پژوهش‌های مدیریت منابع سازمانی
دوره ۱۵، شماره ۲، تابستان ۱٤۰٤

طراحی چارچوب مفهومی سنجش تحقق هوش مصنوعی مسئولیت‌پذیر در مدیریت منابع انسانی

مونا جامی‌پور[۱]، شهناز اکبری امامی[۲*]

۱- دانشیار، گروه مدیریت، دانشکده مدیریت و حسابداری، دانشگاه حضرت معصومه (س)، قم، ایران.
۲- استادیار، گروه مدیریت، دانشکده مدیریت و حسابداری، دانشگاه حضرت معصومه (س)، قم، ایران.

دریافت: ۱٤۰۳/۱۱/۲٤ بازنگری: ۱٤۰٤/۰۵/۲۱
پذیرش: ۱٤۰٤/۰٦/۱٦ انتشار: ۱٤۰٤/۰۷/۰۹

چکیده

فناوری هوش مصنوعی با ایجاد تحول بنیادین در مدیریت منابع انسانی، به ابزاری راهبردی در تصمیم‌گیری‌های کلیدی از جمله استخدام و ارتقا تبدیل شده است. بااین‌حال، گزارش‌های بسیاری از تخلفات و آسیب‌های ناشی از تصمیم‌گیری خودکار حکایت دارند. براین‌اساس، هدف اصلی این پژوهش، طراحی چارچوب مفهومی سنجش تحقق هوش مصنوعی مسئولیت‌پذیر در حوزه مدیریت منابع انسانی است. در راستای تحقق این هدف، از رویکردی کیفی بهره گرفته شده است. در گام نخست، با مرور نظام‌مند ادبیات نظری و تجربی مرتبط، چارچوب مفهومی اولیه استخراج شد. سپس، به‌منظور تعمیق و غنای آن، مصاحبه‌های نیمه‌ساختار یافته‌ای با ۱۰ تن از خبرگان حوزه‌های منابع انسانی و هوش مصنوعی انجام شد. داده‌های گردآوری‌شده با روش تحلیل محتوای کیفی بررسی و تفسیر شد تا ابعاد مفهومی هوش مصنوعی مسئولیت‌پذیر به شکلی دقیق و مبتنی بر شواهد استخراج شود. یافته‌ها نشان داد که ابعاد هوش مصنوعی مسئولیت‌پذیر در شش حوزه اصلی شامل اخلاق با ۷ شاخص، قابلیت اطمینان و اعتمادپذیری با ۵ شاخص، پاسخ‌گویی و مسئولیت‌پذیری با ٦ شاخص، شفافیت با ۷ شاخص، عدالت و عدم تبعیض با ۷ شاخص و همچنین امنیت و حریم خصوصی با ۸ شاخص تعریف می‌شوند که با مجموعه‌ای از ٤۰ شاخص مشخص شده‌اند. این چارچوب می‌تواند مدیران منابع انسانی را در طراحی و پیاده‌سازی سیستم‌های هوش مصنوعی کمک کند تا نه‌تنها این

* نویسنده مسئول مقاله: E-mail: s.emami@hmu.ac.ir

فناوری را به شکلی مسئولانه و منطبق بر ارزش‌های انسانی و سازمانی به‌کار ببرند، بلکه با درک و به‌کارگیری هم‌زمان شش بعد کلیدی مسئولیت‌پذیری از پیامدهای ناخواسته جلوگیری کنند.

واژه‌های کلیدی: مدیریت منابع انسانی، هوش مصنوعی، هوش مصنوعی مسئولیت‌پذیر.

۱- مقدمه

با توجه به قابلیت‌های پیشرفته سیستم‌های نوین در انجام وظایف با دقت، سرعت و اطمینان فراتر از توان انسان و ماشین‌های سنتی، بهره‌گیری از این فناوری‌ها منجر به کاهش چشمگیر هزینه‌ها، افزایش بهره‌وری و ارتقای ایمنی در محیط‌های کاری شده است. این پیشرفت‌ها همچنین امکان آزادسازی نیروی انسانی از فعالیت‌های تکراری و کم‌ارزش را فراهم کرده‌اند [۱]. در این میان، هوش مصنوعی به‌عنوان یکی از مؤلفه‌های بنیادین تحول دیجیتال، در ساختارهای سازمانی جایگاهی راهبردی یافته است؛ به‌گونه‌ای که بسیاری از سازمان‌ها با هدف بهره‌برداری از ظرفیت‌های گسترده آن، فرایند توسعه و استقرار این فناوری را با شتاب بیشتری در اولویت‌های خود قرار داده‌اند [۲].

در این میان، هوش مصنوعی نه‌تنها به بهبود تصمیم‌گیری‌های داده‌محور در سازمان‌ها کمک می‌کند، بلکه با افزایش کارایی نیروی کار و کاهش هزینه‌های نظارت، موجب ارتقای بهره‌وری کلی سازمان‌ها می‌شود [۳؛ ٤]. به‌کارگیری هوش مصنوعی در مدیریت منابع انسانی نیز توانسته است مزایای فراوانی به‌همراه آورد؛ چرا که این فناوری، از یک‌سو ابزارهای پیشرفته‌ای برای هدایت تصمیم‌های استراتژیک منابع انسانی فراهم می‌کند [۵؛ ٦] و از سوی دیگر از طریق بازتعریف اهداف مدیریت منابع انسانی و تغییر ساختار شبکه‌های درون‌سازمانی، فرایندهای نظارت، انگیزش و روابط کارکنان را دگرگون می‌سازد [۷-۹]. هوش مصنوعی با ظرفیت بالای پردازش داده‌ها و پشتیبانی تصمیم‌گیری، قابلیت ایجاد تحولات بنیادین در شیوه‌های متداول مدیریت منابع انسانی را داراست [۱۰].

بااین‌حال، رشد سریع و گسترده سیستم‌های پیچیده هوش مصنوعی، نگرانی‌هایی را دربارهٔ پیامدهای منفی این فناوری بر مشاغل، کسب‌وکارها و جامعه به وجود آورده است [٦؛ ۱۱]. شواهد نشان می‌دهد که بسیاری از سازمان‌ها در تحقق انتظار خود از فناوری‌های دیجیتال،

به‌ویژه هوش مصنوعی، با چالش‌های چشمگیری مواجه‌اند [۵]. با توجه به کاربرد روزافزون هوش مصنوعی و اجتناب‌ناپذیر بودن پیامدهای ناخواسته آن [۱۲]، توجه به مسئولیت‌پذیری این فناوری به‌منظور حفظ ارزش‌های انسانی و حمایت از توسعه پایدار ضرورت پیدا کرده است [۱۳-۱۵].

گسترش هوش مصنوعی به‌ویژه در حوزه مدیریت منابع انسانی، تأثیر عمیقی بر بازطراحی فرایندهای این حوزه داشته است [۳]. حتی در برخی موارد، تصمیم‌گیری‌های کلیدی مانند استخدام و اخراج به سیستم‌های هوش مصنوعی واگذار شده‌اند [۱٦]. اما گزارش‌های بسیاری نشان‌دهنده وجود سوگیری و تبعیض در عملکرد الگوریتم‌های استخدامی است، برای نمونه برنامه کاربردی هوش مصنوعی گوگل نسبت به زنان و متقاضیان آسیایی تبعیض قائل شده است [۱۷] و سامانه استخدامی آمازون نیز دچار سوگیری جنسیتی بوده است [۱۸]. این شواهد در کنار ابهام‌های قابل توجه پیرامون پیامدهای بالقوه هوش مصنوعی در حوزه مدیریت منابع انسانی، بر ضرورت تدوین و پیاده‌سازی چارچوب‌هایی منسجم و مسئولیت‌پذیر برای این فناوری تأکید می‌کند [۱۰؛ ۱۹].

اگرچه استفاده از هوش مصنوعی در مدیریت منابع انسانی رو به افزایش است و پژوهش‌هایی در زمینه هوش مصنوعی مسئولیت‌پذیر شکل گرفته‌اند، بخش عمده این پژوهش‌ها به‌صورت پراکنده و اغلب متمرکز بر مباحث کلی فناوری یا اخلاق بوده‌اند درحالی‌که افزایش نگرانی‌های اجتماعی و مطالبه عمومی برای مواجهه مسئولانه با پیامدهای منفی هوش مصنوعی، توجه به اصول اخلاقی و توسعه پایدار را بیش از پیش ضروری ساخته است [۱٤؛ ۱۵]. به‌رغم رشد پژوهش‌های مرتبط با هوش مصنوعی و کاربرد آن در منابع انسانی، مطالعات موجود اغلب به‌صورت پراکنده و غیرنظام‌مند به برخی مؤلفه‌های اخلاقی یا فنی مسئولیت‌پذیری پرداخته‌اند [۵؛ ٦]، بدون آنکه این مؤلفه‌ها در قالب یک چارچوب نظری منسجم و بومی‌سازی‌شده برای بستر منابع انسانی تجمیع شوند. بااین‌وجود، تاکنون چارچوبی جامع و هدفمند که ابعاد و مؤلفه‌های مسئولیت‌پذیری هوش مصنوعی را به‌طور خاص در زمینه مدیریت منابع انسانی شناسایی و تبیین کند، ارائه نشده است؛ خلأ مفهومی که ضرورت طراحی چارچوبی نظری و عملیاتی در این حوزه را برجسته می‌سازد. پژوهش حاضر با هدف پرکردن این شکاف، می‌کوشد تا با تلفیق پیشینه نظری و داده‌های کیفی حاصل از مصاحبه با

خبرگان، چارچوبی بومی‌شده برای سنجش مسئولیت‌پذیری هوش مصنوعی در مدیریت منابع انسانی ارائه دهد.

در مواجه با چالش‌های ناشی از به‌کارگیری هوش مصنوعی در مدیریت منابع انسانی، پژوهش حاضر با هدف طراحی یک چارچوب مفهومی برای تحقق هوش مصنوعی مسئولیت‌پذیر در این حوزه انجام شده است. در همین راستا، پرسش اصلی پژوهش به‌این‌صورت مطرح می‌شود: ابعاد و مؤلفه‌های هوش مصنوعی مسئولیت‌پذیر در حوزه مدیریت منابع انسانی کدام‌اند؟

۲- مبانی نظری و پیشینه پژوهش

۲-۱- هوش مصنوعی مسئولیت‌پذیر

در سال‌های اخیر، هوش مصنوعی به یکی از فناوری‌های کلیدی در صنایع مختلف تبدیل شده است و سازمان‌ها برای حفظ مزیت رقابتی به آن روی آورده‌اند [۲۰]. این فناوری با بهینه‌سازی عملکرد در حوزه‌هایی مانند منابع انسانی، موجب افزایش بهره‌وری و تولید شده است [۷؛ ۸]. همچنین، تحلیل دقیق داده‌ها به تصمیم‌گیری هوشمندانه‌تر، بهینه‌سازی نیروی کار و کاهش هزینه‌های نظارتی کمک می‌کند [٤؛ ۲۱] و خودکارسازی فرایندهای سازمانی نیز رشد و توسعه کسب‌وکارها را آسان کرده است [۲۲].

گسترش هوش مصنوعی مدیریت منابع انسانی را وارد عصری نوین از تحول و نوآوری کرده است. به‌طوری‌که ابزارهای پیشرفته هوش مصنوعی، فرایندهایی مانند استخدام، آموزش و تصمیم‌گیری را متحول کرده است [٦]. برخلاف فناوری‌های پیشین، هوش مصنوعی نه‌تنها استدلال انسانی را شبیه‌سازی می‌کند، بلکه می‌تواند در برخی موارد جایگزین آن شود. اما، این پیشرفت با چالش‌هایی همراه است، زیرا نتایج غیرمنتظره و غیرقابل توضیح برخی سیستم‌ها، درک فرایند تصمیم‌گیری را دشوار می‌کند [۲۳-۲۵].

