

هوش مصنوعی عامل و پاسخگویی مدیریتی: ارائه چارچوبی مفهومی برای حکمرانی در سازمان های ترکیبی انسان-هوش مصنوعی

مهدی تبرک سوره

دانشجوی دکترای مدیریت دولتی دانشگاه آزاد واحد بین المللی ارس

Mahdi.tabarrok.s@gmail.com

چکیده

پیشرفت های اخیر در حوزه هوش مصنوعی، به ویژه ظهور سیستم های موسوم به «هوش مصنوعی عامل»، مرزهای سنتی مدیریت را به چالش کشیده است. این سیستم ها دیگر صرفاً ابزاری برای اجرای دستورات نیستند، بلکه با قابلیت برنامه ریزی، تصمیم گیری و اقدام مستقل، وارد عرصه ای شده اند که تا پیش از این در انحصار انسان بود. این تحول، مسئله پاسخگویی مدیریتی را با پیچیدگی های بی سابقه ای مواجه ساخته است. به باور نگارنده، آنچه امروز بیش از هر چیز سازمان ها را تهدید می کند، نه نبود فناوری که نبود چارچوبی روشن برای پاسخگویی در برابر اقدامات سیستم های خودمختار است. مقاله حاضر با هدف پر کردن این شکاف نظری و عملی، به دنبال ارائه چارچوبی مفهومی برای حکمرانی هوش مصنوعی عامل در سازمان هاست. این پژوهش با بهره گیری از روش مرور نظام مند ادبیات و تحلیل محتوای کیفی منابع معتبر، به شناسایی ابعاد کلیدی پاسخگویی مدیریتی در مواجهه با عامل های هوشمند پرداخته است. یافته ها نشان می دهد که پاسخگویی در عصر هوش مصنوعی عامل، نیازمند بازتعریف در چهار بعد شفافیت الگوریتمی، قابلیت توضیح پذیری تصمیمات، نظارت انسانی فعال و تعیین مرزهای اختیار است. آنچه این چارچوب را از مدل های پیشین متمایز می سازد، تأکید بر پویایی و تعامل این چهار بعد با یکدیگر و نیز در نظر گرفتن اقتضائات هر سازمان است. چارچوب پیشنهادی این مقاله، می تواند به عنوان نقشه راهی برای مدیران و پژوهشگران در جهت درک و مدیریت چالش های نوین حکمرانی فناوری در سازمان های ترکیبی مورد استفاده قرار گیرد. با این حال، ماهیت کیفی و نظری این پژوهش، تعمیم پذیری نتایج آن را با احتیاط همراه می سازد و پژوهش های تجربی آتی برای اعتبارسنجی چارچوب پیشنهادی ضروری به نظر می رسد.

واژگان کلیدی: هوش مصنوعی عامل، پاسخگویی مدیریتی، حکمرانی فناوری، سازمان های ترکیبی، چارچوب مفهومی

۱. مقدمه

ظهور هوش مصنوعی در عرصه مدیریت و سازمان، شاید یکی از عمیق ترین تحولاتی باشد که بشر در طول تاریخ مدیریت با آن روبرو شده است. اما آنچه این تحول را از موج های قبلی فناوری اطلاعات متمایز می کند، نه صرفاً قدرت محاسباتی بی نظیر این سیستم ها، بلکه تغییر در ماهیت رابطه انسان و ماشین است. تا همین چند سال پیش، مدیران با نرم افزارهایی سروکار داشتند که ابزارهایی منفعل و وابسته به دستورات انسان بودند. اما امروز، نسل جدیدی از سیستم های هوش مصنوعی وارد سازمان ها شده اند که نه تنها دستورات را اجرا می کنند، بلکه خودشان نیز می توانند هدف تعیین کنند، برنامه بریزند و دست به اقدام بزنند.

تجربه سال ها مطالعه و مشاوره در حوزه تحول دیجیتال، به نگارنده نشان داده است که بسیاری از سازمان ها با شور و شوقی وصفناپذیر به استقبال این فناوری می روند، اما غافل از اینکه پیچیده ترین چالش آن، نه در کدنویسی و پردازش داده، که در بازتعریف ساختارهای پاسخگویی و حکمرانی نهفته است. سیستم های هوش مصنوعی عامل که در ادبیات با عنوان «Agentic AI» از آنها یاد می شود، مفهومی فراتر از اتوماسیون ساده را معرفی می کنند. این سیستم ها دارای درجاتی از استقلال، هدف گذاری و کنشگری هستند که پیش تر در انحصار انسان بود.

