



Iranian Independent Islamic Human Rights Commission

کمیسیون حقوق بشر اسلامی ایران
انجمن حقوق انسان اسلامیة فی ایران

Artificial Intelligence and Human Rights

Recommendations for companies



Global Compact
Network Germany



HUMAN RIGHTS &
LABOUR STANDARDS



هوش مصنوعی و حقوق بشر

توصیه‌هایی برای شرکت‌ها

امور پژوهشی، آموزشی و ترویجی کمیسیون حقوق بشر اسلامی ایران

پاییز ۱۴۰۴

توضیحات مقدماتی کمیسیون:

نظر به گسترش فناوریهای نوین و پیامدهای آنها بر ابعاد مختلف حقوق بشر، مراکز و نهادهای متعدد ملی و بین المللی در تلاشند تا از سرعت تحولات علمی و فنی عقب نمانده و قاعده گذاریهای حقوقی مناسبی را در این زمینه فراهم آورند. علیرغم همه تلاشها، روند قاعده گذاری حقوقی یا اخلاقی از شتاب تحولات فناورانه بسیار عقب تر است و به همین علت دغدغه های ناظر به نقض حقوق بشر یا بی توجهی به اصول اخلاقی اهمیت جدی تری پیدا می کند. در این حوزه نه تنها دولتها باید تعهدات خود را دقیق بشناسند و بدان عمل کنند بلکه شرکت های خصوصی و فناورانه که خود بانی تحولات فنی هستند نیز باید به دقت از اصول و معیارهای حقوقی و اخلاقی مطلع شده و نسبت به اجرای آن اهتمام لازم را معمول دارند.

در سطح اروپا، آمریکا و بعضا دیگر مناطق جهان متون متعددی تهیه و منتشر شده با این هدف که شرکت ها و بازیگران خصوصی را آموزش دهند که الزامات حقوق بشری را رعایت نمایند. این متون نه تنها برای شرکت های فعال در کشورهای محل فعالیت بلکه برای همه شرکت های فعال در این موضوع در سراسر جهان کارایی دارد. نوشتاری که هم اینک در معرض استفاده مخاطبان قرار گرفته و از انگلیسی به فارسی برگردان شده است، یکی از متونی است که برای راهنمایی شرکت ها تهیه شده است. امید است مخاطبان فارسی زبان اعم از سیاست گذاران، فعالان در شرکت ها و پژوهشگران و عموم بهره برداران استفاده لازم را از این متن ببرند و با ارتقای آگاهی ها و حساسیت های خود، نسبت به رعایت حقوق انسانی در مجموعه فعالیت های خویش در حوزه هوش مصنوعی بیش از پیش نقش آفرینی مثبتی را نشان دهند. یادآوری می نماید که در متن نوشتار، علاوه بر ارجاعات اعلامی تهیه کنندگان متن که در زیر نویس منعکس شده، مواردی نیز در مرحله نهایی سازی متن فارسی به عنوان توضیح یا معادل انگلیسی واژگان فارسی در زیر نویس هر صفحه افزوده شد تا امکان استفاده بهتری برای مخاطبان فراهم شود.

امور پژوهشی، آموزشی و ترویجی کمیسیون حقوق بشر اسلامی ایران

آذر ماه ۱۴۰۴

◆ نویسنده: دکتر الکساندر کریبیتز و دکتر رافائل مکس

◆ متصدی روند برگردان انگلیسی به فارسی: مهندس مریم صداقتی مدرس دانشگاه در رشته علوم کامپیوتر و همیار

کمیسیون، ویرایش و نهایی سازی متن در امور پژوهشی کمیسیون، آذر ۱۴۰۴

◆ مآخذ: شبکه آلمانی پیمان جهانی سازمان ملل (UN Global Compact Network Germany) آوریل

۲۰۲۴

◆ شناسه متن: ۰۴/۳/۳۰۸

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.



چکیده

چارچوب‌های مختلفی، از جمله « پیمان جهانی سازمان ملل متحد^۱ » و «اصول راهنمای سازمان ملل درباره‌ی کسب و کار و حقوق بشر^۲»، پیش از این تعهدات عمومی مربوط به بررسی‌های لازم در مورد حقوق بشر شرکت‌ها را در بر گرفته‌اند. با این حال، توسعه سریع هوش مصنوعی در تجارت و جامعه، سوالات اخلاقی و نظارتی جدیدی را مطرح می‌کند. نکته قابل توجه این است که قانون هوش مصنوعی اتحادیه اروپا^۳، چالشی را برای شرکت‌های تجاری ایجاد می‌کند، زیرا این قانون آنها را ملزم می‌کند که هنگام اجرای سیستم‌های هوش مصنوعی، خطرات مربوط به حقوق اساسی و حقوق بشر را در نظر بگیرند.

بنابراین، این اثر منتشره به بررسی این موضوع می‌پردازد که چگونه توسعه و پیاده‌سازی هوش مصنوعی بر تعهدات مربوط به [اصل] مراقبت و نظارت مقتضی برای رعایت حقوق بشر^۴ در شرکت‌ها تأثیر خواهد گذاشت. قانون هوش مصنوعی، در هماهنگی با سایر تحولات بین‌المللی در راستای هوش مصنوعی و حقوق بشر، مبنای اولیه‌ای را برای ارزیابی خطرات حقوق بشر مرتبط با هوش مصنوعی فراهم می‌کند.

با این حال، با گسترش جهانی هوش مصنوعی، خطرات حقوق بشری، به‌ویژه در مورد سوءاستفاده از راهکارهای هوش مصنوعی توسط اشخاص ثالث همچنان پابرجا خواهند بود. علاوه بر این، وجود چارچوب‌های حقوقی متفاوت در کشورهای مختلف نیز عاملی پیچیده محسوب می‌شود.

برای محافظت از حقوق بشر در کل زنجیره تأمین و ارزش^۵، شرکت‌ها باید به طور فوری خطرات خاص [ناشی از] هوش مصنوعی را در فرآیند « مراقبت و نظارت مقتضی [برای رعایت حقوق بشر] » خود بگنجانند. همچنین باید اقدامات مناسبی را برای کاهش پیشگیرانه خطرات متوجه حقوق بشر انجام و تدابیر تعدیل‌کننده مناسب را [در فعالیت‌های خود] بسط دهند.

^۱ The UN Global Compact

توضیح: پیمان جهانی سازمان ملل متحد یک ابتکار برای جلب مشارکت داوطلبانه فعالان تجاری است که توسط سازمان ملل متحد راه‌اندازی شده و شرکت‌ها را تشویق می‌کند تا شیوه‌ها و راهبردهای تجاری خود را با ده اصل اعلام شده در زمینه‌های حقوق بشر، کار، محیط زیست و مبارزه با فساد هماهنگ کنند. یکی از محورهای موضوعی پیمان جهانی سازمان ملل متحد، جلوگیری از کمک و تشویق شرکت‌ها به نقض حقوق بشر است.

