبانی بزرگ را برای تسهیل تخصیص مســـئولیت‌پذیری در قبال خرابی‌های سیســـتم طراحی می‌کنند. ســـازگاری مدل‌های

زبانی بزرگ در پایین‌دست، ارائه‌دهندگان فناوری را از هرگونه مسئولیتی مبرا نمی‌کند. برخی از ریسک‌ها «به‌طور معقولی قابل پیش‌بینی» هستند. ترسیم نقش‌ها و مسئولیت‌های ذینفعان مختلف، پاسخگویی را بهبود می‌بخشد و احتمال «ساختارمند بودن» ارزیابی‌ها را افزایش داده، و به شناسایی و کاهش پیشگیرانه آسیب‌ها کمک می‌کند (بولامووینی و گبرو[۶۶]، ۲۰۱۸).

برای انجام این سه وظیفه، حسابرسان لازم است تا با حسابرسی جعبه سفید آشنا باشند. حسابرسی جعبه سفید، بالاترین سطح دسترسی است و نشان می‌دهد که حسابرس می‌داند چگونه و چرا یک مدل‌های زبانی بزرگ توسعه یافته است. در عمل، این به معنای دسترسی به امکانات، اسناد و مدارک و پرسنل است که رویه استاندارد در حسابرسی‌های راهبری در سایر زمینه‌ها است. به عنوان مثال، حسابرسان فناوری اطلاعات به مطالب و گزارش‌های مربوط به فرآیندهای عملیاتی و معیارهای عملکرد دسترسی کامل دارند. حسابرسی جعبه سفید مستلزم وجود توافق‌نامه‌های عدم افشا و اشتراک‌گذاری داده‌ها است. با این حال، از دیدگاه امنیت هوش مصنوعی، اعطای چنین سطح بالایی از دسترسی بسیار مهم است زیرا علاوه بر حسابرسی مدل‌های زبانی بزرگ قبل از استقرار در بازار، حسابرسی‌های راهبری باید حفاظت‌های سازمانی مربوط به پروژه‌های پرریسک را نیز ارزیابی کنند که ارائه‌دهندگان ممکن است ترجیح دهند در مورد آنها به صورت عمومی بحث نکنند (سنفت و گالیگوز[۶۷]، ۲۰۰۸).

نتایج حسابرسی‌های راهبری باید در قالب‌هایی متناسب با مخاطبان مختلف ارائه شود. مخاطب اصلی، مدیریت و مدیران ارائه‌دهنده‌ی مدل‌های زبانی بزرگ هستند. حسابرسان باید گزارش کاملی ارائه دهند که به‌طور مستقیم و شفاف، آسیب‌پذیری‌های ساختارهای راهبری موجود را فهرست کرده و مورد بررسی قرار دهد. چنین گزارش‌هایی ممکن است اقداماتی را توصیه کنند، اما انجام اقدامات همچنان بر عهده‌ی ارائه‌دهنده است. معمولاً چنین گزارش‌های حسابرسی، عمومی نمی‌شوند. با این حال، برخی از شواهد به‌دست‌آمده در طول حسابرسی‌های راهبری را می‌توان برای دو مخاطب ثانویه گردآوری کرد: مجریان قانون و توسعه‌دهندگان برنامه‌های پایین‌دستی. در برخی از حوزه‌های قضایی، قوانین سختگیرانه ممکن است از ارائه‌دهندگان فناوری بخواهند که از الزامات خاصی پیروی کنند.گزارش‌های حاصل از حسابرسی‌های راهبری شرکتی همچنین به توسعه‌دهندگان برنامه‌های کاربردی پایین‌دستی کمک می‌کند تا هدف، ریسک‌ها و محدودیت‌های مورد نظر یک مدل‌های زبانی بزرگ را درک کنند (موکاندر و همکاران، ۲۰۲۴).

حسابرسی مدل

از آنجایی که حسابرسی‌های راهبری تنها تأثیر غیرمستقیم دارند، قبل از استقرار، مدل‌های زبانی بزرگ باید تحت حسابرسی‌های مدل قرار گیرند تا قابلیت‌ها و محدودیت‌های آنها را ارزیابی کند. حسابرسی‌های مدل دارای برخی ویژگی‌های مشترک با حسابرسی‌های راهبری است. به عنوان مثال، هر دو قبل از تطبیق مدل‌های زبانی بزرگ برای کاربردهای خاص انجام می‌گیرند. با این حال، حسابرسی‌های مدل بر روی رویه‌های سازمانی تمرکز نمی‌کنند، بلکه بر روی قابلیت‌ها و ویژگی‌های مدل‌های زبانی بزرگ تمرکز دارند. به طور خاص، آنها باید محدودیت‌های مدل زبانی

بزرگ را شناسایی کنند تا (۱) به طراحی مجدد مداوم سیستم کمک کنند، و (۲) قابلیت‌ها و محدودیت‌های آن را به ذینفعان خارجی اطلاع دهند. این دو وظیفه از روش‌های مشابهی استفاده می‌کنند، اما مخاطبان متفاوتی را هدف قرار می‌دهند (موکاندر و همکاران، ۲۰۲۴).

با وجود اینکه ارزیابی ویژگی‌های یک مدل زبانی بزرگ مستقل چالش‌برانگیز است، اما غیرممکن نیست. برای انجام این کار، حسابرسان می‌توانند از دو رویکرد متمایز استفاده کنند. رویکرد اول شامل شناسایی و ارزیابی ویژگی‌های ذاتی است. به عنوان مثال، مجموعه داده‌های آموزشی را می‌توان بدون ارجاع به موارد استفاده خاص، از نظر کامل بودن و انطباق ارزیابی کرد (فلوریدی و همکاران[۶۸]، ۲۰۲۲). با این حال، بررسی مجموعه داده‌های بزرگ اغلب گران و از نظر فنی چالش‌برانگیز است. رویکرد دوم شامل استفاده از یک رویکرد غیرمستقیم است که مدل را در چندین مورد استفاده بالقوه پایین‌دستی آزمایش می‌کند، نتایج را به ویژگی‌های مختلف پیوند می‌دهد و نتایج تجمیع شده را با استفاده از تکنیک‌های وزن‌دهی مختلف ارزیابی می‌کند. رویکرد دوم ممکن است هنگام ارزیابی عملکرد یک مدل زبانی بزرگ مثمر ثمرتر باشد (پائولادا[۶۹]، ۲۰۲۱).

