

تکنولوژی اخلاقی

نویسنده: پائولا بادینگتون^۱

مترجم: ملیکا خوش نژاد

استفاده از هوش مصنوعی کار انسان‌ها را راحت‌تر از گذشته کرده است. می‌شود با داشتن نمونه‌های بیشتر و بررسی کردن شرایط مختلف، سریع‌تر و بهتر تصمیم گرفت. شاید در دهه‌ی گذشته این‌که هوش مصنوعی بتواند قهرمان شطرنج جهان را برده اتفاق عجیبی بود، اما امروزه بخش زیادی از تصمیم‌گیری‌های انسانی (چه به صورت فردی و چه جمعی) به کمک هوش مصنوعی انجام می‌گیرد. اما آیا هوش مصنوعی باعث ابله‌تر شدن انسان‌ها نشده است؟ حالا که دیگر لازم نیست خودمان فکر کنیم، به این فکر کرده‌ایم که چقدر به هوش مصنوعی اعتماد داریم؟ آیا ممکن است هوش مصنوعی شر باشد؟ و شاید مهم‌تر از همه، آیا مدل‌های اندیشیدن انسانی، همه قابل واگذاری به ربات‌های هوشمند است؟ یا چیزی انسانی هم در کار است.

اتومبیل‌های خودران^۲ در مستی رانندگی نمی‌کنند و هوش مصنوعی‌های مورد استفاده در پزشکی خستگی نمی‌شناسند. با در نظر گرفتن خطاهای آشکارمان، پس ما انسان‌ها چه کاری را بهتر انجام می‌دهیم؟

پسر بچه‌ی پنج ساله‌ای در حال کمک کردن به مادر بزرگش در آشپزی است. او تکه‌های بیسکویت را از خمیری که مادر بزرگش آماده کرده، جدا می‌کند. او در این کار هیچ مهارتی ندارد. بنابراین به ربات خانوادگی‌شان دستور می‌دهد تا به جای او این کار را انجام دهد و گرچه آن ربات تا به حال چنین کاری نکرده است، به سرعت چگونگی انجامش را یاد می‌گیرد و بیسکویت‌ها را به نحو احسن از خمیر جدا می‌کند. مادر بزرگ حسابی ناامید می‌شود، چون از ته قلب به یاد روزهایی می‌افتد که پسرش با لذت تمام در همان سن و سال کنونی نوه‌اش، برای او بیسکویت‌های

¹ محقق دانشگاه کاردیف Paula Boddington

² autonomous car

نازکی درست می‌کرد که جای انگشتانش روی سطح آن باقی مانده بود. نوهی او برای انجام دادن چنین کارهایی باز هم از ربات‌شان استفاده خواهد کرد و بدون بهره‌وری از مهارت‌های دست‌ورزی بزرگ خواهد شد.

هنگامی که والدین پسر بچه به خانه بیایند، او می‌گوید: «نگاه کنید، این بیسکویت‌ها را برای شما درست کردم.» یکی از والدین می‌گوید: «اوه، چقدر خوب. اجازه دارم یکی بردارم؟» دیگری در سکوت با خود می‌اندیشد: «نه، تو این‌ها را خودت درست نکرده‌ای، ای متقلب کوچک!»

هوش مصنوعی (ای‌آی)^۳ ممکن است پتانسیل آن را داشته باشد که چگونگی انجام وظایف مان و ارزش‌گذاری کارها را دگرگون کند. استفاده از هوش مصنوعی به منظور آن که به جای ما فکر کند، ممکن است مهارت‌های اندیشیدن ما را تضعیف کند.

آیا هوش مصنوعی باعث ابله شدن انسان می‌شود؟ جاروبرقی و ماشین لباس‌شویی و سایر اختراعاتی که به انسان در کارهای روزمره کمک می‌کنند، باعث کم شدن تحرک بدنی و تنبل شدن انسان‌های معاصر شده‌اند. ما - آدم‌های دوران تکنولوژی‌ها - بیشتر از همیشه تاریخ از مشکل اضافه‌وزن رنج می‌بریم. آیا علاوه بر چاق شدن، به لطف هوش مصنوعی ابله هم شده‌ایم؟