پژوهش‌ها نشان می‌دهند که استفاده از هوش مصنوعی در محیط‌های کاری نه‌تنها باعث تحول در فرایندهای سازمانی می‌شود، بلکه پیامدهای اجتماعی و سیاسی عمیقی نیز به‌دنبال دارد [۹؛ ۲٦- ۲۷]. این پیامدها، ضرورت توجه به خطرهای بالقوه ناشی از توسعه این فناوری

را برجسته می‌سازند و بر اهمیت تدوین چارچوب‌های مدیریتی و اصول راهبردی برای هدایت و نظارت مسئولانه بر کاربرد آن تأکید دارند [۲۸؛ ۲۹].

در همین راستا، گسترش هوش مصنوعی باعث برجسته‌تر شدن برخی پیامدهای منفی آن شده است، برای مثال مواردی چون برچسب‌گذاری نادرست افراد [۳۰]، عملکرد مغرضانه الگوریتم‌های گوگل در تقویت تبعیض‌های جنسیتی و نژادی [۱۷] و نیز سوگیری سیستم‌های استخدامی نظیر آمازون علیه زنان [۳۱]، همگی نشان می‌دهند که دقت در طراحی الگوریتم‌ها و پردازش داده‌ها امری حیاتی است. به‌همین‌دلیل، پژوهشگران بر بازبینی دقیق فرایندهای داده‌کاوی و توسعه سامانه‌های هوش مصنوعی تأکید کرده و از مدیران می‌خواهند در بهره‌برداری از این فناوری، با احتیاط و مسئولیت‌پذیری بیشتری عمل کنند [۳۲؛ ۳۳].

برخی از این خطرها به‌طور مستقیم به ویژگی‌های منحصربه‌فرد هوش مصنوعی بازمی‌گردند، برای مثال سیستم‌های نظارتی که با هدف کاهش جرم و افزایش امنیت طراحی شده‌اند، گاهی به نقض حریم خصوصی و محدودیت آزادی‌های فردی منجر می‌شوند [٤]. علاوه‌بر این، وابستگی شدید هوش مصنوعی به داده‌ها، نگرانی‌های جدی درباره نشت اطلاعات شخصی و سوءاستفاده از تحلیل‌های کاربران را به‌همراه داشته است [۱۰]. این نوع پیامدها با عنوان «هوش مصنوعی اشتباه شده»، زمینه‌ساز تدوین اصولی برای توسعه مسئولیت‌پذیر این فناوری شده است [۳٤].

اگرچه هوش مصنوعی با وعده‌های چشمگیر در حوزه منابع انسانی وارد شده است، اما واقعیت‌های اجرایی آن همچنان با چالش‌های مهمی روبه‌روست. بررسی ادبیات نشان می‌دهد که شکاف معناداری میان ظرفیت‌های تبلیغ‌شده این فناوری و پیامدهای واقعی آن در عمل وجود دارد. به‌طور خاص، چهار چالش اساسی مانع بهره‌برداری مسئولانه و اثربخش از هوش مصنوعی در مدیریت منابع انسانی هستند: نخست، پیچیدگی و ظرافت پدیده‌های انسانی در محیط کار که به‌سختی در قالب الگوهای عددی گنجانده می‌شوند؛ دوم، محدودیت‌های ناشی از کمبود داده‌های بزرگ، دقیق و باکیفیت؛ سوم، دشواری در تعیین مسئولیت و پاسخ‌گویی دربرابر پیامدهای تصمیم‌های الگوریتمی، به‌ویژه در ابعاد اخلاقی و حقوقی و چهارم، واکنش‌های منفی کارکنان نسبت به تصمیم‌هایی که از شفافیت و انسان‌محوری کافی برخوردار نیستند [۱۰].

این ملاحظه‌ها به‌روشنی نشان می‌دهند که بهره‌گیری مؤثر از هوش مصنوعی در منابع انسانی تنها زمانی ممکن است که اصول بنیادینی مانند عدالت، شفافیت و مسئولیت‌پذیری اخلاقی در طراحی، استقرار و استفاده از این فناوری‌ها لحاظ شود. براین‌اساس، پژوهش حاضر با هدف پرکردن خلأ نظری موجود، به ارائه چارچوبی برای سنجش هوش مصنوعی مسئولیت‌پذیر در مدیریت منابع انسانی می‌پردازد تا گامی در جهت کاربرد اخلاق‌محور و پاسخ‌گو از این فناوری در سازمان‌ها برداشته شود.

۲-۲- پیشینه پژوهش

پیشینه پژوهش در حوزه کاربرد هوش مصنوعی در مدیریت منابع انسانی بیشتر بر مزایای عملکردی آن، ازجمله بهبود کارایی، دقت در تصمیم‌گیری و افزایش شفافیت تمرکز داشته است [۳۵]. در سال‌های اخیر، جریان نوظهوری از پژوهش‌ها با رویکردی انتقادی، به بررسی پیامدهای منفی، مخاطره‌های اخلاقی و پیچیدگی‌های اجتماعی ناشی از استقرار این فناوری پرداخته‌اند [٦].

پژوهش‌هایی چون بنکینز[1] (۲۰۲۱) و دلکراز[2] و همکاران (۲۰۲۲) تلاش کرده‌اند با رویکرد «طراحی منصفانه»، نقش انسان و ماشین را در فرایندهای استخدام متعادل کرده و با ارائه الگوریتم‌های حفاظتی، از سوگیری‌های تصمیم‌گیری بکاهند [۳٦؛ ۳۷]. ادامه این خط پژوهش در مطالعه آزمایشی بانکینز و همکاران (۲۰۲۲) نشان داد که ادراک کارکنان از عدالت و اعتماد تحت تأثیر نوع تصمیم‌گیرنده (انسان یا هوش مصنوعی) و ماهیت تصمیم (مثبت یا منفی) قرار دارد. یافته‌ها نشان می‌دهد که حتی تصمیم‌های مثبت اگر از سوی هوش مصنوعی اتخاذ شوند، ممکن است در کارکنان واکنش‌های منفی ایجاد کنند، مگر آنکه تجربه، شفافیت و اعتماد تقویت شده باشد [۳۸].

دراین‌راستا، مجموعه‌ای از پژوهش‌ها به طراحی چارچوب‌هایی برای استقرار هوش مصنوعی مسئولیت‌پذیر پرداخته‌اند. چانگ[3] و همکاران (۲۰۲٤) با تمرکز بر مسئولیت‌پذیری

1. Bankins
2. Delecraz
3. Chang

اجتماعی و آگاستانو[1] و همکاران (۲۰۲۳) با رویکرد تلفیقی اخلاق کاربردی و تناسب وظیفه-فناوری، بر ضرورت مشارکت فعال کارکنان و نقش انسان در حلقه تصمیم‌گیری تأکید کرده‌اند [۳۶؛۳۷؛۳۹؛۴۰]. این پژوهش‌ها، ضمن عبور از نگاه فناورمحور، به بُعد انسانی و ادراکی توجه دارند و مشارکت کاربر را شرط موفقیت سامانه‌های هوش مصنوعی می‌دانند.

چن[2] (۲۰۲٤) نیز با بهره‌گیری از رویکردهای میان‌رشته‌ای روان‌شناختی، اقتصادی و سیستمی، به بررسی هوش مصنوعی در آموزش سازمانی پرداخته و ضرورت طراحی مسئولانه سامانه‌های آموزشی را مطرح کرده است؛ سامانه‌هایی که نه‌تنها کارامد، بلکه عادلانه، شفاف و پاسخ‌گو باشند [٤١].

از سوی دیگر، ماتیاس[3] (۲۰۰٤) و پژوهش‌های جدیدتر در زمینه «هوش مصنوعی توضیح‌پذیر[4] [٤٢] بر نقش انسان در پاسخ‌گویی اخلاقی و نیاز به ارائه دلایل قابل فهم برای تصمیم‌های هوش مصنوعی تأکید کرده‌اند. این پژوهش‌ها، شفافیت و تعامل انسان– فناوری را شرط کلیدی برای اعتمادپذیری و پذیرش اجتماعی این سامانه‌ها می‌دانند [۱۱].

پژوهش‌هایی با رویکرد پارادوکسی، مانند کورلت[5] و همکاران (۲۰۲۱) به ماهیت دوگانه استفاده از هوش مصنوعی اشاره دارند؛ جایی‌که هم‌زمان فرصت‌هایی نظیر بهره‌وری و شخصی‌سازی و تهدیدهایی چون کاهش اعتماد و افزایش ابهام پدیدار می‌شوند. این پژوهش‌ها، ضمن تأکید بر مدیریت فعالانه پیامدهای متضاد، بر نقش حیاتی ارتباطات شفاف و پیاده‌سازی تدریجی تأکید می‌ورزند [۱۹؛ ۲۲].

مرور انتقادی ادبیات نشان می‌دهد که اگرچه چارچوب‌های اخلاق‌محور و اصول حقوق بشری برای هدایت توسعه هوش مصنوعی در حال گسترش‌اند، هنوز پژوهش‌های محدودی به بررسی عمیق، مقایسه‌ای و زمینه‌مند این اصول پرداخته‌اند [۳۰]. ازاین‌رو، پژوهش‌های آینده می‌توانند با تمرکز بر ارزیابی تجربی اصول اخلاقی، تحلیل شکاف‌های میان نظر و عمل و مطالعه واکنش‌های کاربرمحور، به درک عمیق‌تر از استقرار مسئولانه هوش مصنوعی در منابع انسانی یاری رسانند [٦].

1. Agustono
2. Chen
3. Matthias
4. Explainable AI
5. Charlwood

با توجه به شتاب فزاینده به‌کارگیری هوش مصنوعی در مدیریت منابع انسانی، چالش‌های آن بیش از پیش مورد توجه قرار گرفته‌اند. بااین‌حال، بررسی‌ها نشان می‌دهد که ابعاد اخلاقی و مسئولیت‌پذیری این فناوری بیشتر به‌صورت پراکنده و غیرنظام‌مند مطالعه شده‌اند. ازهمین‌رو، پژوهش حاضر با هدف پاسخ به این خلأ مفهومی و کاربردی، به‌دنبال ارائه چارچوبی جامع برای سنجش مسئولیت‌پذیری هوش مصنوعی در حوزه منابع انسانی است.

۳- روش پژوهش

پژوهش حاضر در گروه پژوهش‌های کیفی قرار دارد. به‌منظور بررسی نظام‌مند داده‌های کیفی و آشکار کردن الگوهای معنایی پنهان در آنها از روش تحلیل محتوای کیفی استفاده شده است، زیرا پژوهشگران، تحلیل محتوای کیفی را به منزلهٔ روشی انعطاف‌پذیر به‌ویژه برای داده‌های متنی در نظر می‌گیرند [۵۳]. تحلیل محتوای کیفی را می‌توان پژوهشی برای تفسیر ذهنی محتوای داده‌های متنی از راه فرایند نظام‌مند، کدبندی و تم‌سازی یا طراحی الگوهای شناخته شده دانست [٤۳]. تحلیل محتوای کیفی به پژوهشگران اجازه می‌دهد اصالت و حقیقت داده‌ها را به‌گونه‌ای ذهنی ولی به روش علمی تفسیر کنند. تحلیل محتوای کیفی به فراتر از کلمه‌ها یا محتوای عینی متون می‌رود و تم‌ها یا الگوهایی را که آشکار یا پنهان هستند، به‌صورت محتوای آشکار نمایان می‌سازد [٤٤].