با این وجود، انتخاب ویژگی‌هایی که باید در طول حسابرسی‌های مدل بر آنها تمرکز شود، همچنان چالش‌برانگیز است. با توجه به هدف چنین حسابرسی‌هایی، توصیه می‌شود ویژگی‌هایی بررسی شوند که (۱) از نظر اجتماعی و اخلاقی مبسوط باشند، یعنی بتوان چنین ویژگی‌هایی را مستقیماً با ریسک های اجتماعی و اخلاقی ناشی از مدل‌های زبانی بزرگ مرتبط ساخت؛ (۲) به طور قابل پیش‌بینی قابل انتقال باشند، یعنی بر ماهیت کاربردهای پایین‌دستی تأثیر بگذارند؛ و (۳) به طور معناداری قابلیت عملیاتی داشته باشند، یعنی بتوانند با ابزارها و روش‌های موجود ارزیابی شوند. با در نظر گرفتن این معیارها، حسابرسی‌های مدل باید (حداقل) بر عملکرد، استحکام، امنیت اطلاعات و صداقت مدل‌های زبانی بزرگ تمرکز کنند. اکنون به بحث در مورد چگونگی ارزیابی چهار ویژگی نمونه در طول حسابرسی‌های مدل می‌پردازیم:

(۱) عملکرد، یعنی اینکه مدل زبانی بزرگ در وظایف مختلف چقدر خوب عمل می‌کند. معیارهای استاندارد می‌توانند با مقایسه آن با یک مبنای انسانی، به ارزیابی عملکرد مدل زبانی بزرگ کمک کنند.

(۲) استحکام، یعنی اینکه مدل چقدر خوب به درخواست‌های غیرمنتظره یا موارد حاشیه‌ای واکنش نشان می‌دهد. در یادگیری ماشین، استحکام نشان می‌دهد که یک الگوریتم هنگام مواجهه با داده‌های ورودی جدید و بالقوه غیرمنتظره (یعنی خارج از دامنه) چقدر خوب عمل می‌کند. مدل زبانی بزرگ فاقد استحکام، دو ریسک متمایز را ایجاد می‌کنند. اول، ریسک خرابی‌های بحرانی سیستم. دوم، ریسک حملات تهاجمی. بنابراین، محققان و توسعه‌دهندگان ابزارها و روش‌هایی را برای ارزیابی استحکام مدل‌های زبانی بزرگ ایجاد کرده‌اند، از جمله روش‌های تهاجمی مانند تیم قرمز، ابزارهای ارزیابی مانند Robustness Gym[۷۰]. بنابراین، حسابرسی‌های مدل باید از ابزارها و روش‌های متعددی برای ارزیابی استحکام مدل، استفاده کنند (گول و همکاران[۷۰]، ۲۰۲۱).

(۳) امنیت اطلاعات، یعنی اینکه استخراج داده‌های آموزشی از مدل زبانی بزرگ چقدر دشوار

اســـت. چندین ریســـک مرتبط با مدل زبانی بزرگ را می‌توان به عنوان «ریسک‌های اطلاعاتی» درک کرد، از جمله ریســـک مرتبط با حریم خصوصی که منجر به نشر داده‌های شخصی می‌شود. ارزیابی امنیت اطلاعات مدل زبانی بزرگ در طول حسابرسی‌های مدل، تمام ریسک‌های اطلاعاتی را برطرف نمی‌کند (ویدینگر و همکاران[۷۱]، ۲۰۲۲).

(۴) صداقت، یعنی اینکه مدل زبانی بزرگ تا چه حد می‌تواند بین دنیای واقعی و دنیای ممکن تمایز قائل شود.

برخی از ریســـک‌های مربوط به مدل‌های زبانی بزرگ ناشـــی از ظرفیت آنها در ارائه اطلاعات نادرست یا گمراه‌کننده است که به طور بالقوه اعتماد عمومی به اطلاعات مشترک را از بین می‌برد. روش‌های آماری در تشـــخیص بیـــن اطلاعات واقعاً صحیح در مقابل اطلاعـــات محتمل اما واقعاً نادرست مشکل دارند. این مشکل با این واقعیت تشدید می‌شود که بسیاری از شیوه‌های آموزشی مدل‌های زبانی بزرگ، مانند تقلید از متن انسانی در وب یا بهینه‌سازی برای کلیک‌ها، بعید است که هوش مصنوعی صادق ایجاد کنند. با این حال، در طول حسابرسی‌های مدل، نگرانی ما توسعه هوش مصنوعی صادق نیســـت، بلکه ارزیابی صداقت اســـت. چنین حسابرسی‌هایی باید بر ارزیابی صداقـــت کلی تمرکز کنند، نه صداقت یک جمله منفرد. هنـــگام ارزیابی یک مدل زبانی بزرگ با کمک TruthfulQA، حسابرســـان امتیاز درصدی در مورد میزان صداقت مدل دریافت می‌کنند. با این حال، حتی عملکرد قوی در TruthfulQA به این معنی نیســـت که یک مدل زبانی بزرگ در یـــک حوزه تخصصی صادق خواهد بود. با ایـــن وجود، چنین معیارهایی ابزارهای مفیدی برای حسابرسی مدل ارائه می‌دهند (ایوانز و همکاران[۷۲]، ۲۰۲۱).

ایـــن چهار ویژگـــی مربوط به مدل‌های زبانی بزرگ از پیش آموزش داده شـــده اســـت. با این حال، حسابرســـی مدل باید مجموعه داده‌های آموزشـــی را نیز بررســـی کند.کاملاً مشخص است که شـــکاف‌ها یا ســـوگیری‌های داده‌های آموزشـــی، مدل‌هایی را ایجاد می‌کنند که در مجموعه داده‌هـــای مختلف عملکرد ضعیفی دارند. آموزش مدل‌های زبانی بزرگ با داده‌های ســـوگیرانه یا ناقص می‌تواند باعث آسیب‌های بازنمایی و تخصیصی شود. حسابرسی‌های مدل مستلزم دسترسی ممتاز حسابرســـان به مدل‌های زبانی بزرگ و مجموعه داده‌های آموزشی آنها هستند. با این حال، برخلاف حسابرسی‌های جعبه سفید، دسترسی حسابرسان مدل به سیستم فنی محدود می‌شود و به فرآیندهای سازمانی ارائه‌دهندگان فناوری گسترش نمی‌یابد. برخی از ویژگی‌های مورد آزمایش در طول حسابرسی‌های مدل، مســـتقیماً با ریسک‌های اجتماعی و اخلاقی مدل‌های زبانی بزرگ (LLMs) مطابقت دارند (کالیسکان و همکاران[۷۳]، ۲۰۱۷).