اهمیت این پرسش از آن جهت دوچندان است که واگذاری اختیار به سیستم های خودمختار، بدون طراحی سازوکارهای پاسخگویی شفاف، می تواند سازمان ها را با خطرات جدی از جمله تصمیمات غیراخلاقی، تبعیض آمیز، یا حتی پرخطر مواجه سازد. از سوی دیگر، بررسی ادبیات موجود نشان می دهد که بیشتر پژوهش های انجام شده تا به امروز، بر جنبه های فنی هوش مصنوعی عامل متمرکز بوده اند و به ندرت به ابعاد مدیریتی و به ویژه مسئله پاسخگویی پرداخته اند. این شکاف پژوهشی، انگیزه اصلی شکل گیری مقاله حاضر بوده است.

بر این اساس، مقاله پیش رو با هدف ارائه چارچوبی مفهومی برای حکمرانی هوش مصنوعی عامل در سازمان های ترکیبی نگاشته شده است. پرسش اصلی پژوهش این است که ابعاد و مؤلفه های پاسخگویی مدیریتی در مواجهه با هوش مصنوعی عامل کدامند و چگونه می توان چارچوبی نظام مند برای حکمرانی این سیستم ها در سازمان طراحی کرد؟

۲. پیشینه تحقیق و مبانی نظری

۲-۱. سیر تحول هوش مصنوعی در مدیریت: از ابزار تا عامل

هوش مصنوعی در طول یک دهه اخیر، تحولی بنیادین از یک «ابزار تحلیلی» به «عاملی مستقل» را تجربه کرده است. در نسل اول، هوش مصنوعی عمدتاً به عنوان ابزاری در اختیار مدیران قرار می گرفت که به آنها در پردازش اطلاعات کمک می کرد و تصمیم گیرنده نهایی، انسان بود. اما نسل جدید، هوش مصنوعی عاملی است که نه فقط تحلیل گر که کنشگر است و می تواند به طور مستقل تصمیم گیری و اقدام کند.

پژوهش های جدید نشان می دهد که هوش مصنوعی عامل در حوزه های مختلفی از جمله مدیریت عملیات، خدمات مشتری و حتی حکمرانی دولتی در حال نفوذ است. برای نمونه، سیستم های هوش مصنوعی عامل در مراکز عملیات امنیتی (SOC) می توانند به طور خودکار تهدیدات امنیتی را شناسایی، پاسخ دهی و کاهش دهند که پیش بینی می شود کارایی عملیاتی را تا ۴۰ درصد افزایش دهد. همچنین در بخش دولتی، نمونه هایی مانند سیستم پاسخگویی به بحران های بهداشتی در هند نشان داده است که چگونه عامل های هوشمند با نظارت انسانی می توانند میلیون ها تعامل با شهروندان را مدیریت کنند .

از نگاه نگارنده، آنچه امروز سازمان ها را تهدید می کند، نبود چارچوبی روشن برای حکمرانی این سیستم هاست. پژوهش ها نشان می دهد که سازمان هایی که پیش از استقرار فناوری، به بازطراحی نظام های پاسخگویی خود پرداخته اند، عملکرد بهتری در مدیریت ریسک های هوش مصنوعی داشته اند .

۲-۲. مروری بر پژوهش های پیشین در حوزه هوش مصنوعی و پاسخگویی

پژوهش های انجام شده در حوزه هوش مصنوعی و مدیریت را می توان به چند دسته کلی تقسیم کرد:

دسته اول: پژوهش های فنی و پیاده سازی . این پژوهش ها که بیشتر در مجلات مهندسی و علوم کامپیوتر منتشر شده اند، به معماری سیستم های هوش مصنوعی عامل، الگوریتم های یادگیری و بهینه سازی عملکرد پرداخته اند. برای نمونه، مطالعه مروری نظام مند بر ۴۲ مقاله از پایگاه های Scopus ، ProQuest و ScienceDirect به بررسی پیشرفت های فناوری هوش مصنوعی عامل در حوزه های سلامت، مالی، آموزش و سایر حوزه های کسب و کار پرداخته است .

دسته دوم: پژوهش های مرتبط با پیامدهای سازمانی و اخلاقی . این پژوهش ها به نقش های سازمانی، تأثیرات و چالش های اخلاقی هوش مصنوعی عامل پرداخته اند. مطالعه ای با مرور ۱۵ مقاله منتشر شده بین ۲۰۲۰ تا ۲۰۲۵ نشان داده است که هوش مصنوعی عامل نقش های تحول آفرینی در تصمیم گیری استراتژیک، کارایی عملیاتی و نوآوری ایفا می کند، اما در عین حال چالش های اخلاقی مانند سوگیری الگوریتمی، نگرانی های حریم خصوصی و شکاف های پاسخگویی را نیز به همراه دارد .

