



Iranian Independent Islamic Human Rights Commission
کمیسیون حقوق بشر اسلامی ایران
انجمن حقوق انسان اسلامیة فی ایران



چرا حقوق بشر باید در هسته مدیریت و حکمرانی هوش مصنوعی قرار بگیرد؟

امور پژوهشی، آموزشی و ترویجی کمیسیون حقوق بشر اسلامی ایران

پاییز ۱۴۰۴

توضیحات مقدماتی کمیسیون:

پیشرفت های حاصله در حوزه فناوریهای متنوع اگرچه دستاوردهای مهم در زندگی بشر داشته اما مسائل جدید و چالشهای خاص خود را نیز به همراه داشته است. در میان فناوریهای نوین، هوش مصنوعی در روندی رو به رشد به همه عرصه های زندگی بشر ورود کرده و قطعا در آینده جلوه های جدیدتری از آن نیز ظهور و بروز خواهد یافت. در قبال هوش مصنوعی، رویکردهای مختلفی را می توان ملاحظه کرد. برخی عمدتا بر نتایج مثبت کاربرد آن تاکید دارند. بعضی ابراز نگرانی های مختلف در خصوص پیامدهای هوش مصنوعی را مطرح می کنند و طیف سوم نیز رویکرد میانه ای اتخاذ نموده و بر این باورند که هوش مصنوعی نیز مانند هر دستاورد دیگر بشری می تواند کاربردهای مثبت و یا منفی داشته باشد، مهم این است که بدرستی از آن بهره برداری شود و روندهای توسعه فنی آن با لحاظ دفع ابعاد منفی صورت پذیرد. معتقدین به رویکرد اخیر تلاش می کنند تا هر گونه توسعه فنی هوش مصنوعی یا بهره برداری های متنوع از آن با لحاظ موازین اخلاقی، انسانی و حقوقی باشد. در این زمینه آثار فنی متعددی در سطح جهانی تالیف و انتشار یافته است. نوشتار حاضر یکی از قلم زنی های فشرده و مختصری است که در این خصوص ارائه شده، فلذا مفید دیده شد تا به فارسی برگردان شود. نویسنده نوشتار حاضر ضمن تحلیل اینکه چرا باید در هر سیاست گذاری و برنامه ریزی مربوط به هوش مصنوعی، حقوق بشر محور قرار گیرد، داده های مفیدی را در خصوص برخی پدیده های منفی پیرامون هوش مصنوعی از جمله سیطره جمع محدودی از شرکت ها و دولتها بر طرح های مربوط به هوش مصنوعی در سطح جهانی نیز ارائه کرده که جای تامل جدی دارد.

در کشور ما طی سنوات اخیر محدود ضوابطی در شورای عالی انقلاب فرهنگی در عرصه سیاست گذاری هوش مصنوعی تصویب شده و هم اینک نیز طرحی در این زمینه در قوه مقننه در دست بررسی است تا قانونی در این مورد به تصویب رسد. سیاستگذاران ملی حتما لازم است از دغدغه ها و نگرانی های مربوط به رعایت حقوق انسانی در هر سیاست گذاری و برنامه ریزی مربوط به هوش مصنوعی مطلع باشند و نکات مفید را در تلاشهای خود مد نظر قرار دهند. امید است برگردان فارسی این قبیل آثار برای نهادهای سیاست گذار ما و هم چنین عموم مردم مفید باشد.

◆ نویسنده: دکتر دانیل روفر (Daniel Leufer) دانش آموخته فلسفه با تمرکز بر فلسفه فناوریهای نوین، کارشناس ارشد سیاست گذاری حقوق دیجیتال و فناوری های نوظهور در دفتر بروکسل سازمان غیر دولتی Access Now.

◆ متصدی روند برگردان انگلیسی به فارسی: امیراحسان بی غم، دانش پژوه رشته حقوق در دانشگاه شهید بهشتی و همیار کمیسیون/ ویرایش و نهایی سازی متن در امور پژوهشی کمیسیون حقوق بشر اسلامی ایران

◆ مآخذ: سامانه سازمان غیردولتی بین المللی Access Now، منتشره در ۲۴ سپتامبر ۲۰۲۴

◆ شناسه متن: ۰۴/۳/۳۰۱

هر گونه نقل مطالب گزارش حاضر یا باز نشر جزئی یا کلی آن با ذکر مآخذ و رعایت استناد به امور پژوهشی کمیسیون حقوق بشر اسلامی ایران، مجاز است.