حسابرسی کاربردی

محصـــولات و خدماتی که با اســـتفاده از مدل‌های زبانی بزرگ ســـاخته می‌شـــوند، باید تحت حسابرســـی‌های کاربردی قرار گیرند که قانونی بودن عملکردهای مورد نظر آنها و چگونگی تأثیر آنها بـــر کاربران و جوامع را ارزیابی می‌کند. حسابرســـی‌های کاربردی برخلاف حسابرســـی‌های

راهبـــری و مدل، بر کاربرانـــی که مدل‌های زبانی بـــزرگ را در کاربردهای پایین‌دســـتی به کار می‌گیرند، تمرکز دارند. چنین حسابرسی‌هایی برای اطمینان از انطباق با قوانین ملی و منطقه‌ای، استانداردهای خاص بخش و اصول اخلاق سازمانی مناسب است. حسابرسی‌های کاربردی دو جزء دارند: حسابرســـی‌های عملکردی، که برنامه‌های کاربردی را با استفاده از مدل‌های زبانی بزرگ بر اســـاس اهداف مورد نظر و عملیاتی آنها ارزیابی می‌کنند، و حسابرســـی‌های تأثیر، که برنامه‌ها را بر اســـاس تأثیرات آنها بر کاربران، گروه‌ها و محیط طبیعـــی مختلف ارزیابی می‌کنند (موکاندر و همکاران، ۲۰۲۴).

در طول حسابرســـی‌های عملکرد، حسابرســـان باید بررســـی کنند که آیا هدف مورد نظر یک برنامه خاص (۱) به خودی خود قانونی و اخلاقی اســـت و (۲) با اســـتفاده مورد نظر از مدل زبانی بزرگ مورد نظر همسو است یا خیر. اولین مورد بررسی، برای انطباق قانونی و اخلاقی است؛ یعنی پایبندی به قوانین، مقررات، دستورالعمل‌ها و مشخصات مربوط به یک برنامه خاص، اصول اخلاقی یا ضوابط رفتاری. هدف بررســـی‌های انطباق ساده اســـت: اگر یک برنامه غیرقانونی یا غیراخلاقی باشـــد، عملکرد مؤلفه مدل‌های زبانی بزرگ آن غیرمرتبط است و برنامه نباید در بازار مجاز باشد. دومین مورد بررسی در حسابرســـی‌های عملکرد، با هدف رسیدگی به ریسک‌های ناشی از اغراق یا نادرســـت جلوه دادن قابلیت‌های یک برنامه خاص توسط توسعه‌دهندگان انجام می‌شود (راجی و همکاران[۷۴]، ۲۰۲۲).

برای انجام این کار، حسابرسی‌های عملکرد بر اساس حسابرسی‌های سطوح دیگر انجام می‌شوند و خروجی‌های حاصل از آنها را در نظر می‌گیرند. در طول حسابرســـی‌های راهبری، ارائه‌دهندگان فناوری موظفند موارد اســـتفاده مورد نظر و غیرمجاز مدل‌های زبانی بزرگ خود را تعریف کنند. در طول حسابرســـی‌های مدل، محدودیت‌های مدل‌های زبانی بزرگ مستندســـازی می‌شـــوند تا انطباق آنها را در ســـطوح پایین‌تر اطلاع دهند. با اســـتفاده از چنین اطلاعاتی، حسابرســـی‌های عملکرد باید اطمینان حاصل کنند که برنامه‌های پایین‌دســـتی با موارد اســـتفاده مورد نظر مدل زبانی بزرگ مشخص به‌گونه‌ای همسو هستند که محدودیت‌های مدل را در نظر بگیرند. بنابراین، حسابرســـی‌های عملکرد، عناصر حسابرسی انطباق و ریســـک‌های مورد نیاز برای ارائه اطمینان برای مدل‌های زبانی بزرگ را ترکیب می‌کنند. برای انجام حسابرسی‌های کاربردی، سطوح پایین دسترســـی کافی است همچنین، برای حسابرســـی برنامه‌های کاربردی مبتنی بر مدل‌های زبانی بزرگ برای انطباق با قانون و همســـویی اخلاقی، حسابرسان نیازی به دسترسی مستقیم به مدل زیربنایـــی ندارند، اما می‌توانند بـــه اطلاعات عمومی موجود - از جملـــه ادعاهای ارائه دهندگان فناوری و توسعه دهندگان پایین‌دستی در مورد سیستم‌های خود و دستورالعمل‌های کاربر متصل به آنها - تکیه کنند (موکاندر و همکاران، ۲۰۲۴).

۳- بحث و نتیجه‌گیری

مدل‌های زبانی بزرگ به دلیل توانایی‌شـــان در پاســـخگویی به سوالات کاربران و کمک به آنها

در انجام وظایف، بسیار فراگیر شده و محبوبیت خاصی بین کاربران مختلف دارند. این مدل‌ها به بخشی جدایی‌ناپذیر از زندگی مردم تبدیل شده است و مزایای بی‌شماری در طیف وسیعی از برنامه‌ها ارائه می‌کنند. با این حال، عدم وجود نظارت کافی در این حوزه، نگرانی‌های قابل توجهی را ایجاد می‌کند. سوگیری، تولید اطلاعات گمراه‌کننده از نگرانی‌های قابل توجه است. مطالعه حاضر ضمن طبقه‌بندی انواع مختلف مدل‌های زبانی بزرگ، ویژگی‌ها، فرصت‌ها، و چالش‌های ایجاد شده توسط انواع مختلف مدل‌های زبانی بزرگ را بصورت جامع بیان کرده است که در جداول (۱)، (۲)، (۳) خلاصه شده است