داده‌ها با استفاده از کدگذاری باز تحلیل شد و سپس کدهای مشابه در طبقه‌های مفهومی گروه‌بندی شدند. در گام نهایی، مقولات اصلی برای طراحی چارچوب مفهومی مسئولیت‌پذیری هوش مصنوعی در مدیریت منابع انسانی استخراج شدند [٤۵؛ ٤٦]. براساس دستورالعمل تحلیل محتوای کیفی، نخست ادبیات پیشین و سپس متون مصاحبه‌ها با دقت مطالعه شد تا درک کاملی از عقاید و تجربه‌های مشارکت‌کنندگان حاصل شود. در مرحله بعد، جمله‌ها و عبارت‌های اصلی استخراج و معانی هر جمله دسته‌بندی شدند. این مفاهیم در خوشه‌های مختلف طبقه‌بندی و در قالب کدهای تفسیری قرار گرفتند که حاوی معانی مشخصی هستند.

مشارکت‌کنندگان این پژوهش از میان کارشناسان و خبرگان حوزه‌های فناوری اطلاعات، هوش مصنوعی و مدیریت منابع انسانی انتخاب شدند که حداقل سه سال سابقه فعالیت حرفه‌ای در یکی از این حوزه‌ها را دارا بودند. بااین‌حال فقط سابقه کاری ملاک انتخاب نبوده و معیارهای تکمیلی برای تشخیص خبرگی در نظر گرفته شد. این معیارها شامل:

- ✓ مؤثر در طرح‌های هوش مصنوعی، به‌ویژه در فرایندهای منابع انسانی (مانند استخدام، ارزیابی عملکرد، تحلیل داده‌های کارکنان)،
- ✓ داشتن دانش و تجربه در موضوعات مرتبط با مسئولیت‌پذیری، اخلاق فناوری و تصمیم‌گیری الگوریتمی،
- ✓ و برخورداری از سابقه پژوهشی یا مشاوره‌ای در حوزه کاربرد فناوری‌های هوشمند در سازمان‌ها می‌باشد.

به‌منظور اطمینان از تسلط و صلاحیت علمی مشارکت‌کنندگان در ارتباط با موضوع پژوهش، پیش از آغاز مصاحبه‌های اصلی، با هریک از افراد نشست‌هایی مقدماتی برگزار شد. در این پیش‌نشست‌ها، ضمن تشریح هدف و چارچوب کلی پژوهش، میزان آگاهی، تجربه عملی و قابلیت مشارکت مؤثر افراد ارزیابی شد. تنها افرادی که در این ارزیابی اولیه توان علمی و شناخت کافی از موضوع را نشان دادند، به مرحله اصلی مصاحبه راه یافتند. در ادامه، از مشارکت‌کنندگان خواسته شد در صورت شناخت افراد واجد شرایط دیگر، آنان را معرفی کنند. براین‌اساس، فرایند نمونه‌گیری به‌صورت قضاوتی از نوع گلوله برفی ادامه پیدا کرد [٤٧].

پس از انجام ۱۰ مصاحبه، تحلیل داده‌ها نشان داد که مفاهیم و کدهای جدیدی از داده‌ها استخراج نمی‌شود و پژوهش به نقطه اشباع مفهومی[1] رسیده است؛ مفهومی که به‌معنای کفایت اطلاعات برای پاسخ به پرسش‌های پژوهش تلقی می‌شود [٤٨]. بنابراین، این حجم نمونه برای دستیابی به اهداف پژوهش کفایت داشته است. درنهایت، ده نفر از خبرگان و کارشناسانی که پیشینه مرتبط و تمایل به همکاری داشتند، برای انجام مصاحبه‌های نیمه‌ساختار یافته انتخاب

1. data saturation

شدند. در جدول ۱، مشخصات کلی مصاحبه‌شوندگان به‌گونه‌ای آورده شده است که هویت آنها محفوظ بماند.

جدول ۱. مشخصات مصاحبه‌شوندگان

ردیف	تحصیلات	رشته/ تخصص	سابقه کاری	جنسیت	کد مصاحبه
۱	ارشد	هوش مصنوعی و فناوری اطلاعات	۳	زن	P1
۲	ارشد	سیستم‌های هوشمند و فناوری اطلاعات	٤	مرد	P2
۳	ارشد	توسعه سازمانی و برنامه‌ریزی نیروی کار	٤	زن	P3
٤	دکتری	مهندس داده و طراح الگوریتم‌های هوش مصنوعی	٥	مرد	P4
٥	دکتری	مدیریت منابع انسانی	۷	مرد	P5
٦	ارشد	مهندس داده و طراح الگوریتم‌های هوش مصنوعی	٥	مرد	P6
۷	ارشد	الگوسازی، الگوریتم‌های هوش مصنوعی	٦	مرد	P7
۸	ارشد	مدیریت‌فناوری اطلاعات با تجربه منابع انسانی	٥	مرد	P8
۹	دکتری	متخصص در پیاده‌سازی هوش مصنوعی	٤	زن	P9
۱۰	ارشد	مدیریت منابع انسانی	۳	زن	P10

۳-۱- روایی و پایایی پژوهش کیفی

برای اطمینان از اعتبار (روایی) و قابلیت اعتماد (پایایی) داده‌های حاصل از مصاحبه‌ها، از چارچوب چهار معیاری لینکولن و گوبا[1] (۱۹۸۵) استفاده شد [٤۹]. این چارچوب شامل اعتبار، تأییدپذیری، قابلیت اعتماد، و انتقال‌پذیری است.

در راستای افزایش اعتبار، از راهبرد بازبینی به‌وسیله اعضا[2] استفاده شد؛ به‌این‌صورت که در طول مصاحبه‌ها، برداشت‌های اولیه پژوهشگر به مشارکت‌کنندگان بازگشت داده شد تا آنها صحت درک مطالب خود را تأیید یا اصلاح کنند. همچنین، صرف زمان کافی برای هر مصاحبه و ایجاد فضای تعاملی و عمیق، به افزایش درک زمینه‌ای و صحت داده‌ها کمک کرد [۵۰].

1. Lincoln & Guba
2. member checking

برای تأییدپذیری، فرایند گردآوری و تحلیل داده‌ها با مشارکت چند متخصص حوزه روش‌شناسی بررسی شد. همچنین برای کاهش سوگیری پژوهشگر، بخشی از متن مصاحبه‌ها و کدگذاری‌ها به‌وسیله یک تحلیلگر دوم مرور شد.

در راستای افزایش قابلیت اعتماد تمامی مراحل جمع‌آوری و تحلیل داده‌ها به‌صورت نظام‌مند مستندسازی شد؛ ازجمله یادداشت‌های میدانی، ثبت تصمیم‌های تحلیلی و نگارش یادداشت‌های حاشیه‌ای در زمان کدگذاری داده‌ها. این مستندسازی امکان بازبینی و ممیزی فرایند پژوهش را فراهم می‌سازد.

برای ارتقای انتقال‌پذیری، سعی شد با ارائه توضیحات کامل درباره بستر پژوهش، مشخصات مشارکت‌کنندگان، نحوه گردآوری داده‌ها و زمینه تحلیل، زمینه قضاوت خوانندگان درباره امکان تعمیم یا کاربرد یافته‌ها در سایر محیط‌های مشابه فراهم شود.

٤- یافته‌ها

برای تحلیل داده‌ها در این پژوهش، از روش تحلیل محتوای کیفی با جهت‌گیری قیاسی–اکتشافی استفاده شده است. نخست مجموعه‌ای از پژوهش‌های علمی معتبر در حوزه هوش مصنوعی مسئولیت‌پذیر و کاربرد آن در مدیریت منابع انسانی مرور شد و براساس مفاهیم کلیدی موجود در ادبیات، چارچوبی اولیه متشکل از مؤلفه‌هایی مانند شفافیت، عدالت، مسئولیت‌پذیری، قابلیت اعتماد و امنیت طراحی شد. این مرحله نشان‌دهنده جهت‌گیری قیاسی تحلیل بوده است. در ادامه، با انجام مصاحبه‌های نیمه‌ساختاریافته با خبرگان و تحلیل محتوای آنها از راه کدگذاری باز، مفاهیم جدیدی از دل داده‌ها استخراج شد که بخشی از آنها در ادبیات پیشین به‌طور مستقیم مطرح نشده بودند. سپس مفاهیم جدید با مفاهیم قیاسی تلفیق شده و مؤلفه‌های اولیه توسعه پیدا کردند. این فرایند تا زمان رسیدن به اشباع نظری ادامه یافت، یعنی زمانی که دیگر مفهومی تازه برای افزودن به مؤلفه‌ها شناسایی نشد.

در گام نهایی، مفاهیم مشابه در قالب بحث اصلی دسته‌بندی شدند و ابعاد کلیدی مسئولیت‌پذیری هوش مصنوعی در مدیریت منابع انسانی شکل گرفتند. این فرایند براساس

چارچوب پیشنهادی الو و کینگاس[1] (۲۰۰۸) و گرانهایم و لوندمن[2] (۲۰۰۴) هدایت شده و مبنایی برای طراحی چارچوب مفهومی پژوهش فراهم آورده است.

به‌این‌ترتیب، ابعاد نهایی جدول ۳ حاصل یک رویکرد ترکیبی از تحلیل قیاسی (برگرفته از ادبیات نظری) و تحلیل استقرایی (مبتنی بر داده‌های تجربی) هستند. این تلفیق، شیوه‌ای رایج و معتبر در پژوهش‌های کیفی پیچیده و میان‌رشته‌ای محسوب می‌شود [۵۱؛ ۵۲].

پیش از ارائه یافته‌های پژوهش نمونه‌ای از مراحل تحلیل کیفی در قالب جدول ۲ ارائه می‌شود.

جدول ۲. نمونه‌ای از مراحل کدگذاری و تحلیل داده‌ها

مرحله تحلیل	کدگذاری باز	کدگذاری محوری	کدگذاری انتخابی
شرح فعالیت‌ها	شناسایی، استخراج واژگان و عبارت‌های مرتبط با سوگیری، بی‌عدالتی و تبعیض	دسته‌بندی کدهای باز و ایجاد ارتباط میان آنها	یکپارچه‌سازی و شکل‌دهی به بعد نهایی براساس اهمیت و ارتباط مفاهیم
نمونه کدهای استخراج شده	سوگیری داده‌ها، بی‌عدالتی در جذب، تبعیض در الگوریتم؛	مراقبت از سوگیری داده‌ها، حذف بی‌عدالتی، کاهش تبعیض	ادغام محور مدیریت عدالت در داده‌ها با سایر مفاهیم اخلاقی
دسته‌بندی نهایی	–	محور مفهومی: «مدیریت عادلانه داده‌های مرتبط با جذب استعداد»	بعد نهایی: «بعد اخلاقی» در هوش مصنوعی مسئولیت‌پذیر
شرح فعالیت‌ها	شناسایی عبارات مرتبط با پایداری و صحت عملکرد الگوریتم در مواجهه با داده‌های جدید	گروه‌بندی کدهای باز مرتبط با ثبات عملکرد الگوریتم و سازگاری با داده‌های جدید	ترکیب محورهای مرتبط و شکل‌دهی بعد قابلیت اطمینان و اعتمادپذیری کلی الگوریتم
نمونه کدهای استخراج شده	پایداری عملکرد، ورود داده‌های جدید، قابلیت اطمینان الگوریتم	ثبات عملکرد با داده‌های به‌روزشده، اعتماد به الگوریتم جذب استعدادها	ادغام محور اطمینان از عملکرد پایدار الگوریتم در چارچوب کلی اعتمادپذیری هوش مصنوعی
دسته‌بندی نهایی	–	محور مفهومی: «پایداری و صحت عملکرد الگوریتم جذب استعدادها با داده‌های جدید»	بعد نهایی: «قابلیت اطمینان و اعتمادپذیری» در هوش مصنوعی مسئولیت‌پذیر

1. Elo & Kyngäs
2. Graneheim & Lundman

براساس نتایج به‌دست آمده، ابعاد مسئولیت‌پذیری هوش مصنوعی در مدیریت منابع انسانی در شش محور اصلی طبقه‌بندی می‌شوند. نخست، بُعد اخلاقی با هفت گزاره به اصول بنیادین رفتار مسئولانه در استفاده از فناوری می‌پردازد. دوم، بُعد قابلیت اطمینان و اعتمادپذیری که با پنج گزاره، به سنجش میزان اعتماد سازمان به نتایج حاصل از سیستم‌های هوش مصنوعی اختصاص دارد. سوم، بُعد پاسخ‌گویی و مسئولیت‌پذیری است که با شش گزاره، نقش سازمان در پاسخ به پیامدهای تصمیم‌گیری مبتنی بر هوش مصنوعی را بررسی می‌کند. چهارمین بُعد، شفافیت و عدم تبعیض است که هفت گزاره را دربرمی‌گیرد و به‌وضوح به عملکرد الگوریتم‌ها و حذف سوگیری‌های احتمالی می‌پردازد. پنجمین بُعد، عدالت و انصاف با هفت گزاره که تأکید آن بر برخورد یکسان و منصفانه با کارکنان در فرایندهای مبتنی بر هوش مصنوعی است. درنهایت، بُعد امنیت و حفظ حریم خصوصی با هشت گزاره، به حفظ داده‌های شخصی و محرمانه کارکنان اشاره دارد.

