



## هوش مصنوعی: قابلیت‌های بالقوه و چالش‌های اخلاقی استفاده از آن

ماندانا لواف زاده

دانشجوی کارشناسی ارشد مدیریت فناوری اطلاعات- مدیریت منابع اطلاعاتی، واحد اهواز، دانشگاه آزاد اسلامی، اهواز ایران.

### چکیده:

توانایی سیستم‌های هوش مصنوعی در تبدیل مقادیر زیادی از اطلاعات پیچیده و مبهم موجود به منظور ایجاد بینش است. هرچند هوش مصنوعی این توانایی را دارد که اسرار دیرینه را فاش و به حل برخی از پدیدارترین مشکلات جهان کمک کند؛ با این حال، مانند تمام فناوری‌های قدرتمند، توسعه آن نیازمند دقت زیادی است. برای این منظور لازم است در کنار توجه به مزایای اجتماعی سیستم‌های هوش مصنوعی، چالش‌های مرتبط با اعتماد، اطمینان، پایبندی به اصول و ارزش‌های اخلاقی و هنجارهای انسانی نیز مورد توجه قرار گیرد. از این رو، مطالعه حاضر به روش کتابخانه‌ای با مرور یافته‌های تحقیقات مختلف، به دنبال بررسی ابعاد اخلاقی استفاده از هوش مصنوعی و چالش‌های مرتبط در این زمینه است.

### مقدمه و بیان مساله:

هوش مصنوعی به مانند چتری است که بسیاری از رشته‌ها مانند علوم کامپیوتر، تجارت، مهندسی، زیست‌شناسی، روان‌شناسی، ریاضیات، آمار، منطق، فلسفه و زبان‌شناسی را دربرمی‌گیرد (وانگ و سیائو<sup>۱</sup>، ۲۰۱۸). پیچیدگی و قابلیت هوش مصنوعی آن را منحصر به فرد و بحث برانگیز کرده است (سیائو، ۲۰۱۸). نگرانی محققان در خصوص هوش مصنوعی در خصوص هوش مصنوعی از این حیث است که در آینده منجر به یک ابرهوش شود (مولر و بوستروم ۲۰۱۶)؛ ابرهوشی که می‌تواند به طور ساده به عنوان «یک عقلی که تقریباً در همه حوزه‌های مورد علاقه بسیار فراتر از عملکرد شناختی انسان کار کند» (بوستروم، ۲۰۱۴). برنامه‌های کاربردی هوش مصنوعی برنامه‌های داده‌محور هستند. داده‌های مورد استفاده برای آموزش سیستم هوش مصنوعی و داده‌هایی که پردازش می‌کند، به نوع وظایفی که یک سیستم برای آن طراحی شده است، بستگی دارد (عباسی و تیموری، ۱۴۰۳). جمع‌آوری و پردازش داده‌ها برای توسعه، آموزش و استفاده از سیستم‌های هوش مصنوعی ممکن است مسائل اخلاقی و حقوقی مهمی مانند حریم خصوصی، حفاظت از داده‌ها، سوءگیری و روند قانونی ایجاد کند (گیسیلویچ پاکین و لو-آرتز<sup>۲</sup>، ۲۰۱۸). داده‌های مورد استفاده برای آموزش الگوریتم ممکن است منعکس کننده شیوه‌های تبعیض آمیز و تعصبات ریشه‌ای باشد (اینستنس و همکاران<sup>۳</sup>، ۲۰۱۶).

در مقایسه با بحث داغ و سرمایه گذاری در فناوری هوش مصنوعی، توجه به مباحث و چالش‌های اخلاقی آن در مرحله ابتدائی قرار دارد. هرچند برخی براین باورند که هنوز راه درازی برای مقایسه هوش مصنوعی با توانایی‌های هوشمندانه انسان وجود دارد و برای همین، عجله‌ای برای بررسی چالش‌های اخلاقی وجود ندارد، اما برخی از محققان بر این مساله تاکید دارند که قبل از برجسته شدن مسائل اخلاقی مرتبط با هوش مصنوعی، باید این مسائل زودتر مورد بررسی و مذاقه قرار گیرند (وانگ و سیائو، ۲۰۱۸).

<sup>1</sup> Wang & Siau

<sup>2</sup> Bostrom

<sup>3</sup> Geslevich Packin & Lev-Aretz

<sup>4</sup> Instance et al.