با توجه به تحولات اساسی ایجاد شده توسط مدل‌های زبانی بزرگ در حوزه مختلف کسب و کار و قابلیت‌های منحصربفرد این مدل‌ها در انجام وظایف، چهار مؤسسه بزرگ حسابرسی، سرمایه‌گذاری‌های قابل توجهی در مدل‌های زبانی بزرگ انجام داده و از مدل‌های زبانی بزرگ خارجی به مدل‌های زبانی بزرگ داخلی و نیمه داخلی، گذار کرده‌اند. علت این امر کاهش برخی از چالش‌های مرتبط با مدل‌های زبانی بزرگ خارجی و در عین حال افزایش اعتماد و اطمینان کاربران به حسابرسی‌ها بوده است. نتایج مطالعه نشان‌دهنده پذیرش آهسته مدل‌های زبانی بزرگ داخلی و نیمه داخلی در شرکت‌های حسابرسی است. از جمله موانع استفاده از مدل‌های زبانی بزرگ داخلی و نیمه داخلی هزینه‌های بالا، عدم هماهنگی این مدل‌ها با نیازهای صاحبکاران، و رضایت صاحبکاران از مدل‌های زبانی بزرگ خارجی است. با این حال، یافته‌ها نشان می‌دهد که مدل‌های زبانی بزرگ داخلی و نیمه داخلی مزایای بیشتری برای فرآیند حسابرسی، از طریق تسهیل رویه‌های برنامه‌ریزی حسابرسی، شناسایی و کاهش کنترل‌های داخلی، پشتیبانی از کار میدانی حسابرسی، فراهم می‌کنند. زیرا این مدل‌ها به طور بالقوه کارایی، اثربخشی و کیفیت کلی فرآیند حسابرسی را بهبود می‌بخشند.

با توجه به تفاوت‌های بین مدل‌های زبانی بزرگ داخلی و نیمه داخلی و خارجی در مدیریت داده‌های حساس، محرمانه و اختصاصی صاحبکاران و پتانسیل آنها برای تأثیرگذاری بر نتایج حسابرسی، نهادهای نظارتی ممکن است نیاز به به‌روزرسانی یا تدوین دستورالعمل‌هایی برای اطمینان از استفاده مسئولانه از این فناوری‌ها داشته باشند. به طور مشابه، دستورالعمل‌های اخلاقی حسابرسی ممکن است نیاز به به‌روزرسانی داشته باشند تا به چالش‌های اخلاقی نوظهور در پذیرش مدل‌های زبانی بزرگ بپردازند. همچنین نتایج مطالعه نشان داده است که استفاده از چارچوب حسابرسی سه سطحی شامل: حسابرسی راهبری، حسابرسی مدل، و حسابرسی کاربردی، می‌تواند رویه‌ای جامع برای حسابرسی مدل‌های زبانی بزرگ ارائه دهد؛ بطوریکه سه سطح حسابرسی مطرح در این چارچوب مکمل یکدیگر است. حسابرسی‌های راهبری ، ساختارهای پاسخگویی و سیستم‌های مدیریت کیفیت ارائه‌دهندگان فناوری را از نظر استحکام، کامل بودن و کفایت ارزیابی می‌کنند. حسابرسی‌های مدل، قابلیت‌ها و محدودیت‌های مدل‌های زبانی بزرگ را از چندین بعد، از جمله عملکرد، استحکام، امنیت اطلاعات و صداقت، ارزیابی می‌کنند. در نهایت، در طول حسابرسی‌های کاربردی، محصولات و خدماتی که بر اساس مدل‌های زبانی بزرگ ساخته

می‌شـــوند، ابتدا از نظر انطباق با قوانین ارزیابی می‌شوند و متعاقباً بر اساس تأثیر آنها بر کاربران، گروه‌ها و محیط طبیعی ارزیابی می‌شـــوند. به عبارتی دیگر، می‌توان گفت که حسابرسی کاربردی جز لاینفک فرآیند اطمینان‌بخشی، مدیریت ریسک و رعایت اصول اخلاقی در استقرار سامانه‌های مبتنی بر مدل‌های زبانی بزرگ محســـوب می‌گردد. این ســـطح از حسابرسی، بستر لازم را برای پیشـــگیری از اثرات منفـــی، تضمین حقوق کاربران و حمایت از اصـــول اخلاق حرفه‌ای را فراهم می‌ســـازد و نقش حیاتی در بهبود فرآیندهای توسعه مسئولانه این فناوری ایفا می‌کند. توسعه و عملیاتی‌سازی چارچوب‌های استاندارد و مدون در حوزه حسابرسی کاربردی، در کنار دیگر سطوح حسابرســـی، زمینه را برای بهره‌برداری مطمئن و پایـــدار از فناوری‌های نوظهور فراهم می‌آورد و می‌تواند به ارتقای سطح اعتماد عمومی و تقویت نظارت‌های قانونی کمک کند.

از این رو، حسابرســـی‌ها در ســـه ســـطح باید در یک فرآیند ســـاختاریافته به هم متصل شوند. حسابرســـی‌های راهبری باید اطمینان حاصل کنند که ارائه‌دهندگان، ســـازوکارهایی برای در نظر گرفتن گزارش‌های خروجی تولید شـــده در طول حسابرســـی‌های کاربردی هنگام طراحی مجدد مدل‌های زبانی بزرگ دارند. به طور مشـــابه، حسابرسی‌های کاربردی باید اطمینان حاصل کنند که توسعه‌دهندگان پایین‌دستی، محدودیت‌های شناسایی شده در طول حسابرسی‌های مدل را هنگام ســـاخت بر روی یک مدل زبانی بزرگ خاص در نظر می‌گیرند. دوم، حسابرسی‌ها در هر سطح باید توســـط یک شخص ثالث مستقل انجام شود تا اطمینان حاصل شـــود که مدل زبانی بزرگ از نظر اخلاقی، قانونی و فنی قابل اتکا هســـتند. بنابراین، بررسی مبانی نظری و چارچوب‌های پیشنهادی نشان می‌دهد که مدیریت و ارزیابی سیستم‌های مبتنی بر مدل‌های زبانی بزرگ (LLMs) مستلزم رویکردی سه‌ســـطحی (راهبردی، مدل، و کاربردی) و منسجم است که هر سطح، نقش و مأموریت خاص خود را در تضمین توسعه، استقرار و بهره‌برداری مسئولانه از فناوری‌های مذکور بر عهده دارند. در مجموع، این رویکرد سه‌سطحی و مکمل، هر یک در کنار دیگری، سازوکاری نظام‌مند و کارآمد برای تضمین امنیت، مشروعیت و مسئولیت‌پذیری سامانه‌های هوشمند زبانی فراهم می‌آورند.