هوش مصنوعی در حال حاضر ضعیف است و فقط می‌تواند کارهای مشخص و انتخاب‌شده‌ای را انجام دهد. حتی هنگامی که هوش مصنوعی می‌تواند کارهای خاصی را به خوبی انسان‌ها یا بهتر از آن‌ها انجام دهد، لزوماً به شیوه‌ی انسان‌ها به نتیجه‌ی مدنظر نمی‌رسد. یکی از کارهایی که هوش مصنوعی به خوبی

می‌تواند از پس آن بر بیاید، غربال کردن توده‌ی داده‌ها با سرعت بسیار زیاد است. با استفاده از یادگیری ماشین^۴، هوش مصنوعی‌ای که با هزاران تصویر آموزش دیده است، می‌تواند از توانایی شناسایی تصویر یک گربه بهره‌مند شود (که با توجه به پُر شدن تصاویر گربه‌ها در اینترنت، موفقیت بزرگی به‌شمار می‌رود). اما انسان‌ها این کار را به شیوه‌ی بسیار متفاوتی انجام می‌دهند. یک کودک انسان، اغلب بعد از دیدن فقط یک نمونه، می‌تواند گربه‌ها را شناسایی کند.

از آنجایی که هوش مصنوعی ممکن است، نسبت به انسان‌ها به شیوه‌ی متفاوتی «فکر کند» و به دلیل تمایل عمومی‌ای که نسبت به غرق شدن در جذابیت‌های هوش مصنوعی وجود دارد، استفاده از آن می‌تواند نحوه‌ی رویکرد ما به وظایف مان و شیوه‌ی تصمیم‌گیری مان را دگرگون کند. جذابیت اغواگرانه‌ای که اطراف هوش مصنوعی را

^۳ Artificial intelligence (AI)

^۴ machine learning به عنوان یکی از شاخه‌های وسیع و پرکاربرد هوش مصنوعی، یادگیری ماشین به تنظیم و اکتشاف شیوه‌ها و الگوریتم‌هایی می‌پردازد که بر اساس آن‌ها رایانه‌ها و سامانه‌ها توانایی تعلّم و یادگیری پیدا می‌کنند. - م.

فراگرفته است در واقع یکی از خطرات آن را به تصویر می‌کشد. کمابیش تمام مقاله‌هایی که در رابطه با هوش مصنوعی نوشته می‌شوند، قدرت‌های آن را تبلیغ می‌کنند و حتی آن‌هایی که مربوط به کاربردهای معمولی هوش مصنوعی است با روبات‌های قاتل به تصویر کشیده شده‌اند و این باعث ناامیدی کسانی است که در این زمینه فعالیت می‌کنند.

تأثیر تکنولوژی بر شکل‌دهی ارزش‌های ما بر کسی پوشیده نیست. در یکی از میزگردهای اخیر در رابطه با مباحث اخلاقی مربوط به هوش مصنوعی، در گروهی که من در آن بودم بیشتر وقت‌مان را صرف بحث بر سر نمونه‌ی شناخته‌شده‌ی «ماشین لباس‌شویی» کردیم. ماشین لباس‌شویی صرفاً «جایگزین» شست‌وشوی دستی نشد، بلکه تأثیر عظیمی بر شیوه‌ی نگرش ما درباره‌ی تمیزی، کارهای خانه و حتی تولید لباس‌ها داشت. از آنجایی که هوش مصنوعی نه فقط به منظور کمک‌رسانی به ما در امر شست‌وشو، بلکه برای یاری رساندن به ما در نحوه‌ی اندیشیدن و تصمیم‌گیری درباره‌ی وظایف نامشخص بی‌شماری طراحی شده است، باید به صورت جدی به این مسئله بپردازیم که چگونه ممکن است افکار و رفتار خود ما را دگرگون کند.

آیا روبات‌ها می‌توانند تصمیم‌های اخلاقی بگیرند؟ اگر منظورمان از اخلاق، دستورالعمل‌های صوری اخلاق کانتی یا فرمول‌بندی ریاضیاتی اخلاق دکارتی یا شاخه‌ی هندسی اخلاق اسپینوزایی باشد شاید. اما اگر اخلاق را چیزی انسانی‌تر بدانیم می‌توانیم همچنان به روبات‌های اخلاقی مشکوک باشیم. یک تصمیم اخلاقی ممکن است فارغ از تمام صورت‌بندی‌های ممکن، صرفاً مربوط یک انسان خاص در یک شرایط خاص باشد.