پیشرفت سریع فناوری‌های هوش مصنوعی، به ویژه در چشم‌انداز کامپیوتر، نه تنها بخش‌های مختلف را متحول کرده است، بلکه نگرانی‌های اخلاقی قابل توجهی را نیز برانگیخته است (لی، رويس و لو<sup>۵</sup>، ۲۰۲۲). همانطور که سیستم‌های هوش مصنوعی بیشتر در زندگی ما ادغام می‌شوند، سؤالاتی در مورد پیامدهای اخلاقی آنها نیز مطرح است. مسائلی مانند حریم خصوصی داده‌ها، سوءگیری الگوریتمی و قدرت متمرکز به موضوعات اصلی در بحث‌های اخلاق هوش مصنوعی تبدیل شده‌اند؛ که نگرانی‌ها از سوء استفاده از داده‌ها تا تهدیدات بالقوه برای اصول دموکراتیک را پوشش می‌دهد. در پاسخ، دینفغان، از جمله جامعه علمی، رهبران صنعت، و سیاست‌گذاران، برای هدایت این چالش‌های اخلاقی پیچیده همکاری می‌کنند. استراتژی‌ها و چارچوب‌ها برای اطمینان از همسویی اخلاقی در توسعه، استقرار و کاربرد هوش مصنوعی توسعه داده شده‌اند. رسانه‌ها و مردم که از پتانسیل تحول آفرین هوش مصنوعی آگاه هستند، خواستار شفافیت و فراگیری در این تلاش‌ها هستند (کامیلا و جاسروتیا<sup>۶</sup>، ۲۰۲۳). اخلاق مفهومی پیچیده، ناشناخته و درهم تنیده است. این واقعیت در خصوص هوش مصنوعی وقتی بیشتر جلوه می‌کند که بدانیم وجه عمومی آن هر روز در حال درنوردیدن حوزه‌های جدیدی از فعالیت‌های بشر است. از این رو، مطالعه جنبه‌های اخلاقی آن در همه زمینه‌ها ممکن نیست. با توجه به آنچه گفته شد، امروزه یکی از دغدغه‌های اساسی پژوهشگران این است که قابلیت‌های بالقوه و دغدغه‌های اخلاقی مرتبط با هوش مصنوعی کدامند و اقدامات احتمالی بایستی مبتنی بر چه الزاماتی باشد؟

### هوش مصنوعی و استیلای آن بر گستره‌ای از فعالیت‌ها بشری

به اعتقاد مؤسسه آینده زندگی (نقل از دی‌جی<sup>۷</sup>، ۲۰۲۰)، هوش مصنوعی "نویدبخش تغییرات موفق در حوزه‌های مختلف اقتصادی، اجتماعی، پزشکی، امنیتی و زیست محیطی بزرگ با مزایای بالقوه از جمله کمک به افراد برای کسب مهارت‌ها و آموزش‌های جدید؛ دموکراتیک کردن خدمات؛ طراحی و ارائه زمان‌های تولید سریعتر و چرخه‌های تکرار سریعتر؛ کاهش مصرف انرژی؛ ارائه نظارت بر محیط زیست در زمان واقعی برای آلودگی و کیفیت هوا؛ تقویت دفاع از امنیت سایبری؛ افزایش تولید ملی؛ کاهش ناکارآمدی مراقبت‌های بهداشتی؛ ایجاد انواع جدیدی از تجربیات و تعاملات لذت بخش برای افراد؛ و بهبود خدمات ترجمه بلادرنگ برای ارتباط مردم در سراسر جهان است. مردم برای قرن‌ها نگران جابجایی نقش کارگران توسط فناوری بوده‌اند. پیش‌بینی می‌شود که اتوماسیون و البته خودکارسازی، محاسبات و اخیراً هوش مصنوعی و رباتیک مشاغل را از بین ببرد و آسیب‌های جبران‌ناپذیری به بازار کار وارد کند. لئونتیف (۱۹۸۳)، با مشاهده پیشرفت‌های چشمگیر در قدرت پردازش تراشه‌های کامپیوتری، نگران بود که ماشین‌ها جایگزین مردم شوند، همانطور که اسب‌ها با اختراع موتورهای احتراق داخلی منسوخ شدند. با این حال، در گذشته، اتوماسیون اغلب در کوتاه مدت جایگزین نیروی انسانی شده است، اما منجر به ایجاد شغل در بلند مدت شده است (آتور<sup>۸</sup>، ۲۰۱۵). با این وجود، نگرانی گسترده‌ای وجود دارد که هوش مصنوعی و فناوری‌های مرتبط می‌توانند در طی دو دهه آینده بیکاری گسترده ایجاد کنند. نتایج مطالعه فری و آزبورن (۲۰۱۳) نشان داد که فناوری‌های اطلاعاتی جدید «سهم قابل توجهی از اشتغال را در طیف گسترده‌ای از مشاغل در آینده نزدیک در معرض خطر قرار خواهند داد». هوش مصنوعی در حال حاضر در امور مالی، اکتشاف فضا، تولید پیشرفته، حمل و نقل، توسعه انرژی و مراقبت‌های بهداشتی بطور گسترده مورد استفاده قرار گرفته است.

