Islamic Sciences and Culture Academy Institute of Family and Theology

Islamic Sciences and Culture Academy

Elssn:2251-7898
Issn:2676-6760

Identifying the Challenges of Generative Artificial Intelligence in Organizations: Emphasis on Ethical and Sustainability Dimensions*

Sanaz Shafiee[1]
1- Assistant Professor in Payame Noor University
s.shafiei@pnu.ac.ir

Abstract

With the rapid growth of generative artificial intelligence (AI) technologies and their increasing applications across diverse domains, the need to examine the implications and challenges arising from their adoption has become more pressing. The aim of this study is to identify and analyze the issues associated with the use of generative AI in organizations. This qualitative research employs thematic analysis, and data were collected through semi-structured interviews with experts. The findings indicate that the identified challenges fall into six main themes: transparency and accountability; legal; security and governance; human and cultural; content quality and credibility; and economic and operational dimensions. These six areas encompass a total of twenty-four subthemes that span a wide range of technical, legal, organizational, and ethical concerns. A deeper examination revealed that many of these issues stem from the absence of clear and

*Cite this article: Shafiee, S.(2025). C Identifying the Challenges of Generative Artificial Intelligence in Organizations: Emphasis on Ethical and Sustainability Dimensions,Journal in Applied Ethics Studies, 3(81), 119-156.
https://doi.org/1022081/jf.2025.72615.2076.

▣ **Publisher**: Islamic Propagation Office of the Seminary of Qom (Islamic Sciences and Culture Academy, Isfahan, Iran). ***Type of article**: Research Article
▣ **Received**:2025/08/13 •**Revised**:2025/09/05 •**Accepted**:2025/09/21 •**Published online:**2025/12/05

ethical governance frameworks, poor quality and diversity of training data, ambiguity in model decision-making processes, and the lack of organizational readiness in terms of culture and structure to adopt emerging technologies. The results of this study can serve as a practical guide for managers, policymakers, and developers in pursuing responsible and sustainable implementation of generative AI.

Keywords
generative artificial intelligence, technology ethics, data governance, organizational issues, digital sustainability.

دراسات في الأخلاق التطبيقية
مجلة فصلية محكّمة
معهد العلوم والثقافة الإسلامية
معهد الإلهيات والأسرة
Elssn:۲۲۵۱-۷۸۹۸
Issn:۲۶۷۶-۶۷۶۰

البحث عن تحديد وإدارة تحديات الذكاء الاصطناعي التوليدي في المنظمات مع التركيز على الجوانب الأخلاقية والاستدامة*

ساناز شفيعي [1]

١- أستاذة مساعدة لقسم إدارة تكنولوجيا المعلومات، جامعة بيام نور. طهران ـ إيران.
s.shafiei@pnu.ac.ir

ملخص

مع النمو السريع لتقنيات الذكاء الاصطناعي التوليدي وتوسّع تطبيقاتها في مجالات متعددة، أصبح من الواضح بشكل متزايد ضرورة دراسة العواقب والتحديات الناتجة عن توظيف هذه التكنولوجيا. يهدف هذا البحث إلى تحديد وتحليل التحديات المرتبطة بتطبيق الذكاء الاصطناعي التوليدي في المنظمات. اعتمدت الدراسة منهجاً نوعياً باستخدام التحليل الموضوعاتي، وجُمعت البيانات من خلال مقابلات شبه مهيكلة مع خبراء. أظهرت النتائج أن التحديات المحددة يمكن تصنيفها ضمن ستة محاور رئيسية: الشفافية والمساءلة؛ القضايا القانونية والأمنية والحوكمة؛ الجوانب البشرية والثقافية؛ جودة ومصداقية المحتوى؛

*الاستناد إلى هذه المقالة: شفيعي، ساناز (٢٠٢٥). البحث عن تحديد وإدارة تحديات الذكاء الاصطناعي التوليدي في المنظمات مع التركيز على الجوانب الأخلاقية والاستدامة. دراسات في الأخلاق التطبيقية، ٣(٨١)، صص١١٩-١٥٦.
https://doi.org/1022081/jf.2025.72615.2076.

نوع المقال: بحثي؛ الناشر: مركز الدعوة الإسلامية في حوزة قم (معهد العلوم و الثقافة الإسلامية، أصفهان، إيران) ©المؤلفون

تاريخ الاستلام: ٢٠٢٥/٠٨/١٣ • تاريخ التعديل: ٢٠٢٥/٠٩/٠٥ • تاريخ القبول: ٢٠٢٥/٠٩/٢١ • تاريخ الإصدار: ٢٠٢٥/١٢/٠٥

والاعتبارات الاقتصادية والتشغيلية. وتشمل هذه المحاور الستة أربعة وعشرين محوراً فرعياً تعالج مجموعة واسعة من القضايا التقنية والقانونية والتنظيمية والأخلاقية. وأظهر تحليلٌ أعمق أن العديد من هذه التحديات تنبع من غياب أطر حوكمة واضحة وأخلاقية، وضعف جودة البيانات التعليمية ونقص تنوعها، وغموض عمليات اتخاذ القرار في النماذج، وعدم جاهزية المنظمات من الناحية الثقافية والبنيوية لتبني التقنيات الناشئة. يمكن أن تُعدّ هذه النتائج دليلاً عملياً للمديرين وواضعي السياسات والمطورين لاعتماد الذكاء الاصطناعي التوليدي بشكل مسؤول ومستدام.

الكلمات المفتاحية

الذكاء الاصطناعي التوليدي، أخلاقيات التكنولوجيا، حوكمة البيانات، التحديات التنظيمية، الاستدامة الرقمية.

مطالعات اخلاق کاربردی
فصلنامه علمی-پژوهشی
پژوهشگاه علوم و فرهنگ اسلامی
پژوهشکده ادبیات و خانواده
Elssn:۲۲۵۱-۷۸۹۸
Issn:۲۶۷۶-۶۷۶۰

شناسایی مسائل هوش مصنوعی مولد در سازمان‌ها با تأکید بر ابعاد اخلاقی و پایدار*

ساناز شفیعی
۱- استادیار گروه مدیریت فناوری اطلاعات، دانشگاه پیام نور، تهران، ایران.
s.shafiei@pnu.ac.ir

چکیده

با رشد شتابان فناوری‌های هوش مصنوعی مولد و گسترش روزافزون کاربردهای آن در حوزه‌های گوناگون، ضرورت بررسی پیامدها و مسائل ناشی از به‌کارگیری این فناوری بیش از پیش آشکار شده است. هدف این پژوهش، شناسایی و تحلیل مسائل مربوط به به‌کارگیری هوش مصنوعی مولد در سازمان‌هاست. رویکرد پژوهش، کیفی و مبتنی بر تحلیل مضمون است و داده‌ها از طریق مصاحبه‌های نیمه‌ساختاریافته با خبرگان گردآوری شده است. یافته‌ها نشان می‌دهد مسائل شناسایی‌شده در قالب شش مضمون اصلی شفافیت و پاسخ‌گویی، حقوقی، امنیتی و حاکمیتی، انسانی و فرهنگی، کیفیت و اعتبار محتوا، و

*استناد به این مقاله: شفیعی، ساناز (۱۴۰۴). شناسایی مسائل هوش مصنوعی مولد در سازمان‌ها با تأکید بر ابعاد اخلاقی و پایدار، مطالعات اخلاق کاربردی، ۳ (۸۱)، صص ۱۱۹-۱۵۶.
https://doi.org/1022081/jf.2025.72615.2076.
نوع مقاله: پژوهشی؛ ناشر: دفتر تبلیغات اسلامی حوزه علمیهٔ قم (پژوهشگاه علوم و فرهنگ اسلامی، اصفهان، ایران) © نویسندگان
تاریخ دریافت: ۱۴۰۴/۰۵/۲۲ تاریخ اصلاح: ۱۴۰۴/۰۶/۱۴ تاریخ پذیرش: ۱۴۰۴/۰۶/۳۰ تاریخ انتشار آنلاین: ۱۴۰۴/۰۹/۱۴

اقتصادی و عملیاتی دسته‌بندی می‌شوند. این شش حوزه در مجموع ۲۴ خرده‌مضمون را شامل می‌شوند که طیفی از مسائل فنی، حقوقی، سازمانی و اخلاقی را پوشش می‌دهند. بررسی عمیق‌تر نشان داد بسیاری از این مسائل ناشی از نبود چارچوب‌های حکمرانی اخلاقی و شفاف، ضعف کیفیت و تنوع داده‌های آموزشی، ابهام در فرایند تصمیم‌گیری مدل‌ها، و آمادگی‌نداشتن سازمان‌ها از نظر فرهنگ و ساختار برای پذیرش فناوری‌های نوین است. از نتایج این پژوهش می‌توان به‌عنوان راهنمای عملی برای مدیران، سیاست‌گذاران و توسعه‌دهندگان در مسیر بهره‌گیری مسئولانه و پایدار از هوش مصنوعی مولد استفاده کرد.

کلیدواژه‌ها

هوش مصنوعی مولد، اخلاق فناوری، حکمرانی داده، مسائل سازمانی، پایداری دیجیتال.

مقدمه

هوش مصنوعی در سال‌های اخیر به یکی از مهم‌ترین محرک‌های تحول دیجیتال تبدیل شده است. پیشرفت‌های سریع در الگوریتم‌های یادگیری عمیق، پردازش زبان طبیعی و بینایی ماشین باعث شده سازمان‌ها از این فناوری نه‌تنها برای تحلیل داده و پیش‌بینی، بلکه برای خلق محتوای نوآورانه نیز بهره ببرند. در این میان، هوش مصنوعی مولد[1] به‌عنوان یکی از شاخه‌های برجسته و نوظهور هوش مصنوعی مطرح شده است که توانایی تولید محتوای جدید شامل متن، تصویر، صدا و ویدئو را دارد (Macdonald et al., 2023). ظهور ابزارهایی مانند چت جی‌پی‌تی، میدجرنی[2] و دال ای[3] نشان داد مرزهای خلاقیت دیجیتال و ظرفیت پردازش رایانه‌ای به‌شکل بی‌سابقه‌ای گسترش یافته است (Floridi, 2023).

با توجه به اینکه فناوری‌های مولد می‌توانند انواع گوناگونی از محتوا را ایجاد کنند، سازمان‌ها با پرسش‌های جدی اخلاقی و حقوقی دربارهٔ محدوده و نحوهٔ به‌کارگیری آن مواجه‌اند (Singh et al., 2024). تولید محتوای نادرست، سوگیری الگوریتمی، نقض حریم خصوصی و ابهام در مسئولیت‌پذیری تنها بخشی از مخاطراتی است که می‌تواند پیامدهای منفی قابل توجهی برای اعتبار، امنیت و پایداری سازمان‌ها داشته باشد (Weidinger et al., 2021Jobin et al., 2019)؛ برای مثال، گزارش‌های بین‌المللی هشدار می‌دهند بدون وجود سازِوکارهای حکمرانی شفاف، استفاده از مدل‌های مولد می‌تواند به گسترش اطلاعات نادرست، تبعیض ساختاری و کاهش اعتماد عمومی منجر شود (OECD, 2024).