درمجموع، این شش بُعد شامل چهل گزاره هستند که چارچوبی جامع برای سنجش مسئولیت‌پذیری استفاده از هوش مصنوعی در حوزه منابع انسانی فراهم می‌کنند. جزئیات کامل این ابعاد به شرح جدول ۳ است.

جدول ۳. ابعاد سنجش مسئولیت‌پذیری هوش مصنوعی در مدیریت منابع انسانی

ابعاد	شاخص‌ها	برخی منابع مرتبط
بعد اخلاقی	شناسایی ارزش های اخلاقی تمامی صاحبان منافع (کارفرما، کارمند و متقاضیان استخدام) در توسعه الگوریتم‌های مدیریت منابع انسانی	[۵۳-۵٦]
	مشارکت صاحبان منافع در فرایند طراحی سیستم ارزیابی عملکرد با علایق و ارزش‌های متفاوت	[۵۷؛۵٤]
	همسویی داده‌های استفاده شده در توسعه الگوریتم ارزیابی عملکرد، جذب کارمند، جبران خدمات با ارزش‌های اخلاقی و هنجارهای اجتماعی بستر مورد استفاده	[۵۷؛۵۵؛۳۷]
	استفاده از استدلال‌های ارزشی و اخلاقی در رفتار سیستم‌های هوشمند منابع انسانی	[۵۸] مصاحبه
	کنترل نبود سوگیری و بی‌عدالتی در داده‌های استفاده شده در الگوریتم جذب استعدادها	[۵٦ ؛۵٤]
	شناسایی اصول اخلاقی با توجه به استقلال سیستم در پیاده‌سازی کارکردهای منابع انسانی (توصیفی، پیش‌بینی و تجویزی)	[۵۷]
	درک ارزش‌های اخلاقی و هنجارهای محیط اجتماعی-فنی محیط بزرگتر سیستم هوشمند مدیریت منابع انسانی	[۵۸؛۵۷]

ابعاد	شاخص‌ها	برخی منابع مرتبط
قابلیت اطمینان و اعتمادپذیری	استفاده از رویکرد "آزمون-آزمون مجدد" جهت درستی پیش‌بینی الگوریتم ارزیابی عملکرد کارکنان	[۵۹ ؛۳۷]
	کیفیت و یکپارچگی داده‌های مورد استفاده در توسعه الگوریتم‌های تطابق شغل و غربالگری کاندیدا	[۱۰ ؛۳۷ ؛۵۳]
	استفاده از دیتاست‌های به‌روز برای توسعه الگوریتم تطابق شغل و شاغل	[۳۷؛ ۵۳؛ ۵۹]
	عمکرد قابل اطمینان الگوریتم‌های جذب استعدادها با وجود داده‌های جدید	مصاحبه [۶۰]
	پیش‌بینی قابل اطمینان الگورتیم‌های ارزیابی عملکرد و جبران خدمات برای داده‌های متناقض	[۱۰ ؛۵۳]
پاسخگویی و مسئولیت‌پذیری	مشخص بودن اهداف و مقاصد توسعه الگوریتم‌ها	مصاحبه
	هوش مصنوعی قابل توضیح/ توصیف‌پذیر (سیستم بتواند توضیح دهد چرا یک تصمیم را گرفته؟)	[۵۳؛ ۵۴؛۵۵؛ ۵۷؛ ۵۸؛۶۰؛ ۶۱]
	مشخص‌بودن زنجیره افراد مسئول در توسعه الگوریتم‌های منابع انسانی هوشمند	[۵۴-۵۶؛ ۶۰؛ ۶۲]
	توجیه‌پذیربودن تصمیم‌ها و انتخاب‌های سیستم منابع انسانی هوشمند	[۵۳-۵۵؛ ۵۷؛ ۵۸]
	وجودسازوکارهای حکمرانی داده در تحقق الگوهای سیستم	مصاحبه
	نمایش ارزش‌های اخلاقی و هنجارهای اجتماعی که سیستم در فرایند تصمیم‌گیری استفاده کرده است.	مصاحبه
شفافیت	شفافیت در منابع داده‌های توسعه الگوریتم های غربالگری و ارزیابی عملکرد	[۵۳؛ ۶۲]
	واضح بودن چگونگی گردآوری داده‌های کارکنان و ذخیره‌سازی آنها	[۵۷؛ ۶۲]
	دردسترس بودن داده‌های توسعه و آزمون الگوریتم‌های هوش مصنوعی منابع انسانی برای پژوهش‌های بیشتر	[۵۷؛ ۶۰]
	آشکار بودن فرضیه‌ها و ارزش‌های فرایند طراحی الگوریتم‌های جبران خدمات، ارزیابی عملکرد و غربالگری متقاضیان شغل	[۳۷؛ ۵۴؛ ۵۵]
	قابل فهم بودن الگو تصمیم‌گیری الگوریتم‌های جبران خدمات، ارزیابی عملکرد و غربالگری متقاضیان شغل برای کاربران	[۵۴؛ ۵۸]
	وجود سازوکارهای شفاف در مورد یادگیری سیستم	[۵۷؛ ۵۸]
	شفافیت در الگوریتم‌ها و عوامل تأثیرگذار بر تصمیم‌های الگوریتم‌ها	[۵۵؛۵۷؛۵۸؛ ۶۰]
عدم تبعیض و عدالت	مستندسازی گروه‌ها، نمونه‌ها و افرادی با ویژگی‌های خاص به‌منظور شناسایی اشتباهات سوگیرانه در الگوریتم جذب کارکنان	[۵۳؛ ۵۶]
	فراهم‌کردن دسترسی عادلانه به داده‌ها برای افراد با چت‌بات‌ها	مصاحبه
	ارزیابی/ پیامد مشابه برای افرادی با عملکردی مشابه	مصاحبه
	بی‌طرفی الگوریتمیک و عدم تبعیض‌های جنسیتی، سن، گرایش جنسی، علاقه‌مندی‌ها در غربالگری رزمه‌ها	[۳۷؛۵۳ ؛ ۵۴؛ ۵۶؛ ۵۸؛۶۰؛۶۲]
	آزمون عدالت الگوریتم ارزیابی عملکرد در ارزیابی افراد و رتبه‌بندی گروه‌ها در مقایسه با داده‌های واقعی	[۳۷؛ ۵۴]

ابعاد	شاخص‌ها	برخی منابع مرتبط
	فقدان سوگیری در تشخیص چهره متقاضیان استخدام برمبنای فرضیه‌های نادرست چون رنگ پوست، پوشش و غیره	مصاحبه [٥٦؛ ٦٠]
	استفاده از دیتاست‌های جامع‌تر و معرف برای جلوگیری از کلیشه‌های رایج در استخدام افراد	[١٠؛ ٣٧؛ ٥٦]
امنیت و حریم خصوصی	تبعیت از چارچوب‌های حریم خصوصی و قانونی استفاده از داده‌ها در فرایند بررسی درخواست متقاضیان شغل	[٥٣ ؛ ٥٦؛ ٥٧؛٦٢؛ ٦٣]
	استفاده از روش‌های امنیتی چندگانه برای محافظت از داده‌های کارکنان و مصاحبه‌شوندگان	[٥٣؛ ٦٣]
	محافظت از داده‌های کارکنان زمان تعامل با سیستم	[٥٣ ؛ ٥٦؛ ٥٨؛٦٤]
	اخذ توافق از متقاضیان شغل در استفاده از داده‌های آنها	[٥٣؛ ٦٤]
	آگاهی کارکنان از چگونگی گردآوری، ذخیره‌سازی و استفاده از داده‌های آنها	مصاحبه
	تعریف مالیکت داده‌های منابع انسانی سازمان	مصاحبه
	استفاده از کوکی‌های مورد توافق در گردآوری داده‌های عملکردی کارکنان	[٥٣] مصاحبه
	گمنامی داده‌های ارایانامه‌ها و مستندها در توسعه الگوریتم‌های ارزیابی عملکرد کارکنان	مصاحبه

٤-٢- تشریح یافته‌های پژوهش

٤-٢-١- بعد اخلاقی

یکی از چالش‌های بنیادین در بهره‌گیری از هوش مصنوعی در مدیریت منابع انسانی، به‌ویژه در فرایندهای جذب و استخدام که در مصاحبه‌ها اغلب بر آن تأکید داشتند، مسئله کنترل سوگیری‌ها و حفظ عدالت در الگوریتم‌های تصمیم‌گیری است. الگوریتم‌هایی که جدا از ملاحظه‌های دقیق اخلاقی و تحلیل داده‌های زمینه‌ای توسعه پیدا کرده‌اند، می‌توانند به‌طور ناخواسته الگوهای تبعیض‌آمیز گذشته را بازتولید کنند؛ موضوعی که پیامدهای آن در تجربه زیسته بسیاری از متخصصان منابع انسانی در مصاحبه‌ها نیز انعکاس پیدا کرده است. این مسئله به‌طور دقیق همان چیزی است که برایسون[1] (٢٠٢٣) نیز هشدار می‌دهد که اخلاق هوش مصنوعی نباید تنها به دغدغه‌هایی مثل تعصب داده تقلیل پیدا کند، بلکه باید به شکل بنیادین به آثار اجتماعی، خطرهای نهادینه‌سازی خشونت و تقویت بی‌عدالتی سیستمی نیز توجه شود

1. Bryson

[٦٥]. برای مثال یکی از مدیران منابع انسانی شرکت فناورمحور در این رابطه اینگونه اظهار کرد: «اگر الگوریتم‌ها برمبنای داده‌های تاریخی طراحی شوند، احتمال دارد سوگیری‌های ساختاری گذشته، مثل جنسیت‌گرایی یا تعصب نسبت به دانشگاه‌های خاص، در نتایج بازتاب پیدا کنند».

دراین‌راستا، به باور راغوان[1] و همکاران (۲۰۲۰)، در بسیاری از سامانه‌های استخدامی، داده‌های گذشته می‌توانند به‌عنوان عنصر تبعیض‌آمیز عمل کنند؛ مگر آنکه پیش از تحلیل، فرایندهای پالایش و بازنگری داده‌ها به‌صورت شفاف و هدفمند انجام شود [٦١؛ ٦٦].