در برخی موارد، هوش مصنوعی می‌تواند در راستای بهبود رویکرد به ما یاری برساند. اما در موارد دیگر می‌تواند نگرش ما نسبت به مسائل مهم را تقلیل دهد یا تضعیف کند. حتی ممکن است نحوه‌ی اندیشیدن ما درباره‌ی ارزش‌ها را از مسیر درست منحرف کند.

ناآگاهی نسبت به تغییرات و سازگاری سریع با تکنولوژی می‌تواند بدان معنا باشد که ما کاملاً نسبت به چنین دگرگونی‌هایی در فرهنگ و ارزش‌هایمان آگاه نیستیم. به‌عنوان مثال، نگرش ما نسبت به حریم خصوصی به طرز چشمگیری همراه با دگرگونی‌های عظیم تکنولوژیکی در رابطه با نحوه‌ی ارتباط برقرار کردن‌هایمان و چگونگی به‌اشتراک‌گذاری و پردازش داده‌ها، تغییر کرده است. یکی از مهم‌ترین چیزهایی که منجر به پیشرفت هوش مصنوعی شده، حجم عظیم داده‌هایی است که در حال حاضر در دسترس قرار دارد. داده‌هایی که بیشتر آن‌ها درباره‌ی خودمان است و در طول زندگی روزمره‌مان جمع‌آوری شده‌اند. بسیاری از آدم‌ها به شدت در رابطه با سازماندهی‌هایی که

کنترل داده‌های ما را در دست دارند، محتاط‌اند، باین حال همچنان حجم بالایی از اطلاعات بسیاری شخصی‌شان را پُست می‌کنند، اطلاعاتی که تا همین چند سال پیش به شدت خصوصی در نظر گرفته می‌شدند. پژوهش‌ها حاکی از آن است که نگرانی‌های مردم درباره‌ی حریم شخصی اطلاعات‌شان در موقعیت‌های مختلف متغیر است. این بدان معنا نیست که تکنولوژی «به‌تنهایی» عامل این اتفاق است، چراکه همیشه دگرگونی‌های اجتماعی دیگر هم به صورت هم‌زمان دست‌اندر کار بوده‌اند.

و شاید نسبت به تأثیرات برخی از تکنولوژی‌ها به صورت ویژه‌ای ناآگاهیم، چراکه بر شکل‌دهی نحوه‌ی نگرش ما بر جهان به شدت اثر گذارند. قسمت چالش‌برانگیز هوش مصنوعی این است که می‌تواند به شیوه‌ای عمل کند که ما به طور کامل از آن آگاه نیستیم. پیدا کردن الگوهایی که بر اساس آن‌ها با یکدیگر ارتباط برقرار می‌کنیم، نحوه‌ی اندیشیدن‌مان و چگونگی کشف جهان مفید است. این مسئله‌ای کاملاً جدید نیست. تکنولوژی‌های مربوط به نوشتن، چاپ و تلفن از پیش چگونگی ادراکات و تعاملات ما با جهان و حتی مغزهایمان را متحول کرده‌اند. اما هوش مصنوعی می‌تواند حتی قدرتمندتر از این حرف‌ها باشد. الگوریتم‌های موجود در تکنولوژی‌ها که از طریق آن‌ها به داده‌های بی‌شماری دسترسی داریم، می‌توانند شکل‌دهنده‌ی اطلاعات دریافتی و حتی چگونگی دریافت و واکنش ما باشند. این صرفاً پیامدی ناخواسته در نتیجه‌ی استفاده از ای‌آی نیست، بلکه اصلاً به همین منظور طراحی شده است. تکنولوژی غالباً به کمک هوش مصنوعی در حال بهره‌برداری از روان‌شناسی انسانی است تا شیوه‌ی رفتار ما را شکل دهد. تلفن‌های همراه و شبکه‌های اجتماعی براساس پژوهش‌های صورت گرفته در حوزه‌ی روان‌شناسی درباره‌ی ایجاد واکنش‌های اعتیادآور به استفاده از آن‌ها طراحی شده‌اند.

بنابراین اجازه دهید با تمرکز بر این مسئله که ماشین‌ها و انسان‌ها چگونه از داده‌ها استفاده و آن‌ها را تحلیل می‌کنند، چند نمونه از استفاده و کاربردهای بالقوه‌ی هوش مصنوعی را بررسی کنیم.