در پاسخ به این نگرانی‌ها، نهادهای بین‌المللی مانند اتحادیهٔ اروپا با قانون هوش مصنوعی[4] (European Cmmission, 2021) و یونسکو با توصیه‌نامهٔ اخلاق هوش مصنوعی[5] (UNESCO, 2022) تلاش کرده‌اند چارچوب‌هایی برای توسعه و استفادهٔ

1. Generative Artificial Intelligence.
2. Midjourney.
3. DALL·E.
4. EU AI Act, 2024.
5. UNESCO, 2021.

مسئولانه ارائه کنند. بااین‌حال، اجرای این چارچوب‌ها در سطح سازمانی همچنان با موانع عملی و فرهنگی همراه است (Cath, 2018).

از منظر مدیریتی نیز نبود سیاست‌های داخلی روشن، کمبود آموزش اخلاقی کارکنان و نبود نظام ارزیابی ریسک، از عواملی است که مانع استفادهٔ اثربخش از این فناوری در سازمان‌ها می‌شود (Dwivedi et al., 2023). این شرایط، نیاز به مطالعات عمیق و اکتشافی را برجسته می‌کند تا از طریق شناسایی دقیق مسائل اخلاقی، بتوان مسیر تدوین سیاست‌ها، رویه‌ها و آموزش‌های متناسب را فراهم ساخت.

حوزهٔ هوش مصنوعی در سال‌های اخیر رشد چشمگیری داشته است؛ از توسعهٔ فناوری‌های نو و افزایش سرمایه‌گذاری (بیش از نودمیلیارد دلار در آمریکا در سال ۲۰۲۱) تا جهش در انتشار مقالات و ثبت اختراعات، به‌ویژه در یادگیری ماشین، بینایی کامپیوتر و پردازش زبان. این رونق، علاوه بر مزایا، نگرانی‌های اخلاقی جدی مانند نقض حریم خصوصی، گسترش نظارت، هزینه‌های زیست‌محیطی و تعمیق تبعیض را ایجاد کرده و به بحث دربارهٔ اخلاق هوش مصنوعی منجر شده است. پرسش کلیدی این است که چه اصول اخلاقی باید توسعهٔ هوش مصنوعی را هدایت کنند (Corrêa et al., 2023). براین‌اساس، پژوهش حاضر با رویکرد کیفی و مبتنی بر مصاحبه با خبرگان، در پی شناسایی و تحلیل مسائل چندبعدی استفاده از هوش مصنوعی مولد در سازمان‌ها، با تأکید ویژه بر ابعاد اخلاقی و پایداری، و ارائهٔ راهکارهای مدیریتی برای مواجهه با آن است. نتایج این مطالعه می‌تواند مبنایی برای طراحی چهارچوب‌های بومی‌سازی‌شدهٔ حکمرانی فراهم کند؛ به‌نحوی که ضمن حفظ مزایای نوآوری، از مخاطرات احتمالی آن نیز بکاهد.

پیشینهٔ پژوهش

در سال‌های اخیر، توجه گسترده‌ای به ابعاد اخلاقی و اجتماعی هوش مصنوعی مولد شده است (Weidinger et al., 2021). پژوهشگران، با مرور نظام‌مند، ۲۱ نوع آسیب را شناسایی کردند؛ ازجمله تولید اطلاعات نادرست، تشدید سوگیری‌های اجتماعی، نقض

حریم خصوصی، آسیب‌های زیست‌محیطی و مسائل مالکیت فکری. آنان نتیجه گرفتند حتی مدل‌های پیشرفته، بدون سیاست‌های روشن و نظارت انسانی، می‌توانند پیامدهای منفی گسترده‌ای داشته باشند (Dwivedi et al., 2023)؛ همچنین، پژوهشگران پیامدهای چندبعدی سامانه‌های مولد را در بستر سازمانی تحلیل کردند. یافته‌های آنان نشان داد این فناوری‌ها می‌توانند اعتماد عمومی را تضعیف و امنیت شغلی را تهدید کنند و مشکلات حقوقی پدید آورند؛ ازاین‌رو، ضرورت آموزش کارکنان و تدوین سیاست‌های داخلی آشکارتر می‌شود.

همچنین کات (Cath, 2018) با رویکرد کیفی و مصاحبه با مدیران و سیاست‌گذاران، چارچوبی سه‌لایه برای حکمرانی اخلاقی ارائه کرد که شامل سیاست‌گذاری داخلی، ارزیابی آثار اخلاقی پیش از استقرار و نظارت مستمر است.

در سطح سیاست‌گذاری، دو چارچوب راهبردی تأثیرگذار معرفی شده‌اند: نخست، چارچوب مدیریت خطر هوش مصنوعی (NIST AI RMF 1.0) که با رویکرد داوطلبانه به سازمان‌ها کمک می‌کند مخاطرات فنی و اجتماعی را شناسایی و مدیریت کنند و بر مشارکت ذی‌نفعان در چرخهٔ عمر هوش مصنوعی تأکید دارد؛ دوم، قانون هوش مصنوعی اتحادیهٔ اروپا که با رویکرد مبتنی بر خطر، برخی کاربردها را پرمخاطره دانسته و تکالیفی برای شفاف‌سازی و انطباق ارائه‌دهندگان وضع می‌کند (European AI, 2023 ; Cmmission, 2021). در کنار این چهارچوب‌ها، پژوهش‌های تجربی و مروری همچنان بر تداوم مسائلی چون سوگیری الگوریتمی، نبود شفافیت و ابهام در مسئولیت‌پذیری تأکید می‌کنند (Raghavan et al., 2020Chen, 2023).

مرورهای تحلیلی اخیر (۲۰۲۴–۲۰۲۵)، خوشه‌ای از مخاطرات اصلی را در هوش مصنوعی مولد شناسایی کرده‌اند؛ شامل مسائل حریم خصوصی و امنیت داده، مالکیت فکری و کپی‌رایت، تولید محتوای جعلی و دیپ‌فیک، و تقویت سوگیری‌ها. این مرورها همچنین راهکارهایی مانند شفاف‌سازی منبع محتوا، ممیزی مدل‌ها، حاکمیت داده و ارزیابی تأثیرات اجتماعی را برای کاهش احتمال خطر پیشنهاد داده‌اند

(a; Surbakti, 2025Al-kfairy et al., 2024). از منظر اقتصادی و رقابتی نیز گزارش‌های سازمان‌های بین‌المللی به‌طور هم‌زمان بر ظرفیت‌های بهره‌وری و مخاطرات ناشی از تمرکز داده و منابع محاسباتی تأکید کرده‌اند؛ موضوعی که شاید نیازمند مداخلات تنظیم‌گرانه برای محافظت از منافع عمومی باشد (OECD, 2024).

در ادبیات مدیریتی، پژوهش‌های جدید به بررسی اصول اخلاقی در استقرار هوش مصنوعی مولد پرداخته‌اند. این اصول شامل دقت، شفافیت، انصاف، پاسخ‌گویی، امنیت و احترام به حریم خصوصی است و نشان می‌دهد موفقیت سازمان‌ها در استفاده از این فناوری نیازمند ترکیب سازوکارهای حاکمیتی، کنترل‌های فنی و توانمندسازی سرمایهٔ انسانی است (Rana et al., 2024).

با آنکه پژوهش‌های حوزهٔ هوش مصنوعی در ایران هنوز در مراحل ابتدایی قرار دارند، برخی از پژوهشگران به جنبه‌های اخلاقی و حقوقی هوش مصنوعی پرداخته‌اند؛ برای مثال، فرزین و سمیعی (۱۴۰۲) در مقالهٔ «چالش‌های اخلاقی و حقوقی استفاده از هوش مصنوعی در کسب‌وکارهای دیجیتال»، مسائل اخلاقی و حقوقی کسب‌وکارهای دیجیتال را بررسی و مسائلی مانند نقض حریم خصوصی، تبعیض الگوریتمی و مسئولیت‌پذیری را شناسایی کرده‌اند. عباسی و تیموری (۱۴۰۲) در مقاله‌ای با عنوان «مروری بر چالش‌های اخلاقی و حقوقی کاربرد هوش مصنوعی در نظام سلامت»، در حوزهٔ سلامت به مسائلی چون اعتماد به هوش مصنوعی، کرامت انسانی و استقلال فردی اشاره کرده‌اند. همچنین، قوامی‌پور سرشکه و محمودی (۱۴۰۳) در مقالهٔ «واکاوی چالش‌های پیاده‌سازی هوش اخلاقی در هوش مصنوعی» به دشواری‌های اعمال قواعد اخلاقی در سامانه‌های هوشمند پرداخته‌اند.

علاوه بر این، رحیمی‌اقدم و همکاران (۱۴۰۳)، در مقاله‌ای با عنوان «چالش‌های اخلاقی اتخاذ هوش مصنوعی در مدیریت منابع انسانی»، مسائل اخلاقی به‌کارگیری هوش مصنوعی در مدیریت منابع انسانی را در شش دسته از جمله نقض عدالت سازمانی، کاهش خودمختاری کارکنان و تهدید حریم خصوصی شناسایی کرده‌اند.

رحیم زارعی (۱۴۰۴) نیز در مطالعه‌ای مروری با عنوان «بررسی چالش‌های اخلاقی در سیستم‌های هوشمند و راهکارهای جلوگیری از آن‌ها»، مشکلاتی چون نبود شفافیت، سوگیری و ابهام در مسئولیت‌پذیری را بررسی و راهکارهایی مانند مدل‌های قابل تفسیر، آموزش اخلاقی مدیران و تدوین چارچوب‌های قانونی جهانی پیشنهاد داده است.

جمع‌بندی پیشینه نشان می‌دهد در سطح بین‌المللی چارچوب‌های کلان و مطالعات گسترده‌ای دربارهٔ مسائل اخلاقی هوش مصنوعی مولد انجام شده است؛ اما پژوهش‌های عمیق در بستر سازمانی و بین‌صنعتی هنوز محدودند. در سطح داخلی نیز بیشتر مطالعات به‌طور کلی به اخلاق هوش مصنوعی پرداخته‌اند و پژوهش‌های ویژه و کیفی در زمینهٔ هوش مصنوعی مولد بسیار اندک است. این خلأ پژوهشی، ضرورت مطالعه‌ای جامع و اکتشافی را برجسته می‌سازد؛ لازم است با استفاده از دیدگاه خبرگان حوزه‌های فناوری، مدیریت، حقوق و اخلاق، مسائل اخلاقی هوش مصنوعی مولد در سازمان‌ها شناسایی و تحلیل گردد.

ادبیات نظری پژوهش

تا چندی پیش، هوش مصنوعی بیشتر مفهومی در داستان‌های علمی‌تخیلی مطرح بود؛ اما این فناوری در مدت کوتاهی به یکی از محورهای اصلی تحولات دیجیتال تبدیل شد و سازمان‌ها را واداشت تا به بررسی کاربردهای عملی آن در حوزه‌های مختلف بپردازند. اوج توجه جهانی به این فناوری با معرفی چت جی‌پی‌تی در پایان سال ۲۰۲۲ شکل گرفت.

هوش مصنوعی مولد زیرشاخه‌ای از هوش مصنوعی است که با بهره‌گیری از مدل‌های آماری و شبکه‌های عصبی پیشرفته، قادر به تولید داده‌های جدید و منحصربه‌فردی است که شباهت زیادی به داده‌های واقعی دارند (Cao et al., 2025). این فناوری با استفاده از

مدل‌های بزرگ می‌تواند انواع تبدیل‌ها را انجام دهد؛ مانند متن به متن، متن به تصویر، تصویر به متن، متن به کد، متن به صوت، متن به ویدئو یا حتی متن به مدل سه‌بعدی (Hagendorff, 2024).این فناوری براساس مدل‌های یادگیری عمیق، ازجمله مدل‌های زبانی بزرگ[1]، مدل‌های انتشار پایدار[2] و شبکه‌های مولد تخاصمی[3]، توسعه یافته است و می‌تواند انواع مختلف محتوا از قبیل متن، تصویر، صدا و ویدئو را تولید کند (Floridi, 2023).

