بررسی تأثیرات هوش مصنوعی بر امنیت انسانی: چالش‌ها و راهکارها

محمدامین نقشبند[۱]، اباالفضل بابایی[۲]

[۱]دانشجوی ارشد علوم سیاسی، مطالعات منطقه‌ای، دانشگاه شهید بهشتی، تهران، ایران
naghshbandamin@gmail.com
[۲]دانشجوی دکتری علوم سیاسی، مطالعات منطقه‌ای، دانشگاه تهران، تهران، ایران
abolfazl.babaei@ut.ac.ir

چکیده

مبحث امنیت انسانی از مهم‌ترین مباحث ملی و فراملی در طول تاریخ بوده است؛ این مبحث باهدف شناسایی تهدیدات نوین در هر عصر می‌کوشد تا به راهکارهای دفع یا تقلیل تهدیدات بپردازد. باتوجه‌به اینکه امروزه فناوری هوش مصنوعی به‌عنوان نسل چهارم انقلاب صنعتی امکان ایجاد تهدیدات نوین و غیرقابل‌پیش‌بینی علیه امنیت انسانی را دارد؛ ازاین‌رو در مقاله پیشرو تلاش بر آن است تا ضمن بررسی این فناوری به این پرسش پاسخ داده شود که تهدیدات هوش مصنوعی در ارتباط با امنیت انسانی شامل چه مواردی بوده و راهکارهای مقابله با آن چیست؟ یافته‌های پژوهش نشان می‌دهد که تهدیدات هوش مصنوعی به سبب قابلیت نفوذ و همچنین فریب اطلاعاتی به آسیب‌پذیری امنیت انسانی منجر می‌شود. راهکارهای مقابله با آن می‌تواند از طریق توسعه شفافیت‌ها و ایجاد سازوکارهای قانونی برای مسئولیت‌پذیری در قبال استفاده از آن باشد. بااین‌حال به سبب دسترسی گسترده عمومی به این فناوری هنوز نمی‌توان تضمین کرد که در مقابله با توانایی و امکان تهدید آفرینی آن می‌توان به ضریب اطمینان یقینی دراین‌خصوص دست‌یافت و بنابراین پیش از هر چیز همکاری و آموزش لازم به نظر می‌رسد. این پژوهش به روش توصیفی – تحلیلی و با اسناد کتابخانه‌ای گردآوری و تالیف شده است.

کلمات کلیدی: هوش مصنوعی، امنیت انسانی، فناوری، تهدیدات سایبری.

۱ مقدمه

هوش مصنوعی (AI) به‌عنوان یکی از مهم‌ترین دستاوردهای فناوری در دنیای امروز، نقش بسیار حیاتی در جنبه‌های مختلف زندگی انسانی ایفا می‌کند. این فناوری که به‌سرعت در حال پیشرفت است، قابلیت‌هایی بی‌نظیر در تحلیل داده‌ها، تشخیص الگوها و انجام وظایف پیچیده دارد که تا پیش‌ازاین تنها توسط انسان قابل‌انجام بود. درحالی‌که مزایای هوش مصنوعی در بهبود کیفیت زندگی و افزایش کارایی در بسیاری از

حوزه‌ها قابل‌توجه است؛ اما مثل هر فناوری دیگری تهدیدات بالقوه آن برای امنیت انسانی را نمی‌توان نادیده گرفت. درحالی‌که فناوری‌های جدید و نوآورانه همیشه برای بهبود شرایط زندگی ایجاد شده‌اند بااین‌حال استفاده نادرست و مخرب از هوش مصنوعی توسط افراد، گروه‌ها، دولت‌ها و بازیگران مختلف می‌تواند اثرات جبران‌ناپذیری بر امنیت انسانی بگذارد؛ بنابراین این پژوهش، در تحلیل و شناسایی تهدیدات هوش مصنوعی و ارائه راهکارهای مؤثر برای مقابله با این تهدیدات تصمیم دارد به رویکردهای سیاست‌گذاران، متخصصان امنیتی و توسعه‌دهندگان فناوری‌های هوش مصنوعی بپردازد. پرسش مطرح‌شده این پژوهش آن است که تهدیدات هوش مصنوعی در ارتباط با امنیت انسانی شامل چه مواردی بوده و راهکارهای مقابله با آن چیست؟ جهت بررسی موضوع ابتدا تلاش خواهد شد تا بحث امنیت انسانی و رابطه هوش مصنوعی با آن توضیح داده شده و سپس به بررسی تهدیدات هوش مصنوعی از ابعاد مختلف در ارتباط با امنیت انسانی پرداخته شود. در بخش دیگر به بررسی راهکارهای مهم و ضروری برای ارتقای امنیت انسانی نیز پرداخته خواهد شد. فرضیه مطرح‌شده این است که تهدیدات هوش مصنوعی در خصوص امنیت انسانی شامل به‌خطرافتادن حریم شخصی، سو استفاده از اطلاعات یا فریب افراد توسط بازیگرانی چون تروریست‌ها و مسئله دیت فیک‌ها بوده و راهکار مقابله با آن مسئولیت‌پذیری بیشتر در قبال توسعه هوش مصنوعی و تضمین حریم خصوصی و استانداردسازی توسعه هوش مصنوعی است. این پژوهش به روش توصیفی – تحلیلی و با اسناد کتابخانه‌ای گردآوری و تألیف شده است.