در این میان، توانمندی توسعه‌دهندگان در شناسایی و حذف این سوگیری‌ها و طراحی سیستم‌هایی که اصول اخلاقی را در سه سطح توصیفی (تحلیل داده‌های گذشته)، پیش‌بینی (برآورد رفتارهای آینده) و تجویزی (ارائه تصمیم‌های خودکار یا پیشنهادی) رعایت می‌کنند، نقش کلیدی دارد، زیرا رعایت این اصول می‌تواند حقوق کارکنان را در برابر تصمیم‌های الگوریتمی محافظت کند و اعتماد سازمانی را تقویت کند. از منظر نظری نیز، تحلیل اجتماعی-فنی این سیستم‌ها نشان می‌دهد که درک و تبیین صحیح از هنجارهای اجتماعی مرتبط با عدالت، برابری و پاسخ‌گویی می‌تواند توازن میان پیشرفت فناورانه و مسئولیت‌پذیری اجتماعی را برقرار سازد [٦٧].

٤-۲-۲- قابلیت اطمینان و اعتمادپذیری

اعتمادپذیری یکی از مؤلفه‌های کلیدی در پذیرش و اثربخشی سیستم‌های هوش مصنوعی در مدیریت منابع انسانی است. این اعتماد نه از راه عملکرد ظاهری، بلکه از راه ارزیابی دقیق، پیاپی و اعتبارسنجی خروجی‌ها و پیش‌بینی‌های الگوریتمی سیستم حاصل می‌شود. براین‌اساس می‌توان اینگونه استدلال کرد که اعتمادپذیری را باید فراتر از فقط کارکرد فنی دانست و باید در چارچوبی از تلفیق کیفیت داده، انصاف الگوریتمی و ملاحظه‌های انسانی تحلیل کرد. در مصاحبه‌های انجام‌شده، برخی از مشارکت‌کنندگان بر نقش‌آفرینی کیفیت داده‌ها در فرایندهای جذب تأکید داشتند، برای مثال یکی از مدیران منابع انسانی اظهار داشت: «اگر داده‌هایی که به

1. Raghavan

سیستم می‌دهیم ناقص یا سوگیری داشته باشند، خروجی هم به همان میزان ناکارآمد و به‌خصوص غیرقابل اعتماد خواهد بود».

به‌این‌ترتیب، ارزیابی دوره‌ای و به‌روزرسانی مداوم دیتاست‌ها، همراه با تحلیل بی‌طرفانه داده‌های جدید، ضرورتی اجتناب‌ناپذیر برای جلوگیری از بازتولید سوگیری‌های ساختاری در تصمیم‌گیری است. این امر به‌ویژه در محیط‌های پویا و در حال تغییر اهمیت بیشتری پیدا می‌کند؛ جایی‌که الگوریتم‌ها باید بتوانند با تغییر نیازمندی‌های بازار کار، مهارت‌های جدید و الگوهای رفتاری نوظهور سازگار شوند. در همین رابطه، یکی از مصاحبه‌شوندگان تصریح کرد: «سیستم باید هم‌زمان با تغییرات بازار، خودش را به‌روزرسانی کند. سیستمی بنیادش بر داده‌های دیروز است، نمی‌تواند استعدادهایی برای فردا را بیابد». به همین دلیل، الگوهای پیش‌بینی لازم است از منظر اجتماعی بازنگری مداوم شوند تا اعتماد صاحبان منافع مخدوش نشود [۵۸؛ ٦٥].

٤-۲-۳- پاسخ‌گویی و مسئولیت‌پذیری

در بهره‌گیری از سیستم‌های هوشمند در منابع انسانی، یکی از دغدغه‌های اصلی مصاحبه‌شوندگان وجود نداشتن شفافیت پذیرش مسئولیت در زمان بروز خطاهای الگوریتمی است. از منظر مشارکت‌کنندگان، زمانی‌که سیستم به اشتباه فردی را رد یا تأیید می‌کند، این پرسش اساسی مطرح است که «چه کسی مسئول است و باید پاسخ‌گوی این خطا باشد؟» یکی از مدیران فناوری در این زمینه به نکته جالبی اشاره کرد: «وقتی الگوریتم اشتباه می‌کند، همه می‌گویند که سیستم تصمیم گرفته؛ ولی سیستم که پاسخ‌گو نیست، انسان‌ها باید پاسخ‌گو باشند». این نگاه، اهمیت تعیین زنجیره مسئولیت در طراحی و پیاده‌سازی الگوریتم‌ها را برجسته می‌سازد [۵۳؛ ٦٥]. براین‌اساس، مشخص بودن نهاد یا فرد پاسخ‌گو در فرایند توسعه، آموزش و اعمال الگوریتم‌ها، نقشی کلیدی در اعتماد سازمانی و مشروعیت اجتماعی سیستم‌های منابع انسانی مبتنی بر هوش مصنوعی دارد.

دراین‌راستا، امکان نداشتن توجیه‌پذیری تصمیم‌های سیستم، نقطه ضعف سیستم تلقی می‌شود. در فرایند انجام مصاحبه مشارکت‌کنندگان، به‌ویژه مدیران منابع انسانی خواهان آن بودند که معیارهای الگوریتمی به‌صورت مستند و با قابلیت توضیح دردسترس متقاضیان قرار

گیرد. یکی از آنها تصریح کرد: «اگر متقاضیی مورد تأیید قرار نمی‌گیرد، باید آگاه شویم چرا. این شفاف‌سازی هم به ما کمک می‌کند، هم به متقاضی».

٤-٤- شفاف بودن

شفاف بودن در منابع داده و نحوه پردازش آنها برای توسعه الگوریتم‌های غربالگری و ارزیابی عملکرد، از اهمیت بالایی برخوردار است. این موضوع از تصمیم‌گیری‌های ناعادلانه، نقض حریم خصوصی و استفاده نادرست از اطلاعات جلوگیری می‌کند. شفافیت نداشتن در تصمیم‌های هوش مصنوعی، معروف به «اثر جعبه سیاه» می‌تواند اعتماد کارکنان را کاهش دهد و اعتبار سازمان را زیر پرسش ببرد [٦٨-٧١]. اثر جعبه سیاه به‌عنوان خطر جدی برای مشروعیت تصمیم‌های سازمانی مطرح می‌شود. سیستم‌های فاقد تبیین‌پذیری نه‌تنها کاربران نهایی را در ابهام می‌گذارند، بلکه اعتماد عمومی به عملکرد عادلانه سازمان را تضعیف می‌کنند. به‌این‌ترتیب، می‌توان اینگونه استدلال کرد که تداوم این وضعیت ممکن است به نوعی «فاصله ادراکی» میان ماهیت تصمیم‌های الگوریتمی و نوع انتظارهای انسانی منجر شود؛ جایی‌که کاربر احساس می‌کند سیستمی طوری طرح‌ریزی شده که وی نادیده گرفته شود یا ناعادلانه قضاوت شود. درک این تصمیم‌ها در فرایندهایی مانند استخدام و ارتقا برای کارکنان ضروری است [٧٢]. بنابراین، سازمان‌ها باید الگوهای هوش مصنوعی را تفسیرپذیر کنند و مستندهای روشنی ارائه دهند. شفافیت در نحوهٔ یادگیری و به‌روزرسانی این سیستم‌ها، نظارت‌پذیری را افزایش و سوگیری‌های پنهان را کاهش می‌دهد [٧١].

٤-٥- عدم تبعیض و عدالت

بعد پنجم جدول یافته‌ها به عدم تبعیض و عدالت اشاره دارد. دراین‌خصوص، می‌توان اذعان کرد که مستندسازی دقیق ویژگی‌های جمعیتی و فردی در طراحی و اصلاح الگوریتم‌های جذب منابع انسانی نقش حیاتی ایفا می‌کند. شرکت‌کنندگان در مصاحبه‌ها بر این نکته تأکید داشتند که «اگر سیستم نداند با چه نوع و میزان تنوعی از داوطلبان مواجه است، چگونه می‌تواند بی‌طرف بماند؟». این مسئله نشان می‌دهد که شناسایی نداشتن صحیح زیرگروه‌ها در داده‌های وارد شده به سیستم، می‌تواند به بازتولید سوگیری‌های نهادینه شده منجر شود [٥٧؛

۵۹]. براین‌اساس، مستندسازی دقیق گروه‌ها و ویژگی‌های خاص به اصلاح سوگیری‌های الگوریتم‌های جذب منابع انسانی کمک می‌کند. دسترسی عادلانه به داده‌ها از راه چت‌بات‌ها فرصت‌های برابر را بیش از پیش تقویت می‌کند. برخی از مصاحبه‌شوندگان، این فناوری را به‌عنوان «پنجره‌ای بی‌طرف» توصیف کردند که به افراد با پیشینه‌های مختلف فرصت ابراز توانایی‌هایشان را می‌دهد؛ بدون آنکه تحت تأثیر سوگیری‌های انسانی یا فرم‌های سنتی قرار بگیرند [۵۳؛ ٦۰]. بنابراین حساسیت دست‌اندرکاران نسبت به بی‌طرفی الگوریتم‌ها در غربالگری رزومه و ارزیابی عملکرد ضروری تلقی می‌شود تا افراد با شرایط مشابه نتایج یکسانی دریافت کنند. همچنین، حذف سوگیری‌ها در تشخیص چهره و استفاده از دیتاست‌های متنوع، انتخاب‌های عادلانه‌تر را تضمین می‌کند.

٤-٦- امنیت و حریم خصوصی

بعد ششم یافته‌های پژوهش؛ حفظ امنیت و حریم خصوصی داده‌ها در فرایند بررسی درخواست‌های شغلی است که از محورهای اساسی در مسئولیت‌پذیری دیجیتال سازمان‌ها محسوب می‌شود. داده‌های شخصی داوطلبان، ازجمله سوابق شغلی، تحصیلات، ویژگی‌های روان‌شناختی و حتی رفتارهای دیجیتالی آنها، در صورت مراقبت نداشتن مناسب می‌توانند در معرض استفاده نادرست یا نشت اطلاعات قرار گیرند. این نگرانی در اغلب مصاحبه‌ها به‌وضوح از سوی مشارکت‌کنندگان مطرح شد. یکی از آنها اظهار داشت: «ما نمی‌توانیم به بهانه دیجیتال‌سازی، از مرزهای حریم خصوصی عبور کنیم؛ هر قدم ما باید با رضایت آگاهانه همراه باشد.» این موضوع به مفهوم «حاکمیت داده»[1] اشاره دارد. درواقع، به پیروی از چارچوب‌های قانونی تأکید دارد که ضرورتی اخلاقی برای حفظ کرامت انسانی در تعاملات دیجیتال به‌شمار می‌رود [٥٤؛ ۵۵]. مصاحبه‌شوندگان بارها به نگرانی خود نسبت به تعیین دقیق مالکیت داده‌ها اشاره کردند؛ اینکه مشخص باشد چه نهادی مسئول داده است، چگونه داده‌ها ذخیره می‌شوند و در چه شرایطی قابل حذف یا انتقال هستند و اینکه تا چه زمانی نگهداری می‌شوند.

1. Data Governance

۵- بحث و نتیجه‌گیری

طراحان بر جنبه‌های فنی تمرکز دارند درحالی‌که کاربران بیشتر نگران حریم خصوصی، امنیت و شفافیت تصمیم‌گیری خودکار هستند. این تفاوت‌ها چالش‌هایی در طراحی و استفاده از فناوری ایجاد می‌کند [۷۱].

ازآنجایی‌که سیستم‌های هوش مصنوعی در ذات خود از درک و تفسیر ارزش‌های انسانی در بافت فرهنگی ناتوان هستند [٥٤؛ ۷۳]، ارزش‌های اخلاقی به موضوعی کلیدی در مباحث دانشگاهی و کاربردی تبدیل شده است [۵۷؛ ٧٤]. اگرچه بیش از ۳۰۰ چارچوب سیاستی و دستورالعمل اخلاقی در سطح جهانی برای هم‌راستاسازی هوش مصنوعی با اصول انسانی تدوین شده‌اند [۱۳؛ ۵۳]، پژوهش‌ها نشان می‌دهند که این چارچوب‌ها هنوز از تبدیل شدن به معیارهای عملکردی در بستر سازمانی فاصله دارند [۵۸].