هر دو دسته هوش مصنوعی سنتی و مولد از الگوریتم‌هایی الهام‌گرفته از فرایندهای شناختی انسان استفاده می‌کنند؛ اما اهداف آن‌ها متفاوت است. هوش مصنوعی سنتی عمدتاً برای پیش‌بینی و تحلیل داده‌ها به‌کار می‌رود، درحالی‌که هوش مصنوعی مولد با درک الگوهای داده‌های موجود به خلق محتوای جدید می‌پردازد. به بیان دیگر، هوش مصنوعی سنتی پیش‌بینی می‌کند؛ اما هوش مصنوعی مولد می‌آفریند.

ظهور ابزارهایی مانند چت جی‌پی‌تی، میدجرنی و دال ای باعث گسترش سریع کاربردهای این فناوری در حوزه‌هایی همچون بازاریابی، آموزش، طراحی محصول، تولید محتوای رسانه‌ای و حتی مشاورهٔ حقوقی شده است. براساس گزارش مک‌کینزی[4] در سال ۲۰۲۴، بازار جهانی هوش مصنوعی مولد تا سال ۲۰۳۰ به بیش از ۱/۳ تریلیون دلار خواهد رسید که بخش عمده‌ای از آن محصول بهبود بهره‌وری سازمانی است (Chui et al., 2023).

هوش مصنوعی با سرعت به فناوری مهمی برای توسعهٔ کشورها و سیاست جهانی تبدیل می‌شود و می‌تواند تصمیم‌گیری، سیاست‌گذاری و همکاری بین‌المللی را دگرگون کند. بااین‌حال، مسائلی دربارهٔ رعایت انصاف، تبعیت از قانون، مسئولیت‌پذیری در نتایج و معضلات اخلاقی مطرح است (Iqbal et al., 2024).

1. Large Language Models– LLMs.
2. Stable Diffusion Model.
3. GANs.
4. McKinsey, 2024.

اخلاق در هوش مصنوعی

با پیشرفت سریع فناوری‌های هوش مصنوعی، مسائل اخلاقی نیز هم‌زمان گسترش یافته‌اند. از موفقیت‌های یادگیری عمیق در بینایی ماشین و یادگیری تقویتی تا ظهور مدل‌های زبانی بزرگ با توانایی‌های استدلالی، هر مرحله توسعه، مسائل تازه‌ای را پدید آورده است؛ در نتیجه، اخلاق هوش مصنوعی از رویکردی صرفاً واکنشی به‌سمت رویکردی فعال تغییر یافته و اکنون شامل اقدامات عملی، توسعهٔ رویکردهای فضیلت‌محور و پیوند با مقررات قانونی مانند قانون هوش مصنوعی اتحادیهٔ اروپاست. این حوزه همچنین با مباحثی چون هم‌ترازی و ایمنی هوش مصنوعی پیوند خورده که هدف آن‌ها پیشگیری از آسیب‌ها و حتی خطرات وجودی است (Hagendorff, 2024).

مهم‌ترین نگرانی‌ها شامل سوگیری الگوریتمی، سلاح‌های خودکار، تهدید اشتغال، نابرابری اقتصادی و نقض حریم خصوصی است. این فناوری همچنین پرسش‌های فلسفی دربارهٔ ماهیت انسان و آینده را به ذهن می‌آورد. برای توسعه‌ای مسئولانه باید بر شفافیت، بی‌طرفی، امنیت داده‌ها و توزیع عادلانه منافع تأکید شود و همکاری میان سیاست‌گذاران، شرکت‌ها و جامعه استمرار یابد (Chakraborty, 2023).

مطالعات نشان داده‌اند که اخلاق هوش مصنوعی بر چارچوب‌ها و ارزش‌هایی استوار است که توسعه و استفاده از فناوری‌های هوشمند را هدایت می‌کنند (Jobin et al., 2019). این مطالعات، با بررسی ۸۴ سند راهنمای اخلاقی در سطح جهان، اصول محوری را چنین معرفی کردند: شفافیت، عدالت و نبود تبعیض، مسئولیت‌پذیری، محافظت از حریم خصوصی و امنیت.

- شفافیت[1]: توانایی توضیح فرایندها و خروجی‌ها به ذی‌نفعان؛
- عدالت و نبود تبعیض[2]: جلوگیری از بازتولید یا تشدید سوگیری‌های اجتماعی و فرهنگی؛

1. Transparency.
2. Fairness & Non-Discrimination.

- مسئولیت‌پذیری[1]: مشخص کردن مسئولیت حقوقی و اخلاقی تصمیمات هوش مصنوعی؛
- حریم خصوصی و محافظت از داده‌ها[2]: رعایت استانداردهای امنیتی و حقوق کاربران؛
- امنیت: محافظت از سامانه‌ها در برابر سوءاستفاده یا حملات سایبری.

یونسکو در توصیه‌نامهٔ اخلاق هوش مصنوعی و سازمان همکاری و توسعهٔ اقتصادی، این اصول را ستون‌های اصلی استفادهٔ مسئولانه از هوش مصنوعی معرفی کرده‌اند (UNESCO, 2022OECD, 2024).

مسائل اخلاقی هوش مصنوعی مولد

هوش مصنوعی مولد، در کنار فرصت‌های نوآورانه، مجموعه‌ای از مسائل اخلاقی پیچیده برای سازمان‌ها ایجاد می‌کند که هم ابعاد فنی و هم ابعاد اجتماعی و حقوقی را در بر می‌گیرد. مهم‌ترین مسائل عبارت‌اند از:

– سوگیری الگوریتمی[3]: داده‌های آموزشی مدل‌های مولد اغلب حاوی سوگیری‌های تاریخی، فرهنگی یا جنسیتی هستند و این امر می‌تواند در خروجی‌ها بازتولید یا حتی تشدید شود (Bender et al., 2021)؛ برای نمونه، یک مدل تولید تصویر ممکن است به‌طور پیش‌فرض، مشاغل مدیریتی را مردانه و نقش‌های خدماتی را زنانه نمایش دهد. چنین کلیشه‌هایی نه‌تنها تبعیض را تقویت می‌کنند، بلکه می‌توانند به اعتبار سازمان آسیب بزنند و زمینهٔ اتهام تبعیض را فراهم کنند؛

– تولید محتوای نادرست یا جعلی[4]: توانایی تولید متون و تصاویر مشابه واقعیت، خطر انتشار اطلاعات نادرست و جعل هویت را افزایش می‌دهد (Gupta, 2024) ؛ برای مثال،

1. Accountability.
2. Privacy & Data Protection.
3. Algorithmic Bias.
4. Misinformation & Deepfake.

یک ویدئوی جعلی از مدیر سازمان می‌تواند بحران روابط عمومی و پیامدهای حقوقی جدی ایجاد کند؛

– نقض حریم خصوصی: مدل‌های مولد ممکن است داده‌های حساس را بازتولید کنند؛ به‌ویژه اگر در داده‌های آموزشی موجود باشند (Carlini et al., 2023). بازنمایی ناخواستهٔ اطلاعات داخلی سازمان می‌تواند مقررات حریم خصوصی را نقض کند و به جریمه‌های مالی و خدشه به اعتبار منجر شود؛

– ابهام در مسئولیت حقوقی: در صورت بروز خطا یا خسارت ناشی از خروجی مدل، تعیین مسئولیت حقوقی میان توسعه‌دهندگان، کاربران یا سازمان‌ها دشوار است (Cath, 2018)؛ برای نمونه، اگر محتوای تولیدی ناقض قوانین تبلیغات باشد، مشخص نیست مسئولیت بر عهدهٔ کدام بخش است؛ موضوعی که خطر دعاوی حقوقی را افزایش می‌دهد؛

– مسائل مالکیت فکری: ابهام در مالکیت محتوایی که هوش مصنوعی تولید می‌کند، یکی دیگر از مسائل است؛ برای مثال، تصویر تولیدی ممکن است بسیار شبیه یک اثر هنری ثبت‌شده باشد و باعث شکایت حقوقی و آسیب به برند سازمان شود؛

– نبود شفافیت و قابلیت توضیح: بسیاری از مدل‌های مولد مانند جعبه‌سیاه عمل می‌کنند و توضیح نحوهٔ تولید خروجی دشوار است (Weidinger et al., 2021). این مسئله به‌ویژه در تعامل با مراجع قانونی یا نظارتی ممکن است، اعتماد را کاهش دهد؛

– تأثیرات منفی بر فرهنگ و ارزش‌های سازمانی: استفادهٔ افراطی از این فناوری ممکن است مهارت‌های انسانی مانند تفکر انتقادی را تضعیف کند. اگر کارکنان، بدون تحلیل، خروجی مدل‌ها را بپذیرند، کیفیت تصمیم‌گیری و نوآوری سازمانی در بلندمدت کاهش می‌یابد (Dwivedi et al., 2023)؛

– تأثیرات کلان اجتماعی و اقتصادی: گسترش استفاده از هوش مصنوعی مولد می‌تواند بازار کار، اعتماد اجتماعی و حتی فرایندهای دموکراتیک را تحت‌تأثیر قرار دهد.

جایگزینی نیروی انسانی تولید محتوا بدون آموزش مجدد می‌تواند نارضایتی کارکنان، تنش‌های کارگری و آسیب به تصویر عمومی سازمان را در پی داشته باشد.

روش پژوهش

پژوهش حاضر با رویکرد کیفی‌اکتشافی و بر پایهٔ الگوی تفسیرگرایی انجام شده است. ماهیت نوپای هوش مصنوعی مولد و نبود مدل‌های بومی در حوزهٔ حاکمیت اخلاقی، نیازمند رویکردی است که از داده‌های میدانی و دیدگاه‌های خبرگان استخراج شود. برای تحلیل داده‌ها از روش تحلیل مضمون[1] براساس مدل شش‌مرحله‌ای براون و کلارک (Braun & clarke., 2006) استفاده شد. این مدل، امکان شناسایی الگوهای معنایی و ساخت چارچوب مفهومی را فراهم می‌کند.

مهم‌ترین پرسش‌هایی که از خبرگان پرسیده شد، عبارت بودند از:

۱. مهم‌ترین مسائل اخلاقی سازمان‌ها در کاربست هوش مصنوعی مولد چیست؟

۲. این مسائل در کدام مراحل چرخهٔ عمر سامانه‌های هوش مصنوعی مولد بروز می‌کنند؟

۳. چه سازوکارهایی برای حکمرانی اخلاقی و کاهش این مسائل پیشنهاد می‌شود؟

جامعهٔ آماری پژوهش را خبرگان حوزهٔ هوش مصنوعی، اخلاق حرفه‌ای، مدیریت فناوری اطلاعات و سیاست‌گذاری سازمانی تشکیل دادند. معیارهای ورود خبرگان به پژوهش شامل حداقل سه سال تجربهٔ کاری مرتبط با پروژه‌های هوش مصنوعی یا سیاست‌گذاری فناوری، مشارکت در تصمیم‌گیری یا اجرا در پروژه‌های هوش مصنوعی مولد و تمایل به شرکت در مصاحبه و رضایت برای ضبط صوت بود.