۲ ادبیات پژوهش

مفهوم امنیت انسانی برای اولین‌بار در گزارش توسعه انسانی برنامه توسعه ملل متحد در سال ۱۹۹۴ به‌عنوان «رهایی از ترس» و «رهایی از نیاز» تعریف شد. امنیت انسانی به معنای حفاظت از انسان‌ها در برابر انواع تهدیدات مادی و غیرمادی است. نیازهای اساسی انسان ازجمله امنیت، مراقبت‌های بهداشتی، آموزش و حتی محیط‌زیست سالم می‌تواند در سطح دیگر به مفهوم امنیت اجتماعی همچون ثبات معیشت، اعتماد به جامعه و روابط اجتماعی مناسب، قابل‌ارتقا باشد. بااین‌حال، در هر دوره از حیات بشر تهدیدات می‌تواند متنوع باشد همان‌طور که امروزه رشد فناوری به ظهور انواع تهدیدات ازجمله کلاهبرداری و سرقت‌ها و حمله سایبری منجر شده است. ازاین‌رو، به نظر می‌رسد مفهوم امنیت انسانی در حال رسیدن به سطح جدیدی فراتر از نیازهای اساسی است که نیازمند بازتعریف ارزش‌ها و اخلاقیات در سایه توسعه فناوری جدید است [۱]. البته امنیت انسانی، به‌عنوان یک مفهوم پویا بر توسعه و احترام به حقوق بشر نیز توجه دارد؛ در همین زمینه چالش‌های این حوزه نیز مرتبط با حقوق بشر است؛ امنیت انسانی به‌عنوان یک ابزار برای تحلیل علل درگیری‌ها، ارائه سیاست‌های کارآمد برای حل بحران‌ها و ایجاد صلح پایدار، بسیار ارزشمند است. این سیاست‌ها به مسائل اجتماعی و اقتصادی نیز اثر می‌گذارند [۲]. از این نظر مفهوم امنیت انسانی به‌گونه‌ای گسترش می‌یابد که فراتر از گزارش توسعه انسانی سال ۱۹۹۴ مطرح می‌شود و موارد جدید چون امنیت را هم می‌افزاید. این واقعیت که امروزه اتکا بیش‌ازپیش به پلتفرم‌های دیجیتال، افراد و جوامع را در معرض تهدیدات مختلف و جدید قرار داده موجب شده تا لزوم توجه به امنیت دیجیتال برای حفاظت و انعطاف‌پذیری

زیرساخت‌های حیاتی و رفاه انسان بیش‌ازپیش ضروری به نظر برسد. حال با گسترش مفهوم امنیت انسانی می‌توان بهتر نقاط ضعف و نیازهای جامعه را در استراتژی‌ها و سیاست‌های امنیتی شناسایی کرد. این کار به ایجاد سیاست‌های جامع و هدفمند برای حفاظت از نهادها و شهروندان و حفظ امنیت انسانی در دوران دیجیتال کمک بسیار می‌کند [۳].

با توجه به اینکه جهان در حال تجربه تحولی سریع و چشمگیر با فناوری‌های جدید ازجمله هوش مصنوعی درگیر است ابتدا لازم است تا تعریف آن بیان شود. به‌طورکلی هر سامانه‌ای که بتواند وظایف خود را در شرایط مختلف و بدون نیاز به نظارت مستقیم انسان انجام دهد، یا از تجربیات خود یاد بگیرد و عملکردش را بهبود بخشد تداعی‌گر مفهوم هوش مصنوعی است. این سامانه‌ها ممکن است کارهایی را انجام دهند که نیازمند درک مشابه انسان ازجمله شناخت، برنامه‌ریزی، یادگیری، ارتباطات و فعالیت‌های فیزیکی باشند [۴]؛ بنابراین در پیوند بین هوش مصنوعی و امنیت انسانی باید گفت این موضوع به دلایل مختلف که در ادامه بیان می‌گردد به یکدیگر مرتبط هستند.