پیاده‌سازی جامع رویکرد حسابرســـی سه‌سطحی - شامل حسابرسی راهبری، حسابرسی مدل و حسابرســـی کاربردی - برای مدل‌های زبانی بزرگ (LLMs)، پیامدهای کاربردی قابل توجهی را در راســـتای تضمین مسئولیت‌پذیری، انطباق، امنیت و پذیرش اجتماعی این فناوری به همراه دارد. این نتایج، که از هم‌افزایی ســـه سطح حسابرسی حاصل می‌شوند، در حوزه‌های کلیدی زیر قابل مشاهده است:

۱. تضمین انطباق و مشروعیت عملیاتی

حسابرسی راهبری با تدوین سیاست‌های شفاف و حسابرسی کاربردی با ارزیابی انطباق هدف و عملکرد محصول نهایی، اطمینان حاصل می‌کنند که کلیه محصولات و خدمات مبتنی بر مدل‌های زبانی بـــزرگ (LLMs)، از جمله نحوه جمع‌آوری داده، آموزش مدل، و به‌کارگیری آن، با قوانین و مقـــررات ملی و بین‌المللی (ماننـــد GDPR، AI Act اتحادیه اروپا، قوانین حفاظت از داده‌ها و حقوق مصرف‌کننده) و همچنین اصول اخلاقی همخوانی دارند. این امر منجر به مشروعیت‌بخشی

به استقرار این فناوری‌ها و کاهش ریسک‌های حقوقی و جریمه‌های احتمالی می‌شود.

۲. ارتقاء امنیت و کاهش ریسک‌های چندوجهی

حسابرسی مدل با شناسایی و مستندسازی قابلیت‌ها، محدودیت‌ها، سوگیری‌ها و ریسک‌های فنی ذاتی مدل‌های زبانی بزرگ (LLMs)، و حسابرسی کاربردی با ارزیابی پیامدهای عملیاتی و اجتماعی، به طور مؤثری به کاهش ریسک‌های مرتبط با عملکرد ناصحیح مدل، تولید اطلاعات نادرست (توهم‌زایی)، افشای اطلاعات حساس، سوگیری، یا آسیب‌های اجتماعی-فردی کمک می‌کند. این امر منجر به افزایش امنیت و استحکام فنی محصولات نهایی شده و احتمال بروز خسارات مالی، اعتباری یا آسیب‌های گسترده‌تر را به حداقل می‌رساند.

۳. افزایش شفافیت و اعتمادسازی برای ذینفعان

نتایج حاصل از هر سه سطح حسابرسی، به ویژه مستندسازی دقیق محدودیت‌های مدل (حسابرسی مدل) و شفافیت در مورد عملکرد و اثرات برنامه (حسابرسی کاربردی)، به ایجاد شفافیت در مورد قابلیت‌ها و نحوه عملکرد مدل‌های زبانی بزرگ (LLMs)، و محصولات مبتنی بر آن‌ها کمک می‌کند. این شفافیت، اعتماد کاربران نهایی، نهادهای نظارتی، شرکای تجاری و عموم جامعه را نسبت به این فناوری تقویت نموده و پذیرش اجتماعی آن را تسهیل می‌نماید.

۴. هدایت توسعه مسئولانه و اخلاقی

با در نظر گرفتن الزامات حسابرسی راهبری (اصول اخلاقی سازمانی)، یافته‌های حسابرسی مدل (شناسایی سوگیری‌ها) و تحلیل‌های حسابرسی کاربردی (ارزیابی پیامدهای اخلاقی و اجتماعی)، فرآیند توسعه مدل‌های زبانی بزرگ (LLMs)، و محصولات مرتبط به سمت رعایت اصول اخلاقی مانند انصاف، عدم تبعیض و سوگیری، احترام به حریم خصوصی و پاسخگویی هدایت می‌شوند. این امر تضمین می‌کند که نوآوری‌های فناورانه در راستای منافع بلندمدت جامعه و با در نظر گرفتن ارزش‌های انسانی پیش می‌روند.

۵. توانمندسازی کاربران و ارتقاء سواد دیجیتال

اطلاعات حاصل از ارزیابی‌های شفاف (به ویژه از طریق حسابرسی مدل و کاربردی) به کاربران امکان می‌دهد تا با آگاهی کامل از قابلیت‌ها و محدودیت‌های مدل‌های زبانی بزرگ (LLMs)، از آن‌ها استفاده کنند. این امر به افزایش هوشیاری کاربران در برابر اطلاعات نادرست یا سوءاستفاده‌های احتمالی کمک کرده و در نهایت منجر به توانمندسازی آن‌ها در تعامل با فناوری‌های پیچیده هوش‌مصنوعی می‌شوند.

۶. تدوین استانداردها و ارتقاء شیوه‌های صنعتی

تجمیع نتایج حاصل از حسابرسی‌های متعدد در سطوح مختلف، به تدریج به شکل‌گیری و تکامل استانداردها، چارچوب‌های ارزیابی و بهترین شیوه‌ها برای توسعه، استقرار و مدیریت مدل‌های زبانی بزرگ (LLMs)، در سطح صنعت کمک می‌کند. این امر بستری برای مقایسه، ارزیابی و اطمینان از کیفیت محصولات و خدمات در گستره وسیع‌تری فراهم می‌آورد.

در نهایت، اجرای یک رویکرد حسابرسی سه‌سطحی، اطمینان می‌دهد که مدل‌های زبانی

بزرگ (LLMs)، نه تنها ابزارهای قدرتمندی از نظر فنی هستند، بلکه به صورت مسئولانه، ایمن، اخلاقی و در راســـتای منافع جامعه و کاربران به کار گرفته می‌شوند، که این خود منجر به توسعه پایدار و مورد اعتماد فناوری‌های نوین خواهد شد.

تحقیقات آینده بهتر اســـت بر مطالعات تطبیقی مدل‌های زبانـــی بزرگ داخلی و نیمه داخلی در مقابل مدل‌های زبانی بزرگ خارجی در اشـــکال مختلف قراردادهای حسابرسی، تمرکز کنند. همچنین تحقیقات آینده می‌توانند رویکردهای بالقوه برای تدوین دستورالعمل‌ها و استانداردهایی را بررسی کنند که به اســـتفاده از مدل‌های زبانی بزرگ در حسابرسی می‌پردازد. علاوه بر این، با توجه به اینکه پذیرش مدل‌های زبانی بزرگ بر اعتماد صاحبکاران به موسســـات حسابرسی تأثیر می‌گذارد، بررســـی تأثیر مدل‌های زبانی بزرگ بر پویایی روابط و تعاملات حسابرس-صاحبکاران بسیار مهم است. همچنین به سیاست‌گذاران پیشنهاد می‌شود که مقررات هوش مصنوعی موجود و پیشنهادی را مطابق با چارچوب حسابرسی سه‌سطحی برای رسیدگی به چالش‌ها و ریسک‌های مربوط به مدل‌های زبانی بزرگ به‌روزرسانی کنند.