نمونه‌گیری در این پژوهش به‌صورت هدفمند و با بهره‌گیری از تکنیک گلوله‌برفی برای شناسایی افراد کلیدی انجام شد. خبرگان هفده نفر بودند و مصاحبه‌ها تا رسیدن به

1. Thematic Analysis

اشباع نظری، یعنی زمانی که دیگر کد یا مضمون جدیدی شناسایی نشود، ادامه یافت. ابـزار گردآوری داده، مصاحبهٔ نیمه‌ساختاریافته بود.

تحلیل داده‌ها با روش تحلیل مضمون در شش مرحلهٔ زیر انجام شد: ۱) آشنایی با داده‌ها از طریـق خـوانش چنـدباره؛ ۲) کدگـذاری بـاز بـا اسـتفاده از نرم‌افـزار مکـس کیـودا؛ ۳) گروه‌بندی کدها و استخراج خرده‌مضامین؛ ۴) بازبینی مضامین و پالایش سـاختار آن‌هـا؛ ۵) تعریف و نام‌گذاری مضامین اصلی؛ ۶) تدوین گزارش نهایی همراه با نقل‌قول‌های شـاخص از مشارکت‌کنندگان.

برای اطمینان از اعتبار بازبینی نتایج توسط مشـارکت‌کنندگان و هم‌کـدی بـا پژوهشـگر دوم، برای پایایی از ایجاد ردپای ممیزی و ثبت تمام تصمیمات تحلیلـی، بـرای تأییدپـذیری از مستندسازی فرایند و کنترل سوگیری پژوهشـگر و بـرای انتقال‌پـذیری از ارائـهٔ توصـیف غنی از مشارکت‌کنندگان و زمینهٔ پژوهش استفاده شد. ملاحظات اخلاقی نیز شامل رعایـت رضایت آگاهانه، محرمانگی، حق انصراف و ناشناس‌سازی داده‌ها بوده است.

بـرای گـردآوری داده‌هـا، هفـده مصـاحبهٔ نیمه‌سـاختاریافته بـا خبرگـان حوزه‌هـای مـرتبط بـا هـوش مصـنوعی، اخـلاق فنـاوری، مـدیریت و حقـوق در بـازهٔ زمـانی اردیبهشت تا تیـر۱۴۰۴ انجـام شـد. هـر مصـاحبه بـین چهـل تـا شصـت دقیقـه طـول کشـید و بـا رضـایت آگاهانـهٔ مشـارکت‌کنندگان به‌صـورت صـوتی ضـبط و سـپس، بـه‌طور کامل مکتوب گردید.

در مجموع، حدود ۱۶۰ صفحه متن مصاحبه تولید شد کـه مبنـای تحلیـل مضـمون قـرار گرفـت. بـه‌منظور افـزایش اعتبـار یافتـه‌ها، نسـخهٔ مکتـوب مصـاحبه‌ها در اختیـار مشارکت‌کنندگان قرار گرفت تـا آن را بـازبینی و درسـتی محتـوا را تأییـد کننـد. بـه‌دلیـل ملاحظات اخلاقی و محرمانگی، اطلاعات واقعی هویتی مصاحبه‌شوندگان منتشر نمی‌شـود؛ اما مشخصات جمعیت‌شناختی و تخصصی آنان در جدول ۱ آمـده اسـت و کدگـذاری H1 تا H17 برای شناسایی نقل‌قول‌ها به کار رفته است.

تحلیل یافته‌ها

بـه‌منظور شفاف‌سـازی ویژگـی‌هـای نمونـه و بررسـی تنـوع دیـدگاه‌ها، مشخصـات جمعیت‌شناختی و حرفه‌ای خبرگان شرکت‌کننده در این پـژوهش در جـدول ۱ ارائـه شـده است.

جدول ۱. مشخصات جمعیت‌شناختی خبرگان شرکت‌کننده در پژوهش

کد مشارکت‌کننده	جنسیت	حوزهٔ تخصصی	سابقهٔ کار مرتبط (سال)	نوع سازمان
H1	مرد	توسعهٔ هوش مصنوعی	۱۲	خصوصی- فناوری
H2	زن	حقوق فناوری و مالکیت فکری	۹	خصوصی- مشاوره
H3	مرد	علوم داده و اخلاق هوش مصنوعی	۱۵	دانشگاهی
H4	زن	تحول دیجیتال	۱۰	خصوصی- خدمات
H5	مرد	امنیت سایبری و حاکمیت داده	۸	خصوصی- فناوری
H6	زن	طراحی و اجرای هوش مصنوعی	۷	خصوصی- استارتاپ
H7	مرد	مقررات‌گذاری فناوری	۱۴	دولتی
H8	زن	اجرای هوش مصنوعی در سازمان‌ها	۶	خصوصی - مشاوره
H9	مرد	یادگیری ماشین	۹	خصوصی - فناوری
H10	زن	انطباق داده و مقررات محافظت داده	۸	خصوصی - بین‌المللی
H11	مرد	بازاریابی دیجیتال	۱۰	خصوصی- رسانه
H12	زن	حقوق سایبری	۱۳	خصوصی- حقوقی
H13	مرد	توسعهٔ هوش مصنوعی	۶	خصوصی- فناوری
H14	زن	تحلیل داده و مصورسازی	۵	خصوصی- فناوری
H15	مرد	اخلاق کاربردی و فناوری	۱۱	دانشگاهی
H16	زن	تحول دیجیتال	۹	خصوصی
H17	مرد	استارتاپ‌های هوش مصنوعی	۷	خصوصی- استارتاپ

از نظر جنسیت، ترکیب تقریباً متعادلی از زنان و مردان وجود دارد که این امر می‌تواند به انعکاس دیدگاه‌های متنوع و متوازن کمک کند.

میانگین سابقهٔ کاری مشارکت‌کنندگان در حوزه‌های مرتبط با هوش مصنوعی مولد و فناوری حدود ۹/۵ سال است که نشان‌دهندهٔ تجربهٔ قابل توجه آن‌ها در این حوزه‌هاست. خبرگان از بخش‌های مختلف از جمله خصوصی، دولتی، دانشگاهی، بین‌المللی و استارتاپی انتخاب شده‌اند که این تنوع نهادی باعث می‌شود یافته‌ها از ابعاد مختلف فنی، مدیریتی، حقوقی و راهبردی بررسی شوند.

حوزه‌های تخصصی شرکت‌کنندگان شامل توسعهٔ هوش مصنوعی، علوم داده، امنیت سایبری، تحول دیجیتال، سیاست‌گذاری فناوری، حقوق فناوری و مالکیت فکری، بازاریابی دیجیتال و اخلاق هوش مصنوعی بوده است.

از تحلیل هفده مصاحبهٔ نیمه‌ساختاریافته، در مرحلهٔ کدگذاری باز، ۱۸۶ کد اولیه استخراج شد. در جدول ۲، نمونه‌ای از کدهای باز استخراج‌شده آورده شده است.

جدول ۲. نمونه کدهای باز مسائل اخلاقی استفاده از هوش مصنوعی مولد در سازمان‌ها

کد باز	توضیح کوتاه	نمونهٔ نقل‌قول از مشارکت‌کننده
سوگیری در داده‌های آموزشی	داده‌های آموزشی ناقص یا جانب‌دارانه باعث تصمیمات تبعیض‌آمیز می‌شوند.	مدل در بخش استخدام، مردان را بیشتر انتخاب می‌کرد؛ چون داده‌های قبلی‌اش این‌طور بود.
ابهام در تصمیم‌گیری مدل	توضیح دلیل خروجی مدل ممکن نیست.	وقتی مدیرعامل می‌پرسد چرا این متن یا تحلیل تولید شد، جواب دقیقی نداریم.
نشت احتمالی داده‌های محرمانه	خطر افشای داده‌های مشتری یا سازمان در فرایند آموزش مدل وجود دارد.	یک بار مدل اطلاعات پروژه را تولید کرد که در هیچ سند عمومی نبود.
ابهام در مسئولیت‌پذیری	مسئول خطا یا خسارت مشخص نیست.	وقتی مدل اشتباه کرد، همه گفتند مقصر من نیستم.
تضعیف استقلال کارکنان	وابستگی بیش از حد تصمیمات به خروجی مدل	کارمندان می‌گفتند ما فقط تأییدکنندهٔ تصمیم هوش مصنوعی هستیم.

کد باز	توضیح کوتاه	نمونهٔ نقل‌قول از مشارکت‌کننده
تأثیر منفی بر شهرت برند	مدل، محتوای اشتباه یا توهین‌آمیز تولید کرده است.	یک متن تولیدی باعث اعتراض کاربران شبکهٔ اجتماعی شد.
مسئلهٔ انطباق با قوانین	مدل با مقررات حریم خصوصی یا مالکیت فکری هماهنگ نیست.	برای قوانین محافظت از دادهٔ اتحادیهٔ اروپا باید مدل را چند ماه بازبینی کنیم.
مشکل در کنترل خروجی مدل	خروجی غیرمنتظره و غیرقابل پیش‌بینی است.	مدل گاهی محتوای سیاسی یا نامناسب می‌سازد که به موضوع ما ربط ندارد.
هزینه‌های پنهان استفاده	هزینه‌های امنیتی، حقوقی و بازبینی در ابتدا پیش‌بینی نشده بود.	مجوزها و بازرسی‌ها هزینه‌بر شد.
اثر روانی بر کارکنان	ترس از جایگزینی موجب اضطراب یا کاهش انگیزه می‌شود.	برخی از کارکنان نگران بودند که مدل، کارشان را بگیرد.

پس از انجام کدگذاری باز و مرور دقیق داده‌های حاصل از مصاحبه‌های نیمه‌ساختاریافته، خرده‌مضامین در هفت دستهٔ‌مضمون اصلی قرار گرفتند.

این مضامین در جدول زیر به همراه توضیح کوتاه از هر خرده‌مضمون ارائه شده‌اند.

جدول ۳. خرده‌مضامین استخراج‌شده از تحلیل داده‌ها

ردیف	مضمون اصلی	خرده‌مضمون	توضیح
۱	مسائل عدالت و نبود تبعیض	سوگیری داده‌ای و الگوریتمی	مدل‌های هوش مصنوعی مولد ممکن است از داده‌های جانب‌دارانه آموزش ببینند و تصمیمات تبعیض‌آمیز بگیرند.
	مسائل عدالت و نبود تبعیض	تبعیض غیرمستقیم در فرایند تصمیم‌گیری	حتی بدون قصد قبلی، خروجی مدل می‌تواند به ضرر گروه‌های خاص تمام شود.
	مسائل عدالت و نبود تبعیض	بازتولید کلیشه‌ها و پیش‌داوری‌ها	محتوا یا تحلیل تولیدی ممکن است کلیشه‌های جنسیتی یا فرهنگی را تقویت کند.
۲	مسائل شفافیت و پاسخ‌گویی	ابهام در منطق تصمیم‌گیری مدل	سازمان‌ها قادر به توضیح دلیل خروجی مدل نیستند.
	مسائل شفافیت و پاسخ‌گویی	نامشخص‌بودن مسئول خطا	مشخص نیست مسئولیت خطا با توسعه‌دهنده، کاربر یا سازمان است.