استفاده از هوش مصنوعی به جای تصمیم‌گیری‌های انسانی مزیت‌های فراوانی دارد. برای مثال، پل نیومن^۵، مهندس گروه روباتیک موبایل آکسفورد^۶ خاطر نشان کرده است که یادگیری از تصادفات در وسایل نقلیه‌ای که توسط انسان‌ها رانده می‌شوند، فرآیندی آهسته و پیچیده است. انسان‌های دیگر نمی‌توانند مستقیماً از هر یک از موارد چیزی یاد بگیرند، حتی انسانی که در آن تصادف نقش داشته هم ممکن است چیز بسیار کمی بیاموزد یا اصلاً هیچی یاد

⁵ Paul Newman

⁶ Oxford Mobile Robotics Group

نگیرد. اما هرگاه یک اتومبیل خودران تصادف کند، تمام داده‌های مربوط به آن می‌تواند بلافاصله در اختیار همه‌ی اتومبیل‌های خودران دیگر قرار بگیرد و از آن به منظور کاهش احتمال تصادف‌های آینده استفاده کرد.

این بُعدِ هوش مصنوعی یعنی توانایی به اشتراک گذاشتن اطلاعات مانند یک ذهنِ جمعی^۷ و تجزیه و تحلیل سریع و دقیق داده‌ها، می‌تواند به منزله‌ی پیشرفتی واقعی در توانایی حل مسئله‌ی ما تلقی شود. در حقیقت همین توانایی منجر به رونق گرفتن کنونی هوش مصنوعی شده است.

انسان‌ها به شیوه‌های مختلف در گردآوری و تجزیه و تحلیل داده‌هایی که برای تصمیم‌گیری درست لازم‌اند و در به کار بستن آن‌ها شکست می‌خورند. یک وسیله‌ی نقلیه‌ی خودران هرگز از پذیرش خطاهایش خجالت نمی‌کشد، هرگز آن‌قدر احمق نیست که از گذاشتن عینکش هنگام رانندگی سرباز زند، هیچ‌گاه هنگام خستگی اصراری به رانندگی ندارد و هرگز از گذراندن دوره‌ی رانندگی پیشرفته سرباز نمی‌زند. غلبه بر تعصب، غرض‌ورزی و غیرمنطقی بودن یکی از روش‌های بهبود تصمیم‌گیری‌های انسانی است، به‌ویژه هنگامی که پای مسائل مربوط به ارزش‌ها در میان است. برخی از این تعصبات و غیرمنطقی بودن‌ها شامل رد کردن اطلاعات مرتبط یا شکست در پردازش آن‌هاست. بنابراین این مدل استفاده از هوش مصنوعی برای گردآوری داده‌ها ظاهراً مزیتی است که می‌توانیم در تصمیم‌گیری‌هایمان به کارگیریم.

اما چنین نتیجه‌گیری‌ای ممکن است عجولانه باشد. تمام مشکلات ما نمی‌توانند با رویکردی صرفاً داده‌محور حل شوند. آشکارا اجتناب از تصادفات رانندگی خوب است. در جایی که آن‌چه انجام می‌دهیم غالباً رفع و رجوع مشکلات به کمکِ تکنولوژی است، مسئله‌ای مربوط به ایمنی به‌شمار می‌رود و میزان موفقیت آن به سادگی سنجیده می‌شود. وسیله‌ی نقلیه یا تصادف می‌کند یا نه و مرگ و صدمات را می‌توان تعیین کرد. هم چنین سنجش رویدادهای نزدیک به تصادفات بسیار ساده است. اما در رابطه با مشکلاتی که صرفاً فنی نیستند، چندان آشکار نیست که رویکردی مبتنی بر ذهنِ جمعی همیشه به کار آید.

برای مثال پزشکی یعنی یکی از امیدبخش‌ترین حوزه‌های هوش مصنوعی را در نظر بگیرید. پزشکی در عین حال که یک علم است، هنر نیز به‌شمار می‌رود. در پزشکی علم و تکنولوژی با دنباله‌روی ارزش‌ها با یکدیگر ترکیب می‌شوند: ارزش‌های سلامتی، رابطه‌ی خوب با بیمار، مراقبت فرد-محور؟، خودمختاری بیمار و ... اینجا ما فقط به دنبال رفع و رجوع مشکلات به کمک تکنولوژی نیستیم.