ملاحظه‌های اخلاقی

در پژوهش حاضر اصول اخلاق امانتداری علمی رعایت و حقوق معنوی مؤلفان محترم شـــمرده می‌شود و نویسندگان اعلام می‌دارند که هیچ‌گونه تضاد منافعی وجود ندارد.

منابع:

علوی، سید کمال؛ نعمتی، علی و دارابی، رویا. (۱۴۰۴). الگوی بهینه و کاربردی فناوری اطلاعات در حسابرســـی بـــا توجه به آزمون های محتوا، کنترل و ریســـک های حسابرســـی. پژوهش‌های حسابرسی حرفه‌ای، ۵(۲۰), ۳۰-۵۷.

نقدی، ســـجاد؛ احمدیان، وحید، فضل زاده، علیرضا و ولی پـــور، احمد.(۱۴۰۳) .طراحی مدل استقرار اثربخش حسابرسی فناوری اطلاعات در سازمان تأمین اجتماعی. پژوهش‌های حسابرسی حرفه‌ای، ۵(۱۸), ۵۹-۳۴.

Adler, P., Falk, C., Friedler, S. A., Nix, T., Rybeck, G., Scheidegger, C., ... & Venkatasubramanian, S. (2018). Auditing black-box models for indirect influence. Knowledge and Information Systems, 54(1), 95-122. DOI: 10.1007/s10115-017-1116-3.

AICPA. (1993). Code of Professional Conduct. New York: AICPA. https://aicpa.org/research/standards/code-of-professional-conduct.html.

Alavi, S. K., Nemati, A., & Darabi, R. (2025). An Optimal and Practical Model of Information Technology in Auditing Considering Substantive and Control Testing, and Audit Risks. (In Persian). DOI: 10.22034/jpar.2024.2031862.1327.

Avin, S., Belfield, H., Brundage, M., Krueger, G., Wang, J., Weller, A., ... & Zilberman, N. (2021). Filling gaps in trustworthy development of AI. Science, 374(6573), 1327-1329. DOI: 10.1126/science.abj0689.

Bandy, J. (2021). Problematic machine behavior: A systematic literature review of algorithm audits. Proceedings of the acm on human-computer interaction, 5(CSCW1), 1-34. DOI: 10.1145/3449144.

Bauer, J. (2016). The necessity of auditing artificial intelligence. SSRN, 577, 1-16. DOI: 10.2139/ssrn.2740657.

Berghout, E., Fijneman, R., Hendriks, L., de Boer, M., & Butijn, B. J. (2023). Advanced Digital auditing: Theory and practice of auditing complex information systems and technologies (p. 256). Springer Nature. DOI: 10.1007/978-3-031-34918-7.

Boskou, G., Kirkos, E., Chatzipetrou, E., Tiakas, E., & Spathis, C. (2024). Exploring the boundaries of financial statement fraud detection with large language models. Available at SSRN 4842962. https://ssrn.com/abstract=4842962.

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., ... & Anderljung, M. (2020). Toward trustworthy AI development: mechanisms for supporting verifiable claims. arXiv preprint arXiv:2004.07213. https://arxiv.org/abs/2004.07213.

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency (pp. 77-91). PMLR. https://proceedings.mlr.press/v81/buolamwini18a.html.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334), 183-186. 10.1126/science.aal4230.

Chopra, A. K., & Singh, M. P. (2018, December). Sociotechnical systems and ethics in the large. In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (pp. 48-53). https://dl.acm.org/doi/10.1145/3278721.3278727.

Contractor, D., McDuff, D., Haines, J. K., Lee, J., Hines, C., Hecht, B., ... & Li, H. (2022, June). Behavioral use licensing for responsible AI. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 778-788). DOI: 10.1145/3531146.3533084.

Deloitte. (2024). How we harness Generative AI for a smarter, digital audit. https://www2.deloitte.com/us/en/pages/strategy/articles/harnessing-generative-ai-for-a-smarter-digital-audit.html.

Dong, M. M., Stratopoulos, T. C., & Wang, V. X. (2024). A scoping review of ChatGPT research in accounting and finance. International Journal of Accounting Information Systems, 55, 100715. DOI: 10.1016/j.isa.2024.100715.

Elad, F., & Mugwira, T. (2025). Classifying and Evaluating Large Language Models in External Audits. Available at SSRN 5325170. https://ssrn.com/abstract=5325170.

Emett, S., Eulerich, M., Lipinski, E., Prien, N., & Wood, D. A. (2024). Leveraging ChatGPT for Enhancing the Internal Audit Process—A Real-world Example from Uniper, a Large Multinational Company. Accounting Horizons, 1-11. https://doi.org/10.2308/HORIZONS-2023-094.

Engler, A. (2023). Early thoughts on regulating generative AI like ChatGPT. https://www.brookings.edu/articles/early-thoughts-on-regulating-generative-ai-like-chatgpt/.

Engler, A. C. (2021). Independent auditors are struggling to hold AI companies accountable. Fast Company. https://www.fastcompany.com/90695092/independent-auditors-are-struggling-to-hold-ai-companies-accountable.

Eulerich, M., & Wood, D. A. (2025). A Demonstration of How ChatGPT and Generative AI Can be Used in the Internal Auditing Process. Journal of Emerging Technologies in Accounting, 1-31. https://doi.org/10.2308/JETA-2023-030.

Eulerich, M., Sanatizadeh, A., Vakilzadeh, H., & Wood, D. A. (2024). Is it all hype? ChatGPT's performance and disruptive potential in the accounting and auditing industries.

Review of Accounting Studies, 29(3), 2318-2349. DOI: 10.2308/RET-2349-2349.

Evans, O., Cotton-Barratt, O., Finnveden, L., Bales, A., Balwit, A., Wills, P., ... & Saunders, W. (2021). Truthful AI: Developing and governing AI that does not lie. arXiv preprint arXiv:2110.06674. https://arxiv.org/abs/2110.06674.

Falco, G., Shneiderman, B., Badger, J., Carrier, R., Dahbura, A., Danks, D., ... & Yeong, Z. K. (2021). Governing AI safety through independent audits. Nature Machine Intelligence, 3(7), 566-571. DOI: 10.1038/s42256-021-00383-x.

Flint, D. (1988). Philosophy and principles of auditing: an introduction. (No Title).