متن مقاله:

از زمان آغاز به کار [هوش مصنوعی] چت جی پی تی (ChatGPT) توسط شرکت OpenAI در سال ۲۰۲۲، جهان درگیر موجی از هیاهو و تبلیغات پیرامون آنچه «هوش مصنوعی مُوگد»^۱ نامیده می‌شود، گردیده است؛ پدیده‌ای که موجب شده نهادهای قانون‌گذار و تنظیم‌گر با شتاب به دنبال مهار آسیب‌ها و افراط‌های ناشی از آن باشند. [این امر نشان‌دهنده تکرار چرخه‌ای است که در تاریخ فناوری‌های نوظهور دیده شده است، جایی که هیجان عمومی اغلب از ظرفیت‌های نظارتی پیشی می‌گیرد].

با این حال، این وضعیت در واقع پدیده‌ای نوظهور محسوب نمی‌شود. در سال ۲۰۱۸، هنگامی که پیشرفت‌های قابل ملاحظه‌ای در حوزه‌ی یادگیری ماشین^۲ سبب شکل‌گیری موجی از شور و اشتیاق نسبت به هوش مصنوعی شد، همزمان شاهد ظهور مجموعه‌ای از «دستورالعمل‌های اخلاقی هوش مصنوعی» فاقد ضمانت اجرا^۳ و دیگر پیشنهاد‌های خودتنظیمی داوطلبانه بودیم [که نشان‌دهنده تلاش صنایع و دولت‌ها برای ایجاد چارچوب‌های اخلاقی بدون الزام قانونی بود].

در همان ایام، سازمان Access Now و نهادهای همسو خواستار تدوین ابزارهای حقوقی الزام‌آور مبتنی بر حقوق بشر شدند و از همین رهگذر، اعلامیه تورنتو^۴ را در کنفرانس RightsCon در سال ۲۰۱۸ راه‌اندازی کردند. هدف این اعلامیه، ترویج احترام به حق برابری و منع تبعیض در سامانه‌های یادگیری ماشین بود. [این اقدام تاریخی، نمونه‌ای اولیه از تلاش‌های سازمان‌های حقوق دیجیتال برای پیوند مستقیم توسعه فناوری با مفاهیم حقوق بشر بین‌المللی است].

از آن زمان تاکنون، سازمان‌های مدافع حقوق دیجیتال به‌صورت پیوسته از رویکردی مبتنی بر حقوق بشر در حکمرانی هوش مصنوعی دفاع کرده‌اند. اما این تلاش از جنبه‌های مختلفی با تهدید مواجه شده است؛ و به همراه آن، هرگونه امید به تحقق وضعیتی که در آن توسعه‌ی فناوری‌های هوش مصنوعی در خدمت انسان و نه در خدمت نظارت، بهره‌کشی و استثمار انسانی باشد نیز به خطر افتاده است. [این تهدیدها شامل فشارهای سیاسی، لابی‌های صنعتی و ناهماهنگی میان مقررات بین‌المللی و ملی است].

اکنون و پس از گذشت هفت سال، ممکن است مشاهده‌گر بی‌طرف چنین بیندیشد که مسئله‌ی حکمرانی بر هوش مصنوعی به‌صورت قطعی و نهایی حل‌وفصل شده است. به‌هرحال، تنها در ۲۰۲۴، شاهد تصویب قانون هوش

^۱ Generative Artificial Intelligence – AI

^۲ Machine Learning

^۳ Toothless AI Ethics Guidelines

^۴ The Toronto Declaration

مصنوعی اتحادیه اروپا^۵ [این قانون اولین مقررات جامع اتحادیه اروپا در زمینه هوش مصنوعی است که تلاش دارد میان نوآوری و حقوق بشر تعادل ایجاد کند]، کنوانسیون چارچوب شورای اروپا درخصوص هوش مصنوعی، حقوق بشر، دموکراسی و حاکمیت قانون^۶ و نیز [موارد] بی‌شمار طرحهای تنظیم‌گرانه‌ی جهانی دیگر بوده‌ایم که همگی ظاهراً بر پایه‌ی چارچوب حقوق بشر استوارند [و این روند، نمایانگر تلاش‌های بین‌المللی برای ایجاد استانداردهای مشترک جهانی است].