<sup>5</sup> Li, Ruijs & Lu

<sup>6</sup> Kamila & Jasrotia

<sup>7</sup> DG

<sup>8</sup> Autor



به لحاظ تأثیرات اقتصادی نیز اقتصاددانان عموماً مشتاق چشم‌اندازهای هوش مصنوعی بر رشد اقتصادی هستند. طبق یافته‌های گریتز و مایکلز (۲۰۱۵) رباتیک بین سال‌های ۱۹۹۳ تا ۲۰۰۷ حدود ۰.۴ درصد به رشد تولید ناخالص داخلی و بهره‌وری نیروی کار سالانه برای ۱۷ کشور اضافه کرد که به همان نسبت تأثیر معرفی موتورهای بخار بر رشد در بریتانیا است. نتایج مطالعه اسمیت و اندرسون (۲۰۱۴) با طرح سوال از ۱۸۹۶ متخصص در خصوص تأثیرات هوش مصنوعی بر محیط کار نشان داد که ۴۸ درصد بر این باور بوده‌اند روبات‌ها و عوامل دیجیتال تعداد قابل توجهی از کارگران یقه‌ی «آبی» و «سفید» را جابه‌جا می‌کند و بسیاری ابراز نگرانی کردند که این امر منجر به افزایش گسترده در نابرابری درآمد، بیکاری تعداد زیادی از افراد و فروپاشی نظم اجتماعی خواهد شد.

**جک استیلگو** بزرگترین سوال در مورد هوش مصنوعی را نابرابری می‌داند که به طور معمول در بحث در مورد اخلاق هوش مصنوعی گنجانده نمی‌شود. در حالی که یک موضوع کاملاً اخلاقی - سیاسی است؛ این که چه کسی از هوش مصنوعی سود می‌برد؟ انتظار می‌رود که هوش مصنوعی و فناوری رباتیک به شرکت‌ها اجازه دهد تا کسب‌وکار خود را ساده‌تر و کارآمدتر و نظام‌مند کنند. با این حال، برخی استدلال می‌کنند که این به هزینه نیروی انسانی آنها تمام می‌شود. این بدان معنی است که درآمدها بین افراد کمتری تقسیم می‌شود و نابرابری‌های اجتماعی را افزایش می‌دهد. در نتیجه، افرادی که مالکیت شرکت‌های مبتنی بر هوش مصنوعی را دارند، به طور نامتناسبی سود خواهند برد. **نمیتز** (۲۰۱۸) اعتقاد دارد که نادیده گرفتن این موضوع که هوش مصنوعی قدرت را در دستان چند غول اینترنتی دیجیتال متمرکز می‌کند، ساده لوحانه خواهد بود، زیرا «واقعیت نحوه استفاده [بیشتر جوامع] از اینترنت و آنچه اینترنت به آنها ارائه می‌دهد توسط چند شرکت بزرگ شکل گرفته است. به اعتقاد وی، توسعه هوش مصنوعی دقیقاً تحت تسلط این شرکت‌های بزرگ و اکوسیستم‌های وابسته به آنها است». انباشت قدرت فن آوری، اقتصادی و سیاسی در دست پنج بازیگر برتر - گوگل، فیس بوک، مایکروسافت، اپل و آمازون - به آنها نفوذ نامناسبی را در حوزه‌های اجتماعی مرتبط با تغییر افکار در حوزه‌های مختلف از جمله دولت‌ها، قانون‌گذاران، جامعه مدنی، احزاب سیاسی، مدارس و آموزش، روزنامه‌نگاری و آموزش روزنامه‌نگاری و - مهم‌تر از همه - علم و تحقیق دموکراسی‌ها می‌دهد. در همین حال، **برایسون** (۲۰۱۹)، معتقد است که این تمرکز قدرت می‌تواند نتیجه اجتناب ناپذیر کاهش هزینه‌های فناوری رباتیک باشد؛ جاییکه هزینه‌های بالا می‌تواند تنوع را در سیستم‌های اقتصادی حفظ کند.