دسته سوم: پژوهش های مرتبط با حکمرانی و چارچوب های نظارتی . در این حوزه، چندین چارچوب معتبر توسط نهادهای بین المللی ارائه شده است . از جمله می توان به چارچوب حکمرانی هوش مصنوعی عامل سنگاپور (MGF for Agentic AI) اشاره کرد که در ژانویه ۲۰۲۶ منتشر شد و در ماه مه ۲۰۲۶ با افزودن مطالعه های موردی واقعی به روزرسانی گردید . همچنین راجات پاندی (۲۰۲۶) چارچوبی شش اصل محور را بر اساس مدل های موجود مانند NIST AI RMF و ISO/IEC 42001 ارائه داده است که بر ردیابی، پاسخگویی و انطباق پذیری در استقرار سیستم های عامل تأکید دارد .

با این حال، از نگاه نگارنده، این پژوهش ها علی رغم ارزشمندی، دچار نقص هایی هستند:

نقص اول: رویکرد تک بعدی به پاسخگویی . بسیاری از پژوهش ها، پاسخگویی را به یک یا دو بعد (معمولاً شفافیت و نظارت) تقلیل داده اند، در حالی که پاسخگویی در مواجهه با هوش مصنوعی عامل، پدیده ای چندبعدی و پویاست .

نقص دوم: غفلت از زمینه سازمانی. بیشتر مدل های ارائه شده برای حکمرانی هوش مصنوعی، جنبه های اقتصادی و زمینه ای را نادیده گرفته اند. پاسخگویی در یک سازمان تولیدی با یک سازمان خدماتی، یا در یک شرکت کوچک با یک شرکت چندملیتی، نمی تواند یکسان باشد.

۲-۳. مفهوم پردازی پاسخگویی مدیریتی

پاسخگویی (Accountability) یکی از مفاهیم بنیادین در نظریه های مدیریت و سازمان است. در یک تعریف ساده اما جامع، پاسخگویی به معنای «وظیفه یک فرد یا نهاد در قبال توجیه و توضیح اقدامات خود در برابر ذی نفعان» تعریف می شود. اما آنچه در این تعریف، حضوری پررنگ اما ناگفته دارد، فرض «کنشگر انسانی» است. در نظریه های کلاسیک، پاسخگویی در زنجیره ای خطی از انسان به انسان دیگر تعریف می شد.

اما در عصر هوش مصنوعی عامل، این زنجیره خطی دچار گسست می شود. پرسش اصلی این است که وقتی سیستمی که خودش به طور مستقل تصمیم گرفته است، دچار خطا می شود، چه کسی پاسخگوست؟. چارچوب های جدید مانند ACAAI حساب دهی از طریق طراحی برای هوش مصنوعی عامل) که در آگوست ۲۰۲۶ منتشر شده، به این موضوع از منظر «مهندسی سیستم ها» نگاه می کند و پاسخگویی را به عنوان یک ویژگی عملیاتی تعریف می کند که باید در طراحی، استقرار و بهره برداری از سیستم های خودمختار نهادینه شود.

۲-۴. حکمرانی فناوری در سازمان های ترکیبی

سازمان های امروزی به سرعت در حال تبدیل شدن به نهادهای «ترکیبی» هستند که در آنها عامل های غیرانسان نیز در فرآیندهای تصمیم گیری و اجرا مشارکت دارند. حکمرانی در چنین سازمان هایی، به چارچوب های نوینی نیاز دارد.

دو چارچوب اصلی در این حوزه عبارتند از:

NIST AI RMF: این چارچوب که توسط مؤسسه ملی فناوری و استانداردهای آمریکا منتشر شده، یک راهنمای داوطلبانه برای مدیریت ریسک هوش مصنوعی است که بر چهار کارکرد «حکمرانی، نقشه برداری، اندازه گیری و مدیریت» استوار است. این چارچوب برای تیم های فنی و مهندسی قابل استفاده است و بر آزمون مدل ها از نظر سوگیری، استحکام و عملکرد تأکید دارد.

ISO/IEC 42001: این استاندارد بین المللی که در دسامبر ۲۰۲۳ منتشر شد، اولین استاندارد برای سیستم مدیریت هوش مصنوعی است که قابلیت صدور گواهینامه توسط نهادهای معتبر را دارد. این استاندارد بر مستندسازی، بهبود مستمر و پاسخگویی سازمانی تأکید دارد.

برای حکمرانی هوش مصنوعی عامل، ترکیب این دو چارچوب پیشنهاد می شود: استفاده از NIST AI RMF برای شناسایی و مدیریت ریسک های عملیاتی و ISO/IEC 42001 برای ایجاد یک سیستم مدیریتی قابل ممیزی و پاسخگو.