سازمان ملل متحد اخیراً با دو قطعنامه مکمل، در مجمع عمومی به اجماع رسید که هوش مصنوعی بر حقوق بشر تأثیر می‌گذارد، و حتی با لحنی ستودنی هشدار داد که برخی از کاربردهای هوش مصنوعی با حقوق بشر ناسازگار است.

علاوه بر این، هیئت مشورتی عالی‌رتبه سازمان ملل در زمینه‌ی هوش مصنوعی^۷ در گزارش جدید خود تأکید کرده است که «حکمرانی بر هوش مصنوعی در خلأ اتفاق نمی‌افتد» و «حقوق بین‌الملل، به‌ویژه حقوق بین‌الملل بشر^۸ در رابطه با هوش مصنوعی قابل اعمال است».

با این همه، واقعیت کنونی به هیچ وجه خوشایند نیست [و شکاف میان تعهدات اعلامی و اقدامات عملی، چالش‌های عمیقی را در مسیر تحقق حکمرانی مبتنی بر حقوق بشر در حوزه‌ی هوش مصنوعی پدید آورده است].

چه چیزی حکمرانی مبتنی بر حقوق بشر در هوش مصنوعی را تهدید می‌کند؟

نخستین مشکل اساسی این است که همان ابزارهای حقوقی‌ای که از آن‌ها به‌عنوان دستاوردهای تاریخی یاد می‌شود - همچون قانون هوش مصنوعی اتحادیه اروپا^۹ و معاهده‌ی هوش مصنوعی شورای اروپا^{۱۰} - در عمل چنان آکنده از استثنائات (معافیت‌های قانونی محدود کننده) و معافیت‌ها و عدول‌ها از تعهدات هستند که هرگونه ادعا در خصوص ترویج حکمرانی مبتنی بر حقوق بشر بر هوش مصنوعی را تضعیف می‌کنند. [این نوع معافیت‌ها معمولاً ناشی از فشار لابی‌های صنعتی و توجیهاتی چون ارجاع به امنیت ملی است و سابقه‌ای مشابه در ضوابط حمایت از داده‌های عمومی (GDPR) و موازین ناظر به داده‌های اروپایی دیده شده است].

همچنین قانون هوش مصنوعی اتحادیه اروپا نتوانسته است خطرناک‌ترین کاربردهای هوش مصنوعی را ممنوع کند، این قانون علیه مهاجران تبعیض قائل می‌شود و نهادهای مسئول اجرای قانون و مقامات مهاجرتی را از الزامات

^۵ European Union Artificial Intelligence Act – EU AI Act

^۶ Council of Europe's Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law – CoE AI Treaty

^۷ UN High-Level Advisory Board on AI

^۸ International Human Rights Law

^۹ EU AI Act

^{۱۰} CoE AI Treaty

شفافیت مستثنی می‌کند. [این بخش نشان‌دهنده تضاد آشکار میان اهداف حفاظت از حقوق بشر و اعمال واقعی قانون است].

در همین حال، هر دو قطعنامه‌ی مجمع عمومی سازمان ملل متحد^{۱۱} که به ترتیب توسط ایالات متحده و چین هدایت شدند، به صراحت صرفاً به «حوزه غیرنظامی» پرداخته‌اند؛

قطعنامه چین همچنین تصریح می‌کند که این سند «به توسعه یا استفاده از هوش مصنوعی برای اهداف نظامی نمی‌پردازد»؛ و هیچ‌یک از این دو قطعنامه حتی اشاره‌ای به هزینه‌های زیست‌محیطی هوش مصنوعی نکرده‌اند. [و این نکته نشان‌دهنده محدودیت گستره تعهدات بین‌المللی و کمبود نگرش جامع به تأثیرات اجتماعی و زیست‌محیطی است].

مرکز اروپایی حقوق سازمان‌های غیرانتفاعی^{۱۲} یادآور می‌شود که در جریان مذاکرات معاهده‌ی شورای اروپا، درخواست‌ها برای پایبندی به تعهدات حمایت افقی از حقوق بشر بدون اعمال معافیت‌های فراگیر، پرهیز از استانداردهای دوگانه میان بخش عمومی و خصوصی، و جلوگیری از تعهدات مبهم که حقوق را به اصول کلی تقلیل می‌دهند، نادیده گرفته شد.