Floridi, L., Holweg, M., Taddeo, M., Amaya, J., Mökander, J., & Wen, Y. (2022). CapAI-A procedure for conducting conformity assessment of AI systems in line with the EU artificial intelligence act. Available at SSRN 4064091. https://ssrn.com/abstract=4064091.

Föhr, T. L., Schreyer, M., Juppe, T. A., & Marten, K. U. (2023). Assuring sustainable futures: Auditing sustainability reports using AI foundation models. Available at SSRN 4502549. https://ssrn.com/abstract=4502549.

Fotoh, L. E., & Mugwira, T. (2025). Exploring Large Language Models in external audits: Implications and ethical considerations. International Journal of Accounting Information Systems, 56, 100748. DOI: 10.1016/j.accinfosec.2025.100748.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92. DOI: 10.1145/3442188.

Goel, K., Rajani, N., Vig, J., Tan, S., Wu, J., Zheng, S., ... & Ré, C. (2021). Robustness gym: Unifying the NLP evaluation landscape. arXiv preprint arXiv:2101.04840. https://arxiv.org/abs/2101.04840.

Green, N., Procope, C., Cheema, A., & Adediji, A. (2022). System cards, a new resource for understanding how AI systems work. Meta AI. https://ai.meta.com/blog/system-cards-understanding-ai-systems/.

Green, R. M., & Donovan, A. (2009). The methods of business ethics.

Gu, H., Schreyer, M., Moffitt, K., & Vasarhelyi, M. (2024). Artificial intelligence co-piloted auditing. International Journal of Accounting Information Systems, 54, 100698. https://doi.org/10.1016/j.accinf.2024.100698.

Gu, H., Schreyer, M., Moffitt, K., & Vasarhelyi, M. (2024). Artificial intelligence co-piloted auditing. International Journal of Accounting Information Systems, 54, 100698.

Hodges, C. (2015). Ethics in business practice and regulation. Law and Corporate Behaviour: Integrating Theories of Regulation, Enforcement, Compliance and Ethics, 1-21. https://www.elgaronline.com/abstract/9781784710690.00010.xml.

IESBA. 2023. Handbook of the international code of ethics for professional accountants. New York: International Federation of Accountants, Professional Code. https://www.ifac.org/system/files/publications/files/IESBA-2023-Handbook-Code-of-Ethics.pdf.

Iliescu, F. M. (2010). Auditing IT governance. Informatica Economica, 14(1), 93. http://www.revistaie.ase.ro/content/53/09%20-%20Iliescu.pdf.

Kirchenbauer, J., Geiping, J., Wen, Y., Katz, J., Miers, I., & Goldstein, T. (2023, July). A watermark for large language models. In International Conference on Machine Learning (pp. 17061-17084). PMLR. https://proceedings.mlr.press/v202/kirchenbauer23a.html

LaBrie, R. C., & Steinke, G. (2019). Towards a framework for ethical audits of AI algorithms. https://www.researchgate.net/publication/337828146_Towards_a_framework_for_ethical_

audits_of_AI_algorithms.

Laux, J., Wachter, S., & Mittelstadt, B. (2021). Taming the few: Platform regulation, independent audits, and the risks of capture created by the DMA and DSA. Computer law & Security review, 43, 105613. https://doi.org/10.1016/j.clsr.2021.105613.

Li, H., & Vasarhelyi, M. A. (2024). Applying large language models in accounting: A comparative analysis of different methodologies and off-the-shelf examples. Journal of Emerging Technologies in Accounting, 21(2), 133-152. https://doi.org/10.2308/JETA-2023-043.

Li, H., Freitas, M. M. D., Lee, H., & Vasarhelyi, M. (2024). Enhancing continuous auditing with large language models: AI-assisted real-time accounting information cross-verification. Available at SSRN 4692960. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4692960.

Marda, V., & Narayan, S. (2021). On the importance of ethnographic methods in AI research. Nature Machine Intelligence, 3(3), 187-189. https://doi.org/10.1038/s42256-021-00316-8.

Metaxa, D., Park, J. S., Robertson, R. E., Karahalios, K., Wilson, C., Hancock, J., & Sandvig, C. (2021). Auditing algorithms: Understanding algorithmic systems from the outside in. Foundations and Trends® in Human–Computer Interaction, 14(4), 272-344. https://doi.org/10.1561/1100000077.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019, January). Model cards for model reporting. In Proceedings of the conference on fairness, accountability, and transparency (pp. 220-229). https://doi.org/10.1145/3287560.3287596.

Mökander, J., Axente, M., Casolari, F., & Floridi, L. (2022). Conformity assessments and post-market monitoring: a guide to the role of auditing in the proposed European AI regulation. Minds and Machines, 32(2), 241-268. https://doi.org/10.1007/s11023-022-09598-y.

Mökander, J., Schuett, J., Kirk, H. R., & Floridi, L. (2024). Auditing large language models: a three-layered approach. AI and Ethics, 4(4), 1085-1115. https://doi.org/10.1007/s43681-024-00393-y.

Mökander, J., Sheth, M., Gersbro-Sundler, M., Blomgren, P., & Floridi, L. (2022). Challenges and best practices in corporate AI governance: Lessons from the biopharmaceutical industry. Frontiers in Computer Science, 4, 1068361. https://doi.org/10.3389/fcomp.2022.1068361.

Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. (2020). The ethical implications of using artificial intelligence in auditing. Journal of business ethics, 209-234. https://doi.org/10.1007/s10551-019-04407-1.

Naghdi, S., Fazlzadeh, A., & Valipour, A. (2025). Designing an Effective Deployment Model of Information Technology Audit in the Social Security Organization. Professional Auditing Research, 5(18), 34-59. (In Persian). DOI:10.22034/jpar.2024.2015837.1243.

O'Leary, D. E. (2023). Enterprise large language models: Knowledge characteristics, risks, and organizational activities. Intelligent systems in accounting, finance and management, 30(3), 113-119. https://doi.org/10.1002/isaf.1542.

Otero, A. R., & Agu, M. (2025). Benefits and Drawbacks of Incorporating ChatGPT in Financial Audits. Current Issues in Auditing, 1-8.

Papachristou, G., & Bekiaris, M. (2020). Ethics in auditing. The Routledge Handbook of Accounting Ethics, 169-185. https://doi.org/10.4324/9780429200606-15.

Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its (dis) contents: A survey of dataset development and use in machine learning research. Patterns, 2(11). https://doi.org/10.1016/j.patter.2021.100336.