^۲ UN Guiding Principles on Business and Human Rights

^۳ EU AI Act

^۴ Human rights due diligence obligations

^۵ Supply and value chain

فهرست

۷	مقدمه
۸	هوش مصنوعی و حقوق بشر
۸	تعریف و کاربرد هوش مصنوعی
۹	خطرات کلی هوش مصنوعی
۱۱	انواع خطرات متوجه حقوق بشر در پیاده‌سازی هوش مصنوعی
۱۱	خطر نوع اول: سوءاستفاده از هوش مصنوعی برای نقض حقوق بشر
۱۲	خطر نوع دوم: عدم توجه کافی به حقوق بشر در طراحی راهکارهای هوش مصنوعی
۱۳	خطر نوع سوم: تأثیرات نامطلوب حقوق بشری ناشی از استفاده از هوش مصنوعی
۱۴	خطرات خاص کاربردی
۱۵	مطالعه موردی ۱: استفاده از هوش مصنوعی در پزشکی
۱۵	مطالعه موردی ۲: استفاده از هوش مصنوعی در منابع انسانی
۱۶	الزامات بررسی صلاحیت شرکت‌ها برای استفاده از هوش مصنوعی
۱۷	ارزیابی ریسک‌های متوجه حقوق بشر در سامانه‌های هوش مصنوعی
۱۷	ارزیابی خطر سوء کاربرد
۱۷	ارزیابی خطر عدم توجه کافی به حقوق بشر در طراحی راهکارهای هوش مصنوعی
۱۸	ارزیابی خطر اثرات منفی بر حقوق بشر
۱۸	اقدامات کاهش خطر و ارتباطات خارجی
۱۹	پیشگیری از نقض استانداردهای حداقلی حقوق بشر
۲۰	ارتباطات بیرونی
۲۰	خلاصه و چشم‌انداز
۲۲	برخی از ادبیات تخصصی برای مطالعات بیشتر
۲۲	قوانین و ابتکارات حقوقی درباره‌ی هوش مصنوعی
۲۲	چارچوب‌ها و بیانیه‌های بین‌المللی
۲۳	نهادهای و مراکز اطلاعاتی

۲۴

منابع مرتبط و شناخت رویه های خوب

۲۵

توضیح کوتاه در مورد اصول راهنمای سازمان ملل متحد در مورد کسب و کار و حقوق بشر



مقدمه

سال‌های اخیر به دلیل بهبود دسترسی به داده‌ها و افزایش عملکرد سخت‌افزار و مدل‌های هوش مصنوعی، شاهد توسعه سریع قابلیت‌های سامانه‌های هوش مصنوعی بوده‌ایم. به طور خاص، «انتشار مدل‌های زبان بزرگ»^۶، از جمله Chat GPT و Google Bard، بحث‌های اجتماعی گسترده‌ای را در مورد خطرات اخلاقی و حقوق بشری هوش مصنوعی برانگیخته است.^۷

توسعه و کاربرد هوش مصنوعی در حال حاضر یک تحول اجتماعی بزرگ را تشکیل می‌دهد. این امر به ویژه بر بخش‌های آموزش، تأمین انرژی، بهداشت و حمل و نقل و همچنین بر اجرای قانون و «نظام عدالت کیفری»^۸ تأثیر می‌گذارد.^۹ با این حال، در برخی موارد، فناوری‌های مورد استفاده موضوع بحث‌های اخلاقی و اجتماعی جدی هستند. به عنوان مثال می‌توان به توسعه نرم‌افزار تشخیص چهره، استفاده از دروغ‌سنج مبتنی بر هوش مصنوعی و تجزیه و تحلیل خودکار داده‌های شخصی یا بهداشتی اشاره کرد.^{۱۰} کسب‌وکارها در مرکز این بحث قرار دارند زیرا نه تنها در توسعه این ابزارهای هوش مصنوعی، بلکه در اجرای آنها نقش دارند. سایر حوزه‌های مرتبط شامل تجزیه و تحلیل داده‌های مشتری، ارزیابی چکیده سوابق افراد^{۱۱}، استفاده از هوش مصنوعی در تولید، کاربرد سامانه‌های هوش مصنوعی برای ایمنی کارگاه^{۱۲} و ارزیابی الگوریتمی مبتنی بر تمایل مصرف‌کنندگان به پرداخت^{۱۳} است.

بنابراین، کسب‌وکارهایی که از هوش مصنوعی بهره می‌گیرند، به طور فزاینده‌ای باید به جنبه‌های حقوق بشری این فناوری توجه کنند. این امر ناشی از پذیرش روزافزون استانداردهای بین‌المللی برای کسب‌وکار و حقوق بشر است.

^۶ Large Language Model

^۷ Li, H., Moon, J. T., Purkayastha, S., Celi, L. A., Trivedi, H., & Gichoya, J. W. (۲۰۲۳). Ethics of large language models in medicine and medical research. *The Lancet Digital Health*, ۵(۶), e۳۳۳–e۳۳۵.

^۸ Criminal Justice System

^۹ White Paper on Artificial Intelligence—a European approach to excellence and trust COM/۲۰۲۰/۶۵ final, p. ۱ & p. ۱۱. ‘Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union acts.’ ۲۰۲۱/۰۱۰۶(COD).

^{۱۰} Oravec, J. A. (۲۰۲۲), The emergence of “truth machines”? Artificial intelligence approaches to lie detection. *Ethics and Information Technology*, ۲۴(۱), p. ۶.

^{۱۱} CVs

^{۱۲} توضیح: واژه PLANT که از ایمنی آن در متن صحبت شده چند معنا دارد از جمله کارگاه یا گیاه که در این معنای اخیر منظور کاربرد هوش مصنوعی برای اهداف ایمنی محصولات کشاورزی خواهد بود و اتفاقاً هوش مصنوعی در صنایع غذایی و گیاهی و سایر حوزه‌های صنایع کاربرد روز افزون دارد.

^{۱۳} Consumers' willingness to pay

که از جمله می‌توان به پیمان جهانی سازمان ملل و اصول راهنمای سازمان ملل در مورد کسب و کار و حقوق بشر اشاره کرد. همچنین به دلیل فرایند های قاعده گذاری حقوقی سخت گیرانه تر درباره هوش مصنوعی است. این گزارش، بر اساس چارچوب های حقوق بشر، مقدمه ای بر موضوع هوش مصنوعی و حقوق بشر ارائه می‌دهد و توصیه های اولیه ای را برای شرکت هایی که مایل به بررسی دقیق تر آن هستند، ارائه می‌کند.

هوش مصنوعی و حقوق بشر

چارچوب های بین المللی برای حمایت از حقوق بشر، شرکت ها را موظف می‌سازد تا فرآیندهای خود را از نظر وجود خطرات احتمالی برای حقوق بشر مورد بازبینی قرار دهند. در این زمینه، بسیاری از شرکت ها این پرسش را مطرح می‌کنند که آیا توسعه و به کارگیری هوش مصنوعی می‌تواند خطرات حقوق بشری ایجاد کند و اگر پاسخ مثبت است، تحت چه شرایطی رخ خواهد داد. خطرات خاص مرتبط با استفاده شرکت ها از هوش مصنوعی در قانون «هوش مصنوعی اتحادیه اروپا»^{۱۴}، کنوانسیون پیشنهادی شورای اروپا در مورد هوش مصنوعی، حقوق بشر، دموکراسی و حاکمیت قانون و همچنین قطعنامه های اخیر مجمع عمومی سازمان ملل متحد^{۱۵} مورد بررسی قرار گرفته است.

با این حال، پرداختن به این موضوع مستلزم درک اساسی تری از فناوری و بررسی چالش های عمومی و خاص هر بخش مرتبط با پیاده سازی هوش مصنوعی است.

تعریف و کاربرد هوش مصنوعی

در حال حاضر، تعریف واحد و استاندارد از هوش مصنوعی در ادبیات علمی وجود ندارد. این اصطلاح معمولاً به توانایی سامانه ها یا الگوریتم های کامپیوتری برای شبیه سازی رفتار هوشمند انسانی اشاره دارد. رانندگی با وسیله نقلیه، تشخیص تصاویر یا نوشتن متن، مثال هایی از این رفتارها است.^{۱۶} ماده ۳(۱) قانون هوش مصنوعی اتحادیه اروپا، خروجی های هوش مصنوعی را به عنوان «خروجی هایی مانند محتوا،

^{۱۴} توضیح: قانون هوش مصنوعی، استفاده، استقرار و توسعه هوش مصنوعی را در اتحادیه اروپا تنظیم می‌کند. رویکرد مبتنی بر ریسک، اقدامات احتیاطی ایمنی ویژه ای را برای سیستم های پرخطر تعریف می‌کند و شیوه های مرتبط با هوش مصنوعی را که با حقوق اساسی در تضاد هستند، ممنوع می‌کند.