۳. روش تحقیق

۳-۱. رویکرد و روش پژوهش

پژوهش حاضر با توجه به ماهیت اکتشافی و نظری خود، از رویکرد کیفی و روش مرور نظاممند ادبیات (Systematic Literature Review) بهره گرفته است. انتخاب این روش به دلیل نوظهور بودن موضوع هوش مصنوعی عامل و پاسخگویی مدیریتی است که هنوز نظریه های بالغی درباره آن شکل نگرفته است. مرور نظاممند ادبیات، امکان شناسایی، ارزیابی و ترکیب یافته های پراکنده از پژوهش های پیشین را فراهم می آورد تا بتوان بر اساس آن، چارچوبی مفهومی و منسجم ارائه داد.

روش مرور نظاممند در این پژوهش، بر اساس پروتکل استاندارد PRISMA طراحی و اجرا شده است.

۳-۲. استراتژی جستجو و پایگاه های داده

جستجوی نظاممند در پایگاه های داده معتبر بین المللی شامل 'Google Scholar'، 'Scopus'، 'Web of Science'، 'ProQuest' و 'ScienceDirect' انجام شد. همچنین برای پوشش کامل تر، از پایگاه های داخلی مانند 'Magiran' و 'Noormags' نیز استفاده گردید.

استراتژی جستجو با ترکیب کلیدواژه های زیر و با استفاده از عملگرهای بولی AND، (OR) طراحی شد:

- کلیدواژه های اصلی "Agentic AI" OR "Autonomous AI" OR "AI Agency" AND "Management" OR "Organization"
- کلیدواژه های مرتبط با پاسخگویی "Accountability" OR "Responsibility" OR "Governance" OR "Oversight"
- کلیدواژه های مرتبط با حکمرانی "AI Governance" OR "Technology Governance" OR "Hybrid Organization"

جستجو در بازه زمانی ۲۰۱۸ تا ۲۰۲۶ انجام شد.

۳-۳. معیارهای ورود و خروج

معیارهای ورود:

- مقالات منتشر شده در مجلات علمی-پژوهشی معتبر (دارای نمایه در پایگاه های داده بین المللی)
- مقالات مرتبط با هوش مصنوعی در مدیریت و سازمان
- مقالات مرتبط با پاسخگویی، حکمرانی یا نظارت در حوزه فناوری
- مقالات منتشر شده به زبان فارسی یا انگلیسی

معیارهای خروج:

- مقالات صرفاً فنی و مهندسی که به جنبه های مدیریتی نپرداخته اند
- مقالاتی که صرفاً به هوش مصنوعی به عنوان ابزار پرداخته و به ابعاد عاملیت توجه نکرده اند
- مقالات منتشر شده در کنفرانس های غیرمعتبر

۳-۴. فرآیند غربالگری و انتخاب مقالات

فرآیند جستجو و غربالگری در چهار مرحله انجام شد:

مرحله اول (شناسایی): جستجوی اولیه با کلیدواژه های تعیین شده، منجر به شناسایی ۸۴۷ مقاله در پایگاه های داده مختلف شد. پس از حذف مقالات تکراری (۲۳۱ مقاله)، ۶۱۶ مقاله باقی ماند.

مرحله دوم (غربالگری اولیه): در این مرحله، عنوان و چکیده مقالات بررسی شد. مقالات نامرتبط حذف شدند. ۴۲۳ مقاله حذف و ۱۹۳ مقاله برای مرحله بعدی انتخاب شد.

مرحله سوم (غربالگری ثانویه): متن کامل مقالات باقی مانده مطالعه و بر اساس معیارهای ورود و خروج ارزیابی شد. ۱۲۴ مقاله دیگر حذف و ۶۹ مقاله باقی ماند.

مرحله چهارم (گنجاندن نهایی): در مرحله نهایی، پس از مطالعه عمیق و ارزیابی کیفیت مقالات، ۴۷ مقاله به عنوان جامعه نهایی پژوهش انتخاب شدند که شامل ۳۹ مقاله انگلیسی و ۸ مقاله فارسی بود.

۳-۵. روش تحلیل داده ها و آمار توصیفی

برای تحلیل داده های استخراج شده از مقالات منتخب، از روش تحلیل محتوای کیفی با رویکرد استقرایی استفاده شد. فرآیند تحلیل با استفاده از نرم افزار MAXQDA 2024 انجام شد. برای افزایش اعتبار و پایایی، از روش توافق بین کدگذاران استفاده گردید (نرخ توافق ۸۷ درصد).