Raji, I. D., Xu, P., Honigsberg, C., & Ho, D. (2022, July). Outsider oversight: Designing a third

party audit ecosystem for ai governance. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (pp. 557-571). https://doi.org/10.1145/3514094.3534168.

Ross, M., & Zhang, J. (2025). ChatGPT is Ready for the Profession—But is the Profession Ready for ChatGPT? Accounting Horizons, 39(2), 185-203. https://doi.org/10.2308/HORIZONS-2023-057.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. Data and discrimination: converting critical concerns into productive inquiry, 22(2014), 4349-4357. https://pdfs.semanticscholar.org/1f3d/c9a2dba0d5b3cef9a6a8c4d6c6d2d1e8d3e5.pdf.

Senft, S., & Gallegos, F. (2008). Information technology control and audit. Auerbach publications. https://www.taylorfrancis.com/books/mono/10.1201/9781420065506/information-technology-control-audit-sandra-senft-frederick-gallegos.

Stodt, J., & Reich, C. (2021). Machine learning development audit framework: assessment and inspection of risk and quality of data, model and development process. International Journal of Computer and Information Engineering, 15(3), 187-193. http://www.waset.org/publications/10012046

Street, D., & Wilck, J. (2023). 'Let's Have a Chat': Principles for the Effective Application of ChatGPT and Large Language Models in the Practice of Forensic Accounting. Journal of Forensic and Investigative Accounting, July to December. https://doi.org/10.2308/JFIA-2023-011.

Street, D., Wilck, J., & Chism, Z. (2023). Six Principles for the Effective Use of Artificial Intelligence Large Language Models: How to Leverage ChatGPT, Bard, and Bing Chat in Accounting Work. The CPA Journal, 93(11-12), 50-57. https://www.cpajournal.com/2023/11/13/six-principles-for-the-effective-use-of-artificial-intelligence-large-language-models/.

Tamkin, A., Brundage, M., Clark, J., & Ganguli, D. (2021). Understanding the capabilities, limitations, and societal impact of large language models. https://arxiv.org/abs/2102.02503.

Vasarhelyi, M. A., Moffitt, K. C., Stewart, T., & Sunderland, D. (2023). Large language models: An emerging technology in accounting. Journal of emerging technologies in accounting, 20(2), 1-10. https://doi.org/10.2308/JETA-2023-010.

Wang, Z.J., Choi, D., Xu, S., Yang, D. (2021). Putting Humans in the Natural Language Processing Loop: A Survey. In: Proceedings of the First Workshop on Bridging Human–Computer Interaction and Natural Language Processing, pp. 47–52. https://aclanthology.org/2021.hcinlp-1.7/.

Wei, T., Wu, H., & Chu, G. (2023). Is ChatGPT competent? Heterogeneity in the cognitive schemas of financial auditors and robots. International Review of Economics & Finance, 88, 1389-1396. https://doi.org/10.1016/j.iref.2023.03.018.

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P. S., Mellor, J., ... & Gabriel, I. (2022, June). Taxonomy of risks posed by language models. In Proceedings of the 2022 ACM conference on fairness, accountability, and transparency (pp. 214-229). https://doi.org/10.1145/3531146.3533088.

Zhao, J., & Wang, X. (2024). Unleashing efficiency and insights: Exploring the potential applications and challenges of ChatGPT in accounting. Journal of Corporate Accounting & Finance, 35(1), 269-276. https://doi.org/10.1002/jcaf.22688.

پی‌نویس:

1. Long Language Models
2. Gu et al.
3. Eulerich et al.
4. Emett et al.
5. Ross & Zhang
6. Eulerich & Wood
7. Deloitte
8. American Institute of Certified Public Accountants (AICPA)
9. Papachristou & Bekiaris
10. Natural language processing
11. Machine learning
12. Tamkin et al.
13. Avin et al.
14. Wang et al.
15. Marda & Narayan
16. Mitchell et al.
17. Gebru et al.
18. Green et al.
19. Kirchenbauer et al.
20. Engler
21. Dong et al
22. Föhr et al.
23. BERT که از حروف اول عبارت Bidirectional Encoder Representations from Transformers گرفته شده یک مدل زبانی دوطرفه است، به این معنی که برای پیش‌بینی هر کلمه در جمله، به تمام کلمات قبل و بعد از آن مراجعه می‌کند و از اطلاعات موجود در آن‌ها استفاده می‌کند. این ویژگی باعث می‌شود که مدل BERT برای تمام وظایف پردازش زبان طبیعی، از جمله تشخیص احساسات، ترجمه ماشینی و پاسخ به پرسش‌ها کارآمد باشد. الگوریتم BERT روش ابداعی گوگل برای آموزش پردازش زبان طبیعی (NLP) به کامپیوترها است.
24. Generative Pre-trained Transformer
25. Named Entity Recognition
26. Elad & Mugwira
27. Ernst & Young (EY)
28. PricewaterhouseCoopers (PWC)
29. KPMG
30. Fotoh & Mugwira
31. Li & Vasarhelyi
32. Boskou et al.
33. Zhao & Wang
34. Oteri & Agu
35. Wei et al.
36. O'Leary
37. Vasarhelyi et al.
38. Li et al.
39. Street et al.

40. IESBA
41. Street & Wilck

۴۲. اتاق پژواک، توصیفی کنایی از وضعیتی است که در آن اطلاعات، ایده‌ها، یا باورها با ارتباطات و تکرار درون یک سامانه تعریف شده تقویت می‌شوند. این اتفاق معمولاً به دلیل الگوریتم‌های شبکه‌های اجتماعی، انتخاب‌های شخصی افراد در دنبال کردن صفحات و افراد همفکر، و همچنین تعامل بیشتر با کسانی که نظرات مشابه دارند، رخ می‌دهد.

43. Munoko et al.
44. International Financial Reporting Standards
45. International standards on auditing
46. Generally accepted accounting principles
47. Flint
48. LaBrie & Steinke
49. Sandvig
50. Stodt & Reich
51. Adler et al.
52. Laux et al.
53. Bandy
54. Metaxa et al.
55. Berghout et al.
56. Mökander & Axente
57. Mökander et al.
58. Brundage et al.
59. Hodges
60. Iliescu
61. alco et al.
62. Engler
63. Bauer
64. Chopra & Singh
65. Contractor et al.
66. Buolamwini & Gebru
67. Senft & Gallegos
68. Floridi et al.
69. Paullada et al.
70. Goel et al.
71. Weidinger et al.
72. Evans et al.
73. Caliskan et al.
74. Raji et al.