⁷ hive mind

استفاده از هوش مصنوعی در تشخیص بیماری‌ها بسیار امیدبخش است؛ مثلاً با بررسی سریع داده‌های بسیار زیادی، به تفسیر تصاویر پزشکی کمک می‌کند. ظاهراً شواهد حاکی از آن است که هوش مصنوعی می‌تواند متوجه تفاوت‌های جزئی‌ای میان تصاویر شود که از چشم انسان پوشیده می‌مانند. اما افزون‌براین، از هوش مصنوعی می‌تواند خطاهای فاحشی سر بزند که یک انسان هرگز مرتکب آن‌ها نخواهد شد. بنابراین در حال حاضر، به نظر می‌رسد ترکیب هوش مصنوعی با مهارت‌های انسانی بهترین گزینه‌ای است که برای بهبود تشخیص بیماری‌ها پیش رو داریم. تا بدین جا، این خبر خیلی خوبی است.

اما یادداشتی از دانتون چار^۸، نیکام شا^۹ و دیوید مگنوس^{۱۰} از دانشکده‌ی پزشکی دانشگاه استنفورد در سال ۲۰۱۸ در ژورنال پزشکی نیوانگلند^{۱۱} منتشر شد که پرسش‌هایی بسیار جدی را در رابطه با استفاده از هوش مصنوعی در

هوش مصنوعی به ما در کنترل شبکه‌های اجتماعی کمک می‌کند. کسی نمی‌تواند میلیون‌ها توئیت را بخواند و بررسی‌شان کند. اما همین هوش مصنوعی که در صد شبکه‌های اجتماعی کمک‌مان می‌کند، می‌تواند به تولیدکننده‌ی محتوا در همین شبکه‌ها تبدیل شود. وقتی که کنترل‌کننده و تولیدکننده یک چیز باشد، ماجرا از دست کاربران انسانی خارج می‌شود. کسی در شبکه‌های اجتماعی زورش به رویت‌ها نمی‌رسد.

تشخیص بیماری‌ها و تصمیم‌گیری درباره‌ی شیوه‌های درمانی مطرح کرد. پزشکی را به عنوان یک علم در نظر بگیرید. اگر هوش مصنوعی «مخزنی را برای اذهان جمعی پزشکی» تشکیل دهد، باید پیش از استفاده از آن به شیوه‌ای که پزشکی را به سمت تفکر حرفه‌ای یکپارچه‌ای می‌برد به شدت محتاط

باشیم، چون ممکن است مانع از تفکر مستقل و تجربه‌ی بالینی فردی شود. در حال حاضر، مشخص است که نحله‌های نظری پزشکی متفاوتی در رابطه با تشخیص بیماری‌ها و شیوه‌های درمانی وجود دارند. اگر می‌توانستیم کاملاً مطمئن باشیم که هوش مصنوعی صرفاً در حال بهبود دقت ماست، در آن صورت یکپارچگی تفکر پزشکی خوب بود. اما این خطر وجود دارد که هوش مصنوعی پیش از موعود گزینه‌های پیش رو را از میان بردارد یا ما را به سمت مسیرهای درمانی خاصی هدایت کند. افزون‌براین، نویسندگان این یادداشت هشدار می‌دهند که چنین یادگیری ماشینی به جای تمرکز بر آن‌چه برای بیمار خوب است، می‌تواند برای دستیابی به اهداف یا منافع خاص مورد استفاده قرار بگیرد تا

⁸ Danton Char

⁹ Nigam Shah

¹⁰ David Magnus

¹¹ The New England Journal of Medicine

به روش درمانی مدنظرش برسد) حتی ممکن است به جای تمرکز بر آنچه برای بیمار خوب است طوری استفاده شود که درمان را به سمت اهداف و منافع خاصی هدایت کند).