۳ تهدیدات امنیت انسانی توسط هوش مصنوعی

اگرچه روایت غالب درباره هوش مصنوعی جایگزینی یا شبیه‌سازی با هوش انسان‌ها از طریق ربات‌ها دارد؛ اما امنیت سایبری نیز در رابطه با امنیت انسانی نیز موردتوجه قرار می‌گیرد، زیرا عواملی که از هوش مصنوعی علیه انسان‌ها از طریق انتشار بدافزارها و آلوده‌کردن کدها و داده‌ها، استفاده از سلاح‌ها یا ابزارهای ضدبشری استفاده می‌شود، بازگوکننده خطرات این فناوری علیه انسان‌ها است. به نظر کسانی که خطرات هوش مصنوعی علیه امنیت انسانی را بسیار جدی می‌دانند دامنه تهدیدات این فناوری می‌تواند بسیار گسترده و متنوع باشد بااین‌حال عمده این تهدیدات به شکل زیر مطرح می‌شود.

۱.۳ تهدید امنیتی حریم خصوصی

هوش مصنوعی با توانایی تجزیه‌وتحلیل حجم عظیمی از داده‌ها می‌تواند به ابزار قدرتمندی برای جاسوسی و نفوذ به حریم خصوصی افراد تبدیل شود. این فناوری، باقابلیت جمع‌آوری، پردازش و تحلیل اطلاعات حساس، موجب ایجاد نگرانی‌هایی نسبت به سوءاستفاده‌های احتمالی و نقض حقوق بشر را افزایش می‌دهد؛ بنابراین، در نبود قوانین و مقرراتی محکم برای حفاظت از حریم خصوصی و جلوگیری از سوءاستفاده‌های احتمالی این تهدید جدی وجود دارد.

۲.۳ تهدیدات تروریستی

هوش مصنوعی می‌تواند توسط گروه‌های تروریستی علیه افراد بکار گرفته شود. تبلیغات، به‌عنوان ابزاری اساسی برای گروه‌های تروریستی برای انتشار ارزش‌ها و اعتقادات خود، با استفاده از هوش مصنوعی مولد، می‌تواند به طریقی آسان‌تر و با تأثیر بیشتر انجام شود. این فناوری می‌تواند برای گسترش سخنان نفرت‌انگیز و ایدئولوژی‌های رادیکال بهره‌برداری شده و با استفاده از تصاویر، ویدئوها یا صداگذاری‌هایی که با ارزش‌های این

گروه‌ها همخوانی دارند عاملی در توسعه و افزایش دامنه تأثیرگذاری بر نگرش‌ها و احساسات عمومی داشته باشد. گروه‌های تروریستی اغلب به بخش تأثیرپذیر جمعیت‌ها ازجمله کودکان از طریق شبکه‌های آنلاین رجوع می‌کنند؛ چراکه بسیاری از کودکان اوقات زیادی را در فعالیت‌های آنلاین مانند بازی‌های ویدئویی و تماشای فیلم سپری می‌نمایند. هوش مصنوعی به‌جز کاربردهای خود در تولید محتوای برای مخاطبان در زمینه‌های دیگری ازجمله هک و ساخت سلاح‌ها نیز کاربرد دارد. این فناوری در هواپیماهای بدون سرنشین، بمب‌های خودرویی و پهپادها نیز استفاده می‌شود و قابلیت قابل‌توجهی را می‌تواند برای تروریست‌ها فراهم آورد. زیرا هوش مصنوعی در شناسایی و تشخیص اشیا، هدایت، تصمیم‌گیری و برنامه‌ریزی مأموریت‌ها می‌تواند کارآمدتر و مستقل‌تر عمل کند [۵]. البته گفتنی است که به‌غیراز گروه‌های تروریستی جان انسان‌ها توسط دولت‌های زورگو و پیرو سیاست‌های تهاجمی نیز بیش‌ازپیش در معرض تهدید است.