این استاندارد دوگانه به‌ویژه زمانی روشن می‌شود که اسرائیل معاهده [مزبور] را امضا کرده است، در حالی که هم‌زمان از سامانه‌های هوش مصنوعی برای هدف‌گیری و تولید فهرست‌های کشتار جمعی در غزه استفاده می‌کند. همانطور که مونا شتایا از [شبکه جهانی] Digital Action به درستی اشاره می‌کند، این تناقض بین تعهدات روی کاغذ و واقعیت‌های میدانی نه تنها «یک ریاکاری نگران‌کننده را آشکار می‌کند»، بلکه «سوالات جدی در مورد تعهد جامعه بین‌المللی به پاسخگویی واقعی» را نمودار می‌سازد.

علاوه بر معافیت‌های گسترده در حوزه‌ی امنیت ملی و دفاعی، شاهد فشارهای گسترده‌ی صنایع مختلف و لابی‌های صنعتی برای تضعیف تعهدات قانونی و گسترش معافیت‌ها بوده‌ایم. برخی حتی خواستار خارج کردن خودکار بازیگران بخش خصوصی از حوزه‌ی شمول معاهده‌ی شورای اروپا شده‌اند.

نوآوری، رقابت‌پذیری ملی و سایر مفاهیم تفسیر بردار به عنوان مفاهیمی مطرح شده‌اند که گویا از تلاش‌ها برای حفاظت از حقوق مردم در معرض تهدید قرار گرفته‌اند. این نگرش به «مدل‌های بزرگ‌تر و هوش مصنوعی بیشتر در همه جا به هر قیمتی» تبدیل شده است - حتی زمانی که هزینه‌های انسانی، طبیعی یا زیست‌محیطی در بر داشته باشند.

^{۱۱} UN General Assembly Resolutions

^{۱۲} The European Center for Not-for-Profit Law (ECNL)

یک تهدید پیچیده‌تر و اضافی برای حکمرانی مبتنی بر حقوق بشر، نفوذ فزاینده افراد و سازمان‌های همسو با جریان [مدعی] خیرخواهی مؤثر^{۱۳} است که [اغلب] مورد حمایت میلیاردرها هستند. هر کسی که در بحث‌های مربوط به سیاست‌های هوش مصنوعی شرکت می‌کند، متوجه رشد قارچ گونه سازمان‌های مدافع، اغلب با کلمه "آینده" در نام خود، شده است که یا بر به اصطلاح "ایمنی هوش مصنوعی" یا بر خطرات "بلندمدت" و "وجودی" هوش مصنوعی متمرکز هستند. در حاشیه، برخی از این سازمان‌ها با فلسفه‌های نژادپرستانه و اصلاح نژادی مرتبط شده‌اند، اما حتی میانه‌روترین آنها نیز رویکردی کاملاً فایده‌گرایانه به حکمرانی هوش مصنوعی دارند؛ رویکردی که نه تنها اساساً با حقوق بشر همسو نیست، بلکه با تمرکز توجه عمومی بر خطرات حدسی، حواس را از مضرات واقعی هوش مصنوعی پرت می‌کند.

تمرکز بیشتر بر تأثیرات منفی هوش مصنوعی نیز به دلیل تمایل قابل درک برای اطمینان از توزیع مساوی مزایای آن، از بین می‌رود. اکثر منابع، قدرت و دانش فنی هوش مصنوعی در اختیار چند کشور و شرکت بزرگ قرار دارد؛ به گونه‌ای که تخمین زده می‌شود آمریکای شمالی دست کم ۴۰ درصد درآمد جهانی هوش مصنوعی را کنترل می‌کند. بنابراین، بخش عمده‌ای از بحث‌های سیاستگذاری بین‌المللی بر ظرفیت‌سازی و افزایش دسترسی به زیرساخت‌های فنی، قدرت محاسباتی، داده‌ها و استعدادها، به عنوان بخشی از تلاش‌های گسترده‌تر برای تحول دیجیتال، متمرکز است.