به عبارتی در این صورت، این داده‌ها هستند که پزشک را همراه خود می‌کنند، نه برعکس یعنی جوری که باید باشد. پزشکی را به عنوان یک هنر در نظر بگیرید. پزشک باید با بیمارانش به عنوان افرادی واقعی که زندگی خود را دارند ارتباط برقرار کند. گرچه هوش مصنوعی می‌تواند به ما کمک کند تا به هدف سلامتی به شکل بهتری دست یابیم، با در نظر گرفتن همه چیز، برای بعضی از بیماران روش‌های درمانی‌ای که شانس موفقیت کمتری دارند، انتخاب بهتری است. رویکردی صرفاً داده‌محور نمی‌تواند چنین چیزی را به ما بگوید. و ما باید نسبت به این مسئله محتاط باشیم که قدرت تکنولوژی ما را بیش از حد از خود بی‌خود نکند. چرا که از پیش می‌دانیم رضایت آزادانه و آگاهانه در عمل به‌دشواری به دست می‌آید و مؤسسات پزشکی بر رضایت بیماران تأثیرگذارند. اما با توجه به اضافه شدن جاذبه‌ی تکنولوژی و پوشیده ماندن توافق‌های حرفه‌ای، خطر آن است که با افزودن قدرت موجود در حرفه‌ی پزشکی به قدرت افزوده‌ی هوش مصنوعی، چنین گفته‌ای تبدیل به واقعیت شود که «کامپیوتر گفت که این دارو را مصرف کنیم».

رابطه‌ی میان پزشک و بیمار، در مرکزیت این حرفه و درک ما از اخلاقیات پزشکی جای دارد. اما استفاده از هوش مصنوعی می‌تواند به آرامی، یا حتی به شکلی رادیکال این مسئله را دگرگون کند. به طور ویژه، این که چگونه اهداف از نظر اخلاقی خیرخواهانه‌ی استفاده از هوش مصنوعی را به منظور بهبود مراقبت از بیمار به کار می‌بندیم، نیازمند بررسی‌های دقیق و محتاطانه است.

توانایی هوش مصنوعی در دستکاری و پردازش داده‌های بسیار گسترده ممکن است برای حل مسائل و شناسایی آن‌ها، ما را به سمت برتری دادن بی‌مورد به رویکردهای داده‌محور سوق دهد یا آن‌ها را بدل به تنها رویکردهایی که پیش رو داریم سازد. این می‌تواند به یکپارچگی اندیشیدن بیانجامد، حتی در مواردی که دلایلی برای بهره‌مندی از شیوه‌های تفکر و رویکردهای متنوع در دست داریم. افزون‌براین، می‌تواند عوامل دیگر را نیز تحت الشعاع قرار دهد و با این کار نه تنها شیوه‌ی تفکر ما، بلکه ارزش‌هایمان را نیز تحریف کند.

این که یک تصمیم چگونه و توسط چه کسی گرفته می‌شود و این که یک عمل چطور و توسط چه کسی انجام می‌شود، در بسیاری از موقعیت‌ها مسائلی حیاتی به‌شمار می‌روند؛ به‌ویژه زمانی که پای ارزش‌ها در میان است. آن والدی که نسبت به این موضوع که آیا واقعاً خودِ پسرش آن بیسکویت‌ها را درست کرده است یا نه، مشکوک بود،

حق دارد. شاید اگر پسر بچه بود که در وهله‌ی اول آن ربات را طراحی و ساخته بود، ادعای او اعتبار بیشتری داشت. اهمیت این عوامل، همچنین اهمیت بالقوه‌ی جایگزینی هوش انسانی یا تکمیل آن با هوش مصنوعی، مورد به مورد با یکدیگر متفاوت است.

استفاده از هیئت منصفه را در نظر بگیرید. همه می‌دانیم که هیئت منصفه دچار خطا می‌شوند و گاهی پاسخ‌های اشتباهی می‌دهند. همین الان هم الگوریتم‌ها به دادگاه‌های ایالات متحده کمک می‌کنند تا بر اساس داده‌هایی همچون اطلاعات مربوط به نرخ تکرار جرایم، در رابطه با صدور محکومیت یا عفو مشروط برخی از تصمیم‌گیری‌ها را انجام دهند که در نهایت اختلاف چشمگیری با حکم اولیه دارد. این نگرانی وجود دارد که این مسئله باعث تثبیت تعصبات موجود در برابر گروه‌های خاصی شود. اما فرض کنید به نقطه‌ای برسیم که گذاشتن تمام شواهد موجود در اختیار کامپیوتر منجر به صدور حکم دقیق‌تری نسبت به حکم هیئت منصفه شده است. در چنین موردی، کامپیوتر می‌تواند تمام داده‌ها را با سرعت، دقت و کارایی، با یکدیگر ترکیب و آن‌ها را تجزیه و تحلیل کند. این را مقایسه کنید با نحوه‌ی واقعی کار کردن هیئت منصفه که در طی آن افراد ممکن است درباره‌ی پرونده یادداشت‌های متفاوتی برداشته باشند، چیزهای متفاوتی از آن را به یاد بیاورند و حتی پس از گذشت ساعت‌ها مشورت و تأمل، همچنان دیدگاه‌های متفاوتی در رابطه با شواهد موجود داشته باشند. قدرت هوش مصنوعی در جمع‌آوری و تجزیه و تحلیل داده‌ها باعث می‌شود تا با استفاده از میان‌برهای زیادی زودتر به هدف برسیم.