۳.۳ تهدیدات دیپ‌فیک

دیپ‌فیک[1]، نوعی رسانه است که از هوش مصنوعی برای جایگزینی چهره یا صدای یک شخص با شخص دیگر استفاده می‌کند. این کار به‌گونه‌ای انجام می‌شود که بسیار واقعی و برای چشم غیرمسلح غیرقابل‌تشخیص است. هوش مصنوعی در حال افزایش توانایی هکرها برای انجام حملات سایبری پیچیده و فریبنده‌تر است. این امر به‌ویژه برای مهندسی اجتماعی که در آن هکرها از فریب و دست‌کاری روان‌شناختی برای فریب قربانیان به‌منظور افشای اطلاعات حساس یا انجام اقدامات مضر استفاده می‌کنند، نگران‌کننده است. هوش مصنوعی می‌تواند برای ایجاد ایمیل‌های فیشینگ بسیار متقاعدکننده‌ای استفاده شود که تشخیص آن‌ها از ایمیل‌های واقعی دشوار است. این ایمیل‌ها می‌توانند کارمندان را فریب دهند تا روی پیوندهای مخرب کلیک کنند، فایل‌های آلوده را دانلود کنند یا اطلاعات شخصی خود را فاش کنند. هوش مصنوعی همچنین برای ایجاد دیپ‌فیک‌های ویدئویی، تصویری و صوتی بسیار واقعی استفاده می‌شود که می‌تواند برای جعل هویت افراد، انتشار اطلاعات نادرست و فریب قربانیان استفاده شود [۶]. تهدیدات دیپ‌فیک‌ها می‌تواند موجب اختلال در نظم و امنیت افراد و محیط آن‌ها باشد.

۴ راهکارهای افزایش امنیت انسانی با هوش مصنوعی

افزایش امنیت انسانی با استفاده از هوش مصنوعی می‌تواند از طریق راهکارهای مختلفی صورت گیرد. در زیر به چند راهکار مهم اشاره‌شده است.

۱.۴ توسعه هوش مصنوعی در توسعه خدمات همگانی

امنیت انسانی فقط مختص افراد معدودی نیست، بنابراین قابلیت‌های هوش مصنوعی باید در اختیار همگان قرار گیرد وقتی صحبت از مواردی مانند امدادرسانی در بلایا، جلوگیری از درگیری‌ها، حفاظت از حقوق بشر و اطمینان از اجرای عدالت می‌شود، ضروری است که مردم، گروه‌ها، سازمان‌های غیردولتی و دولت‌ها

[1] DeepFake

اطلاعات را به طور گسترده‌ای به اشتراک بگذارند؛ اما درعین‌حال، حفظ امنیت و خصوصی‌بودن این داده‌ها نیز بسیار مهم است؛ بنابراین باید قوانین و مقررات روشنی در مورد نحوه استفاده و محافظت از داده‌ها وضع شود. هوش مصنوعی می‌تواند تأثیر مثبتی بر حقوق افراد داشته باشد، ازجمله حق زندگی و امنیت شخصی که توسط اعلامیه جهانی حقوق بشر نیز به رسمیت شناخته می‌شود. این فناوری می‌تواند در موارد مختلف نه‌تنها به حفظ جان انسان‌ها کمک کند؛ بلکه قادر است خطاهای انسانی را نیز کاهش دهد. مثلاً عملکرد تشخیص پزشکی و همچنین تخصیص منابع در مؤسسات مراقبت‌های بهداشتی مانند بیمارستان‌ها را بهبود بخشد. ضمناً در راستای تقویت خدمات عمومی هوش مصنوعی از طریق تحلیل کلان‌داده‌ها اقدامات پیشگیرانه انجام دهد؛ مثلاً در حوزه بهداشت می‌تواند به شناسایی مناطق شیوع بیماری ناشی از غذا بپردازد و با اطلاع‌رسانی به بازرسان بهداشت کمک مهمی در خدمت به‌سلامتی انجام دهد. به‌طورکلی، هوش مصنوعی در شناسایی و پاسخ به آسیب‌پذیری‌هایی امنیت انسانی بسیار سودمند می‌تواند عمل کند.

یکی از مهم‌ترین موارد کاربرد این فناوری در خدمت همگانی زمانی است که باوجود موانع سیاسی ازجمله تحریم‌ها تأمین امنیت انسانی دشوار شده است. پیشرفت‌های اخیر هوش مصنوعی در برنامه‌ریزی، به‌ویژه در موقعیت‌های اضطراری، بسیار امیدوارکننده به نظر می‌رسد به‌عنوان مثال، برنامه‌ای تحت عنوان «برنامه‌ریزی لجستیک اضطراری» با استفاده از هوش مصنوعی به مدیریت منابع در هنگام بلایای طبیعی کمک می‌کند. این برنامه می‌تواند بهترین مکان‌ها برای راه‌اندازی ایستگاه‌ها، کارآمدترین مسیرها برای توزیع کمک و تخلیه افراد، میزان امدادرسانی موردنیاز و زمان‌بندی وظایف را نیز تعیین کند.