ردیف	مضمون اصلی	خرده‌مضمون	توضیح
	مسائل شفافیت و پاسخ‌گویی	دسترسی‌نداشتن به فرایند و داده‌های آموزشی	بسته‌بودن داده‌ها و معماری مانع ارزیابی مدل می‌شود.
۳	مسائل حقوقی و امنیتی	نقض حریم خصوصی و نشت داده‌ها	اطلاعات حساس ممکن است در فرایند آموزش یا تولید محتوا افشا شود.
	مسائل حقوقی و امنیتی	انطباق‌نداشتن با مقررات و استانداردها	مدل ممکن است با قوانین حفاظت داده یا مالکیت فکری هم‌خوانی نداشته باشد.
	مسائل حقوقی و امنیتی	مسائل مالکیت محتوا	مالکیت محتوای تولیدشده توسط هوش مصنوعی مولد مشخص نیست.
	مسائل حقوقی و امنیتی	تهدیدات امنیت سایبری	مدل‌ها می‌توانند برای حملات مهندسی اجتماعی یا تولید محتوای مخرب استفاده شوند.
۴	مسائل انسانی و فرهنگی	کاهش استقلال و عاملیت کارکنان	وابستگی بیش از حد به مدل، خلاقیت و قضاوت انسانی را کاهش می‌دهد.
	مسائل انسانی و فرهنگی	اثر روانی و اضطراب شغلی	ترس از جایگزینی یا کاهش اهمیت نقش شغلی کارکنان.
	مسائل انسانی و فرهنگی	کاهش تعامل انسانی در سازمان	جایگزینی ارتباطات انسانی با تعاملات ماشینی.
۵	مسائل کیفیت و اعتبار محتوا	تولید محتوای نادرست یا گمراه‌کننده	مدل ممکن است اطلاعات غلط با ظاهر معتبر ایجاد کند.
	مسائل کیفیت و اعتبار محتوا	خروجی نامناسب یا توهین‌آمیز	محتوای غیرقابل‌قبول که می‌تواند به اعتبار سازمان آسیب بزند.
	مسائل کیفیت و اعتبار محتوا	نداشتن کنترل کامل بر خروجی مدل	پیش‌بینی‌ناپذیری خروجی مدیریت خطر را دشوار می‌کند.
	مسائل کیفیت و اعتبار محتوا	وابستگی به داده‌های خارجی با کیفیت نامشخص	مدل برای آموزش یا تکمیل اطلاعات از منابع خارجی استفاده می‌کند که کیفیت و صحت آن‌ها تضمین‌شده نیست.
۶	مسائل اقتصادی و عملیاتی	هزینه‌های پنهان استفاده	شامل هزینه‌های بازبینی، امنیت و انطباق حقوقی.

ردیف	مضمون اصلی	خرده‌مضمون	توضیح
	مسائل اقتصادی و عملیاتی	نیاز به نیروی انسانی متخصص	کمبود مهارت در سازمان برای کار با هوش مصنوعی مولد.
	مسائل اقتصادی و عملیاتی	وابستگی بیش از حد به فناوری خارجی	خطر اتکا به پلتفرم‌ها و سرویس‌های بیرونی.
	مسائل اقتصادی و عملیاتی	اثر زیست‌محیطی مدل‌های بزرگ	مصرف زیاد انرژی و منابع محاسباتی در مدل‌های بزرگ می‌تواند آثار زیست‌محیطی منفی ایجاد کند.
۷	مسائل راهبردی و حاکمیتی	نبود چارچوب حاکمیت اخلاقی مشخص	نبود سیاست‌های روشن برای استفادهٔ مسئولانه از هوش مصنوعی مولد.
	مسائل راهبردی و حاکمیتی	مقاومت سازمانی در برابر تغییرات	مقاومت کارکنان یا مدیران در پذیرش دستورالعمل‌های اخلاقی.
	مسائل راهبردی و حاکمیتی	استفادهٔ غیرمجاز کارکنان از ابزارهای هوش مصنوعی مولد	کارکنان بدون اطلاع یا مجوز سازمان از ابزارهای هوش مصنوعی مولد استفاده می‌کنند.

در فرایند تحلیل مضمون، پس از انجام کدگذاری باز و محوری، ابتدا هفت مضمون اصلی متمایز از داده‌ها استخراج شد. بااین‌حال، در مرحلهٔ بازبینی و پالایش مضامین که مطابق با رویکرد براون و کلارک (Braun & Clarke, 2006) انجام شد، مشخص گردید دو مضمون مسائل راهبردی و حاکمیتی و مسائل حقوقی و امنیتی هم‌پوشانی مفهومی و مصداقی زیادی دارند.

به‌طور خاص، زیرمضامین مرتبط با نبود چارچوب حاکمیت اخلاقی و مسائل انطباق با مقررات، هر دو به حوزهٔ سیاست‌گذاری، الزامات قانونی و حاکمیت فناوری مرتبط بودند؛ بنابراین، این دو مضمون در قالب یک مضمون جامع‌تر با عنوان مسائل حقوقی، امنیتی و حاکمیتی ادغام شدند. این ادغام باعث شد مدل نهایی پژوهش (جدول ۴) شامل شش مضمون اصلی و ۲۴ خرده‌مضمون باشد که از انسجام و تمایز مفهومی بیشتری برخوردارند.

جدول ۴. مضامین اصلی و خرده‌مضامین نهایی پژوهش

ردیف	مضمون اصلی	خرده‌مضامین
۱	مسائل عدالت و نبود تبعیض	سوگیری داده‌ای و الگوریتمی تبعیض غیرمستقیم در فرایند تصمیم‌گیری بازتولید کلیشه‌ها و پیش‌داوری‌ها
۲	مسائل شفافیت و پاسخ‌گویی	ابهام در منطق تصمیم‌گیری مدل نامشخص‌بودن مسئولیت خطا دسترسی‌نداشتن به فرایند و داده‌های آموزشی
۳	مسائل حقوقی، امنیتی و حاکمیتی	نقض حریم خصوصی و نشت داده‌ها انطباق‌نداشتن با مقررات و استانداردها مسائل مالکیت محتوا تهدیدات امنیت سایبری نبود چارچوب حاکمیت اخلاقی مشخص مقاومت سازمانی در برابر تغییرات استفادهٔ غیرمجاز کارکنان از ابزارهای هوش مصنوعی مولد
۴	مسائل انسانی و فرهنگی	کاهش استقلال و عاملیت کارکنان اثر روانی و اضطراب شغلی کاهش تعامل انسانی در سازمان
۵	مسائل کیفیت و اعتبار محتوا	تولید محتوای نادرست یا گمراه‌کننده خروجی نامناسب یا توهین‌آمیز نبود کنترل کامل بر خروجی مدل وابستگی به داده‌های خارجی با کیفیت نامشخص
۶	مسائل اقتصادی و عملیاتی	هزینه‌های پنهان استفاده نیاز به نیروی انسانی متخصص وابستگی بیش از حد به فناوری خارجی اثر زیست‌محیطی مدل‌های بزرگ

در مرحلهٔ آخر برای ارزیابی و اولویت‌بندی ۲۴ خرده‌مضمون شناسایی‌شده، از دو معیار اثر بالقوه بر عملکرد و پایداری سازمان و شدت پیامدها استفاده شد. منظور از اثر بالقوه بر عملکرد و پایداری سازمان، میزان تأثیرگذاری هر مسئله بر کارایی، بهره‌وری، اعتبار و دوام

بلندمدت سازمان در صورت بروز یا تداوم آن است. منظور از شدت پیامدها، گستره و عمق آثار منفی احتمالی هر مسئله بر جنبه‌های کلیدی فعالیت سازمان (مانند منابع انسانی، امنیت اطلاعات، کیفیت خدمات و هزینه‌ها) می‌باشد.

برای سنجش این دو معیار، پرسش‌نامه‌ای طراحی شد که شامل توضیح روشن هر خرده‌مضمون و دو مقیاس پنج‌درجه‌ای (از ۱= کمترین تا ۵= بیشترین) برای ارزیابی است. این پرسش‌نامه در اختیار خبرگان قرار گرفت و از میانگین امتیازات دریافتی برای هر معیار به‌عنوان شاخص نهایی استفاده شد.

این رویکرد باعث شد وزن نسبی هر خرده‌مضمون از منظر اثرگذاری واقعی و اهمیت عملیاتی آن مشخص و زمینه برای رتبه‌بندی و تحلیل اولویت‌ها فراهم شود. امتیازدهی ۲۴ خرده‌مضمون که خبرگان شناسایی کردند، براساس تأثیر بالقوه بر عملکرد و پایداری سازمان در جدول ۵ آورده شده است.

جدول ۵. رتبه‌بندی خرده‌مضامین مسائل هوش مصنوعی مولد براساس اهمیت سازمانی

ردیف	مضمون اصلی	خرده‌مضمون	اهمیت نسبی	امتیاز میانگین خبرگان
۱	حقوقی، امنیتی و حاکمیتی	نقض حریم خصوصی و نشت داده‌ها	زیاد	۴/۹
۲	مسائل عدالت و نبود تبعیض	سوگیری داده‌ای و الگوریتمی	زیاد	۴/۸
۳	حقوقی، امنیتی و حاکمیتی	انطباق‌نداشتن با مقررات و استانداردها	زیاد	۴/۸
۴	مسائل عدالت و عدم تبعیض	تبعیض غیرمستقیم در فرایند تصمیم‌گیری	زیاد	۴/۷
۵	حقوقی، امنیتی و حاکمیتی	تهدیدات امنیت سایبری	زیاد	۴/۷
۶	کیفیت و اعتبار محتوا	تولید محتوای نادرست یا گمراه‌کننده	زیاد	۴/۷

ردیف	مضمون اصلی	خرده‌مضمون	اهمیت نسبی	امتیاز میانگین خبرگان
۷	مسائل عدالت و عدم تبعیض	بازتولید کلیشه‌ها و پیش‌داوری‌ها	زیاد	۴/۶
۸	شفافیت و پاسخ‌گویی	نامشخص‌بودن مسئولیت خطا	زیاد	۴/۶
۹	حقوقی، امنیتی و حاکمیتی	نبود چارچوب حاکمیت اخلاقی مشخص	زیاد	۴/۶
۱۰	اقتصادی و عملیاتی	وابستگی بیش‌ازحد به فناوری خارجی	زیاد	۴/۶
۱۱	شفافیت و پاسخ‌گویی	ابهام در منطق تصمیم‌گیری مدل	زیاد	۴/۵
۱۲	کیفیت و اعتبار محتوا	نبود کنترل کامل بر خروجی مدل	متوسط	۴/۵
۱۳	حقوقی، امنیتی و حاکمیتی	مسائل مالکیت محتوا	متوسط	۴/۴
۱۴	کیفیت و اعتبار محتوا	خروجی نامناسب یا توهین‌آمیز	متوسط	۴/۴
۱۵	شفافیت و پاسخ‌گویی	دسترسی‌نداشتن به فرایند و داده‌های آموزشی	متوسط	۴/۳
۱۶	حقوقی، امنیتی و حاکمیتی	استفادهٔ غیرمجاز کارکنان از ابزارهای هوش مصنوعی مولد	متوسط	۴/۳
۱۷	کیفیت و اعتبار محتوا	وابستگی به داده‌های خارجی با کیفیت نامشخص	متوسط	۴/۳
۱۸	حقوقی، امنیتی و حاکمیتی	مقاومت سازمانی در برابر تغییرات	متوسط	۴/۲
۱۹	انسانی و فرهنگی	کاهش استقلال و عاملیت کارکنان	متوسط	۴/۲
۲۰	اقتصادی و عملیاتی	هزینه‌های پنهان استفاده	متوسط	۴/۲
۲۱	اقتصادی و عملیاتی	اثر زیست‌محیطی مدل‌های بزرگ	متوسط	۴/۲
۲۲	انسانی و فرهنگی	اثر روانی و اضطراب شغلی	متوسط	۴/۱
۲۳	اقتصادی و عملیاتی	نیاز به نیروی انسانی متخصص	متوسط	۴/۱
۲۴	انسانی و فرهنگی	کاهش تعامل انسانی در سازمان	متوسط	۴/۰

بحث

یافته‌های این پژوهش نشان می‌دهد به‌کارگیری هوش مصنوعی مولد در سازمان‌ها با مجموعه‌ای پیچیده از مسائل و مخاطرات همراه است که درهم‌تنیده و چندبُعدی‌اند. در بُعد عدالت و نبود تبعیض، یافته‌ها نشان دادند سوگیری داده‌ای و الگوریتمی و تبعیض غیرمستقیم در فرایند تصمیم‌گیری، بیشترین میانگین اهمیت را در کل جدول رتبه‌بندی دارند و این نشان می‌دهد دغدغهٔ برابری و جلوگیری از سوگیری، از دید خبرگان، مهم‌ترین اولویت مدیریتی در مواجهه با هوش مصنوعی مولد است. مدلی که با داده‌های جانب‌دارانه آموزش دیده باشد، حتی بدون قصد قبلی، ممکن است تصمیم‌هایی تولید کند که به زیان گروه‌های خاص تمام شود.