^{۱۵} توضیح: چند سازوکار تابعه سازمان ملل قطعنامه هایی را در موضوع فناوریهای دیجیتال و حقوق بشر تصویب کرده اند که مباحث هوش مصنوعی نیز در آنها مورد اشاره قرار گرفته است. هم چنین در سال ۲۰۲۴ برای اولین بار مجمع عمومی ملل متحد قطعنامه مستقلی در مورد هوش مصنوعی ایمن برای توسعه پایدار تصویب کرد که نکات مهمی در حوزه رعایت حقوق بشر در آن درج شده است.

^{۱۶} MerriamWebster.com Dictionary, MerriamWebster, <https://www.merriamwebster.com/dictionary/artificial%20intelligence>. See also Alan Turing. Computing machinery and intelligence. S.I.: Mind, ۱۹۵۰;

Gablers Wirtschaftslexikon: <https://wirtschaftslexikon.gabler.de/definition/kuenstliche-intelligenz-40285> (in German).

پیش‌بینی‌ها، توصیه‌ها یا تصمیماتی که بر محیط‌هایی که با آنها تعامل دارند تأثیر می‌گذارند» توصیف می‌کند.^{۱۷}

توسعه هوش مصنوعی مبتنی بر رویکردها و مفهومی مختلف است.^{۱۸} در حال حاضر تمرکز ویژه‌ای بر «یادگیری ماشین»^{۱۹} و دو شکل مهم آن یادگیری عمیق^{۲۰} و یادگیری تقویتی^{۲۱} وجود دارد. هر دو رویکرد، سامانه‌های هوش مصنوعی را قادر می‌سازند تا به صورت خودکار یاد بگیرند و بر اساس تجربه و بازخورد از محیط، تصمیمات مستقلی بگیرند.

یادگیری عمیق در تشخیص تصویر و اشیاء بسیار مورد استفاده قرار می‌گیرد. یکی از کاربردهای خاص آن الگوریتم‌هایی در پزشکی است که تصاویر اشعه ایکس یا سی‌تی‌اسکن را برای یافتن الگوهای مشکوک، مانند متاستازهای سرطانی، تجزیه و تحلیل می‌کنند.

یادگیری تقویتی به سامانه‌های هوش مصنوعی اشاره دارد که راهبردهای حل مسئله مستقلی را توسعه می‌دهند. این فرآیند معمولاً مبتنی بر سیستم پاداش یا مجازات است و در رانندگی خودکار و رباتیک بسیار مورد استفاده قرار می‌گیرد.

این مثال‌ها نشان می‌دهند که هوش مصنوعی می‌تواند در حوزه‌های کاملاً متفاوتی از زندگی مورد استفاده قرار گیرد. از جمله این موارد می‌توان به سیستم‌های زبانی، ربات‌ها و ابزارهای تحلیلی اشاره کرد. همچنین کاربردهای نظامی، اطلاعاتی و پلیسی برای این فناوری وجود دارد. این امر هوش مصنوعی را به یک فناوری با کاربرد دوگانه تبدیل می‌کند که ممکن است اغلب به روش‌های غیرقابل پیش‌بینی به کار گرفته شود. یک مثال، استفاده از برنامه‌های تحلیل متن یا ارزیابی داده‌های روان‌سنجی توسط مجریان قانون است.

خطرات کلی هوش مصنوعی

خطرات [متوجه] حقوق بشر ناشی از توسعه یا استفاده از هوش مصنوعی، ارتباط نزدیکی با حوزه‌ی کاربرد آن (مانند پزشکی یا امور مالی) دارد. با این حال، برخی ویژگی‌های ذاتی سامانه‌های هوش مصنوعی وجود دارد که صرف‌نظر از نوع کاربردها، می‌تواند منجر به خطرات [متوجه] حقوق بشر شود.

^{۱۷} به قانون هوش مصنوعی مراجعه کنید. ماده ۳؛ همچنین به سازمان همکاری اقتصادی و توسعه (OECD)، «توصیه شورای هوش مصنوعی» ۲۰۱۹:

(<https://legalinstruments.oecd.org/en/instruments/>) مراجعه کنید.

^{۱۸} به پیوست «تکنیک‌ها و رویکردهای هوش مصنوعی» در قانون هوش مصنوعی [اروپا] مراجعه کنید.

^{۱۹} machine learning (ML)

^{۲۰} Deep Learning

^{۲۱} Reinforcement Learning

یکی از ویژگی‌های کلیدی هوش مصنوعی، استفاده هم‌زمان حجم عظیمی از داده‌ها و الگوریتم‌های یادگیرنده‌ی پیچیده است.^{۲۲} این امر چالش‌های متعددی در زمینه‌ی حقوق بشر ایجاد می‌کند زیرا استفاده از داده‌های پزشکی، بیومتریک و اطلاعات استخراج‌شده از شبکه‌های اجتماعی، مسائل حساس حریم خصوصی و حفاظت از داده‌ها را دربردارد.

برای مثال، سامانه‌هایی که قادرند افراد را از طریق صدای آنان شناسایی کنند یا تحلیل‌های روان‌شناختی دقیقی از شخصیت افراد ارائه دهند، به‌شدت به حریم خصوصی آسیب می‌زنند.

علاوه بر این، هوش مصنوعی می‌تواند به نتایجی برسد که در طول مرحله توسعه نمی‌توان جزئیات آنها را پیش‌بینی کرد. دلیل این امر آن است که هر داده جدیدی که یک سامانه هوش مصنوعی دریافت می‌کند، ممکن است منجر به تغییر مدل درونی آن گردد. ابهام یا عدم قابلیت توضیح و تبیین^{۲۳}، درک تصمیمات گرفته شده توسط هوش مصنوعی را برای افراد دشوار می‌کند.^{۲۴} در عین حال، عدم نمایش مجموعه داده‌ها یا ویژگی‌های مدل هوش مصنوعی می‌تواند باعث خطاهای ناخواسته‌ای شود که پیامدهای قابل توجهی برای حقوق بشر به همراه دارد. از آنجایی که داده‌های با کیفیت کافی اغلب برای گروه‌های خاصی در دسترس نیست - یا اصلاً در دسترس نیست - ورودی ضعیف داده‌ها می‌تواند منجر به رفتار نابرابر با افراد شود. در گذشته، «سوگیری‌های الگوریتمی»^{۲۵} به ویژه اغلب بر گروه‌های آسیب‌پذیر (مانند افراد با پیشینه مهاجرت) تأثیر گذاشته است. این می‌تواند بدان معنا باشد که گروه‌های جمعیتی خاصی نسبت به جمعیت عمومی، قرار ملاقات‌های دیرتری برای درمان در بیمارستان دریافت می‌کنند یا مجبور به پرداخت قیمت‌های بالاتر برای بلیط هواپیما می‌شوند. یکی دیگر از جنبه‌های پیاده‌سازی هوش مصنوعی، گرایش به سمت استانداردسازی است. به عنوان یک قاعده، هوش مصنوعی جایگزین یک فرد خاص، مانند پزشکی که بیماری را تشخیص می‌دهد، نمی‌شود. در عوض، کل زیست بوم ها، مانند فرآیندهای یک بیمارستان، یا کل رشته‌ها، به عنوان نمونه تشخیص سرطان، را تغییر می‌دهد.^{۲۶}

^{۲۲} DKE German Commission for Electrical, Electronic & Information Technologies of DIN and VDE: White paper on Ethik und Künstliche Intelligenz ('Ethics and Artificial Intelligence', in German), p. ۹.

^{۲۳} opacity or lack of explainability

توضیح: فقدان قابلیت توضیح و تبیین به چالش درک و تفسیر فرآیندهای تصمیم‌گیری سیستم‌های هوش مصنوعی اشاره دارد

^{۲۴} RAI, Arun. Explainable AI: From black box to glass box. Journal of the Academy of Marketing Science, ۲۰۲۰, ۴۸, pp. ۱۳۷-۱۴۱. Virginia Dignum, Responsible Artificial Intelligence (Springer Nature Switzerland, ۲۰۱۹). Tan Ming, J. (۲۰۲۲). Nondeterministic artificial intelligence systems and the future of the law on unilateral mistakes in Singapore. Singapore Academy of Law Journal, ۳۴(۱), pp. ۹۱-۱۲۴.