اما این مثال به راحتی نشان می‌دهد که ما به چیزی بیش از راست و ریس کردن اوضاع نیازمندیم. حتی اگر با استفاده از یک ماشین به پاسخ دقیق‌تری برسیم، هنوز هم دلایل دیگری برای بها دادن به مشارکت متفاوت انسان‌ها به عنوان هیئت منصفه در اختیار داریم. تاریخ چگونگی جنگیدن برای اصلاحات قانونی و حقوق فردی و ارزش مورد قضاوت قرار گرفتن توسط «هیئت منصفه‌ای متشکل از هم‌تایان خود فرد» را در نظر بگیرید. شاید بخواهیم این وظیفه را به یک هوش مصنوعی محول کنیم، اما شاید هم نه.

البته که در مسیر رسیدن به حکم، تعصبی که انسان‌ها می‌توانند از خود نشان دهند و گیر کردن در نوسان میان احساسات ضعیفی؟ بالقوه به‌شمار می‌رود؛ اما همچنین نیروی محرک تحولات قانونی نیز هستند. نمونه‌هایی از «ابطال حکم هیئت منصفه»^{۱۲} وجود دارند، مواردی که هیئت منصفه تحت سلطه‌ی همان احساسات آزاردهنده‌ی انسانی درباره‌ی بی‌عدالتی به سادگی نتوانسته است مجرم را محکوم کند، علی‌رغم آن که متهم در چارچوب اعمال دقیق

¹² jury nullification

قانون، به وضوح گناه کار است. مهم نیست که یک ماشین چقدر خوب شواهد را ارزیابی می کند، هنوز تا توسعه‌ی ماشین‌هایی با فهم درستی از مفهوم عدالت، قابلیت توجه به مظلومان و بهره‌مند از استحکام اخلاقی برای به مبارزه طلبیدن دستگاه نظام حقوقی، فاصله‌ی زیادی داریم.

اما ایده‌ی کلی‌تر مبنی بر این که هیئت منصفه نقش منبع مستقل قضاوت به مثابه مخالفت با منافع اهدا شده به قدرتمندترین افراد را بر عهده دارد، به قوه‌ی خود باقی می ماند. چنان که لرد دولین^{۱۳} در مجلس اعیان بریتانیا^{۱۴} در سال ۲۰۰۴ خاطر نشان کرد: «محاكمه توسط هیئت منصفه چیزی بیش از ابزار [برقراری] عدالت و گرداننده‌ی قانون اساسی است؛ چراغی ست است که نشان می دهد آزادی زنده است.» هم چنین به این مسئله توجه داشته باشید که ویژگی بسیار مهم هوش مصنوعی یعنی جمع‌آوری اطلاعات و ترکیب بینش‌ها در بستر قانون به صورت مستقیم اصل پایه‌ای اخلاقی مربوط به استقلال هیئت منصفه را تضعیف می کند. این استقلال مقابله‌ای با احتمال همیشگی منافع شخصی قدرتمند است و دلیلی ارائه می دهد تا محاکمه را توسط هیئت منصفه ادامه دهیم و با چیزی بیش از صرف پردازش اطلاعاتی که در دادگاه مطرح می شوند، درگیر باشیم.