کاربرد این فناوری در سطح فرا دولتی چون سازمان‌های بین‌المللی می‌تواند زمان پیش‌بینی و برنامه‌ریزی در اقدامات جهانی را کاهش داده و در هزینه منابع صرفه‌جویی ایجاد کند. همچنین، دولت‌ها، سازمان‌های غیردولتی و گروه‌های مدنی می‌توانند از قابلیت برنامه‌ریزی هوش مصنوعی استفاده کنند. این نهادها، با توجه به انعطاف‌پذیری بیشتر و موانع کمتر، می‌توانند فناوری‌های جدید را آزمایش و بهبود بخشند و در بحران‌های بشردوستانه کارایی بیشتری داشته باشند.

۲.۴ تضمین حریم خصوصی و استانداردسازی توسعه هوش مصنوعی

تضمین حریم خصوصی و استانداردسازی توسعه ایمن و قابل‌اعتماد هوش مصنوعی در جهت جلوگیری از سوءاستفاده‌ها و حفظ اعتماد عمومی به این فناوری پیشرفته امری بسیار حیاتی است. برای اطمینان از حفظ حریم خصوصی، اقدام در جهت استانداردسازی هوش مصنوعی در سال ۲۰۱۸ تلاش داشت تا تمرکز بر ایجاد اصول اجتماعی اولیه مانند انصاف، ایمنی و شفافیت را تقویت کند و توسعه فناوری‌های هوش مصنوعی با ملاحظات اخلاقی انجام شود. دستورالعمل‌هایی نیز برای جلوگیری از تصمیم‌گیری مغرضانه و افزایش شفافیت فرایندهای هوش مصنوعی اتخاذ شد. در ادامه این روند تا سال ۲۰۲۰، توسعه چارچوب‌های ملی و ساختارهای حکومتی مانند مدیریت ریسک هوش مصنوعی NIST (RMF) در ایالات متحده و ارائه پیشنهاد چارچوب مقررات اتحادیه اروپا انجام شد که هدف آن‌ها تضمین توسعه مسئولانه و اخلاقی هوش مصنوعی بود. در سال ۲۰۲۲، ایجاد دستورالعمل‌های فنی و استانداردسازی رویه‌ها در صنعت، به‌ویژه از سوی ISO/IEC، بر جنبه‌های فنی هوش مصنوعی متمرکز شد تا قابلیت همکاری، اطمینان و امنیت سیستم‌های هوش مصنوعی

را تضمین کند و از سال ۲۰۲۳ تاکنون، تمرکز بر هماهنگ‌سازی و گسترش استانداردها و مقررات موجود در سطح جهانی در حال انجام است [۷]. علی‌رغم این اقدامات به نظر برخی اولین چالشی که باید به آن پرداخته شود نیاز به نظارت مجدد بر زیرساخت‌های هوش مصنوعی برای جلوگیری از آسیب‌پذیری‌ها اساسی است؛ چراکه احتمال درز داده‌ها می‌تواند برای سازمان‌ها و افراد پیامدهای مخرب ازجمله خسارات مالی و آسیب به شهرت، به همراه داشته باشد. این مسئله با درنظرگرفتن توانایی هکرها و افزایش جرائم سایبری اهمیت بسیاری می‌یابد به‌عنوان‌مثال در اوت ۲۰۲۳، تعداد زیادی از رکوردهای حساس اینترنتی توسط هکرها به خطر افتادند. همچنین بیم نفوذ به چت‌بات‌ها توسط آن‌ها یک تهدید دیگر است موردی که می‌توان اشاره کرد اینم است که بیش از ۱۰۰۰۰۰ حساب چت جی‌پی‌تی بین سال‌های ۲۰۲۲ تا ۲۰۲۳ در معرض نقض حریم خصوصی کاربران قرار گرفته است. مسمومیت داده‌ها نیز یک موضوع جدی است که ممکن است توسط عوامل خطر سوءاستفاده گرفته شود، به‌عنوان‌مثال، این عمل شامل داده‌های مخرب مدل‌های GenAI است که می‌تواند نتایج را تحت‌تأثیر قرار دهد و به سوگیری تبدیل شود [۸].