^{۲۵} Algorithmic Biases

^{۲۶} Samorani, M., Harris, S. L., Blount, L. G., Lu, H., & Santoro, M. A. (۲۰۲۲). Overbooked and overlooked: Machine learning and racial bias in medical appointment scheduling. Manufacturing & Service Operations Management, ۲۴(۶), pp. ۲۸۲۵-۲۸۴۲. <https://doi.org/10.1287/msom.2021.0999>.

در یادداشت توضیحی در مورد قانون هوش مصنوعی اتحادیه اروپا، مشکلات اساسی هوش مصنوعی را با استفاده از کلمات کلیدی ابهام^{۲۷}، پیچیدگی^{۲۸}، وابستگی به داده‌ها^{۲۹} و رفتار خودمختار^{۳۰} خلاصه می‌کند. هر چهار جنبه می‌توانند توسط انسان‌ها تحت تأثیر قرار گیرند، به عنوان مثال از طریق انتخاب داده‌های خاص برای ورودی، میزان خودمختاری افراد تحت تأثیر در فرآیندهای تصمیم‌گیری و میزان استانداردسازی یا شخصی‌سازی تصمیمات.

انواع خطرات متوجه حقوق بشر در پیاده‌سازی هوش مصنوعی

با توجه به تأثیر قابل توجه توسعه‌دهندگان (Developers)^{۳۱} بر طراحی و پیاده‌سازی سامانه‌های هوش مصنوعی، چالش‌های مطرح شده در بخش پیشین، منجر به بروز سه نوع خطر [علیه] حقوق بشر در چرخه‌ی عمر این فناوری می‌شود.

خطر نوع اول: سوءاستفاده از هوش مصنوعی برای نقض حقوق بشر

شرکت‌هایی که هوش مصنوعی را توسعه می‌دهند یا از آن استفاده می‌کنند، ممکن است به طور غیرمستقیم در نقض حقوق بشر (مثلاً محدود کردن آزادی بیان، تبعیض علیه اقلیت‌ها) همدست باشند، چه از طریق همکاری‌های بین‌المللی و چه از طریق ارائه فناوری، داده یا راه‌حل‌های نرم‌افزاری به بازیگران خصوصی و دولتی. شرکت‌هایی که سامانه‌های هوش مصنوعی را طراحی یا عرضه می‌کنند، ممکن است به صورت غیرمستقیم در نقض حقوق بشر (مثلاً محدود کردن آزادی بیان، تبعیض علیه اقلیت‌ها) مشارکت داشته باشند. این خطر ممکن است از طریق همکاری بین‌المللی یا ارائه فناوری، داده یا نرم‌افزار به نهادهای خصوصی یا دولتی پدید آید.^{۳۲}

یکی از سوابق شناخته‌شده [در این خصوص]، مربوط به منطقه خودمختار اویغور در سین‌کیانگ غرب چین^{۳۳} است. در چارچوب نقض‌های همه‌جانبه حقوق بشر توسط دولت چین و مدیریت منطقه‌ای، [در منطقه یاد شده] سامانه‌های هوش مصنوعی برای رصد و نظارت و کنترل هدفمند اقلیت‌های قومی بکار گرفته می‌شوند.^{۳۴} گزارش‌های

^{۲۷} Opacity

^{۲۸} Complexity

^{۲۹} Dependence on Data

^{۳۰} Autonomous Behaviour

^{۳۱} توضیح: دولوپر مسئول ساخت و توسعه‌ی خدمات یا محصول نرم‌افزاری می‌باشد

^{۳۲} UN General Assembly. (۲۰۱۸). Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. A/HRC/۳۸/۳۵۲.

^{۳۳} Xinjiang Uyghur Autonomous Region

^{۳۴} StarkTu, R. (۲۰۲۱). China's use of artificial intelligence in their war against Xinjiang. Tul. J. Int'l & Comp. L., ۲۹(۱۵۳).

سازمان‌های بین‌المللی، از جمله سازمان ملل متحد، استفاده گسترده از هوش مصنوعی برای شناسایی زیستی (بیومتریک) اویغورها و سایر اقلیت‌های قومی را مورد انتقاد قرار می‌دهند.^{۳۵}

مشکل اینجا استفاده از سیستم‌های هوش مصنوعی خاصی که ممنوع یا فی‌نفسه غیراخلاقی هستند، نیست، بلکه زمینه‌ای است که در آن از آنها استفاده می‌شود. بدین صورت که هنگام همکاری با سازمان‌های امنیتی در رژیم‌های اقتدارگرا، شرکت‌ها [در بخش خصوصی] می‌توانند در این جرم همدست باشند. با این حال، خطراتی از این نوع نه تنها شامل همکاری با دولت‌ها می‌شود، بلکه با راه‌حل‌های هوش مصنوعی که پروفایل‌های شخصیتی جامعی از کارمندان برای شرکت‌ها ایجاد می‌کنند یا با الگوریتم‌هایی که هدفشان دستکاری [ذهن و رفتار] مصرف‌کنندگان یا کاربران شبکه‌های اجتماعی است، نیز رخ می‌دهد.

خطرات سوءاستفاده از سامانه‌های هوش مصنوعی که در بالا توضیح داده شد، موضوع بحث‌های گسترده به عنوان بخشی از روند قانونگذاری فعلی است. به عنوان مثال، قانون هوش مصنوعی اتحادیه اروپا شامل فهرستی از اقدامات ممنوع هوش مصنوعی است که صرفاً به دلیل هدفشان، مانند دستکاری [ذهنی و رفتاری] افراد، بر حقوق بشر تأثیر منفی می‌گذارند. انتقال سامانه‌های هوش مصنوعی به اشخاص ثالث که به دلیل شرایط استفاده، پیامدهای نامطلوبی برای حقوق بشر دارند، هنوز در این قانون مورد توجه قرار نگرفته است و بنابراین ریسک بالایی را به همراه دارد.

خطر نوع دوم: عدم توجه کافی به حقوق بشر در طراحی راهکارهای هوش مصنوعی

واگذاری تصمیم‌گیری از انسان‌ها به سامانه‌های هوش مصنوعی می‌تواند از نظر حقوق بشر و اصول قانون اساسی یک چالش اساسی ایجاد کند. این امر به ویژه مربوط به حقوقی است که از مشارکت و رضایت افراد و ذینفعان در فرآیندهای سیاسی، اجتماعی و اقتصادی، مانند شورای کار یک شرکت، محافظت می‌کند.^{۳۶} اگر سامانه‌های هوش مصنوعی بدون نظارت انسانی و بدون رضایت یا آگاهی افراد تحت تأثیر استفاده شوند، یا اگر به اندازه کافی شفاف نباشند، ممکن است این حقوق نقض شوند. این امر در صورتی رخ می‌دهد که یک شرکت از یک سامانه هوش مصنوعی برای ارزیابی کارمندان بدون مشارکت نماینده کارگران استفاده کند. همچنین در زمینه‌هایی که با عدم

^{۳۵} Smith, M. & Miller, S. (۲۰۲۲). The ethical application of biometric facial recognition technology. AI & Society, p. ۱.

^{۳۶} WHITE PAPER On Artificial Intelligence—A European approach to excellence and trust COM (۲۰۲۰) ۶۵ final, p. ۱۴.

تقارن قدرت مشخص می‌شوند، نیاز فوری به اقدام وجود دارد. این امر به ویژه در مورد استفاده از هوش مصنوعی در زمینه‌های حیاتی مانند مدیریت عمومی، سیستم دادگاه یا قوه قضائیه صدق می‌کند.

در اینجا، ارائه توضیح کافی به افراد آسیب‌دیده راجع به اینکه اگر تصمیمی اتخاذ شود که به نفع آنها نباشد، از نظر قانونی قابل اعتراض خواهد بود، بسیار مهم است.^{۳۷} «حق بهره‌مندی از مدیریت خوب» (ماده ۴۱ منشور حقوق اساسی اتحادیه اروپا)^{۳۸} و حقوق قضایی مندرج در فصل ششم [سند یاد شده]، زیربنای این حق هستند.