یافته‌های پژوهش حاضر نیز این فاصله را در بستر سازمانی ایران تأیید می‌کند، جایی‌که «اخلاق»، اغلب در سطح شعارهای رسمی یا منشورهای ارزش سازمانی باقی می‌ماند. یکی از مشارکت‌کنندگان در مصاحبه‌ها تأکید کرد: «برای ما اخلاق یک واژه تزئینی است؛ وقتی بربهره‌وری بالا به شدت تأکید شود، تصمیم با عددها گرفته می‌شود نه ارزش‌ها».

این در حالی است که پژوهش‌هایی نظیر [۵۵؛ ٥٦] هشدار می‌دهند که بدون تعهد اخلاقی واقعی، الگوریتم‌ها ممکن است تبعیض‌های پنهان را تقویت کنند. این موضوع به‌ویژه در فرایندهایی همچون غربالگری اولیه رزومه، ارزیابی خودکار عملکرد و تخصیص منابع انسانی بحرانی‌تر می‌شود [۳۷].

از سوی دیگر، ادبیات جدید مانند پژوهش‌های [۵۷؛ ٦۰] بر لزوم مشارکت فعال کاربران انسانی در تعریف اخلاقی بودن سیستم تأکید می‌کنند. این پژوهش نیز نشان داد که اخلاق در هوش مصنوعی زمانی معنا پیدا می‌کند که همه صاحبان منافع- از توسعه‌دهندگان گرفته تا تحلیلگران منابع انسانی و خود کارکنان در فرایند طراحی، آزمون و بازبینی الگوریتم‌ها مشارکت داده شوند. این مهم می‌تواند با تشکیل کمیته‌های اخلاقی میان‌بخشی و بررسی آثار اخلاقی آن بر حق وحقوق کارکنان و گروه‌های اقلیت در فرایند طراحی لحاظ کنند.

یافته‌های این پژوهش نشان می‌دهد که اعتمادپذیری و قابلیت اطمینان هوش مصنوعی در مدیریت منابع انسانی به‌ویژه در فرایندهای تطابق شغل و شاغل، نیازمند عملکرد پایدار، دقیق و قابل پیش‌بینی الگوریتم‌هاست [۳٦]. در گفتگو با یکی از مدیران منابع انسانی، وی تأکید داشت؛ «ما به سیستمی نیاز داریم که حتی نسبت به رزومه‌های ناقص یا حتی داده‌های متناقض، همچنان تصمیم‌هایی منطقی اتخاذ کند».

یافته‌های پژوهشی نیز بر این موضوع که الگوهای هوش مصنوعی باید به‌گونه‌ای طراحی شوند که نه‌تنها در شرایط ایدئال بلکه در مواجهه با داده‌های متنوع و ناسازگار نیز دقت خود را حفظ کنند؛ صحه می‌گذارند [۳۷؛ ۵۹]. از منظر توسعه سیستم‌ها، این موضوع به‌معنای ضرورت به‌روزرسانی مستمر الگوریتم‌ها، بازآموزی با داده‌های جدید و آزمون‌های اعتبارسنجی است [۱۰؛ ۵۳].

شرکت‌های تخصصی می‌توانند با بررسی این الگوریتم‌ها و صدور گواهی‌های تأیید مبنی بر قابلیت اعتماد به آنها، در اطمینان‌بخشی به این فناوری‌ها نقش‌آفرینی کنند. همچنین، مستندسازی پیامدهای تصمیم‌های الگوریتمی در سیستم‌های پرخطر- نظیر ارزیابی صلاحیت یا اخراج کارکنان- یکی از پیشنهادهایی است که از دل مصاحبه‌ها نیز برآمده است؛ به‌ویژه آنجاکه یکی از پاسخ‌دهندگان اشاره کرد؛ اگر تصمیم بر این است که سیستم تصمیم بگیرد، پس باید بتوانیم بفهمم بر چه اساس و معیاری این تصمیم گرفته شده است.

یافته‌های پژوهش نشان می‌دهد که یکی از چالش‌های اساسی در به‌کارگیری هوش مصنوعی در منابع انسانی، نبود نظام پاسخ‌گویی روشن و مؤثر در فرایند تصمیم‌گیری است. وقتی تصمیم‌های مهمی نظیر رد یا تأیید استخدام، ارتقا یا تعدیل نیرو به‌وسیله سیستم‌های الگوریتمی اتخاذ می‌شود، ولی مسئولیت مشخصی برای آن تعریف نشده، کارکنان دچار بی‌اعتمادی، احساس ناتوانی و حتی بیگانگی سازمانی می‌شوند [۷۵]. همان‌گونه که یکی از مشارکت‌کنندگان مصاحبه بیان کرد؛ وقتی مشخص نیست چه کسی پشت تصمیم است، اعتراض به نتیجه را کجا و به چه کسی باید برد؟!!

در این رابطه، پژوهشگران نیز بر اهمیت پاسخ‌گویی در تمامی فرایندهای سیستم هوش مصنوعی از طراحی و توسعه گرفته تا پیاده‌سازی و نظارت تأکید دارند [٥٤؛ ٧٦؛ ٧٧]. این درحالی است که بیانیه‌های مختلفی در این رابطه تدوین شده است؛ اما مشکل اصلی «شکاف

در اجرا» است؛ به‌عبارتی فقدان مکانیسم‌هایی در بطن سیستم برای تضمین رعایت این اصول [۵۷؛ ۵۸].

یافته‌های پژوهش حاضر، همچنین بیان می‌کند که بیشتر کاربران سیستم‌های منابع انسانی، اعم از مدیران، کارشناسان یا متقاضیان شغل، از منطق تصمیم‌گیری الگوریتم‌ها اطلاع دقیقی ندارند. همان‌طورکه یکی از مدیران در مصاحبه بر آن اشاره داشت؛ سیستم امتیاز می‌دهد، ولی معلوم نیست چطور و چرا. فقط می‌دانیم چه کسی رد شده، نه اینکه چرا. این ابهام، همان چیزی است که پژوهش‌ها با عنوان «اثر جعبه سیاه» شناخته و آن را خطر و تهدیدی برای پاسخ‌گویی و شفافیت می‌داند [۶۱؛ ۷۷]. به همین دلیل، مدافعان، سیاست‌گذاران و پژوهشگران حقوقی خواستار ماشین‌هایی هستند که بتوانند خود را توضیح دهند [۷۸].

در این میان، حکمرانی داده‌ها، ازجمله کنترل کیفیت داده‌ها، ثبت و مستندسازی تصمیم‌ها و تعریف مالکیت و مسئولیت‌های حقوقی، به‌عنوان راهکارهای ضروری در پژوهش‌های نوین و یافته‌های میدانی شناخته شده‌اند [۵۵؛ ۵۶؛ ۶۲]. براین‌اساس، برای هر تصمیم شغلی اتخاذ شده به‌وسیله سیستم‌های هوش مصنوعی، مسئولیت انسانی برای آن تعریف شود. درضمن برای کاربران گزارش‌های شفاف و قابل فهم تهیه شود و فرایندهای شکایت یا بازبینی را امکان‌پذیر کند. چراکه پاسخ‌گویی تنها به‌معنای شفافیت فنی نیست، بلکه مستلزم تعیین بازیگران مسئول، معیارهای حساب‌کشی و سازوکارهای واکنش‌پذیر به خطاهای سیستم است [۷۹].

شفافیت به‌عنوان یکی از ارکان اصلی مسئولیت‌پذیری هوش مصنوعی در مدیریت منابع انسانی، نقش کلیدی در پذیرش نتایج و تعامل اثربخش بین کاربران و سیستم‌های الگوریتمی ایفا می‌کند. یافته‌های پژوهش نشان می‌دهد که برخی از سیستم‌های هوش مصنوعی برای افزایش شفافیت، اطلاعات را به‌صورت بیانیه‌های توضیح‌پذیر در اختیار کاربران، صاحبان منافع و عموم قرار می‌دهند [۸۰] تا بتوانند عملکرد این سیستم‌ها را درک کرده و نسبت به پذیرش یا رد نتایج، تصمیم‌گیری آگاهانه‌تری داشته باشند [۸۱؛ ۸۲] و در صورت لزوم، تصمیم‌های مضر را شناسایی و به حداقل برسانند [۸۳].

توضیح‌پذیری نتایج به این معناست که سیستم‌های هوش مصنوعی باید به‌گونه‌ای طراحی شوند که تصمیم‌ها و خروجی‌های آنها برای کاربران قابل فهم باشند؛ یعنی مشخص باشد بر چه اساسی، با چه داده‌هایی و به کمک چه الگوریتمی تصمیم گرفته شده است [۸٤].

به‌این‌ترتیب، یک فرایند شفاف در تصمیم‌گیری الگوریتمی به بهبود درک روند استخدام کمک می‌کند [۸۵]؛ اما در نبود آن، نتایج تبعیض‌آمیز و غیر قابل دفاع بروز می‌کند. نمونه شاخص آن، الگوریتم استخدام آمازون است که به‌دلیل استفاده از داده‌های تاریخی مردسالارانه، به حذف نظام‌مند متقاضیان زن منجر شد [۸٦]. این نمونه، مصداق بارز بازتولید تبعیض از راه سیستم‌های به‌ظاهر بی‌طرف است [۸۷].

مصاحبه‌ها با مشارکت‌کنندگان پژوهش نیز نشان‌دهنده این امر است؛ به‌طوری‌که یکی از متخصصان منابع انسانی اشاره می‌کند؛ ما نمی‌دانیم این سیستم به‌طور دقیق چطور انتخاب می‌کند. فقط یک فهرست از اسامی داریم. وقتی دلیل رد یا پذیرش را نمی‌دانیم، شفافیت زیر سؤال می‌رود و قادر به پاسخ‌گویی نیستیم.

پژوهش‌های پیشین نیز به‌طور مستقیم بر همین موضوع تأکید می‌کنند که فقدان شفافیت در سامانه‌های استخدامی می‌تواند موجب بی‌اعتمادی و احساس بی‌عدالتی در میان متقاضیان شود [۵۳؛ ۵۵؛ ۵۸؛ ٦۰]. دراین‌راستا، طراحی سیستم‌های هوش مصنوعی مبتنی بر الگوهایی چون «هوش مصنوعی توضیح‌پذیر امکان تحلیل، نقد و بازنگری انسانی را فراهم می‌کند [۵٤]. دستورالعمل‌های دیجیتالی دقیق، کامل، روشن و کاربرپسند به‌عنوان ابزار تسهیل‌کننده در تعامل کاربران با این نوع سیستم‌ها راه‌گشا خواهد بود.

یافته‌های پژوهش حاضر، هم‌راستا با ادبیات پیشین نشان می‌دهد که بهره‌گیری از هوش مصنوعی در فرایندهای استخدام ممکن است به بازتولید تبعیض‌های اجتماعی موجود منجر شود، به‌ویژه زمانی‌که ورودی‌های الگوریتم‌ها حاوی سوگیری‌های تاریخی نسبت به ویژگی‌های فردی مانند جنسیت، نژاد یا سن باشند [۵۳؛ ۵٦؛ ۸۸]. این موضوع نه‌تنها اصل عدالت را نقض می‌کند، بلکه می‌تواند پیامدهای نامطلوبی در محیط‌های کاری به‌همراه داشته باشد [۵٤؛ ۵۸].