این وضعیت، تبعیض غیرمستقیم را در فرایند تصمیم‌گیری تقویت می‌کند و به بازتولید کلیشه‌های جنسیتی، فرهنگی یا قومی می‌انجامد. چنین خروجی‌هایی نه‌تنها بر اعتماد عمومی اثر منفی می‌گذارند، بلکه می‌توانند به‌شکل سیستماتیک، نابرابری را بازتولید کنند. این یافته‌ها با مطالعات الکفیری (Al-kfairy et al 2024) و فرارا (Ferrara, 2023) هم‌سو است که نشان داده‌اند الگوریتم‌های بزرگ زبانی، حتی بدون قصد قبلی، ممکن است بر گروه‌های خاص اثر منفی بگذارند و در نتیجه، نیازمند راهبردهای کاهش سوگیری هستند.

ابعاد شفافیت و پاسخ‌گویی نیز به‌شکل جدی در معرض تهدید قرار می‌گیرند. در این بُعد، ابهام در منطق تصمیم‌گیری مدل و نامشخص‌بودن مسئولیت خطا، در رتبه‌های میانی قرار دارند؛ اما اهمیت بسیار آن‌ها در رتبه‌بندی نشان می‌دهد نبود شفافیت نه‌تنها مشکل فنی بلکه یک مسئلهٔ مدیریتی و حکمرانی است.

ابهام در منطق تصمیم‌گیری مدل، کار را برای درک نحوهٔ رسیدن به خروجی دشوار می‌کند و نبود شفافیت در این مسیر، اعتماد ذی‌نفعان را کاهش می‌دهد. افزون‌بر این، مشخص‌نبودن مسئولیت خطا (اینکه توسعه‌دهنده، کاربر یا سازمان کدام‌یک باید پاسخ‌گو باشند)، به خلأ پاسخ‌گویی دامن می‌زند.

محدودیت دسترسی به فرایندها و داده‌های آموزشی نیز ارزیابی بی‌طرفانه و دقیق مدل را مختل کرده، مانع پایش کیفی آن می‌شود. همان‌طور که گارسیا لوپز و همکاران (García-López et al., 2025) و مارشال و همکاران (Marchal et al., 2024) توضیح داده‌اند، ماهیت جعبه‌سیاه مدل‌های مولد و محرمانگی داده‌های آموزشی، ارزیابی و اعتماد کاربران را دشوار می‌سازد و شفافیت را به یکی از پیش‌شرط‌های پاسخ‌گویی تبدیل می‌کند.

از منظر حقوقی، امنیتی و حاکمیتی، تهدیدات چندلایه‌ای شناسایی شده‌اند. نقض حریم خصوصی و نشت داده‌ها، یکی از بحرانی‌ترین خطرهاست که می‌تواند پیامدهای حقوقی و اعتباری جبران‌ناپذیر داشته باشد. انطباق‌نداشتن با مقررات حفاظت داده و مالکیت فکری، مسائل حقوقی تازه‌ای ایجاد می‌کند و مسائل مالکیت محتوای تولیدشده توسط مدل‌ها همچنان مبهم باقی می‌ماند. در سطح امنیت سایبری، استفاده از مدل‌ها در حملات مهندسی اجتماعی یا تولید محتوای مخرب، تهدیدی جدی است.

در بُعد حقوقی، امنیتی و حاکمیتی، نقض حریم خصوصی و نشت داده‌ها و انطباق‌نداشتن با مقررات و استانداردها از نظر خبرگان در صدر مخاطرات امنیتی قرار گرفته‌اند. در بُعد حاکمیتی، نبود چارچوب اخلاقی روشن، مقاومت سازمانی در برابر تغییرات و حتی استفادهٔ غیرمجاز کارکنان از ابزارهای هوش مصنوعی مولد، اجرای سیاست‌های کنترلی را دشوار می‌سازد.

این موارد با نتایج پژوهش ویلیامز (Williams, 2024) و نینگ و همکاران (Ning et al., 2024) هم‌راستا است که تأکید کرده‌اند نبود خط‌مشی‌های روشن و کنترل سازمانی، مخاطرات حقوقی و امنیتی را تشدید می‌کند و نیاز به قانون‌گذاری و راهبردهای حاکمیتی جامع را ضروری می‌سازد.

در بُعد انسانی و فرهنگی، رتبه‌بندی نشان داد کاهش استقلال و عاملیت کارکنان، بیشترین اهمیت را در این دسته دارد و از نگاه خبرگان، این عامل به‌طور مستقیم بر انگیزش، رضایت شغلی و نوآوری اثر می‌گذارد. همچنین، اثر روانی و اضطراب شغلی

به‌ویژه در میان کارکنانی که مشاغلشان در معرض جایگزینی با سیستم‌های هوش مصنوعی است، یک خطر مهم تلقی می‌شود. کاهش تعامل انسانی در سازمان نیز در رتبهٔ پایین‌تری قرار گرفت؛ اما همچنان تهدیدی قابل توجه برای کیفیت روابط سازمانی و سرمایهٔ اجتماعی داخلی محسوب می‌شود. جایگزینی بخشی از ارتباطات انسانی با تعاملات ماشینی نیز کیفیت روابط سازمانی را تحت‌تأثیر قرار داده و خطر ایجاد محیط کاری غیرشخصی را افزایش می‌دهد.

این نتایج با مطالعات حوزهٔ آموزش و محیط‌های کاری گارسیا لوپز و همکاران (García-López et al., 2025) و ویلیامز (Williams, 2024) تطابق دارد که هشدار داده‌اند فناوری‌های مولد ممکن است جایگزین بخش‌هایی از روابط انسانی و خلاقیت فردی شوند.

مسئلهٔ کیفیت و اعتبار محتوا نیز به‌عنوان یکی از مسائل اصلی مطرح است. در محور کیفیت و اعتبار محتوا، رتبهٔ بالای تولید محتوای نادرست یا گمراه‌کننده نشان می‌دهد تهدیدهای اعتبار سازمان، از نگاه خبرگان، به همان اندازهٔ تهدیدهای امنیتی اهمیت دارد. علاوه بر این، وابستگی به داده‌های خارجی با کیفیت نامشخص می‌تواند دقت و اعتبار خروجی مدل را بسیار کاهش دهد و به تصمیم‌گیری‌های اشتباه منجر شود.

پژوهش مارشال و همکاران (Marchal et al., 2024) و گزارش کِی و همکاران (Kay et al., 2024) نیز بر این موضوع تأکید دارند که ماهیت پیش‌بینی محور مدل‌ها می‌تواند خروجی‌هایی ظاهراً معتبر اما نادرست تولید کند که پیامدهای جدی برای اعتبار سازمان دارد.

در بُعد اقتصادی و عملیاتی، وابستگی بیش از حد به فناوری خارجی، بالاترین رتبه را در این دسته به خود اختصاص داده است. در این بُعد، استفاده از فناوری با هزینه‌های پنهان متعددی همراه است؛ از هزینه‌های پایش و امنیت گرفته تا انطباق حقوقی و آموزش کاربران. نیاز به نیروی انسانی متخصص برای کار با این فناوری و وابستگی بیش‌ازحد به پلتفرم‌ها و سرویس‌های خارجی، استقلال فناورانهٔ سازمان را تهدید می‌کند.

همچنین، اثر زیست‌محیطی مدل‌های بزرگ، ناشی از مصرف بسیار زیاد انرژی در آموزش و اجرا، با اهداف پایداری و مسئولیت اجتماعی سازمان‌ها در تضاد است. کِی و همکاران (Kay et al., 2024) و گارسیا لوپز و همکاران (García-López et al., 2025) اشاره کرده‌اند که مدل‌های مولد علاوه بر هزینه‌های مستقیم، مصرف زیاد انرژی و ردپای کربنی قابل توجهی دارند و این موضوع باید در ارزیابی اقتصادی و مسئولیت‌پذیری محیط‌زیستی لحاظ شود.

در مجموع، این یافته‌ها نشان می‌دهد بهره‌گیری از هوش مصنوعی مولد بدون درنظرگرفتن ملاحظات چندبُعدی اخلاقی، حقوقی، فنی، انسانی و محیط‌زیستی می‌تواند سازمان‌ها را در معرض مخاطرات گسترده و پیچیده قرار دهد. مدیریت این مخاطرات مستلزم ایجاد رویکردی یکپارچه‌ای است که شامل تدوین سیاست‌های شفاف، آموزش و توانمندسازی کارکنان، پایش مستمر خروجی‌ها، بهبود کیفیت داده‌های آموزشی و ارزیابی آثار زیست‌محیطی باشد تا استفاده از این فناوری در راستای منافع پایدار سازمان هدایت شود.

نتیجه‌گیری

بررسی انجام‌شده نشان داد بهره‌گیری از هوش مصنوعی مولد در سازمان‌ها با مجموعه‌ای پیچیده از مسائل چندبُعدی همراه است. این مسائل در قالب شش مضمون اصلی و ۲۴ خرده‌مضمون قابل دسته‌بندی هستند و طیفی از مسائل مرتبط با عدالت و تبعیض، شفافیت و پاسخ‌گویی، الزامات حقوقی و امنیتی، ابعاد انسانی و فرهنگی، کیفیت و اعتبار محتوا، ملاحظات اقتصادی و زیست‌محیطی را در بر می‌گیرند.

یافته‌ها تأیید می‌کند استفاده از این فناوری بدون چارچوب‌های نظارتی و اخلاقی شفاف می‌تواند موجب بازتولید سوگیری‌ها، تضعیف اعتماد کاربران، بروز مخاطرات حقوقی و امنیتی، کاهش عاملیت انسانی و واردآوردن آسیب‌های اقتصادی و زیست‌محیطی شود؛ در نتیجه، به‌کارگیری مؤثر و پایدار هوش مصنوعی مولد نیازمند اتخاذ رویکردی جامع و بین‌رشته‌ای است که هم‌زمان به ابعاد فنی، اخلاقی، حقوقی و اجتماعی توجه کند.