به طور مشخص، این بدان معناست که شرکت‌هایی که هوش مصنوعی را توسعه داده و از آن استفاده می‌کنند، باید راهبردهایی را تدوین کنند تا اطلاعات کافی در مورد نحوه کار سامانه‌های هوش مصنوعی و خطرات ناشی از آن را در اختیار افراد تحت تأثیر قرار دهند. آنها همچنین باید به حقوق افراد و نهادها در مورد رضایت و تصمیم‌گیری مشترک، به ویژه هنگامی که تصمیمات به صورت خودکار گرفته می‌شوند، احترام بگذارند. اگرچه قانون هوش مصنوعی اتحادیه اروپا از قبل تعهدات اولیه اطلاعاتی را فراهم کرده است، اما [وضعیت] اجرا همچنان وابسته به شرایط است. این قانون ممکن است هنوز نیاز به توجه ویژه به گروه‌های آسیب‌پذیر، مانند افراد دارای معلولیت، داشته باشد.

خطر نوع سوم: تأثیرات نامطلوب حقوق بشری ناشی از استفاده از هوش مصنوعی

استفاده از هوش مصنوعی زمانی اثرات نامطلوبی ایجاد می‌کند که سامانه‌های هوش مصنوعی نرخ خطای بالایی داشته باشند یا بین گروه‌های مختلف تبعیض قائل شوند و بنابراین مانع از بهره‌مندی افراد از حقوق خود (مثلاً دسترسی به خدمات عمومی) یا آزادی‌های مدنی (مثلاً حق آزادی بیان) شوند.^{۳۹}

از آنجایی که برخی از این تأثیرات نامطلوب با دسته‌های محافظت‌شده مانند سن، رنگ پوست، قومیت، جنسیت یا مذهب مرتبط هستند، استفاده از هوش مصنوعی می‌تواند منجر به تبعیض ساختاری یا ترجیح بی‌مورد برای گروه‌های

^{۳۷} Opinion of the Data Ethics Commission. p. ۱۷۴; EU AI Act, Article v. European Parliament – Draft report on artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters, ۲۰۲۰/۲۰۱۶ (INI); Cf. also: Wolswinkel, Johan (December ۲۰۲۲). Artificial Intelligence and Administrative Law. Publication of the Council of Europe. <https://www.coe.int/en/web/cdcj/newreportontheinterplaybetweenartificialintelligenceandadministrativelaw>.

^{۳۸} Charter of Fundamental Rights of the European Union

^{۳۹} Obermeyer, Ziad, et al. (۲۰۱۹). Dissecting racial bias in an algorithm used to manage the health of populations. Science. ۳۶۶(۶۴۶۴), pp. ۴۴۷-۴۵۳.

خاص شود. خطر دیگر برای تحقق حقوق اساسی و حقوق بشر، عدم شمول در استفاده از اجزای سخت‌افزاری است که می‌تواند بخش‌های خاصی از جمعیت را از استفاده از سامانه هوش مصنوعی محروم کند (به عنوان مثال، عدم شمول افراد دارای معلولیت در وسایل نقلیه خودران).

بخش‌های سلامت (ماده ۱۲) دسترسی به اشتغال (ماده ۶ و ماده ۷) و آموزش (ماده ۱۳) در «میثاق بین‌المللی حقوق اقتصادی، اجتماعی و فرهنگی»^{۴۰} به طور ویژه مورد حمایت قرار می‌گیرند. معایب ساختاری در اینجا را می‌توان به سوگیری‌های الگوریتمی مبتنی بر مجموعه‌ای از داده‌های نادرست و غیرمبتنی بر نمایندگی [حقایق و خواست مردم] نسبت داد. این موارد همچنین به عنوان حوزه‌های حمایتی توسط قانونگذاری [ناظر به نفی] عدم تبعیض به طور کلی و همه جانبه، مورد تأکید قرار گرفته‌اند.

با توجه به این خطر که استفاده از هوش مصنوعی ممکن است منجر به رفتار غیرقانونی و نابرابر با افراد شود، الزامات ویژه‌ای برای کاربردهای هوش مصنوعی که به حقوق اساسی مربوط می‌شوند، مورد بحث قرار گرفته است. به عنوان مثال [مقرر شده] «مجموعه داده‌های آموزش، اعتبارسنجی و آزمون باید مرتبط، مبتنی بر نمایندگی، عاری از خطا و کامل باشند»^{۴۱}. علاوه بر این، بند ۱ ماده ۱۵ قانون هوش مصنوعی اتحادیه اروپا «سطح مناسبی از دقت، استحکام و امنیت سایبری» را به عنوان پیش‌نیاز ضروری برای سامانه‌های هوش مصنوعی پرخطر در نظر می‌گیرد.

خطرات خاص کاربردی

علاوه بر خطرات عمومی حقوق بشر، تعدادی از کاربردهای هوش مصنوعی وجود دارند که باید از منظر حقوق بشر در اولویت قرار گیرند. اجرای اصل مراقبت و نظارت مقتضی برای رعایت حقوق بشر فقط مربوط به جزئیات فنی نیست بلکه تفاسیر خاص کاربردی حقوق بشر را نیز در برمی‌گیرد. دو مثال زیر نشان می‌دهد که این موضوع می‌تواند به طور مشخص برای شرکت‌ها چه معنایی داشته باشد.

^{۴۰} International Covenant on Economic, Social and Cultural Rights (ICESCR)

^{۴۱} Keates, Simeon; Clarkson, P. John. (۲۰۰۳). Countering design exclusion: bridging the gap between usability and accessibility. Universal Access in the Information Society (UAIS) ۲, pp. ۲۱۵–۲۲۵. Cf. also Article ۱۰(۳) of the AI Act.

مطالعه موردی ۱: استفاده از هوش مصنوعی در پزشکی

هوش مصنوعی طیف گسترده‌ای از کاربردها را در نظام مراقبت‌های بهداشتی، به ویژه در پشتیبانی تشخیصی، پزشکی شخصی سازی شده^{۴۲}، توسعه دارو و مدیریت داده‌های سلامت دارد. بنابراین، استفاده از هوش مصنوعی در بخش مراقبت‌های بهداشتی، بر حوزه‌های مختلف حمایت از حقوق بشر از جمله «حق سلامت»، که در میثاق بین‌المللی حقوق اقتصادی، اجتماعی و فرهنگی (ماده ۱۲) و اساسنامه سازمان جهانی بهداشت^{۴۳} گنجانده شده است، تأثیر می‌گذارد.

در صورت استفاده‌ی نادرست، هوش مصنوعی می‌تواند این حق را محدود کند، برای مثال اگر سبب شود گروه‌هایی از دسترسی به درمان یا داروهای خاص محروم شوند یا کیفیت مراقبت‌های پزشکی به سطحی پایین‌تر از روش‌های سنتی تنزل یابد. این امر همچنین در صورتی صادق است که استفاده از هوش مصنوعی باعث عوارض جانبی روانی نامطلوب، مانند انزوای ساختاری بیماران یا کیفیت پایین‌تر مراقبت باشد. همچنین، عدم مشارکت بیماران در تصمیم‌گیری یا شفاف نبودن فرایندهای هوش مصنوعی می‌تواند به نقض حق تمامیت جسمی و روانی افراد منجر شود.^{۴۴}

مطالعه موردی ۲: استفاده از هوش مصنوعی در منابع انسانی

در زمینه استخدام، هوش مصنوعی برای ارزیابی متقاضیان بر اساس معیارهای خاص یا پیشنهاد کاندیداهای مناسب برای موقعیت‌های [شغلی] خالی نقش فزاینده‌ای دارد.