نکته قابل‌توجه این است که هرچند برخی تهدیدات هوش مصنوعی را فراتر از حد کنترل می‌دانند؛ اما برخی از افراد معتقدند که خطر هوش مصنوعی را نباید بزرگ کرد؛ ولی باید در استفاده از آن به یک سری قوانین و استانداردهایی رسید؛ آن‌ها معتقدند که هوش مصنوعی تابع تفکر انسانی است که برای همه کاربردهای هوش مصنوعی صدق می‌کند. این به این معناست که گرچه امنیت سایبری در حوزه هوش مصنوعی می‌تواند تهدید زا باشد؛ اما واقعیت این است که این فناوری هنوز نتوانسته زمینه‌ها، فرهنگ‌ها و معنای انسانی را به طور کامل درک کند. این ایده نشان می‌دهد که دیدگاه اینکه هوش مصنوعی در آینده ممکن است از ما پیشی گرفته و تبدیل به خطر شود نیز زیاد دقیق نیست [۹]. پس ازآنجایی‌که هر فناوری ساخته دست انسان بوده و توسط افرادی توسعه پیدا می‌کند تهدیدات قابل‌کنترل نیز خواهند بود. خطر دیگری که مطرح می‌شود دسترسی همگانی به این فناوری است و برخی معتقدند که در این راستا، سیاست‌گذاران باید اطمینان حاصل کنند که این فناوری‌ها و مزایای آن برای همه افراد در چارچوب قوانین است. دراین‌خصوص استانداردهای بین‌المللی هوش مصنوعی، چارچوبی برای استفاده مسئولانه و اخلاقی از فناوری‌های هوش مصنوعی ارائه می‌دهند و حوزه‌هایی مانند حریم خصوصی، سوگیری، شفافیت و پاسخگویی را پوشش می‌دهند. این استانداردها به سازمان‌ها کمک می‌کنند تا سیستم‌های هوش مصنوعی منصفانه و شفاف را پیاده‌سازی کنند. یکی از این استانداردها ISO/IEC 23894 است که بر مدیریت ریسک در سیستم‌های هوش مصنوعی تمرکز دارد. این سند به سازمان‌ها راهنمایی ارائه می‌کند که چگونه ریسک‌های مرتبط با هوش مصنوعی را در توسعه، تولید، استقرار و استفاده از محصولات، سیستم‌ها و خدمات هوش مصنوعی مدیریت کنند. هدف اصلی آن کمک به سازمان‌ها برای ادغام مدیریت ریسک در فعالیت‌ها و عملکردهای مرتبط با هوش مصنوعی است. همچنین، فرایندهایی را برای پیاده‌سازی و ادغام مؤثر مدیریت ریسک هوش مصنوعی توصیف می‌کند. در کل هدف این استاندارد، قابل‌فهم بودن الگوریتم‌ها و مدل‌های هوش مصنوعی و بررسی آن‌ها است تا اعتماد و اطمینان به سیستم‌های هوش مصنوعی افزایش یابد [۱۰].

برخی وجود تدابیر از پیش طراحی‌شده را بدون درنظرگرفتن درست یا نادرست بودن تهدید آفرینی هوش مصنوعی یک امر ضروری می‌دانند تا برای تقویت نقش مثبت هوش مصنوعی اقداماتی انجام شود؛

مثلاً در ۵ دسامبر ۲۰۲۲، کمیسیون اروپا پیش‌نویس درخواست استانداردسازی را منتشر کرد که در آن از کمیته اروپایی استاندارد (CEN) و کمیته اروپایی استانداردسازی الکتروتکنیکی (CENELEC) یک سری از استانداردهایی را باهدف حمایت از توسعه ایمن و قابل‌اعتماد ارائه کرد [۱۱].