یک نظرسنجی از مشتریان آی.پی.ای. نشان داد که بزرگترین نگرانی مردم در مورد حفظ حریم خصوصی، هک شدن دستگاه آنها (۶۳٪/۶۸)، پس از آن جمع آوری اطلاعات شخصی آنها (۱۶٪)، و گوش دادن به مکالمات آنها (۲۴٪/۷)، ضبط مکالمات خصوصی (۱۲٪) و رعایت نکردن حریم خصوصی (۶٪) است (**مانیکوندا و همکاران**<sup>۹</sup>، ۲۰۱۸). یکی دیگر از جنبه‌های هوش مصنوعی که بر حریم خصوصی تأثیر می‌گذارد، داده‌های بزرگ است. فناوری اکنون در مرحله‌ای است که می‌توان سوابق طولانی‌مدت را در مورد هر کسی که داده‌های قابل ذخیره‌سازی تولید می‌کند - هر کسی که صورت‌حساب، قرارداد، دستگاه‌های دیجیتال یا سابقه اعتباری دارد، بدون استفاده از نوشته‌های عمومی و استفاده از رسانه‌های اجتماعی، حفظ کرد. رکوردهای دیجیتال را می‌توان با استفاده از الگوریتم‌هایی برای تشخیص الگو جستجو کرد، به این معنی که ما فرض پیش فرض ناشناس بودن را با ابهام از دست داده ایم (**سلینگر و هارتزوگ**، ۲۰۱۷). هر یک از ما را می‌توان با نرم افزار تشخیص چهره یا داده کاوی عادات خرید یا رسانه‌های اجتماعی



خود شناسایی کرد (پاسکواله<sup>۱۰</sup>، ۲۰۱۵). این عادات آنلاین ممکن است نه تنها هویت ما، بلکه استعدادهای سیاسی یا اقتصادی ما را نشان دهد، و اینکه چه استراتژی هایی ممکن است برای تغییر آنها موثر باشد (کادوالادر<sup>۱۱</sup>، ۲۰۱۴).

هوش مصنوعی توسط انسان ایجاد شده است، به این معنی که می تواند در معرض سوگیری باشد. سوگیری نظام مند ممکن است در نتیجه داده های مورد استفاده برای آموزش سیستم ها، یا در نتیجه ارزش هایی که توسط توسعه دهندگان و کاربران سیستم نگهداری می شود، ایجاد شود. اغلب زمانی اتفاق می افتد که برنامه های یادگیری ماشین بر روی داده هایی آموزش داده می شوند که فقط گروه های جمعیتی خاصی را منعکس می کنند یا سوگیری های اجتماعی را منعکس می کنند. تمرکز قدرت تکنولوژیکی، اقتصادی و سیاسی در میان چند شرکت بزرگ می تواند به آنها اجازه نفوذ نابجا بر دولت ها را بدهد - اما پذیرش و اجرای هوش مصنوعی می تواند دموکراسی را از طرق دیگر نیز تهدید کند. با ظهور کلان داده ها، سیاستمداران به حجم عظیمی از اطلاعات دسترسی دارند که به آنها امکان می دهد دسته های خاصی از رای دهندگان را هدف قرار دهند و پیام هایی را توسعه دهند که بیشتر مورد توجه آنها باشد.

هوش مصنوعی در مدل سازی افکار، تجربه، عمل، گفتگو و روابط انسانی روز به روز بهتر می شود. در عصری که ما به طور مکرر با ماشین ها به گونه ای در تعامل هستیم که گویی انسان هستند، تأثیر آن بر روابط واقعی انسان ها خود محل چالش است. روابط با دیگران هسته وجودی انسان را تشکیل می دهد. انتظار می رود در آینده، روبات ها در نقش های اجتماعی مختلف مانند پرستاری، خانه داری، مراقبت از کودکان و سالمندان، آموزش و غیره به انسان ها خدمت کنند. این ربات ها ممکن است طوری طراحی شده باشند که دقیقاً شبیه انسان ها به نظر برسند و صحبت کنند. مردم ممکن است برای ایجاد وابستگی عاطفی به روبات ها، شاید حتی احساس عشق به آنها شروع کنند؛ در این صورت اگر این اتفاق بیفتد، چه تأثیری بر روابط انسانی و روان انسان قابل تصور است؟ چندین مسئله اخلاقی وجود دارد که توسط انسان ها در ایجاد روابط با روبات ها ایجاد می شود:

- ۱) آیا روبات ها می توانند باورها، نگرش ها و/یا ارزش های ما را در مورد روابط انسان با انسان را تغییر دهند؟ زمانی که افراد بتوانند با یک ربات «کامل» ارتباط داشته باشند و از این چالش ها اجتناب کنند، ممکن است بی صبر شده و تمایلی به تلاش برای کار روی روابط انسان و انسان نداشته باشند.
- ۲) آیا «روبات های صمیمی» می توانند منجر به افزایش رفتار خشونت آمیز شوند؟ اگر کاربر مکرراً به ربات همراه مشت بزند، آیا این غیراخلاقی است (لالجی، ۲۰۱۵)؟ آیا خشونت علیه روبات ها الگوی رفتاری را عادی می کند که در نهایت بر انسان های دیگر تأثیر می گذارد؟

شخصیت بخشی و تعریف هویت برای هوش مصنوعی به منظور شناسایی، مسئولیت پذیر کردن و پاسخگو نمودن چالش دیگری است. از آنجایی که ماشین ها به طور فزاینده وظایف و تصمیماتی را که به طور سنتی توسط انسان انجام می شد را بر عهده می گیرند، چالش دیگر در خصوص شخصیت بخشی به هوش مصنوعی است. این که آیا، ضرورت دارد که به سیستم های هوش مصنوعی «شخصیت» و عاملیت اخلاقی یا قانونی داده شود؟ یکی از راه های برنامه نویسی سیستم های هوش مصنوعی «یادگیری تقویتی» است که در آن بهبود عملکرد با پاداش مجازی تقویت می شود. وقتی ماشین ها را به عنوان موجوداتی در نظر بگیریم که می توانند درک، احساس و عمل کنند، جهش بزرگی نیست که به وضعیت قانونی آنها فکر کنیم. آیا باید با آنها مانند حیواناتی با هوش مشابه رفتار کرد؟ آیا ما رنج ماشین های هوشمند «احساس رنج ماشینی» را باید در نظر بگیریم؟ ایجاد ماشین های هوش مصنوعی و استفاده از آنها در جامعه

<sup>10</sup> Pasquale

<sup>11</sup> Cadwalladr



می‌تواند تاثیر زیادی بر قوانین کیفری و مدنی داشته باشد. کل تاریخ قوانین بشری بر این پیش فرض استوار است که مردم، و نه روبات‌ها، تصمیم می‌گیرند. در جامعه‌ای که در آن تصمیمات پیچیده و مهم به طور فزاینده‌ای به الگوریتم‌ها واگذار می‌شود، این خطر وجود دارد که چارچوب‌های قانونی ما برای مسئولیت ناکافی باشد.

چالش مرتبط با منابع طبیعی نیز مطرح است. استخراج نیکل، کبالت و گرافیت برای استفاده در باتری‌های یونی لیتیوم - که معمولاً در خودروهای الکتریکی و تلفن‌های هوشمند یافت می‌شود - قبلاً به محیط زیست آسیب رسانده است و هوش مصنوعی احتمالاً این تقاضا را افزایش خواهد داد. با کاهش منابع موجود، اپراتورها ممکن است مجبور شوند در محیط‌های پیچیده‌تری کار کنند که برای اپراتورهای انسانی خطرناک است (کاکورل و همکاران<sup>۱۲</sup>، ۲۰۱۸). این امر باعث افزایش بازده و نرخ تخلیه فلزات کمیاب خاکی می‌شود و باعث تخریب بیشتر محیط زیست خواهد شد. در پایان چرخه محصول، کالاهای الکترونیکی معمولاً دور ریخته می‌شوند که منجر به تجمع فلزات سنگین و مواد سمی در محیط می‌شود (اودونگونه<sup>۱۳</sup>، ۲۰۱۰). افزایش تولید و مصرف دستگاه‌های فن‌آوری مانند روبات‌ها این مشکل مرتبط با ضایعات را تشدید می‌کند، به‌ویژه اینکه دستگاه‌ها احتمالاً با «کهنگی داخلی» طراحی می‌شوند - فرآیندی که در آن محصولات به گونه‌ای طراحی می‌شوند که «به‌طور زودرس» فرسوده شوند به طوری که مشتریان مجبور به خرید اقلام جایگزین شوند - که منجر به تولید مقادیر زیادی زباله الکترونیکی می‌شود (کاکورل و همکاران<sup>۱۴</sup>، ۲۰۱۸). هوش مصنوعی قرار است زندگی روزمره ما را در حوزه‌هایی مانند حمل و نقل، صنعت، خدمات، بهداشت و درمان؛ آموزش و پرورش، امنیت فردی و امنیت عمومی و سرگرمی تغییر دهد. با این وجود، این سیستم‌ها باید به گونه‌ای معرفی شوند که اعتماد و تفاهم ایجاد کند و به حقوق بشر و مدنی احترام بگذارد (دیگنوم<sup>۱۵</sup>، ۲۰۱۸). لذا آنها باید از اصول و ارزش‌های اساسی انسانی پیروی کنند و از رفاه مردم و کره زمین محافظت کنند. در این خصوص، اتفاق نظر قاطع در میان جامعه محققان این است که اعتماد به هوش مصنوعی تنها با انصاف، شفافیت، مسئولیت‌پذیری و مقررات قابل دستیابی است. مسائل دیگری که بر اعتماد تأثیر می‌گذارد این است که چقدر می‌خواهیم کنترل روی ماشین‌های هوش مصنوعی اعمال کنیم؟ چالش بعدی که با اعتماد ارتباط تنگاتنگی دارد، بحث انصاف است. برای اعتماد به هوش مصنوعی باید اطمینان حاصل کنیم که تصمیمات آن منصفانه و بی‌طرفانه است. از آنجایی که تصمیمات بیشتر و بیشتری به هوش مصنوعی واگذار می‌شود، باید اطمینان حاصل کنیم که این تصمیمات عاری از تعصب و تبعیض هستند؛ چیزی که کوربت دیویس و همکاران (۲۰۱۷) از این تحت عنوان عدالت الگوریتمی نام می‌برند.