استانداردهای بین‌المللی مانند اصول بنیادین سازمان بین‌المللی کار^{۴۵} بر عدم تبعیض در اشتغال و استخدام تأکید دارند. با این حال، استفاده از هوش مصنوعی خطراتی را نیز به همراه دارد، زیرا داده‌های آموزشی مورد استفاده برای آموزش این سامانه‌ها معمولاً بر اساس تصمیمات پیشین مدیران انسانی است. فلذا، این خطر وجود دارد که هوش مصنوعی در واقع بتواند الگوهای تبعیض موجود را تقویت کند. این امر به ویژه زمانی صادق است که داده‌هایی

^{۴۲} Person-alised medicine

^{۴۳} Constitution of the World Health Organization

^{۴۴} Faden, R. R. (۲۰۰۴). Bioethics: A field in transition. Journal of Law, Medicine & Ethics, ۳۲(۲), pp. ۲۷۶-۲۷۸.

This is particularly relevant to vulnerable groups, such as people with physical, mental and cognitive disabilities, children or ethnic and social minorities (see: UN Convention on the Rights of Persons with Disabilities).

^{۴۵} International Labor Organization (ILO)

مانند محل سکونت، پیشینه تحصیلی یا مهارت‌های زبانی ممکن است به صورت غیرمستقیم به تبعیض طبقاتی یا قومی منجر شود.^{۴۶} خطرات بیشتری نیز در رابطه با تجزیه و تحلیل شبکه‌های اجتماعی افراد ایجاد می‌شود. تحلیل شبکه‌های اجتماعی داوطلبان ممکن است اطلاعاتی درباره‌ی عقاید دینی، عضویت در اتحادیه‌های کارگری یا وضعیت بارداری آشکار کند که همه با قوانین حفاظت از داده در تضاد است.

برخلاف وضعیت حقوقی در آلمان، که موضوع استخدام ممکن است از قبل تحت پوشش قانون هوش مصنوعی یا قانون عمومی [لزوم رعایت] رفتار برابر [با همه شهروندان] باشد، هیچ استاندارد بین‌المللی در مورد استفاده از هوش مصنوعی در استخدام وجود ندارد.^{۴۷} در تعدادی از کشورها، از جمله کشورهای عضو سازمان همکاری اقتصادی و توسعه (OECD) مانند ژاپن و کره جنوبی، تقریباً هیچ قانونی برای جلوگیری از تبعیض تصویب نشده است. بنابراین، اعمال پیمان جهانی سازمان ملل [ناظر به لزوم رعایت ۱۰ اصل در فعالیت‌های تجاری] و اصول راهنمای سازمان ملل [کسب و کار و حقوق بشر] در راستای ایجاد حداقل استانداردها به ویژه در بستر بین‌المللی برای کاهش اختلافات در وضعیت‌های حقوقی مختلف بسیار مهم است.

الزامات بررسی صلاحیت شرکت‌ها برای استفاده از هوش مصنوعی

شرکت‌ها هنگام اجرای تعهدات مربوط به اصل رعایت مراقبت و نظارت مقتضی برای رعایت حقوق بشر، با چالش یافتن واکنشی مناسب به خطرات ناشی از هوش مصنوعی مواجه‌اند. این خطر وجود دارد که شرکت‌های اروپایی با نادیده گرفتن استانداردهای کلی کیفیت هنگام استفاده از هوش مصنوعی در کشورهای ثالث، یا با توسعه سامانه‌های هوش مصنوعی که در وهله اول نقض حقوق بشر را ممکن می‌سازند، استانداردهای حقوق بشر را نقض کنند. در ماه‌های گذشته، مباحث ملی و بین‌المللی درباره‌ی حکمرانی هوش مصنوعی^{۴۸} بر ضرورت انجام ارزیابی اثرات حقوق بشری و بنیادی در کل چرخه‌ی عمر سامانه‌های هوش مصنوعی تأکید کرده‌اند.

تعهدات مربوط به اصل مراقبت و نظارت مقتضی برای رعایت حقوق بشر فراتر از الزامات قانون هوش مصنوعی است و در خدمت تضمین حداقل استانداردهای کیفیت هوش مصنوعی برای رعایت اصول اولیه اخلاق داده‌ها و جلوگیری از موارد جدی تبعیض قرار می‌گیرند. مجمع عمومی سازمان ملل متحد اخیراً به اهمیت اصول راهنمای

^{۴۶} Zuiderveen Borgesius, F. J. (۲۰۲۰). Strengthening legal protection against discrimination by algorithms and artificial intelligence. *The International Journal of Human Rights*, ۲۴(۱۰), pp. ۱۵۷۲-۱۵۹۳.

^{۴۷} یک نمونه خاص از این مورد، قانون حسابرسی سوگیری نیویورک است.

^{۴۸} AI Governance

سازمان ملل [در زمینه کسب و کار و حقوق بشر] به عنوان ابزاری برای تحقق حقوق بشر در زمینه هوش مصنوعی اشاره کرده است.

ارزیابی ریسک [های متوجه] حقوق بشر در سامانه‌های هوش مصنوعی

اصول راهنمای سازمان ملل و چارچوب‌های تخصصی هوش مصنوعی، شرکت‌ها را ملزم به تجزیه و تحلیل خطرات فعالیت‌ها و روابط تجاری خود می‌کنند. ترکیب دو رویکرد «تحلیل مبتنی بر حقوق بشر» و «تحلیل خاص هوش مصنوعی» به شرکت‌ها کمک می‌کند تا چالش‌ها را جامع‌تر شناسایی کنند.

ارزیابی خطر سوء کاربرد

برای سنجش احتمال سوءاستفاده از سامانه، لازم است خطرات مرتبط با شرکای تجاری، ماهیت کاربرد و ظرفیت بالقوه سوءاستفاده‌ی فناوری بررسی شود. به‌ویژه، همکاری با شرکایی که ارتباط نزدیکی با سازمان‌های امنیتی یا نهادهای دولتی در رژیم‌های اقتدارگرا دارند، خطر بالایی دارد. همچنین شرکت‌ها باید تغییرات محیط تجاری و وضعیت حقوق بشر در کشورهای مقصد را در نظر بگیرند، زیرا چنین تغییراتی می‌تواند بر نحوه‌ی استفاده‌ی مستمر از سامانه اثر بگذارد. چارچوب‌های قانونی، مانند قانون هوش مصنوعی اتحادیه اروپا، فهرستی از سامانه‌هایی را ارائه می‌دهد که از نظر خطرات [متوجه] حقوق بشر، در رده‌ی حساس و پرخطر قرار دارند (از جمله سامانه‌های تشخیص چهره و سامانه‌های تحلیل شبکه‌های اجتماعی)^{۴۹}.

بنابراین، هنگام ارزیابی خطر، شرکت‌ها باید ورودی داده‌ها (مانند تحلیل شبکه‌های اجتماعی)، کاربرد بالقوه مرتبط با آن (مانند شناسایی یک فرد) و ویژگی‌های کاربر نهایی (مؤسسات دولتی یا خصوصی) را نیز در نظر بگیرند.

ارزیابی خطر عدم توجه کافی به حقوق بشر در طراحی راهکارهای هوش مصنوعی

خطرات حقوق بشر به ویژه در مواردی که حقوق مشارکت و اطلاعات توسط استانداردهای کلی حقوق بشر و حقوق کار یا چارچوب‌های قانونی محافظت می‌شوند، بروز پیدا می‌کنند. برخی از این موارد قبلاً توسط قانون هوش مصنوعی اتحادیه اروپا پوشش داده شده‌اند که تمرکز ویژه‌ای بر شفافیت و نظارت انسانی دارد. با این حال، همه خطرات خاص کاربردی را نمی‌توان توسط این قانون برطرف کرد. برای نمونه، در حوزه‌ی سلامت باید رضایت آگاهانه‌ی بیماران در تصمیمات مرتبط با درمان مبتنی بر هوش مصنوعی لحاظ شود. علاوه بر این، هنگامی

^{۴۹} Annex III to the AI Act. Also: UN General Assembly. (۲۰۱۸), p. ۲۰.