همان‌طورکه یافته‌های پژوهش حاضر نیز نشان می‌دهد، باید اطلاعات دقیقی از تدوین‌کنندگان الگوریتم‌های هوش مصنوعی در مورد نحوه طراحی سیستم‌ها دریافت و

بررسی‌های لازم به‌دقت انجام شود [۸۹]. علاوه‌بر این، اطمینان حاصل شود که ابزار استخدام هوش مصنوعی با داده‌های صحیح و بی‌طرفانه تغذیه می‌شود، زیرا این موضوع برای اخذ تصمیمات منصفانه‌تر راه گشا است [۹۰]. تأکید مصاحبه‌شوندگان بر لزوم بهره‌گیری از گروه‌های چندرشته‌ای در توسعه این سیستم‌ها نشان می‌دهد که مداخله‌های انسانی در طراحی، پایش و بازبینی هوش مصنوعی نقش کلیدی در کاهش تبعیض دارد [٥٦؛ ٦۰].

در این راستا، استفاده از داده‌های یک‌دست یا جانب‌دارانه چالش‌برانگیز است. اگر این داده‌ها اصلاح نشوند، ممکن است سوگیری‌های موجود در آنها در تصمیم‌گیری‌های الگوریتمی نیز بازتولید شوند [۱۰؛ ۲۰؛ ۳۷]. بنابراین، به‌کارگیری داده‌های متنوع و استفاده از روش‌هایی چون «ممیزی الگوریتمی» برای شناسایی و کاهش تبعیض، توصیه می‌شود [۵۸؛ ۹۱].

از منظر نظری همان‌طورکه بینز[1] (۲۰۱۸) نیز نشان می‌دهد، آنچه محل مناقشه است، مفهوم «انصاف» در یادگیری ماشینی است: آیا باید بر تساوی فرصت‌ها تمرکز کرد یا بر به‌حداقل رساندن آسیب به گروه‌های آسیب‌پذیر؟ آیا می‌توان تعریفی از انصاف ارائه داد که هم با اصول فلسفه اخلاق سازگار باشد و هم در قالب‌های الگوریتمی پیاده‌سازی‌پذیر باشد [۹۲] ؟

بنابراین، طراحی نظام‌های هوش مصنوعی عادلانه باید مبتنی بر اصول برابری فرصت، انصاف در فرایند باشد [٥٤؛ ۹۲]. همچنین، به‌طور عملی، سازمان‌ها باید شیوه‌نامه‌هایی برای پایش مستمر، ارزیابی سوگیری، رسیدگی به شکایات کاربران و اصلاح نواقص طراحی کنند. طراحی «داشبوردهای شفافیت‌ساز» برای مدیران منابع انسانی می‌تواند به شناسایی تبعیض‌ها در زمان واقعی کمک کند. درمجموع، تحقق عدالت در استفاده از هوش مصنوعی تنها از راه رویکردی چندلایه و نظام‌مند امکان‌پذیر است؛ رویکردی که توسعه الگوریتم‌ها، مدیریت داده‌ها و به ویژه نظارت انسانی را در اولویت قرار می‌دهد.

حریم خصوصی و امنیت از دیگر دغدغه‌های اصلی در به‌کارگیری هوش مصنوعی در مدیریت منابع انسانی است. یافته‌های این پژوهش همسو با پژوهش‌های پیشین نشان می‌دهد که استفاده از سیستم‌های هوش مصنوعی بدون رعایت و احترام به استانداردهای حریم خصوصی می‌تواند به نقض حقوق کارکنان و متقاضیان شغلی منجر شود [٤۰؛ ٥۳؛ ٥۷].

1. Binns

از منظر نظری، مفهوم «حریم خصوصی اطلاعاتی» که بر اصل «محرمانگی» تأکید دارد؛ یعنی جلوگیری از دسترسی غیرمجاز به داده‌ها با چالش‌هایی گره خورده است؛ چالش‌هایی که بیشتر به شیوه‌های جمع‌آوری، ذخیره‌سازی، پردازش و به‌اشتراک‌گذاری اطلاعات شخصی بازمی‌گردند [۹۳]. براساس مصاحبه‌ها و یافته‌های پیشین، استفاده از سخت‌افزارهای ایمن، رمزگذاری پیشرفته، کنترل دسترسی دقیق و پردازش محلی داده‌ها می‌توانند نقش مؤثری در کاهش خطرپذیری نقض امنیت اطلاعات ایفا کنند [٦٣؛ ٦٤؛ ٦٧؛ ٩٤]. درواقع، محدود کردن دسترسی به داده‌ها تنها به افراد مجاز، طراحی سیستم‌های «ناشناسی پیش‌فرض»[1] و مدیریت چرخه عمر داده‌ها از جمع‌آوری تا حذف، ازجمله الزام‌های ضروری برای صیانت از حریم خصوصی هستند [۵۸]. همچنین یافته‌های میدانی نشان می‌دهد که ضعف مهارت‌های فنی در گروه‌های منابع انسانی، چالشی اساسی برای تضمین امنیت داده‌ها و رعایت حریم خصوصی محسوب می‌شود. مصاحبه‌شوندگان بر لزوم آموزش کارکنان، تدوین شیوه نامه‌های استاندارد و استفاده از داشبوردهای کنترلی برای نظارت انسانی بر تصمیم‌های هوش مصنوعی تأکید کرده‌اند [۵۳]. درهمین‌راستا، مشارکت با فروشندگان معتبر هوش مصنوعی و طراحی چارچوب‌های حاکمیت داده، همچون الگوهای مسئولانه ذخیره‌سازی و تبادل داده، از توصیه‌های کلیدی برای کاهش خطرپذیری‌های امنیتی به‌شمار می‌روند [٥٦؛ ٦٢؛ ۹۵]. از منظر عملی، سازمان‌ها باید به‌طور مداوم تهدیدهای امنیتی را ارزیابی، سیستم‌های خود را در برابر حملات سایبری مقاوم‌سازی و بازبینی‌های منظم را سرلوحه فعالیت‌های خود قرار دهند.

نتایج این پژوهش با ادغام شواهد نظری و تجربی، نشان داد که تحقق هوش مصنوعی مسئولیت‌پذیر در مدیریت منابع انسانی مستلزم برخورداری از چارچوبی منسجم شامل شش بعد کلیدی و ٤٠ مؤلفه است. این چارچوب، علاوه‌بر بر تقویت بنیان نظری می‌تواند نقشه راهی برای مدیران منابع انسانی، طراحان الگوریتم و نهادهای ناظر باشد. به‌ویژه در بسترهایی که پذیرش اجتماعی، شفافیت تصمیم‌ها و پاسخ‌گویی سازمانی در استفاده از هوش مصنوعی حیاتی است، این چارچوب می‌تواند نقشی محوری ایفا کند. پژوهش حاضر با یکپارچه‌سازی عناصر پراکنده در ادبیات موجود و افزودن شاخص‌هایی برگرفته از مصاحبه، گامی مهم در

1. privacy by design

توسعه دانش میان‌رشته‌ای در تقاطع اخلاق، فناوری و مدیریت منابع انسانی محسوب می‌شود. براین‌اساس، پیشنهاد می‌شود که پژوهش‌های آینده با آزمون تجربی این چارچوب در سازمان‌های مختلف و همچنین توسعه ابزارهای ارزیابی مبتنی بر آن، گام‌های بعدی را در راستای سنجش و پیاده‌سازی هوش مصنوعی مسئولیت‌پذیر بردارند.

٦- محدودیت‌های پژوهش

علیرغم تلاش برای طراحی چارچوبی جامع برای هوش مصنوعی مسئولیت‌پذیر در مدیریت منابع انسانی، این پژوهش نیز همچون سایر پژوهش‌های کیفی با محدودیت‌هایی همراه است که در تفسیر و تعمیم یافته‌ها بهتر است درنظر گرفته شود. ابتدا، ماهیت کیفی پژوهش و استفاده از تحلیل محتوای کیفی باعث می‌شود که نتایج در بسترهای خاص اجتماعی و سازمانی معنا پیدا کند و تعمیم مستقیم آنها به سایر محیط‌ها باید با احتیاط انجام شود. دوم، داده‌های مصاحبه با نمونه‌گیری قضاوتی و گلوله‌برفی از میان خبرگان محدود و خاصی گردآوری شده‌اند. بنابراین ممکن است دیدگاه‌های مطرح شده بازتاب نظر تمام صاحبان نظران حوزه هوش مصنوعی در مدیریت منابع انسانی نباشند. سوم، با وجود بررسی به‌نسبت جامع پیشینه نظری و تلاش برای ادغام یافته‌های متنوع، امکان نادیده ماندن برخی پژوهش‌های مهم یا دیدگاه‌های نوظهور همچنان وجود دارد.

۷- منابع

[1] Conradie N, Kempt H, Königs P. Introduction to the topical collection on AI and responsibility. Philos Technol. 2022;35(4):97.
[2] Helo P, Hao Y. Artificial intelligence in operations management and supply chain management: An exploratory case study. Prod Plan Control. 2022;33(16):1573-1590.
[3] Agrawal A, McHale J, Oettl A. Finding needles in haystacks: Artificial intelligence and recombinant growth. In: Agrawal A, Gans J, Goldfarb A, editors. The economics of artificial intelligence: An agenda. Chicago: University of Chicago Press; 2018. p. 149-174.

[4] Tong S, Jia N, Luo X, Fang Z. The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance. Strateg Manag J. 2021;42(9):1600-1631.
[5] Minbaeva D. Disrupted HR? Hum Resour Manag Rev. 2021;31(4):100820.
[6] Vrontis D, Christofi M, Pereira V, Tarba S, Makrides A, Trichina E. Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review. Artificial intelligence and international HRM. 2023;172-201.
[7] Beane M. Shadow learning: Building robotic surgical skill when approved means fail. Adm Sci Q. 2019;64(1):87-123.
[8] Sergeeva AV, Faraj S, Huysman M. Losing touch: An embodiment perspective on coordination in robotic surgery. Organ Sci. 2020;31(5):1248–1271.
[9] Xie X, Liu Y, Liu J, Zhang X, Zou J, Fontes-Garfias CR, et al. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. Nat Med. 2021;27(4):620-621.
[10] Tambe P, Cappelli P, Yakubovich V. Artificial intelligence in human resources management: Challenges and a path forward. Calif Manag Rev. 2019;61(4):15–42.
[11] Baum K, Mantel S, Schmidt, Speith T. From responsibility to reason-giving explainable artificial intelligence. Philos Technol. 2022;35(1):12.
[12] Fuchs DJ. The dangers of human-like bias in machine-learning algorithms. Missouri S&T's Peer to Peer. 2018;2(1). Available from: https://scholarsmine.mst
[13] Constantinescu M, Voinea C, Uszkai R, Vică C. Understanding responsibility in Responsible AI. Dianoetic virtues and the hard problem of context. Ethics Inf Technol. 2021;23:803-814.
[14] Cheng MM, Hackett RD. A critical review of algorithms in HRM: Definition, theory, and practice. Hum Resour Manag Rev. 2021;31(1).
[15] Köchling A, Wehner M. Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. Bus Res. 2020;13(3):795-848.
[16] Kim PT, Bodie MT. Artificial intelligence and the challenges of workplace discrimination and privacy. ABAJ Lab & Emp L. 2020;35:289.
[17] Weston M, Sun H, Herman GL, Benotman H, Alawini A. Echelon: An AI tool for clustering student-written SQL queries. In: 2021 IEEE Frontiers in Education Conference (FIE); 2021 Oct; IEEE. p. 1-8.