جدول ۱: توزیع فراوانی مقالات بر اساس سال انتشار

درصد	تعداد مقالات	سال انتشار
۴.۳٪	۲	۲۰۱۸
۶.۴٪	۳	۲۰۱۹
۸.۵٪	۴	۲۰۲۰
۱۰.۶٪	۵	۲۰۲۱

18 September 2026 -TBILISI GEORGIA

۲۰۲۲	۶	۱۲.۸٪
۲۰۲۳	۸	۱۷.۰٪
۲۰۲۴	۹	۱۹.۱٪
۲۰۲۵	۷	۱۴.۹٪
۲۰۲۶	۳	۶.۴٪
مجموع	۴۷	۱۰۰٪

توضیح: بیشترین تعداد مقالات در سال های ۲۰۲۳ و ۲۰۲۴ منتشر شده اند که نشان دهنده رشد فزاینده توجه پژوهشگران به این موضوع است.

جدول ۲: توزیع فراوانی مقالات بر اساس نوع منبع

نوع منبع	تعداد مقالات	درصد
مجلات علمی- پژوهشی بین المللی	۲۸	۵۹.۶٪
مجلات علمی- پژوهشی داخلی	۶	۱۲.۸٪
مقالات کنفرانس های معتبر	۷	۱۴.۹٪
گزارش های سازمانی/ نهادی	۴	۸.۵٪
پیش چاپ ها (SSRN)	۲	۴.۲٪
مجموع	۴۷	۱۰۰٪

جدول ۳: توزیع فراوانی مقالات بر اساس حوزه موضوعی

حوزه موضوعی اصلی	تعداد مقالات	درصد
هوش مصنوعی در مدیریت و سازمان	۱۸	۳۸.۳٪
پاسخگویی و مسئولیت پذیری در فناوری	۱۲	۲۵.۵٪
حکمرانی فناوری و هوش مصنوعی	۱۰	۲۱.۳٪
سازمان های ترکیبی و تعامل انسان- ماشین	۷	۱۴.۹٪
مجموع	۴۷	۱۰۰٪

18 September 2026 -TBILISI GEORGIA

جدول ۴: توزیع فراوانی مقالات بر اساس روش پژوهش

روش پژوهش	تعداد مقالات	درصد
مروری/نظری	۱۹	۴۰.۴٪
کمی/پیمایشی	۱۱	۲۳.۴٪
کیفی/مطالعه موردی	۹	۱۹.۲٪
ترکیبی (آمیخته)	۵	۱۰.۶٪
روش های دیگر	۳	۶.۴٪
مجموع	۴۷	۱۰۰٪

جدول ۵: خلاصه فرآیند غربالگری مقالات بر اساس پروتکل (PRISMA)

مرحله	شرح	تعداد مقالات
مرحله ۱: شناسایی	تعداد کل مقالات شناسایی شده در پایگاه های داده	۸۴۷
	حذف مقالات تکراری	-۲۳۱
	مقالات باقی مانده	۶۱۶
مرحله ۲: غربالگری اولیه	بررسی عنوان و چکیده	۶۱۶
	مقالات حذف شده	-۴۲۳
	مقالات انتخاب شده	۱۹۳
مرحله ۳: واجدسازی	بررسی متن کامل	۱۹۳
	مقالات حذف شده	-۱۲۴
	مقالات واجد شرایط	۶۹
مرحله ۴: گنجاندن	ارزیابی نهایی کیفیت روش شناختی	۶۹
	مقالات حذف شده	-۲۲
	مقالات نهایی وارد شده به تحلیل	۴۷

جدول ۶: خلاصه کدهای استخراج شده از تحلیل محتوا

مرحله کدگذاری	تعداد
کدهای اولیه (کدگذاری باز)	۵۷
کدهای محوری (پس از ادغام)	۲۶
مقوله های نهایی	۴
زیرمقوله ها	۱۳

۴. یافته ها

تحلیل محتوای کیفی ۴۷ مقاله منتخب، منجر به شناسایی ۵۷ کد اولیه گردید که پس از ادغام و دسته بندی، در قالب ۴ مقوله اصلی و ۱۳ زیرمقوله سازماندهی شدند.

جدول ۷: توزیع فراوانی ابعاد پاسخگویی در مقالات مورد تحلیل

بعد پاسخگویی	تعداد مقالات دارای این بعد	درصد
شفافیت الگوریتمی	۴۲	۸۹.۴٪
نظارت انسانی فعال	۳۹	۸۳.۰٪
قابلیت توضیح پذیری تصمیمات	۳۷	۷۸.۷٪
تعیین مرزهای اختیار	۳۴	۷۲.۳٪

توضیح: شفافیت الگوریتمی با ۸۹ درصد، بیشترین فراوانی را در میان مقالات داشته است و «تعیین مرزهای اختیار» با ۷۲ درصد، کمترین فراوانی را داشته است. این موضوع نشان دهنده آن است که پژوهشگران توجه بیشتری به شفافیت داشته اند و ابعاد دیگر مانند تعیین مرزهای اختیار، نیازمند پژوهش های بیشتری هستند.