که هوش مصنوعی در سراسر جهان گسترش می‌یابد، باید توجه ویژه‌ای به تعریف حداقل استاندارد های لازم برای تضمین حقوق اطلاع و مشارکت افراد در تمامی زمینه‌های کاربردی مشخص شود.^{۵۰}

ارزیابی خطر اثرات منفی بر حقوق بشر

ویژگی‌های فنی سامانه‌های هوش مصنوعی، از جمله انتخاب داده‌های آموزشی و نوع سخت‌افزار، می‌تواند پیامدهای جدی برای حقوق بشر داشته باشد. این در شرایطی است که سامانه‌های مبتنی بر هوش مصنوعی در دسترسی به منابع اجتماعی خاص (بازار کار، نظام مراقبت‌های بهداشتی، آموزش، حمل و نقل عمومی محلی و بین شهری) نقش داشته باشند.

از آنجایی که تبعیض الگوریتمی ممکن است متفاوت از تبعیض انسانی عمل کند، تجزیه و تحلیل ریسک قبلی نه تنها باید طیف‌های سنتی تبعیض را پوشش دهد، بلکه باید تحقیقات تجربی قبل از توسعه را نیز در نظر بگیرد تا تصویر دقیق‌تری از تعصبات یا الگوهای تبعیض خاصی که یک مدل هوش مصنوعی می‌تواند بر آنها تأثیر بگذارد، به دست آورد.

این بدان معناست که الگوهای کاملاً جدیدی از تبعیض نیز ممکن است پدیدار شوند که به طور کافی توسط قوانین فعلی مورد توجه قرار نگرفته‌اند. بنابراین انتظار می‌رود در آینده تعریف‌های قانونی تبعیض گسترش یابند تا ویژگی‌هایی مانند سوابق ژنتیکی یا پیشینه‌ی اجتماعی را نیز شامل شوند.

اقدامات کاهش خطر و ارتباطات خارجی

علاوه بر نیاز به ارزیابی خطر در حوزه حقوق بشر، چارچوب‌های موجود در حوزه تجارت و حقوق بشر مجموعه‌ای از اقدامات را برای جلوگیری از نقض حقوق بشر تعریف می‌کنند. بر اساس اصول راهنمای سازمان ملل [ناظر به کسب و کار و حقوق بشر]، نتایج ارزیابی خطر باید در فرآیندهای داخلی شرکت گنجانده شود. علاوه بر این، اثربخشی اقدامات انجام شده باید بررسی شود، اقدامات جبرانی برای رفع تأثیرات منفی انجام پذیرد و ارزیابی ریسک و اقدامات شرکت باید به طور عمومی اطلاع‌رسانی شود.

^{۵۰} برای مطالعه بیشتر به این منبع مراجعه کنید:

کاهش خطرات ناشی از همکاری با شرکای خارجی

با در نظر گرفتن خطر سوءاستفاده، لازم است استانداردهای حقوق بشری در [معیارها و فرایندهای] انتخاب شرکای تجاری و فناوری گنجانده شوند، به ویژه در مواردی که سامانه‌های هوش مصنوعی که قرار است توسعه داده شوند، حاوی عناصری باشند که می‌توانند برای اهداف ناقض حقوق بشر، استفاده شوند. این امر در درجه اول مربوط به شیوه‌های ممنوعه و فناوری‌های پرخطر است که در قانون هوش مصنوعی [اتحادیه اروپا] تعریف شده‌اند. ایجاد فرآیندهای تأیید داخلی (شامل واحدهای ارزیابی تطابق و اخلاق) و درج آن در ضوابط رفتاری و آیین‌نامه‌های شرکتی توصیه می‌شود.

برای جلوگیری مؤثر از هرگونه کمک و تشویق به نقض حقوق بشر، منطقی است که علائم هشدار دهنده‌ای که به خطرات ترکیبی اشاره دارند نیز ارسال گردد (مثل مورد بکارگیری هوش مصنوعی برای تشخیص چهره توسط یک آژانس امنیتی یک کشور سرکوبگر).

پیشگیری از نقض استانداردهای حداقلی حقوق بشر

پیشگیری مؤثر از خطرات [متوجه] حقوق بشر هنگام اجرای هوش مصنوعی مستلزم ایجاد یک نماگر^{۵۱} الزامات اخلاقی و حقوق بشری است که یافته‌های ارزیابی ریسک (عوامل خاص گروه هدف، خطرات امنیتی) را بازتاب دهد. اقداماتی که در این نوشتار مورد بحث قرار گرفته است شامل آموزش توسعه‌دهندگان و همچنین گرد هم آوردن تیم‌های بین رشته‌ای و متنوع هنگام طراحی مدل‌ها میشود، به ویژه هنگام طراحی سیستم‌های سخت‌افزاری و رابط‌های انسان و ماشین، مهم است که نهادی با تخصص [و تجربه] در مورد نیازهای افراد دارای معلولیت را مشارکت دهید تا از دسترسی‌پذیر بودن سیستم‌های هوش مصنوعی اطمینان حاصل شود. همچنین توصیه می‌شود سامانه‌های هوش مصنوعی مربوطه توسط اشخاص مستقل و خارجی حسابرسی شوند.^{۵۲}

^{۵۱} Profile

^{۵۲} استانداردهای اخلاقی موسسه مهندسان برق و الکترونیک (IEEE) یکی از نمونه‌های این امر است.

<https://standards.ieee.org/industryconnections/ec/autonomoussystems/>. When using audits and standards, however, attention must always be paid to the applicationspecific risks.

ارتباطات بیرونی

انتشار عمومی ارزیابی خطر و اقدامات کاهش آن در حال حاضر طبق قانون^{۵۳}، الزامی نیست^{۵۴}. با این حال، در وضعیت کنونی یک چارچوب قانونی برای تعهدات شرکت‌ها در خصوص افشا[و آشکار سازی] وجود دارد^{۵۵}. در اینجا، از شرکت‌ها انتظار می‌رود که توسعه و استفاده از هوش مصنوعی در مناطق پرخطر را توصیف و تبیین و خطرات مستقیم[متوجه] حقوق بشر را طبق قانون هوش مصنوعی اتحادیه اروپا گزارش دهند. این مربوط به استفاده از راه کارهای هوش مصنوعی در استخدام یا قرار دادن سامانه های هوش مصنوعی در بازار کشورهای ثالث خارج از اتحادیه اروپا است که کنوانسیون‌های مربوطه سازمان ملل در مورد حمایت از حقوق بشر را امضا نکرده‌اند.

خلاصه و چشم‌انداز

سامانه‌های هوش مصنوعی بدون شک چالشی را برای تحقق تعهدات مربوط به رعایت اصل مراقبت و نظارت مقتضی شرکت‌ها جهت رعایت حقوق بشر ایجاد می‌کنند. این به دلیل خطرات خاص [متوجه] حقوق بشر است که با توسعه و استفاده از هوش مصنوعی مرتبط هستند و از پیچیدگی این فناوری ناشی می‌شوند. این خطرات ارتباط نزدیکی با پردازش خودکار داده‌ها، طراحی راه‌حل‌های هوش مصنوعی، سوگیری‌ها، ابهام سامانه های هوش مصنوعی و ماهیت دوگانه آنها دارند.

مهم‌ترین این خطرات شامل سوءاستفاده از فناوری توسط بازیگران دولتی یا خصوصی، نادیده گرفتن حقوق مشارکت افراد در طراحی سامانه‌ها و در رابط‌های انسان و ماشین و در نهایت، در تأثیرات منفی مستقیم خطاهای الگوریتمی بر زندگی افراد است. خروجی نادرست یا غیردقیق از سامانه‌های هوش مصنوعی می‌تواند به ویژه برای افراد در مناطق پرخطر مضر باشد. این امر می‌تواند بر مشارکت آنها در جامعه و در زندگی اقتصادی و سیاسی و همچنین حق آنها برای دریافت مزایای دولتی تأثیر منفی بگذارد.

^{۵۳} توضیح: چون این اثر توسط یک نهاد آلمانی تهیه شده است احتمالاً منظور نویسنده متن، قانون آلمان است.