شفافیت موضوع دیگری است که دستیابی به آن با سیستم‌های هوش مصنوعی مدرن، به‌ویژه سیستم‌هایی که مبتنی بر سیستم‌های یادگیری عمیق هستند، می‌تواند بسیار دشوار باشد. سیستم‌های یادگیری عمیق مبتنی بر شبکه‌های عصبی مصنوعی، گروهی از گره‌های به هم پیوسته هستند که از ساده‌سازی نحوه اتصال نورون‌ها در مغز الهام گرفته شده‌اند. یکی از ویژگی‌های شبکه‌های عصبی مصنوعی این است که پس از آموزش شبکه‌های عصبی مصنوعی با مجموعه داده‌ها، هرگونه تلاش برای بررسی ساختار داخلی آنها به منظور درک اینکه چرا و چگونه توانسته است یک تصمیم خاص بگیرد، کم و بیش غیرممکن است. چنین سیستم‌هایی به عنوان "جعبه سیاه" نامیده می‌شوند. مشکل دیگر این است که چگونه اطمینان حاصل کنیم که الزامات طراحی مشخص شده توسط سیستم برآورده می‌شود. رویکردهای راستی‌آزمایی کنونی معمولاً فرض می‌کنند که سیستمی که تأیید می‌شود هرگز رفتار خود را

<sup>12</sup> Khakurel et al.

<sup>13</sup> O'Donoghue

<sup>14</sup> Khakurel et al.

<sup>15</sup> Dignum

تغییر نمی‌دهد، با این حال سیستم‌های مبتنی بر یادگیری ماشینی-طبق تعریف- رفتار خود را تغییر می‌دهند، بنابراین هر تأییدی احتمالاً پس از یادگیری سیستم نامعتبر می‌شود (واینفیلد و جیروتکا<sup>۱۶</sup>، ۲۰۱۸).