۴-۱. شفافیت الگوریتمی (Algorithmic Transparency)

این مقوله که در ۸۹ درصد از مقالات مورد تحلیل (۴۲ مقاله) شناسایی شد، شامل سه زیرمقوله است:

۴-۱-۱. **شفافیت در سطح کد و الگوریتم:** به معنای دسترسی به کد منبع و ساختار الگوریتم ها. چارچوب حکمرانی سنگاپور بر شفافیت در سطح کد به عنوان یکی از مؤلفه های ایمنی و قابلیت اطمینان تأکید دارد.

۴-۱-۲. **شفافیت در سطح داده:** به معنای شناخت داده هایی که سیستم بر اساس آنها آموزش دیده و تصمیم می گیرد. یافته ها نشان می دهد که سازمان ها در این سطح از شفافیت، ضعیف ترین عملکرد را دارند.

18 September 2026 -TBILISI GEORGIA

۴-۱-۳. شفافیت در سطح فرآیند: به معنای درک گام‌هایی که سیستم برای رسیدن به تصمیم طی می‌کند. چارچوب ACAAI بر شفافیت در سطح فرآیند به عنوان یکی از شش حوزه کنترل تأکید دارد.

۴-۲. قابلیت توضیح پذیری تصمیمات (Explainability)

این مقوله که در ۷۸ درصد از مقالات (۳۷ مقاله) شناسایی شد، شامل سه زیرمقوله است:

۴-۲-۱. توضیح پذیری فنی: توانایی سیستم در ارائه دلایل فنی و ریاضی برای تصمیمات خود.

۴-۲-۲. توضیح پذیری مدیریتی: توانایی ترجمه دلایل فنی به زبانی که برای مدیران و ذی‌نفعان غیرفنی قابل درک باشد. چارچوب حکمرانی سنگاپور بر این سطح از توضیح پذیری برای نظارت معنادار انسانی تأکید دارد.

۴-۲-۳. تعارض دقت و توضیح پذیری: یکی از یافته‌های مهم، شناسایی تعارض میان دقت و توضیح پذیری است که در ادبیات با عنوان "Accuracy-Explainability Trade-off" شناخته می‌شود.

۴-۳. نظارت انسانی فعال (Human-in-the-Loop Oversight)

این مقوله که در ۸۳ درصد از مقالات (۳۹ مقاله) شناسایی شد، شامل چهار زیرمقوله است:

۴-۳-۱. نظارت سطح اول (مشاهده‌گرانه): مدیر صرفاً اقدامات سیستم را مشاهده می‌کند.

۴-۳-۲. نظارت سطح دوم (مشورتی): مدیر می‌تواند پیشنهاداتی ارائه دهد اما سیستم الزامی به پذیرش ندارد.

۴-۳-۳. نظارت سطح سوم (تصمیم‌ساز): مدیر می‌تواند تصمیمات سیستم را و تو یا اصلاح کند. در چارچوب حکمرانی سنگاپور، برای اقدامات پرریسک (سطح ۳)، سیستم به کلی از اقدام منع می‌شود.

۴-۳-۴. نظارت پس از وقوع در مقابل لحظه‌ای: یافته‌ها نشان می‌دهد که نظارت پس از وقوع در عصر هوش مصنوعی عاملی که تصمیمات در کسری از ثانیه گرفته می‌شوند، کارایی خود را از دست داده است.

۴-۴. تعیین مرزهای اختیار (Boundary Setting)

این مقوله که در ۷۲ درصد از مقالات (۳۴ مقاله) شناسایی شد، شامل سه زیرمقوله است:

۴-۴-۱. مرزهای مبتنی بر سطح ریسک: حوزه‌های تصمیم‌گیری بر اساس سطح ریسک دسته‌بندی می‌شوند. در مطالعه موردی شرکت Dayos، هر نوع درخواست IT بر اساس شدت تأثیر، قابلیت بازگشت و امکان نظارت انسانی ارزیابی شد.

۴-۴-۲. مرزهای مبتنی بر حوزه تصمیم: برخی حوزه‌ها به کلی از محدوده اختیار سیستم خارج می‌شوند.

18 September 2026 -TBILISI GEORGIA

۴-۴-۳. مرزهای مبتنی بر سطح شفافیت و توضیح پذیری: هرچه سیستم در یک حوزه خاص شفاف تر و قابل توضیح تر باشد، اختیار بیشتری به آن واگذار می شود .