برای مدیریت مسائل شناسایی‌شده در به‌کارگیری هوش مصنوعی مولد، رویکردی چندبُعدی و اولویت‌بندی‌شده ضروری است که براساس میزان اهمیت و اثرگذاری بر عملکرد و پایداری سازمان تنظیم گردد.

نخست، در حوزهٔ نقض حریم خصوصی و نشت داده‌ها، لازم است سامانه‌های پایش پیشگیرانه برای شناسایی به‌موقع نشت اطلاعات، حملات مهندسی اجتماعی و خروجی‌های مخرب راه‌اندازی شود. به‌کارگیری سازوکارهای ناشناس‌سازی داده‌ها پیش از آموزش مدل و انطباق کامل با قوانین ملی و فراملی نظیر مقررات حفاظت دادهٔ اتحادیهٔ اروپا، از الزامات حیاتی برای پیشگیری از پیامدهای حقوقی و اعتباری است.

در گام بعد، برای مقابله با مسائل عدالت و نبود تبعیض ناشی از سوگیری داده‌ای و الگوریتمی، باید فرایند ممیزی مستمر داده‌های آموزشی به‌منظور حذف محتوای تبعیض‌آمیز و افزایش تنوع منابع داده اجرا شود. استفاده از شیوه‌های توضیح‌پذیری مدل و

انتشار مستندات شفاف دربارهٔ منطق تصمیم‌گیری، می‌تواند اعتماد ذی‌نفعان را تقویت کرده، خطر بازتولید کلیشه‌ها را کاهش دهد.

ابهام در منطق تصمیم‌گیری و نبود شفافیت، همراه با نامشخص‌بودن مسئولیت خطا، لزوم ایجاد چارچوب‌های روشن پاسخ‌گویی را برجسته می‌کند. تدوین سیاست‌های داخلی دربارهٔ استفادهٔ مجاز و ممنوع از ابزارهای مولد، تعیین صریح مسئولیت‌ها بین توسعه‌دهنده، کاربر و سازمان، و ایجاد نهادهای نظارت مستقل می‌تواند شفافیت و پاسخ‌گویی را تقویت کند. همچنین، فراهم‌کردن دسترسی کنترل‌شده به داده‌ها و فرایندهای آموزشی برای ارزیابی بی‌طرفانهٔ مدل‌ها ضروری است.

برای کاهش مخاطرات امنیتی، انطباق با استانداردهای امنیت سایبری، استقرار لایه‌های چندگانهٔ دفاعی و آموزش کاربران برای شناسایی تهدیدات محتمل اهمیت دارد. نبود چارچوب حاکمیت اخلاقی مشخص باید با تدوین خط‌مشی‌های جامع و اجرای برنامه‌های آگاهی‌بخشی سازمانی برطرف شود تا مقاومت در برابر تغییرات و استفادهٔ غیرمجاز کارکنان به حداقل برسد.

در بُعد انسانی و فرهنگی، طراحی و اجرای برنامه‌های آموزشی ترکیبی که هم مهارت‌های فنی و هم ملاحظات اخلاقی و حقوقی را پوشش دهد، در کنار ایجاد سازوکارهای حمایت روانی برای کاهش اضطراب شغلی، ضروری است. ترویج فرهنگ همکاری انسان–ماشین به‌جای جایگزینی کامل و حفظ فضاهای تعاملی انسانی، می‌تواند پویایی و خلاقیت محیط کار را حفظ کند.

در زمینهٔ کیفیت و اعتبار محتوا، باید فیلترهای چندمرحله‌ای و بازبینی انسانی اجباری برای محتوای حساس به کار گرفته شود. پایش مداوم مدل‌ها، به‌روزرسانی داده‌های آموزشی و کنترل کیفیت داده‌های خارجی، نقش مهمی در جلوگیری از تولید محتوای نادرست یا توهین‌آمیز دارد.

در نهایت، برای پایداری اقتصادی و زیست‌محیطی لازم است هزینه‌های پنهان مانند امنیت، بازبینی محتوا و انطباق حقوقی در بودجه لحاظ شود. استفاده از زیرساخت‌های

پردازشی کم‌مصرف، مدل‌های فشرده‌تر و کاهش وابستگی به پلتفرم‌های خارجی از طریق سرمایه‌گذاری در توسعهٔ داخلی یا بهره‌گیری از مدل‌های متن‌باز، علاوه بر کاهش خطر وابستگی، به کاهش آثار منفی زیست‌محیطی کمک خواهد کرد.

این پژوهش، با ارائهٔ چارچوبی جامع از مضامین و خرده‌مضامین مسائل اخلاقی، فنی و اجتماعی مرتبط با هوش مصنوعی مولد، تصویری یکپارچه و عملیاتی از این حوزه ارائه داده است. ترکیب شش مضمون اصلی و ۲۴ خرده‌مضمون براساس تحلیل کیفی، امکان اولویت‌بندی مخاطرات و تدوین راهبردهای مقابله را برای سازمان‌ها فراهم می‌سازد. علاوه بر این، ادغام ابعاد کمتر بررسی‌شده‌ای همچون تأثیرات زیست‌محیطی و وابستگی به داده‌های خارجی با کیفیت نامشخص، این مطالعه را از پژوهش‌های پیشین متمایز کرده است.

با وجود جامعیت تحلیل، این پژوهش با محدودیت‌هایی همراه بوده است. نخست، داده‌ها عمدتاً از طریق منابع کیفی و دیدگاه‌های خبرگان جمع‌آوری شده و ممکن است بازتاب‌دهندهٔ سوگیری‌های ادراکی آن‌ها باشد. دوم، سرعت تحول فناوری‌های مولد موجب می‌شود برخی از یافته‌ها در بازهٔ زمانی کوتاهی نیازمند به‌روزرسانی باشند. سوم، این پژوهش عمدتاً بر بسترهای سازمانی متمرکز بوده و نتایج آن ممکن است به‌طور کامل به حوزه‌های مصرف‌کننده یا آموزشی قابل تعمیم نباشد.

کتابنامه

AI, N. (2023). "Artificial intelligence risk management framework (AI RMF 1.0)". *URL: https://nvlpubs. nist. gov/nistpubs/ai/nist. ai*, 100-101. https://doi.org/10.6028/NIST.AI.100-1

Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M. & Alfandi, O. (2024a) "A Systematic Review and Analysis of Ethical Challenges of Generative Ai: An Interdisciplinary Perspective". *Available at SSRN 4833030*.

Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M. & Alfandi, O. (2024b). "Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective". *Informatics, 11*(3): 58. https:// doi.org/10.1016/j.orgdyn.2024.101032

Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell, S. (2021). "*On the dangers of stochastic parrots: Can language models be too big?* " Paper presented at the Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. https://doi.org/10.1145/3442188.3445922

Braun, V. & Clarke, V. (2006). "Using thematic analysis in psychology". *Qualitative research in psychology, 3*(2), 77-101. DOI: 10.1191/1478088706qp063oa

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. & Sun, L. (2025). "A Survey of AI-Generated Content (AIGC)". *ACM Comput. Surv., 57*(5), Article 125. doi: 10.1145/3704262

Carlini, N., Hayes, J., Nasr, M., Jagielski, M., Sehwag, V., Tramer, F., . . . Wallace, E. (2023). "*Extracting training data from diffusion models"*. Paper presented at the 32nd USENIX security symposium (USENIX Security 23). https://doi.org/10.48550/arXiv.2301.13188

Cath, C. (2018). "Governing artificial intelligence: ethical, legal and technical opportunities and challenges". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 20180080. https:// doi.org/ 10.1098/rsta.2018.0080

Chakraborty, S. (2023). "AI and ethics: Navigating the moral landscape", *Investigating the Impact of AI on Ethics and Spirituality* (pp. 25-33): IGI Global. DOI: 10.4018/978-1-6684-9196-6

Chen, Z. (2023). "Ethics and discrimination in artificial intelligence-enabled recruitment practices". *Humanities and Social Sciences Communications, 10*(1): 567. doi: 10.1057/s41599-023-02079-x

Chui, M., Hazan, E., Roberts, R., Singla, A., & Smaje, K. (2023). "The economic potential of generative AI".

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., . . . Terem, E. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance". *Patterns, 4*(10). https://doi.org/10.1016/j.patter.2023.100857

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., . . . Wright, R. (2023). "Opinion Paper: "So what if

ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642. doi: https://doi.org/10.1016/j.ijinfomgt.2023.102642

European Cmmission, F. (2021). "Regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts".

Ferrara, E. (2023). "Should chatgpt be biased? challenges and risks of bias in large language models". *arXiv preprint arXiv:2304.03738*. https://doi.org/10.48550/arXiv.2304.03738

Floridi, L. (2023). "AI as agency without intelligence: On ChatGPT, large language models, and other generative models". *Philosophy & technology*, *36*(1), 15. https://doi.org/10.1007/s13347-023-00621-y

García-López, I. M. & Trujillo-Liñán, L. (2025). "Ethical and regulatory challenges of Generative AI in education: a systematic review". *Paper presented at the Frontiers in Educatio*n. https://doi.org/10.3389/feduc.2025.1565938

Gupta, D. (2024). "Generative AI and Deep fake s: Ethical Implications and Detection Techniques". *Journal of Science, Technology and Engineering Research, 1*(1): 45-56. DOI**:** https://doi.org/10.64206/21rgkc40

Hagendorff, T. (2024). "Mapping the ethics of generative AI: A comprehensive scoping review". *Minds and Machine*s, 34(4), 39. https://doi.org/10.1007/s11023-024-09694-w

Iqbal, Q., Khan, D. & Salis, M. (2024). "Navigating the Ethical and Legal Landscape of Artificial Intelligence in Global Governance: A Comprehensive Analysis". *Journal of Asian Development Studies, 13*(2): 945-954. DOI: https://doi.org/10.62345/jads.2024.13.2.75

Jobin, A., Ienca, M., & Vayena, E. (2019). "The global landscape of AI ethics guidelines". *Nature machine intelligence,* 1(9): 389-399. https://doi.org/10.1038/s42256-019-0088-2

Kay, J., Kasirzadeh, A., & Mohamed, S. (2024). "Epistemic injustice in generative AI". P*aper presented at the Proceedings of the AAAI/ACM Conference on AI, Ethics, and Societ*y. DOI: https:// doi.org/10.1609/aies.v7i1.31671

Macdonald, C., Adeloye, D., Sheikh, A. & Rudan, I. (2023). "Can ChatGPT draft a research article? An example of population-level vaccine effectiveness analysis". *Journal of global health, 13*, 01003. doi: 10.7189/jogh.13.01003.

Marchal, N., Xu, R., Elasmar, R., Gabriel, I., Goldberg, B. & Isaac, W. (2024). "Generative AI misuse: A taxonomy of tactics and insights from real-world data". *arXiv preprint arXiv:2406.13843*. https://doi.org/10.48550/arXiv.2406.13843

Ning, Y., Teixayavong, S., Shang, Y., Savulescu, J., Nagaraj, V., Miao, D., . . . Liu, M. (2024). "Generative artificial intelligence and

ethical considerations in health care: a scoping review and ethics checklist". *The Lancet Digital Health*, *6*(11), e848-e856. doi: 10.1016/S2589-7500(24)00143-2

OECD. (2024). *Facts Not Fakes: Tackling Disinformation, Strengthening Information Integrity*: OECD Publishing.