چالش دیگر توانایی تمایز بین تصمیماتی است که انسان و یا هوش مصنوعی می‌گیرد و پذیرش مسئولیت مترتب بر پیامدهای آن است. نکته این است که هوش مصنوعی تصمیم‌گیری‌ها را متفاوت از انسان‌ها می‌گیرد و گاهی اوقات ما این تفاوت‌ها را درک نمی‌کنیم. ما نمی‌دانیم چرا یا چگونه این تصمیم را می‌گیرد (جک استیلگو). روش دیگری برای اطمینان از اعتماد هوش مصنوعی از طریق پاسخ‌گویی است. مسئولیت‌پذیری تضمین می‌کند که اگر یک هوش مصنوعی اشتباهی انجام دهد یا به کسی آسیب برساند، کسی وجود دارد که می‌تواند مسئول شناخته شود، خواه طراح، توسعه دهنده یا شرکتی باشد که هوش مصنوعی را می‌فروشد. در صورت بروز خسارات، باید مکانیزمی برای جبران خسارت وجود داشته باشد تا به قربانیان به اندازه کافی غرامت پرداخت شود. موضوع دیگری که بر اعتماد عمومی نسبت به هوش مصنوعی تأثیر می‌گذارد، کنترل است که به ترس پیرامون ایده "فوق هوش" مربوط می‌شود. افزایش هوش مصنوعی به حدی که از توانایی‌های انسانی پیشی بگیرد، ممکن است سبب شود تا کنترل منابع ما را به دست گرفته و از گونه‌های ما پیشی گرفته و منجر به انقراض انسان شود. یک ترس مرتبط این است که، حتی اگر یک عامل هوش مصنوعی به دقت طراحی شده باشد تا اهدافی همسو با نیازهای انسان داشته باشد، ممکن است اهداف فرعی پیش‌بینی نشده‌ای برای خود ایجاد کند که چنین نیستند. ابهام الگوریتم‌ها اغلب چالش‌هایی را در درک عمومی ایجاد می‌کند. تشویق توسعه فناوری‌های هوش مصنوعی قابل تفسیر بسیار مهم است. توضیحات دقیق، به زبان غیر فنی، در مورد تصمیمات الگوریتمی ضروری است. کارگاه‌های عمومی تعاملی، به رهبری کارشناسان هوش مصنوعی و مربیان اخلاق، بایستی الگوریتم‌های پیچیده را برای عموم مردم رمزگشایی کنند. هوش مصنوعی شفاف اعتماد را تقویت، شکاف دانش را پر و جامعه را آگاه‌تر می‌سازد (لی<sup>۱۷</sup>، ۲۰۲۳).

نتیجه‌گیری و پیشنهادات

مرور مبانی نظری نشان داد که بحث هوش مصنوعی دارای وجوه و ابعاد مختلفی است. از این رو، تصمیم‌گیری در خصوص خوب یا بد بودن آن شاید درست نباشد. بخشی از دشواری تصمیم و شناخت در خصوص هوش مصنوعی بدین خاطر است که این فناوری به شکل عمومی آن هنوز تازه است؛ ضمن این که، بخش دیگری از نگرانی‌ها و دغدغه‌های مرتبط با آن به دلیل شناخت اندک عموم نسبت به وجوه مختلف آن است. از این رو، نیازمند یک نگاه جامع در خصوص هوش مصنوعی هستیم. این که هوش مصنوعی هم در حوزه‌های مختلف اقتصادی، اجتماعی، آموزشی، محیط زیست، صنعت و خدمات و غیره کمک شایانی به جامعه بشری کرده است؛ هم این که بی‌توجهی ابعاد مختلف و پیامدهای آن می‌تواند نگرانی‌های بشری را از یک سو تعمیق ببخشد و از سوی دیگر به غفلت در خصوص اقدامات پیشگیرانه منجر شده و رنگ واقعیت به آن ببخشد. لذا بیشتر محققان با چالش‌ها و ابتکارات فوق‌الذکر موافق هستند. این که هوش مصنوعی باید به روشی اخلاقی مورد تحقیق، توسعه، طراحی، استقرار، نظارت و استفاده قرار گیرد – اما هر کدام دارای اولویت‌های متفاوتی هستند. با توجه به توضیحات پیشین، تعدادی از مسائل کلیدی که لازم است در تصمیم‌گیری‌ها و اقدامات اخلاقی مرتبط با هوش مصنوعی بایستی مورد توجه قرار گیرند، عبارتند از:

۱. حقوق بشر و رفاه: آیا هوش مصنوعی به نفع بشریت و رفاه انسان است؟

<sup>16</sup> Winfield and Jirotk

<sup>17</sup> Li



۲. آسیب عاطفی: آیا هوش مصنوعی یکپارچگی تجربه عاطفی انسان را کاهش می‌دهد یا آسیب عاطفی یا ذهنی را تسهیل می‌کند؟

۳. پاسخگویی و مسئولیت: چه کسی مسئول هوش مصنوعی است و چه کسی مسئول اعمال آن خواهد بود؟

۴. امنیت، حریم خصوصی، دسترسی و شفافیت: چگونه بین دسترسی و شفافیت با حریم خصوصی و امنیت تعادل برقرار کنیم، به خصوص وقتی صحبت از داده‌ها و شخصی‌سازی می‌شود؟

۵. ایمنی و اعتماد: اگر هوش مصنوعی توسط مردم غیرقابل اعتماد تشخیص داده شود، یا به گونه‌ای عمل کند که ایمنی خود یا دیگران را تهدید کند، چه اقداماتی لازم خواهد بود؟

۶. آسیب اجتماعی و عدالت اجتماعی: چگونه اطمینان حاصل کنیم که هوش مصنوعی فراگیر، عاری از تعصب و تبعیض، و همسو با عموم است.