۵. بحث و نتیجه گیری

۵-۱. بحث و تفسیر یافته ها

چارچوب مفهومی استخراج شده در این پژوهش، نشان می دهد که پاسخگویی مدیریتی در مواجهه با هوش مصنوعی عامل، پدیده ای چندبعدی و به هم پیوسته است. این یافته با چارچوب های جدید حکمرانی مانند MGF برای هوش مصنوعی عامل سنگاپور که بر چهار بعد «ارزیابی و محدودسازی ریسک»، «تضمین پاسخگویی معنادار انسانی»، «اجرای کنترل های فنی و فرآیندها» و «مسئولیت پذیری کاربر نهایی» تأکید دارد، همخوانی دارد .

نکته حائز اهمیت، تعامل پویا میان چهار بعد شناسایی شده است. افزایش شفافیت الگوریتمی می تواند به قابلیت توضیح پذیری بیشتر بینجامد و این خود، امکان نظارت انسانی فعال تری را فراهم می آورد و در نهایت، تعیین مرزهای اختیار را دقیق تر می سازد. چارچوب ACAAI نیز بر این نگاه سیستمی تأکید دارد و پاسخگویی را به عنوان یک «ویژگی مهندسی سیستم» تعریف می کند که از پیاده سازی هماهنگ کنترل های سازمانی و فنی در سراسر چرخه عمر سیستم های هوش مصنوعی عامل حاصل می شود .

تعیین مرزهای اختیار با ۷۲ درصد فراوانی، کمترین توجه را در میان پژوهشگران به خود جلب کرده است. این در حالی است که چارچوب های جدید حکمرانی بر این بعد تأکید ویژه ای دارند. در مطالعه موردی شرکت Dayos، سیستم از سطوح ریسک طبقه بندی شده برای هدایت و محدودسازی اقدامات عامل استفاده کرد و برای اقدامات با شدت بالا، عامل به کلی از اقدام منع شد . همچنین چارچوب پانندی (۲۰۲۶) بر تعیین مرزهای اختیار از طریق مکانیسم های عملیاتی مانند شاخص Log Retention و معماری حکمرانی زمان اجرا تأکید دارد .

یافته های این پژوهش همچنین نشان می دهد که چارچوب های موجود مانند NIST AI RMF و ISO/IEC 42001 اگرچه زیرساخت های لازم را فراهم می کنند، اما برای حکمرانی هوش مصنوعی عامل نیازمند تکمیل و عملیاتی سازی هستند. همان طور که در منابع معتبر اشاره شده، NIST AI RMF به عنوان یک «دفترچه راهنمای عملیاتی» برای مدیریت ریسک و ISO/IEC 42001 به عنوان یک «سیستم مدیریت» قابل ممیزی عمل می کنند و ترکیب این دو، رویکرد مطلوبی برای سازمان ها محسوب می شود .

۵-۲. محدودیت های مطالعه

محدودیت اول: ماهیت کیفی و نظری پژوهش. این پژوهش بر اساس مرور نظام مند ادبیات و تحلیل محتوای کیفی صورت گرفته است و فاقد آزمون تجربی چارچوب پیشنهادی است.

18 September 2026 -TBILISI GEORGIA

محدودیت دوم: وابستگی به منابع موجود. با توجه به سرعت بالای تحولات در حوزه هوش مصنوعی، ممکن است برخی از یافته ها با ظهور فناوری های جدید نیازمند بازنگری باشند.

محدودیت سوم: غلبه منابع انگلیسی زبان. ۸۳ درصد منابع مورد تحلیل به زبان انگلیسی بودند که ممکن است باعث غفلت از برخی بافت های فرهنگی خاص جوامع غیرانگلیسی زبان شده باشد.

محدودیت چهارم: عدم لحاظ تنوع صنایع. حجم ناکافی پژوهش های صنعت محور، امکان تحلیل دقیق اقتضائات هر صنعت را فراهم نیاورد.

۳-۵. میزان تعمیم پذیری یافته ها

تعمیم پذیری نظری: چارچوب مفهومی ارائه شده از قابلیت تعمیم پذیری نظری برخوردار است و می تواند به عنوان نقشه راهی برای پژوهش های آتی استفاده شود.

تعمیم پذیری زمینه ای: با توجه به اینکه منابع مورد تحلیل عمدتاً از بافت سازمان های غربی بوده اند، تعمیم پذیری به سازمان های با فرهنگ های متفاوت نیازمند مطالعه های تکمیلی است.

تعمیم پذیری عملی: چارچوب پیشنهادی می تواند به عنوان ابزار خودارزیابی برای مدیران استفاده شود، اما باید با در نظر گرفتن اقتضائات خاص هر سازمان بومی سازی شود.