Raghavan, M., Barocas, S., Kleinberg, J. & Levy, K. (2020). "*Mi*tigating bias in algorithmic hiring: Evaluating claims and practices". *Paper presented at the Proceedings of the 2020 conference on fairness, accountability, and transparency*. https:// doi.org/10.1145/3351095.3372828

Rana, N. P., Pillai, R., Sivathanu, B. & Malik, N. (2024). "Assessing the nexus of Generative AI adoption, ethical considerations and organizational performance". *Technovation*, *135*, 103064. doi: https://doi.org/10.1016/j.technovation.2024.103064

Singh, K., Chatterjee, S. & Mariani, M. (2024). "Applications of generative AI and future organizational performance: The mediating role of explorative and exploitative innovation and the moderating role of ethical dilemmas and environmental dynamism". *Technovation, 133*, 103021. doi: https:// doi.org/ 10.1016/j.technovation.2024.103021

Surbakti, F. P. S. (2025). "Systematic Literature Review on Generative AI: Ethical Challenges and Opportunities. International J*ournal of Advanced Computer Science and Applications (IJACSA)*, 16 (5).

UNESCO. (2022). *Recommendation on the ethics of artificial intelligence*: United Nations Educational, Scientific and Cultural Organization.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., . . . Kasirzadeh, A. (2021). "Ethical and social risks of harm from language models". *arXiv preprint arXiv:2112.04359*. https://doi.org/10.48550/arXiv.2112.04359.https://doi.org/10.3389/feduc.2023.1331607

Williams, R. T. (2024). "The ethical implications of using generative chatbots in higher education". P*aper presented at the Frontiers in Education*. ttps://doi.org/10.3389/feduc.2023.1331607

منابع

١. AI, N. (٢٠٢٣). "Artificial intelligence risk management framework (AI RMF ١.٠)". *URL: https://nvlpubs. nist. gov/nistpubs/ai/nist. ai*, ١٠٠-١٠١. https://doi.org/١٠.٦٠٢٨/NIST.AI.١٠٠-١

٢. Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M. & Alfandi, O. (٢٠٢٤a) "A Systematic Review and Analysis of Ethical Challenges of Generative Ai: An Interdisciplinary Perspective". *Available at SSRN 4833030*.

٣. Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M. & Alfandi, O. (٢٠٢٤b). "Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective". *Informatics, 11*(٣): ٥٨. https://doi.org/١٠.١٠١٦/j.orgdyn.٢٠٢٤.١٠١٠٣٢

٤. Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell, S. (٢٠٢١). "*On the dangers of stochastic parrots: Can language models be too big?* " Paper presented at the Proceedings of the ٢٠٢١ ACM conference on fairness, accountability, and transparency. https://doi.org/١٠.١١٤٥/٣٤٤٢١٨٨.٣٤٤٥٩٢٢

٥. Braun, V. & Clarke, V. (٢٠٠٦). "Using thematic analysis in psychology". *Qualitative research in psychology, 3*(٢), ٧٧-١٠١. DOI: ١٠.١١٩١/١٤٧٨٠٨٨٧٠٦qp٠٦٣oa

٦. Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. & Sun, L. (٢٠٢٥). "A Survey of AI-Generated Content (AIGC)". *ACM Comput. Surv., 57*(٥), Article ١٢٥. doi: ١٠.١١٤٥/٣٧٠٤٢٦٢

٧. Carlini, N., Hayes, J., Nasr, M., Jagielski, M., Sehwag, V., Tramer, F., . . . Wallace, E. (٢٠٢٣). "*Extracting training data from diffusion models".* Paper presented at the ٣٢nd USENIX security symposium (USENIX Security ٢٣). https://doi.org/١٠.٤٨٥٥٠/arXiv.٢٣٠١.١٣١٨٨

٨. Cath, C. (٢٠١٨). "Governing artificial intelligence: ethical, legal and technical opportunities and challenges". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* ٣٧٦(٢١٣٣), ٢٠١٨٠٠٨٠. https://doi.org/١٠.١٠٩٨/rsta.٢٠١٨.٠٠٨٠

٩. Chakraborty, S. (٢٠٢٣). "AI and ethics: Navigating the moral landscape", *Investigating the Impact of AI on Ethics and Spirituality* (pp. ٢٥-٣٣): IGI Global. DOI: ١٠.٤٠١٨/٩٧٨-١-٦٦٨٤-٩١٩٦-٦

١٠. Chen, Z. (٢٠٢٣). "Ethics and discrimination in artificial intelligence-enabled recruitment practices". *Humanities and Social Sciences Communications, 10*(١): ٥٦٧. doi: ١٠.١٠٥٧/s٤١٥٩٩-٠٢٣-٠٢٠٧٩-x

١١. Chui, M., Hazan, E., Roberts, R., Singla, A., & Smaje, K. (٢٠٢٣). "The economic potential of generative AI".

١٢. Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., . . . Terem, E. (٢٠٢٣). Worldwide AI ethics: A review of ٢٠٠ guidelines and recommendations for AI governance". *Patterns, 4*(١٠). https://doi.org/١٠.١٠١٦/j.patter.٢٠٢٣.١٠٠٨٥٧

١٣. Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., . . . Wright, R. (٢٠٢٣). "Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, ١٠٢٦٤٢. doi: https://doi.org/١٠.١٠١٦/j.ijinfomgt.٢٠٢٣.١٠٢٦٤٢

١٤. European Cmmission, F. (٢٠٢١). "Regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts".

١٥. Ferrara, E. (٢٠٢٣). "Should chatgpt be biased? challenges and risks of bias in large language models". *arXiv preprint arXiv:2304.03738*. https://doi.org/١٠.٤٨٥٥٠/arXiv.٢٣٠٤.٠٣٧٣٨

١٦. Floridi, L. (٢٠٢٣). "AI as agency without intelligence: On ChatGPT, large language models, and other generative models". *Philosophy & technology, 36*(١), ١٥. https://doi.org/١٠.١٠٠٧/s١٣٣٤٧-٠٢٣-٠٠٦٢١-y

١٧. García-López, I. M. & Trujillo-Liñán, L. (٢٠٢٥). "Ethical and regulatory challenges of Generative AI in education: a systematic review". *Paper presented at the Frontiers in Educatio*n. https://doi.org/١٠.٣٣٨٩/feduc.٢٠٢٥.١٥٦٥٩٣٨

١٨. Gupta, D. (٢٠٢٤). "Generative AI and Deep fake s: Ethical Implications and Detection Techniques". *Journal of Science, Technology and Engineering Research, 1*(١): ٤٥-٥٦. **DOI:** https://doi.org/١٠.٦٤٢٠٦/٢١rgkc٤٠

١٩. Hagendorff, T. (٢٠٢٤). "Mapping the ethics of generative AI: A comprehensive scoping review". *Minds and Machine*s, ٣٤(٤), ٣٩. https://doi.org/١٠.١٠٠٧/s١١٠٢٣-٠٢٤-٠٩٦٩٤-w

٢٠. Iqbal, Q., Khan, D. & Salis, M. (٢٠٢٤). "Navigating the Ethical and Legal Landscape of Artificial Intelligence in Global Governance: A Comprehensive Analysis". *Journal of Asian Development Studies, 13*(٢): ٩٤٥-٩٥٤. DOI: https://doi.org/١٠.٦٢٣٤٥/jads.٢٠٢٤.١٣.٢.٧٥

٢١. Jobin, A., Ienca, M., & Vayena, E. (٢٠١٩). "The global landscape of AI ethics guidelines". *Nature machine intelligence,* ١(٩): ٣٨٩-٣٩٩. https://doi.org/١٠.١٠٣٨/s٤٢٢٥٦-٠١٩-٠٠٨٨-٢

٢٢. Kay, J., Kasirzadeh, A., & Mohamed, S. (٢٠٢٤). "Epistemic injustice in generative AI". P*aper presented at the Proceedings of the AAAI/ACM Conference on AI, Ethics, and Societ*y. DOI: https://doi.org/١٠,١٦٠٩/aies.v٧i١,٣١٦٧١

٢٣. Macdonald, C., Adeloye, D., Sheikh, A. & Rudan, I. (٢٠٢٣). "Can ChatGPT draft a research article? An example of population-level vaccine effectiveness analysis". *Journal of global health, 13*, ٠١٠٠٣. doi: ١٠,٧١٨٩/jogh.١٣,٠١٠٠٣.

٢٤. Marchal, N., Xu, R., Elasmar, R., Gabriel, I., Goldberg, B. & Isaac, W. (٢٠٢٤). "Generative AI misuse: A taxonomy of tactics and insights from real-world data". *arXiv preprint arXiv:2406.13843*. https://doi.org/١٠,٤٨٥٥٠/arXiv.٢٤٠٦,١٣٨٤٣

٢٥. Ning, Y., Teixayavong, S., Shang, Y., Savulescu, J., Nagaraj, V., Miao, D., . . . Liu, M. (٢٠٢٤). "Generative artificial intelligence and ethical considerations in health care: a scoping review and ethics checklist". *The Lancet Digital Health, 6*(١١), e٨٤٨-e٨٥٦. doi: ١٠,١٠١٦/S٢٥٨٩-٧٥٠٠(٢٤)٠٠١٤٣-٢

٢٦. OECD. (٢٠٢٤). *Facts Not Fakes: Tackling Disinformation, Strengthening Information Integrity*: OECD Publishing.

٢٧. Raghavan, M., Barocas, S., Kleinberg, J. & Levy, K. (٢٠٢٠). "*Mi*tigating bias in algorithmic hiring: Evaluating claims and practices". *Paper presented at the Proceedings of the 2020 conference on fairness, accountability, and transparency.* https://doi.org/١٠,١١٤٥/٣٣٥١٠٩٥,٣٣٧٢٨٢٨

٢٨. Rana, N. P., Pillai, R., Sivathanu, B. & Malik, N. (٢٠٢٤). "Assessing the nexus of Generative AI adoption, ethical considerations and organizational performance". *Technovation, 135*, ١٠٣٠٦٤. doi: https://doi.org/١٠,١٠١٦/j.technovation.٢٠٢٤,١٠٣٠٦٤

٢٩. Singh, K., Chatterjee, S. & Mariani, M. (٢٠٢٤). "Applications of generative AI and future organizational performance: The mediating role of explorative and exploitative innovation and the moderating role of ethical dilemmas and environmental dynamism". *Technovation, 133*, ١٠٣٠٢١. doi: https://doi.org/١٠,١٠١٦/j.technovation.٢٠٢٤,١٠٣٠٢١

٣٠. Surbakti, F. P. S. (٢٠٢٥). "Systematic Literature Review on Generative AI: Ethical Challenges and Opportunities. International J*ournal of Advanced Computer Science and Applications (IJACSA),* ١٦ (٥).

٣١. UNESCO. (٢٠٢٢). *Recommendation on the ethics of artificial intelligence*: United Nations Educational, Scientific and Cultural Organization.

٣٢. Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., . . . Kasirzadeh, A. (٢٠٢١). "Ethical and social risks of harm from language models". *arXiv preprint arXiv:2112.04359*. https://doi.org/١٠,٤٨٥٥٠/arXiv.٢١١٢,٠٤٣٥٩. https://doi.org/١٠,٣٣٨٩/feduc.٢٠٢٣,١٣٣١٦٠٧

٣٣. Williams, R. T. (٢٠٢٤). "The ethical implications of using generative chatbots in higher education". P*aper presented at the Frontiers in Education*. ttps://doi.org/١٠,٣٣٨٩/feduc.٢٠٢٣,١٣٣١٦٠٧