به نظر برخی گنجاندن ملاحظات حقوق بشر در منشور حقوق هوش مصنوعی، قانون هوش مصنوعی و قانون خدمات دیجیتال نقطه عطف مهمی در قبال حقوق بشر است. از مصادیق حقوق بشری شاید کاهش تلفات انسانی در توسعه تجهیزات یا عملیات نظامی است؛ چراکه جایگزینی نیروهای انسانی با سیستم‌های هوش مصنوعی در مأموریت‌های پرخطر، می‌تواند تضمین حقوق بشر از جمله کار در محیط‌های خطرناک و مانند مین‌زدایی و شناسایی مواد باشد. به عبارتی با جایگزینی نیروهای انسانی با سیستم‌های هوش مصنوعی در مأموریت‌های پرخطر، میزان تلفات انسانی کاهش می‌یابد [۱۲]. جالب است که در اولین اجلاس بین‌المللی درباره هوش مصنوعی نظامی، بیانیه‌ای توسط بیش از ۶۰ کشور ازجمله چین و ایالات متحده امضا شد که بر شفافیت در اصول هدایت فناوری‌های هوش مصنوعی نظامی، نظارت دقیق بر سیستم‌های تسلیحاتی هوش مصنوعی، کاهش تعصب و تعهد به قوانین بین‌المللی نیز تأکید داشت. بااین‌حال، منتقدان بر این باورند که این اسناد برای محدودکردن توسعه انواع سلاح‌های هوش مصنوعی کافی نیستند و هشدار می‌دهند که بدون چارچوب‌های اخلاقی واضح‌تر، خطر توسعه سلاح‌های مخرب مانند «ربات‌های قاتل» نیز وجود دارد [۱۳]. به‌طورکلی امنیت انسانی بر محافظت از جان، حقوق و کرامت انسان‌ها تأکید دارد و توسعه فناوری‌های هوش مصنوعی نظامی بدون چارچوب‌های اخلاقی و نظارتی ممکن است تهدیدی جدی برای امنیت انسانی ایجاد کند. نبود شفافیت و نظارت کافی می‌تواند منجر به توسعه سلاح‌هایی شود که توانایی آسیب‌رساندن گسترده و غیرقابل‌پیش‌بینی به انسان‌ها و جوامع را دارند؛ بنابراین، ادغام اصول اخلاقی و چارچوب‌های نظارتی محکم در توسعه و استفاده از هوش مصنوعی نظامی برای حفظ امنیت انسانی ضروری است.

به‌عنوان اقدام فراملی باید گفت در نشست رسمی شورای امنیت درباره هوش مصنوعی نیز سخنرانان بر لزوم تصمیم فوری جامعه بین‌المللی در برخورد با واقعیت جدید هوش مصنوعی، خطرات و محیط‌های این فناوری را انجام دادند. جک کلارک[۲]، یکی از بنیان‌گذاران Anthropic، بیان می‌کند که هوش مصنوعی باوجود فراوان، به دلیل استفاده از پتانسیل و غیرقابل‌پیش‌بینی بودن، تهدیدی برای آرامش، امنیت و ثبات جهانی است. او خواستار توسعه سیستم‌های ارزیابی قوی و قابل‌اعتماد توسط دولت‌ها شد تا از سوءاستفاده شود. عمران شرف از امارات، خواستار ایجاد قوانین مشترک برای جلوگیری از استفاده از هوش مصنوعی قبل از دیر شدن است. نماینده غنا نیز برای محدودکردن اماکن نظامی و توسعه‌های صلح‌آمیز برای هوش مصنوعی سخنرانی کرد. سخنران اکوادور با رد نظامی‌سازی هوش مصنوعی، به خطرات سلاح‌های خودمختار مرگبار اشاره کرد. نماینده چین هوش مصنوعی را یک شمشیر دو لبه دانست و خواستار اولویت‌دادن به اخلاق و ایمنی و فراگیری آن شد. جیمز کلورلی[۳] از بریتانیا نیز به پتانسیل هوش مصنوعی برای تقویت یا مختل کردن ثبات جهانی و چالش‌های اخلاقی آن اشاره کرد و فرصت‌های مهمی را برای جامعه بین‌المللی متصور کرد [۱۴]. همه این موارد بیانگر تلاش‌هایی برای رفع نگرانی از تأثیرات هوش مصنوعی بر امنیت انسانی و

[2] Jack Clark
[3] James Cleverly

جوامع است.