۴-۵. نتیجه گیری نهایی

این مقاله با هدف پر کردن شکاف نظری موجود در حوزه پاسخگویی مدیریتی در مواجهه با هوش مصنوعی عامل، چارچوبی مفهومی برای حکمرانی در سازمان های ترکیبی ارائه داد. یافته ها نشان داد که پاسخگویی مدیریتی در عصر هوش مصنوعی عامل، مفهومی چندبعدی است که بر چهار بعد شفافیت الگوریتمی، قابلیت توضیح پذیری تصمیمات، نظارت انسانی فعال و تعیین مرزهای اختیار استوار است. چارچوب های جدید حکمرانی مانند MGF سنگاپور و ACAAI نیز بر همین ابعاد تأکید دارند و نشان می دهند که پاسخگویی باید به عنوان یک ویژگی عملیاتی در طراحی، استقرار و بهره برداری از سیستم های خودمختار نهادینه شود.

در جهانی که هر روز ماشین ها مستقل تر می شوند، طراحی سازوکارهای پاسخگویی، نه یک انتخاب که یک ضرورت انکارناپذیر برای بقا و مشروعیت سازمان هاست.

۵-۵. پیشنهاد های پژوهش های آتی

پیشنهاد اول: پژوهش های تجربی و کمی، انجام پژوهش های تجربی برای آزمون چارچوب پیشنهادی و توسعه مقیاسی برای سنجش آمادگی پاسخگویی سازمانی در مواجهه با هوش مصنوعی عامل.

18 September 2026 -TBILISI GEORGIA

پیشنهاد دوم: پژوهش های تطبیقی و زمینه ای، انجام پژوهش های تطبیقی میان صنایع مختلف برای شناسایی اقتضات خاص هر صنعت.

پیشنهاد سوم: پژوهش های کیفی مبتنی بر مطالعه موردی، بررسی تجربه زیسته مدیرانی که با سیستم های هوش مصنوعی عامل کار کرده اند.

پیشنهاد چهارم: پژوهش های آینده پژوهانه، استفاده از روش هایی مانند دلفی یا سناریونگاری برای شناسایی چالش های پیش روی پاسخگویی مدیریتی در افق های زمانی بلندمدت.

پیشنهاد پنجم: پژوهش های بین رشته ای، همکاری میان رشته های مدیریت، حقوق، فلسفه اخلاق و علوم کامپیوتر برای تدوین چارچوب های جامع تر.

فهرست منابع

1. **IMDA (2026).** *Updated Model AI Governance Framework for Agentic AI*. Infocomm Media Development Authority, Singapore. Available at: <https://www.imda.gov.sg/>
2. **IMDA (2026).** *Discussion Paper on Legal Responsibility for AI Agents*. Infocomm Media Development Authority, Singapore.
3. **Pandey, R. (2026).** *The Agentic AI Governance Framework: A Universal Model for Risk, Accountability, and Compliance in Autonomous Systems*. SSRN. Available at: <https://ssrn.com/>
4. **ACAAI (2026).** *Accountability by Design for Agentic AI: A Lifecycle Framework for Engineering Accountability in Autonomous AI Systems*. Zenodo. Available at: <https://zenodo.org/>
5. **IDC (2026).** *Autonomy on a Leash: A Framework for Agentic AI Governance*. IDC Perspective, Document #EUR154859826.
6. **NIST (2023).** *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. National Institute of Standards and Technology, U.S. Department of Commerce.
7. **ISO/IEC (2023).** *ISO/IEC 42001:2023 - Artificial Intelligence - Management System*. International Organization for Standardization.
8. **Tran, G.T.C., Le Dinh, T., & Le, T.D. (2025).** Agentic AI Across Technologies, Applications, and Development Domains: A Systematic Literature Review. *7th International Conference on Data Analytics & Management (ICDAM-2026)*, pp. 692.

18 September 2026 -TBILISI GEORGIA

9. **Abror, M.F., Ilyas, M.M.O., Tediantoro, M.R., et al. (2025).** Agentic Artificial Intelligence in Business: A Systematic Literature Review of Roles, Impacts, and Ethical Challenges. *International Conference on Digital Business Innovation and Technology Management (ICONBIT)*, 1(2).
10. **Anghel, I., & Cioara, T. (2026).** Generative AI for IT Project Management: A Systematic Review and Future Research Agenda. *Systems*, 14(6), 722. DOI: 10.3390/systems14060722.
11. **Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al. (2021).** The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. DOI: 10.1136/bmj.n71.
12. **Tranfield, D., Denyer, D., & Smart, P. (2003).** Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), 207-222.
13. **Snyder, H. (2019).** Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333-339.
14. **Krippendorff, K. (2018).** *Content analysis: An introduction to its methodology* (4th ed.). Sage Publications.
15. **Yin, R. K. (2014).** *Case study research: Design and methods* (5th ed.). Sage Publications.