[18] Jeffrey D. Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. 2018. Available from: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G
[19] Charlwood A, Guenole N. Can HR adapt to the paradoxes of artificial intelligence? Hum Resour Manag J. 2022;32(4):729-742.
[20] Hedlund M, Persson E. Expert responsibility in AI development. AI & Soc. 2022;1-12.
[21] Wang H, Fu T, Du Y, Gao W, Huang K, Liu Z, et al. Scientific discovery in the age of artificial intelligence. Nature. 2023;620(7972):47-60.
[22] Mikalef P, Conboy K, Lundström JE, Popovič A. Thinking responsibly about responsible AI and 'the dark side' of AI. Eur J Inf Syst. 2022;31(3):257-268.
[23] Davenport A, Gefflot C, Beck C. Slack-based techniques for robust schedules. In: Sixth European conference on planning; 2014 May.
[24] Hakli R, Mäkelä P. Moral responsibility of robots and hybrid agents. The Monist. 2019;102(2):259-275.
[25] Loh F, Loh J. Autonomy and Responsibility in Hybrid Systems. In: Lin P, Abney K, Jenkins R, editors. Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence. Oxford: Oxford University Press; 2017. p. 35–50.
[26] Goyal A, Aneja R. Artificial intelligence and income inequality: Do technological changes and worker's position matter? Public Affairs. 2020;20(4):e2326.
[27] Servoz M. AI, the future of work? Work of the future! On how artificial intelligence, robotics and automation are transforming jobs and the economy in Europe. Publications Office; 2019.
[28] Barredo-Arrieta A, Del Ser J. Plausible counterfactuals: Auditing deep learning classifiers with realistic adversarial examples. In: 2020 International joint conference on neural networks (IJCNN); 2020 Jul. p. 1-7. IEEE.
[29] Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M. Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication. 2020;(2020-1).
[30] Wiggers G, Verberne S, van Loon WS, Zwenne G. Bibliometric-enhanced legal information retrieval: combining usage and citations as flavors of impact relevance. [Journal name missing]. 2023;(8):1010-1025.

[31] Jeffrey T. Understanding college student perceptions of artificial intelligence. Systemics, Cybernetics and Informatics. 2020;18(2):8-13.
[32] Osoba O, Welser IV W. An intelligence in our image. Santa Mônica: RAND Corporation; 2017.
[33] Rana NP, Chatterjee S, Dwivedi YK, Akter S. Understanding dark side of artificial intelligence (AI) integrated business analytics: assessing firm's operational inefficiency and competitiveness. Eur J Inf Syst. 2022;31(3):364-387.
[34] Desouza KC, Dawson GS, Chenok D. Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. Bus Horiz. 2020;63(2):205–213.
[35] Agustono DO, Nugroho R, Fianto AYA. Artificial Intelligence in Human Resource Management Practices. KnE Soc Sci. 2023;958-970.
[36] Bankins S. The ethical use of artificial intelligence in human resource management: a decision-making framework. Ethics Inf Technol. 2021;23(4):841-854.
[37] Delecraz S, Eltarr L, Becuwe M, Bouxin H, Boutin N, Oullier O. Responsible Artificial Intelligence in Human Resources Technology: An innovative inclusive and fair by design matching algorithm for job recruitment purposes. J Responsible Technol. 2022;11:100041.
[38] Bankins S, Formosa P, Griep Y, Richards D. AI decision making with dignity? Contrasting workers' justice perceptions of human and AI decision making in a human resource management context. Inf Syst Front. 2022;24(3):857-875.
[39] Sargiotis D. Harnessing Digital Twins in Construction: A Comprehensive Review of Current Practices, Benefits, and Future Prospects. 2024.
[40] Chang YL, Ke J. Socially responsible artificial intelligence empowered people analytics: a novel framework towards sustainability. Hum Resour Dev Rev. 2024;23(1):88-120.
[41] Chen Z. Responsible AI in Organizational Training: Applications, Implications, and Recommendations for Future Development. Hum Resour Dev Rev. 2024;23(4):498-521.
[42] Matthias A. The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata. Ethics Inf Technol. 2004;6(3):175–183.
[43] Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res. 2005;15(9):1277-1288.
[44] Vaismoradi M, Jones J, Turunen H, Snelgrove S. Theme development in qualitative content analysis and thematic analysis. 2016.

[45] Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs. 2008;62(1):107-115.
[46] Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today. 2004;24(2):105-112.
[47] Patton MQ. Qualitative research and evaluation methods. 3rd ed. Sage; 2002.
[48] Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. Field Methods. 2006;18(1):59-82.
[49] Patton C, Sawicki D, Clark J. Basic methods of policy analysis and planning—Pearson eText. Routledge; 2015.
[50] Norris LS, White DE, Moules NJ. Thematic analysis: Striving to meet the trustworthiness criteria. Int J Qual Methods. 2017;16(1):160940691773384.
[51] Maxwell JA. Qualitative research design: An interactive approach. 3rd ed. Sage; 2013.
[52] Creswell JW, Poth CN. Qualitative inquiry and research design: Choosing among five approaches. 4th ed. Sage; 2016.
[53] Duke T. Building responsible AI algorithms. Springer; 2023. Available from: https://link.springer.com/book/10.1007/978-1-4842-9306-5
[54] Dignum V. Responsible artificial intelligence: how to develop and use AI in a responsible way. Vol. 1. Cham: Springer; 2019.
[55] Martin K. Ethical implications and accountability of algorithms. J Bus Ethics. 2019;160(4):835-850.
[56] Yam J, Skorburg JA. From human resources to human rights: Impact assessments for hiring algorithms. Ethics Inf Technol. 2021;23(4):611-623.
[57] Bujold A, Roberge-Maltais I, Parent-Rocheleau X, Boasen J, Sénécal S, Léger PM. Responsible artificial intelligence in human resources management: a review of the empirical literature. AI Ethics. 2023;1-16.
[58] Rodgers W, Murray JM, Stefanidis A, Degbey WY, Tarba SY. An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. Hum Resour Manag Rev. 2023;33(1):100925.
[59] Koenig N, Tonidandel S, Thompson I, Albritton B, Koohifar F, Yankov G, et al. Improving measurement and prediction in personnel selection through the application of machine learning. Pers Psychol. 2023;76(4):1061-1123.

[60] Tippins NT, Oswald FL, McPhail SM. Scientific, legal, and ethical concerns about AI-based personnel selection tools: a call to action. Pers Assess Decis. 2021;7(2):1.
[61] Masood A. Responsible AI in the Enterprise: Practical AI Risk Management for Explainable, Auditable, and Safe Models with Hyperscalers and Azure OpenAI. 2023.
[62] Varma A, Dawkins C, Chaudhuri K. Artificial intelligence and people management: A critical assessment through the ethical lens. Hum Resour Manag Rev. 2023;33(1):100923.
[63] Manoharan P. A Review on Cybersecurity in HR Systems: Protecting Employee Data in the Age of AI. Regul. GDPR. 2024; 4:605-612.
[64] Rocha JF. Ethical implementation of AI in job candidate recruitment: insights on data protection, Artificial Intelligence and legal perspectives. Braz J Law Technol Innov. 2024;2(1):120-139.
[65] Bryson JJ. Human Experience and AI Regulation: What European Union Law Brings to Digital Technology Ethics. Weizenbaum J Digit Soc. 2023;3(3).
[66] Raghavan M, Barocas S, Kleinberg J, Levy K. Mitigating bias in algorithmic hiring: Evaluating claims and practices. In: Proceedings of the 2020 conference on fairness, accountability, and transparency; 2020. p. 469-481.
[67] Rakova B, Yang J, Cramer H, Chowdhury R. Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. Proc ACM Hum-Comput Interact. 2021; 5:1-23.
[68] Kim PT, Bodie MT. Artificial intelligence and the challenges of workplace discrimination and privacy. ABAJ Lab & Emp L. 2020; 35:289.
[69] Lacroux A, Martin-Lacroux C. Should I trust the artificial intelligence to recruit? Recruiters' perceptions and behavior when faced with algorithm-based recommendation systems during resume screening. Front Psychol. 2022; 13:895997.
[70] Islam M, Mamun AA, Afrin S, Ali Quaosar GA, Uddin MA. Technology adoption and human resource management practices: the use of artificial intelligence for recruitment in Bangladesh. South Asian J Hum Resour Manag. 2022;9(2):324-349.
[71] London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hastings Cent Rep. 2019;49(1):15-21.
[72] Golbin I, Rao AS, Hadjarian A, Krittman D. Responsible AI: a primer for the legal community. In: 2020 IEEE international conference on big data (Big Data); 2020. p. 2121-2126. IEEE.

[73] Sharkey A. Can robots be responsible moral agents? And why should we care? Connection Sci. 2017;29(3):210–216.
[74] Coeckelbergh M. Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence. AI Ethics. 2023;3(4):1341-1350.
[75] Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. 2017.
[76] Theodorou A, Dignum V. Towards ethical and socio-legal governance in AI. Nat Mach Intell. 2020;2(1):10–12.
[77] European Commission. General Data Protection Regulation (GDPR). Official Journal of the European Union. 2016;L119:1-88.
[78] Giermindl LM, Strich F, Christ O, Leicht-Deobald U, Redzepi A. The dark sides of people analytics: reviewing the perils for organisations and employees. Eur J Inf Syst. 2022;31(3):410-435.
[79] Selbst AD, Barocas S. The intuitive appeal of explainable machines. Fordham L Rev. 2018;87:1085.
[80] Wieringa M. What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In: Proceedings of the 2020 conference on fairness, accountability, and transparency; 2020. p. 1-18.
[81] Simmons & Simmons and Jacob Turner. HireVue AI Explainability Statement. HireVue; 2022. Available from: https://www.hirevue.com/ai-in-hiring
[82] Application of AI in Recruitment and Selection. Artificial Intelligence in Industry 4.0: The Future that Comes True. 2024;234.
[83] Du J. Ethical and Legal Challenges of AI in Human Resource Management. J Comput Electron Inf Manag. 2024;13(2):71-77.
[84] Konidena BK, Malaiyappan JNA, Tadimarri A. Ethical Considerations in the Development and Deployment of AI Systems. Eur J Technol. 2024;8(2):41–53.
[85] Gunning D, Aha D. DARPA's explainable artificial intelligence (XAI) program. AI Mag. 2019;40(2):44-58.
[86] Chiwara JR, Mjoli TQ, Chinyamurindi WT. Factors that influence the use of the Internet for job-seeking purposes amongst a sample of final-year students in the Eastern Cape province of South Africa. SA J Hum Resour Manag. 2017;15(1):1-9.
[87] Dastin J. Amazon scraps secret AI recruiting tool that showed bias against women. In: Ethics of data and analytics; 2022. p. 296-299.
[88] Fernández-Martínez C, Fernández A. AI and recruiting software: Ethical and legal implications. Paladyn J Behav Robot. 2020;11(1):199-216.

[89] Chen Z. Collaboration among recruiters and artificial intelligence: removing human prejudices in employment. Cogn Technol Work. 2023;25(1):135-149.
[90] Yadav S, Kapoor S. RETRACTED ARTICLE: Adopting artificial intelligence (AI) for employee recruitment: the influence of contextual factors. Int J Syst Assur Eng Manag. 2024;15(5):1828-1840.
[91] Yarger L, Cobb Payton F, Neupane B. Algorithmic equity in the hiring of underrepresented IT job candidates. Online Inf Rev. 2020;44(2):383-395.
[92] Binns R. Fairness in machine learning: Lessons from political philosophy. In: Conference on fairness, accountability and transparency; 2018 Jan. p. 149-159. PMLR.
[93] Solove DJ. Artificial intelligence and privacy. Fla L Rev. 2025;77:1.
[94] Brundage M, Avin S, Wang J, Belfield H, Krueger G, Hadfield G, et al. Toward trustworthy AI development: mechanisms for supporting verifiable claims. arXiv preprint arXiv:2004.07213. 2020.
[95] Khoa BQ. Influential factors of Artificial Intelligence (AI) in the digital transformation of the human resources recruitment process sector in Vietnam. Int J Multidiscip Res Growth Eval. 2024;5(6):1118-1193.