۷. آسیب مالی: چگونه هوش مصنوعی را کنترل خواهیم کرد که بر فرصت‌های اقتصادی و اشتغال تأثیر منفی نگذارد و یا شغل را از کارگران انسانی نگیرد یا فرصت و کیفیت این مشاغل را کاهش ندهد؟

۸. قانونمندی و عدالت: چگونه می‌توانیم اطمینان حاصل کنیم که هوش مصنوعی - و داده‌هایی که جمع‌آوری می‌کند - به روشی عادلانه، منصفانه، و قانونی و تابع حاکمیت و مقررات مناسب استفاده، پردازش و مدیریت می‌شوند؟ چنین مقرراتی چگونه خواهد بود؟ آیا باید به هوش مصنوعی «شخصیت» اعطا شود؟

۹. کنترل و استفاده اخلاقی - یا سوء استفاده - از هوش مصنوعی: چگونه ممکن است هوش مصنوعی به صورت غیراخلاقی مورد استفاده قرار گیرد - و چگونه می‌توانیم در برابر این امر محافظت کنیم؟ چگونه می‌توانیم اطمینان حاصل کنیم که هوش مصنوعی تحت کنترل کامل انسان باقی می‌ماند، حتی زمانی که توسعه می‌یابد و «یاد می‌گیرد»؟

۱۰. آسیب‌های زیست محیطی و مرتبط با توسعه پایدار: چگونه از آسیب‌های محیطی بالقوه مرتبط با آن محافظت کنیم؟ چگونه توسعه و استفاده از هوش مصنوعی را به روشی پایدار تولید کنیم؟

۱۱. استفاده آگاهانه: چه کاری باید انجام دهیم تا اطمینان حاصل کنیم که عموم مردم در مورد استفاده و تعامل با هوش مصنوعی آگاه، آموزش و مطلع هستند؟

۱۲. خطر موجودیت نسل انسان: چگونه از مسابقه تسلیحاتی هوش مصنوعی اجتناب کنیم، آسیب‌های احتمالی را پیشگیرانه کاهش دهیم و تنظیم کنیم، و اطمینان حاصل کنیم که یادگیری ماشینی پیشرفته هم پیشرو و هم قابل مدیریت است؟

به طور کلی، هدف از طرح مسئله و چالش‌های اخلاقی مرتبط با هوش مصنوعی، شناسایی و شکل‌گیری چارچوب‌ها و سیستم‌های اخلاقی است که منافع انسانی را در بالاترین سطوح ایجاد کند، منافع را هم برای جامعه انسانی و هم برای محیط زیست در اولویت قرار دهد (بدون اینکه این دو هدف در تضاد قرار گیرند)، و خطرات و اثرات منفی مرتبط با هوش مصنوعی را کاهش دهد.



عباسی، محمود و تیموری، مهرداد. (۱۴۰۳). چالش‌های اخلاق رقومی هوش مصنوعی و امکان‌سنجی بیمه مسئولیت برای سامانه مبتنی بر هوش مصنوعی. فصلنامه حقوق و دولت، ۵(۱۵): ۳۱-۴۴.

DG, E. (2020). The ethics of artificial intelligence: Issues and initiatives.

Geslevich Packin, N and Lev-Aretz, Y (2018). *Learning algorithms and discrimination*, in: W. Barfield and U. Pagallo (eds.), Research Handbook on the Law of Artificial Intelligence,

Instance, B.D. Mittelstadt, P. Allo, M. Taddeo, S (2016). Wachter and Floridi, L. *The ethics of algorithms: Mapping the debate*, Journal of Big Data & Society, 2016, 3 (2).

Li, N. (2023). Ethical considerations in artificial intelligence: A comprehensive discussion from the perspective of computer vision. In *SHS Web of Conferences* (Vol. 179, p. 04024). EDP Sciences.

Li, F., Ruijs, N., & Lu, Y. (2022). Ethics & AI: A systematic review on ethical concerns and related strategies for designing with AI in healthcare. *Ai*, 4(1), 28-53.

Kamila, M. K., & Jasrotia, S. S. (2023). Ethical issues in the development of artificial intelligence: recognizing the risks. *International Journal of Ethics and Systems*, (ahead-of-print).

Wang, W. & Siau, K. (2018). Ethical and moral issues with AI. In: Proceedings of the 24th Americas Conference on Information Systems, Louisiana, USA, 2018 August 16-18. 1-5. Available at: [https://ink.library.smu.edu.sg/sis\\_research/9404](https://ink.library.smu.edu.sg/sis_research/9404).