۵ نتیجه‌گیری

هوش مصنوعی به‌عنوان یکی از پیشرفت‌های برجسته فناوری در دنیای امروز، تأثیرات عمیقی بر امنیت انسانی دارد. این فناوری که به‌سرعت در حال تکامل است، در بسیاری از جوانب امنیت ملی و شخصی نقش مهمی ایفا می‌کند. این پژوهش نشان داد که در زمینه امنیت شخصی هوش مصنوعی می‌تواند به‌عنوان یک ابزار قدرتمند از طریق نقض حریم شخصی، حملات تروریستی و دیپ‌فیک‌ها تهدیدآفرین باشد؛ بنابراین منطقی است که به‌عنوان پیشنهاد برای استفاده مؤثر از هوش مصنوعی در تأمین امنیت انسانی، باید اصول شفافیت، مسئولیت‌پذیری، برابری و جهانی بودن رعایت شود. این اصول تضمین می‌کنند که مزایای هوش مصنوعی به‌صورت منصفانه توزیع شود و همه افراد از آن بهره‌مند شوند. همچنین، مهم است که داده‌های مورداستفاده برای طراحی الگوریتم‌ها شناخته‌شده و اثرات احتمالی آن‌ها بررسی شود. برای اطمینان از حفاظت داده‌ها و جلوگیری از سوءاستفاده، وضع قوانین و مقررات مناسب ضروری است. هدف استفاده از هوش مصنوعی باید کاهش ناامنی‌های انسانی و افزایش توانمندی‌های انسانی باشد و این کار باید به‌صورت عادلانه، شفاف و پاسخگو انجام شود. البته باید گفت مقابله با چالش‌های امنیت انسانی نیازمند اقداماتی فراتر از مواجهه با عوامل ایجادکننده بحران‌ها است. سه مانع اصلی در این زمینه شامل عدم آگاهی از تهدیدات قبل از وقوع، ناتوانی در برنامه‌ریزی برای مقابله با این تهدیدات و عدم توانمندسازی افراد برای پاسخگویی مؤثر است؛ بنابراین اقدامات مطرح‌شده تا حد زیادی شاید بتواند امنیت انسانی را تضمین کند؛ ولی به معنای رفع کامل تهدیدات نیست. شاید درست است که مطرح گردد که برای مقابله با تهدیدات هوش مصنوعی بهتر است کشورها بر روی استفاده از هوش مصنوعی به‌عنوان کمک‌کننده توجه داشته باشند. البته رسیدن به این مرحله نیازمند توسعه و توجه این فناوری است. به‌هرحال دفاع در برابر حملات امنیت انسانی مبتنی بر هوش مصنوعی نیز نیازمند رویکردی چندوجهی است که از قدرت هوش مصنوعی برای شناسایی و خنثی‌کردن تهدیدات در زمان واقعی و از هوش و تفکر انتقادی انسان برای تأیید اطلاعات و گزارش فعالیت‌های مشکوک استفاده کند. با ترکیب راه‌حل‌های فناوری پیشرفته با آگاهی و آموزش کارکنان، سازمان‌ها می‌توانند به طور مؤثرتری در برابر این تهدیدات پیچیده و در حال تکامل محافظت کرد.

مراجع

[1] A. Chugh, How to build a model for human security in the Fourth Industrial Revolution, no. https://www.weforum.org/agenda/2018/09/how-to-build-a-model-for-human-security-in-the-fourth-industrial-revolution/, 2018.

[2] Reliefweb, Human Security: Concepts and Implications with an Application to Post-Intervention Challenges in Afghanistan, no. 2005.

[3] T. Nair, Digital Security's Place in Human Security, no. https://www.rsis.edu.sg/rsis-publication/nts/digital-securitys-place-in-human-security, 2023.

[4] K. M. Sayler, Artificial intelligence and national security. Congressional Research Service, vol. ,45 p. 178, 2020.

[5] C. Nelu, "Exploitation of Generative AI by Terrorist Groups, 2024. [Online]. Available: https://www.icct.nl/publication/exploitation-generative-ai-terrorist-groups.

[6] C. Avey, The Impact of AI on Social Engineering Cyber Attacks, 2023. [Online]. Available: https://www.secureworld.io/industry-news/impact-ai-social-engineering-attacks.

[7] Modulos, The Evolution and Importance of AI Standards, 2024. [Online]. Available: https://www.modulos.ai/blog/the-evolution-and-importance-of-ai-standards/.

[8] WIZ, AI Security Explained: How to Secure AI, no. https://www.wiz.io/academy/ai-security, 2023.

[9] M. Stone, AI Security: How Human Bias Limits Artificial Intelligence, no. https://securityintelligence.com/articles/ai-security-human-bias-artificial-intelligence/, 2021.

[10] ISO, Artificial intelligence: What it is, how it works and why it matters, 2023. [Online]. Available: https://www.iso.org/artificial-intelligence.

[11] E. Gaumond and C. Régis, Assessing Impacts of AI on Human Rights: It's Not Solely About Privacy and Nondiscrimination, no. https://www.lawfaremedia.org/article/assessing-impacts-of-ai-on-human-rights-it-s-not-solely-about-privacy-and-nondiscrimination, 2023.

[12] M. Horowitz, Artifcial Intelligence, International Competition, and the Balance of Power, vol. 3, p. Texas National Security Review, 2018.

[13] Thewarhorse, US Military Pledges Principled, Responsible Approach to Artificial Intelligence, 2023. [Online]. Available: https://thewarhorse.org/us-military-pledges-principled-response-artificial-intelligence/.

[14] webtv, Artificial Intelligence: Opportunities and Risks for International Peace and Security - Security Council, 9381st Meeting, no. http://webtv.un.org/en/asset/k1j/k1ji81po8p, 2023.