این عوامل در کنار هم تلاش‌ها برای اتخاذ یک رویکرد حکمرانی قوی و مبتنی بر حقوق در حوزه هوش مصنوعی را تضعیف می‌کنند، در عین حال که برای بازیگران صنعتی که ترجیح می‌دهند از مداخلات نظارتی قوی بدور باشند، مناسب [و به نفع آنها] است. حتی استانداردهای مؤثر، اما عمدتاً داوطلبانه‌ای که طی دهه‌ها توسط [کارشناسان] جامعه نرم‌افزار آزاد و متن‌باز^{۱۴} توسعه یافته‌اند، توسط شرکت‌هایی که مشتاق عرضه محصولات آزمایش نشده خود در بازار جهانی هستند، تحریف می‌شوند.

در همین حال، انرژی و تمرکز [فرایندهای نظارت و] تنظیم‌گری همچنان با ترس پراکنی در مورد خطرات علمی تخیلی که هرگز تحقق نخواهند یافت، تحلیل می‌رود، حتی در حالی که نقض حقوق بشر بسیار واقعی که توسط نظارت مبتنی بر هوش مصنوعی انجام می‌شود، تحریف زیست بوم اطلاعاتی ما و افزایش شدید خشونت مبتنی بر جنسیت آنلاین و سوءاستفاده جنسی از کودکان (CSAM) نادیده گرفته می‌شوند.

^{۱۳} Effective altruist (EA)

^{۱۴} Open Source Software Community

[توضیح: مهندسان و کارشناسان جامعه مزبور معمولاً نرم‌افزارهایی را شکل می‌دهند و سپس حق تکثیر، استفاده، مطالعه، تغییر و توزیع نرم‌افزار و کد منبع آن را به هر کسی و برای هر هدفی اعطا می‌کنند.]

چه اتفاقی باید بیفتد و چرا مهم است؟

به عبارت ساده، اگر حقوق بشر رعایت نشود، پیشرفت‌های هوش مصنوعی نمی‌تواند به نفع مردم باشد. نهادهای بین‌المللی مانند سازمان ملل متحد و کشورهای عضو آن باید هنجارهای جهانی حکمرانی هوش مصنوعی را که ریشه در موازین بین‌المللی حقوق بشر داشته باشند، شکل دهند و این کار را با اقدام حول آنچه در یافته‌های گزارش جدید هیئت مشورتی عالی رتبه سازمان ملل متحد در مورد هوش مصنوعی منعکس شده، آغاز کنند. حقوق بشر باید در مذاکرات پیرامون اسناد و مجامع بین‌المللی مانند پیمان جهانی دیجیتال جدید یا اجلاس آتی اقدام هوش مصنوعی، محور قرار گیرد و دولت‌ها باید تضمین کنند که مدافعان حقوق بشر، جامعه مدنی و نهادهای حقوقی در بحث‌های حکمرانی جهانی هوش مصنوعی جایگاهی دارند.

در سطح داخلی سازمان ملل، این نهاد باید نظارت و مراقبت مقتضی^{۱۵} رعایت حقوق بشر را هنگام خرید، پیاده‌سازی یا ادغام فناوری‌های نو و نوظهور [درون بخشهای مختلف خود] الزامی کند.

دولت‌های عضو باید از نظر مالی و فنی دفتر کمیساریای عالی حقوق بشر سازمان ملل را تقویت کنند تا [این نهاد سازمان ملل] بتواند با جامعه مدنی، کشورهای عضو و شرکت‌های توسعه‌دهنده هوش مصنوعی همکاری کند؛ از جمله ارائه خدمات مشاوره‌ای در حوزه حقوق بشر در فضای دیجیتال که در پیمان دیجیتال جهانی نیز مورد تأیید قرار گرفته است.

در بُعد بین‌المللی، سازمان ملل باید از قطعنامه‌های اجماعی که برخی کاربردهای هوش مصنوعی را با حقوق بشر ناسازگار دانسته‌اند، فراتر رود و به‌طور صریح مشخص کند که این کاربردها کدام‌اند (مواردی چون نظارت بیومتریک گسترده^{۱۶} یا رویکرد پلیس پیش‌بینی‌کننده^{۱۷}) و اقدامات قانونی و اجرایی برای ممنوعیت این کاربردها اتخاذ کند.