^{۵۴} Directive (EU) ۲۰۲۲/۲۴۶۴ of the European Parliament and of the Council of ۱۴ December ۲۰۲۲ amending Regulation (EU) No ۵۳۷/۲۰۱۴, Directive ۲۰۰۴/۱۰۹/EC, Directive ۲۰۰۶/۴۳/EC and Directive ۲۰۱۳/۳۴/EU, as regards corporate sustainability reporting. See also: Directive ۲۰۱۴/۹۵/EU of the European Parliament and of the Council of ۲۲ October ۲۰۱۴ amending Directive ۲۰۱۳/۳۴/EU as regards disclosure of nonfinancial and diversity information by certain large undertakings and groups.

^{۵۵} Office of the United Nations High Commissioner for Human Rights (OHCHR). A/۷۳/۳۴۸, p. ۲۱.

نگاهی به ادبیات مربوط به هوش مصنوعی نشان می‌دهد که ماهیت خطرات [متوجه] حقوق بشر بسته به [نوع] کاربرد، متفاوت است. در اینجا ما با تفاسیر خاصی از حقوق بشر مواجه می‌شویم که الزامات اضافی را بر پیاده‌سازی سامانه‌های هوش مصنوعی اعمال می‌کنند.

یک مثال، مفهوم محرمانگی پزشکی است که می‌تواند توسط سامانه‌های هوش مصنوعی دور زده شود و بنابراین مستلزم اقدامات احتیاطی ویژه هنگام توسعه آنها است. مورد دیگر، حفاظت از اصل رفتار برابر در استخدام کارمندان است.

تنش بین خطرات خاص هوش مصنوعی و تعهدات کلی مربوط به لزوم اجرای اصل مراقبت و نظارت مقتضی شرکت‌ها برای رعایت حقوق بشر می‌تواند حل و فصل شود. پیمان جهانی سازمان ملل و اصول راهنمای سازمان ملل [ناظر به کسب و کار و حقوق بشر]، رویه‌های کلی برای مدیریت تعهدات مربوط به اجرای اصل مراقبت و نظارت مقتضی برای رعایت حقوق بشر را مشخص می‌کنند. این موارد شامل تعهد به انجام ارزیابی‌های خطرات حقوق بشر در رابطه با فعالیت‌های تجاری یک شرکت می‌شود. این چارچوب‌ها به ویژه برای گسترش بین‌المللی هوش مصنوعی مهم هستند، زیرا خطرات حقوق بشر در ارتباط با رژیم‌های اقتدارگرا را مورد نظرمی‌دهند و بنابراین می‌توانند شکاف‌هایی را که نمی‌توانند در محدوده قانون هوش مصنوعی اتحادیه اروپا پوشش داده شوند، مثلاً در مناطق جنگی، پُر کنند. بر این اساس، اهمیت اجرای اصل مراقبت و نظارت مقتضی شرکت‌ها برای رعایت حقوق بشر به طور ویژه توسط مجمع عمومی سازمان ملل مورد تأکید قرار گرفته است.

بنابراین، پیاده‌سازی هوش مصنوعی منطبق با حقوق بشر، مستلزم آن است که شرکت‌ها نگاهی یکپارچه و جامع به تجارت و حقوق بشر و همچنین مقررات خاص هوش مصنوعی داشته باشند. این امر می‌تواند شکاف‌های موجود در رویکردهای موجود را پُر کند و از گسترش توصیه‌های بیشتر برای اقدام و اقدامات جبرانی و اصلاحی برای شرکت‌ها پشتیبانی کند. در نتیجه، شرکت‌ها نیازی به معرفی فرآیندهای کاملاً جدید برای خطرات هوش مصنوعی نخواهند داشت. یک راهبرد مناسب‌تر، اهمیت دادن بیشتر به بین‌رشته‌ای بودن در توسعه مدل‌های هوش مصنوعی و تطبیق فرآیندهای موجود برای گنجاندن تمرکز بر هوش مصنوعی و حقوق بشر است.

نویسندگان متن در پایان اثر خود فهرستی از منابع قابل رجوع برای مطالعات بیشتر را بیان داشته اند که برگردان فارسی و عین عنوان انگلیسی آن در زیر ارائه میشود:

برخی از ادبیات تخصصی برای مطالعات بیشتر؛

الف. قوانین و ابتکارات حقوقی درباره‌ی هوش مصنوعی:

- چارچوبی برای قانون حقوق هوش مصنوعی - آمریکا
- مقررات عمومی حفاظت از داده‌ها
- قانون هوش مصنوعی اتحادیه اروپا
- قانون رانندگی خودکار آلمان
- قانون ممیزی و نظارت بر سوگیری - ایالت نیویورک
- Blueprint for an AI Bill of Rights
- General Data Protection Regulation (GDPR)
- EU Artificial Intelligence Act (EU AI Act)
- German Act on Autonomous Driving
- New York Bias Audit Law

ب. چارچوب‌ها و بیانیه‌های بین‌المللی:

- چارچوب هوش مصنوعی برای مردم
- اصول ناظر به هوش مصنوعی موسوم به اصول آزیلومار
- معاهده شورای اروپا درباره‌ی هوش مصنوعی و حقوق بشر
- قانون اخلاقیات آلمانی برای رانندگی خودکار
- دستورالعمل‌های اخلاقی برای هوش مصنوعی قابل اعتماد
- نظریه کمیسیون اخلاق داده دولت آلمان

- گزارش کمیسیون اروپا در مورد هوش مصنوعی
- اصول سازمان همکاری اقتصادی و توسعه (OECD) درباره‌ی هوش مصنوعی
- توصیه‌نامه یونسکو درباره‌ی اخلاق هوش مصنوعی

- AI4People framework (Floridi et al. ۲۰۱۸)
- Asilomar AI Principles
- Council of Europe AI Treaty
- German Ethics Code for Automated and Connected Driving
- Ethics Guidelines for Trustworthy AI
- Opinion of the Data Ethics Commission of the German Government
- European Commission White Paper on AI
- OECD AI Principles
- UNESCO Recommendation on the Ethics of Artificial Intelligence

ج. نهادها و مراکز اطلاعاتی:

- پلتفرم سیستم‌های یادگیرنده آلمان (Lernende Systeme)

<https://www.plattform-lernende-systeme.de/startseite.html>

- دیده بان الگوریتم (Algorithm Watch)

<https://algorithmwatch.org/en/>

- انجمن هوش مصنوعی آلمان (KI-Verband)

<https://ki-verband.de/>

- سازمان فدرال ضد تبعیض آلمان

<https://www.antidiskriminierungsstelle.de/DE/startseite/startseite-node.html>

د. منابع مرتبط و شناخت رویه های خوب:

- Association of German Chambers of Commerce and Industry (DIHK) guidelines on the

management of generative AI applications (German only)

- AI in companies – A practical guide on legal issues, published by the Bertelsmann

Stiftung (German only)

- Algorithmic decision making systems – Human rights requirements and trends at the

international level, published by the German Institute for Human Rights (German only)

- AI in the corporate context – Literature analysis and thesis paper, published by the Fraun

hofer Center for International Management and Knowledge Economy IMW (German only)

- Practical guide for the Algo.Rules – Guidance for developers and their managers, published by the Bertelsmann Stiftung (German only)

توضیح کوتاه در مورد اصول راهنمای سازمان ملل متحد در مورد کسب و کار و حقوق بشر:

اصول راهنمای سازمان ملل متحد در مورد کسب و کار و حقوق بشر در قطعنامه‌ای توسط شورای حقوق بشر سازمان ملل متحد در سال ۲۰۱۱ میلادی تصویب شد و تعهدات مربوط به اجرای اصل مراقبت و نظارت مقتضی شرکت‌ها برای رعایت حقوق بشر در فعالیت‌های خود را مشخص کرد. این موارد شامل "انجام ارزیابی خطر، اجرای اقدامات برای جلوگیری از تأثیرات نامطلوب بر حقوق بشر، ارزیابی اقدامات جبرانی و اصلاحی در قبال موارد نقض حقوق بشر و اقداماتی برای ایجاد شفافیت" است.