نقش شرکت‌ها و بخش خصوصی

شرکت‌ها نیز در این فرایند نقش [جدی] برعهده دارند. بسیاری از بزرگ‌ترین توسعه‌دهندگان و به‌کارگیرندگان هوش مصنوعی متعهد شده‌اند که اصول راهنمای سازمان ملل در زمینه کسب‌وکار و حقوق بشر^{۱۸} را رعایت کنند.

^{۱۵} Due diligence

^{۱۶} Biometric Mass Surveillance

[توضیح: نظارت بیومتریک یعنی دولت‌ها یا نهادها از فناوری‌های تشخیص هویت بیومتریک برای پایش مداوم و گسترده مردم در فضاهای عمومی استفاده کنند. بیومتریک یعنی ویژگی‌های زیستی و بدنی افراد، مثل: چهره، اثر انگشت، صدا، راه رفتن یا الگوی حرکت.]

^{۱۷} Predictive Policing

[توضیح: رویکرد پلیس پیش‌بینی‌کننده یعنی استفاده از الگوریتم‌ها و هوش مصنوعی برای پیش‌بینی این‌که جرم کجا و چه زمانی ممکن است رخ دهد یا چه افرادی احتمال دارد مرتکب جرم شوند و سپس پلیس بر اساس این پیش‌بینی‌ها اقدام کند.

اگر گفته شود که مشکل این نوع رویکرد چیست؟ پاسخش این است که می‌تواند باعث برجسب‌زنی ناعادلانه به افراد شود. معمولاً بر داده‌های قبلی متکی است و اگر آن داده‌ها تبعیض‌آمیز باشند، تبعیض را تقویت می‌کند. می‌تواند به نقض حقوق بشر و نظارت بیش از حد منجر شود.]

^{۱۸} UN Guiding Principles on Business and Human Rights – UNGPs

شرکت‌ها باید به چارچوب یادشده پیوندند به گونه ای که با نظارت و مراقبت مقتضی رعایت حقوق بشر در فعالیت های خود^{۱۹}، بتوانند خطرات مرتبط با هوش مصنوعی را به درستی شناسایی و کاهش دهند. این موضوع همچنین به معنای مشارکت دادن جامعه مدنی، مدافعان حقوق بشر و جوامع تحت تأثیر در گفت‌وگوهای مربوط به حکمرانی هوش مصنوعی از همان آغاز است، نه صرفاً به عنوان اقدامی صوری و سطحی پس از پایان کار.

[پیامدهای نادیده گرفتن حقوق بشر]

غیر قابل انکار است که استفاده از هوش مصنوعی به طور نامتناسبی به گروه‌های پیش‌تر به حاشیه رانده شده از جمله زنان، افراد غیر باینری^{۱۹} و جامعه [دگرباشان جنسی] LGBTQ+ و جوامع نژادی آسیب می‌رساند و علیه آنها تبعیض‌هایی را معمول می‌دارد.

ایجاد معافیت‌ها و اولویت دادن به سود، نقض حقوق بشر را تشدید می‌کند، پیروزی‌های به سختی به دست آمده در زمینه حریم خصوصی و حفاظت از داده‌ها را از بین می‌برد، فاجعه اقلیمی را تسریع می‌کند، نابرابری‌های اقتصادی ملی و جهانی را گسترش می‌دهد و در نهایت به هوش مصنوعی اجازه می‌دهد تا به جای سود رساندن به بشریت، بی‌عدالتی را تسریع کند.

در چنین سناریویی، بحث‌ها پیرامون طراحی، دامنه و تنظیم گری هوش مصنوعی در [انحصار] اقلیت جهانی متمرکز خواهد ماند، حتی در حالی که مردم اکثریت جهان برای ساخت و آزمایش سامانه‌های هوش مصنوعی، بدون ضمانت‌های اولیه حقوق بشر یا سازوکارهای پاسخگویی، مورد سوءاستفاده قرار می‌گیرند. این آینده واقعاً ویران‌شهری^{۲۰} است که هوش مصنوعی بدون کنترل و بدون نظارت وعده می‌دهد - و ما نمی‌توانیم اجازه دهیم این کابوس به واقعیت تبدیل شود.

^{۱۹} Non-Binary

افرادی هستند که هویت جنسیتی خود را منحصرأ یکی از دوگانه مرد یا زن نمی‌دانند.

^{۲۰} dystopian future