یک منتقد می تواند بگوید ما فقط به خاطر این که انسان‌ها به شدت غیر قابل اعتمادند، به این استقلال نیاز داریم. حرفه‌ی قانونی نمی تواند به تنهایی مسئول عدالت باشد، بلکه به هوش مصنوعی دقیقی نیاز داریم تا مشکل را حل کند و شاید با گذر زمان به این ایده عادت و نظام عدالت مان را به ماشین‌ها واگذار کنیم. اما کاملاً خیال پردازانه است که گمان کنیم فرار از عدم تعادل قدرت و منافع شخصی که دلیل نیاز به داشتن هیئت منصفه است، اصلاً ممکن است. در رابطه با بی حاصلی مشکل عدالت راه‌حل‌های دیگری به جز تقدیم کردن عدالت به دست ماشین‌ها وجود دارد؛ مثلاً یک نظام استیناف در دسترس و سریع. شاید در آینده هوش مصنوعی بتواند به قاضی‌ها و هیئت منصفه برای تصمیم‌گیری کمک کند، اما این فرق دارد با این انتظار که هوش مصنوعی ممکن است جایگزین انسان‌ها در تصمیم‌گیری‌های قانونی شود. حتی در این جا هم باید با احتیاط در رابطه با تأثیر هوش مصنوعی و این نکته بیاوریم که آیا در حال سوق دادن ما به سمت رویکرد فن‌سالارانه تری است یا نه. قانون از طریق مبارزات زیادی به منزله‌ی یک نظام سیاسی و اجتماعی انسانی توسعه یافته است. اما استفاده از هوش مصنوعی در قانون می تواند در طول زمان به تغییر این وضعیت کمک کند. ما باید با احتیاط این مسئله را با آگاهی کامل از مفاهیم زیادی که برای عدالت و دموکراسی موجودند، در نظر بگیریم.

¹³ Lord Devlin

¹⁴ House of Lords

یکی از جذابیت‌های بزرگ استفاده از هوش مصنوعی صرفاً مربوط به سرعت بالای آن در تجزیه و تحلیل داده‌هاست. کارآیی نیز یک حُسن است، اما بستگی به اهدافی دارد که براساس آن‌ها مورد استفاده قرار می‌گیرد. و البته به هیچ‌وجه تنها حُسن موجود نیست. اگر یک ماشین بتواند با سرعت و کارآمدی بیشتری کاری را انجام دهد، ممکن است وسوسه شویم تا بیش از آنچه سزاوارش است، از آن استفاده کنیم. و سرعت آن در به انجام رسیدن وظایفش ممکن است باعث شود تا مشکلات موجود در نحوه‌ی رسیدن به اهدافش را نادیده بگیریم. در نهایت احتمالاً برای چنین کارآیی‌ای که منجر به رسیدن به نتایج می‌شود ارزش زیادی قائل خواهیم شد.

اتحادیه‌ی ضدافترا^{۱۵} به کمک آزمایش‌دی^{۱۶} در دانشگاه کالیفرنیا، برکلی در حال درست کردن یک شاخص تنفر آنلاین^{۱۷} است که با استفاده از یادگیری ماشین تلاش می‌کند، نفرت پراکنی‌های^{۱۸} آنلاین را تشخیص دهد. این نمونه‌ای جالب است چون از خود تکنولوژی به منظور حل مشکلات تکنولوژی استفاده می‌کند؛ تکثیر نمونه‌ای جالب است چون از خود تکنولوژی به منظور حل مشکلات تکنولوژی استفاده می‌کند؛ تکثیر «نفرت پراکنی‌ها»ی آنلاین (ظاهراً) محصول تکنولوژی کامپیوتری است. اما سوءاستفاده از رقبا به هیچ‌وجه امر تازه‌ای نیست. خیلی پیش‌تر از آن‌که حتی ایده‌ی وب جهان‌گستر^{۱۹} در ذهن تیم برنرز لی^{۲۰} جرقه بزند، فیلسوف قرن هفدهمی فرانسوی، رنه دکارت، کار

تکنولوژی راه خودش را باز می‌کند. این جمله‌ی ترسناک شاید زیادی بدبینانه به نظر برسد؛ اما به عنوان مثال، حالا که تکنولوژی می‌تواند داده‌های زیاد را پردازش کند، پس به داده‌های بیشتری نیاز دارد. پس باید مدل زندگی ما تغییر کند و داده‌های شخصی بیشتری را به صورت عمومی به اشتراک بگذاریم.

ریاضیدانی که رقیب او بود، یعنی پیر دو فرما^{۲۱} را «آشغال» نامید. خاطر نشان کردن دودمان طولانی توهین‌ها خالی از لطف نیست، به‌ویژه با توجه به این‌که حرف‌هایی که در مورد هوش مصنوعی زده می‌شوند، ما را تشویق می‌کنند تا چنین مسائلی را بسیار جدید یا به صورت منحصر به فردی خطرناک قلمداد کنیم. راهی که از طریق آن بتوانیم در جهت وابستگی بیش از حد به طیف قابلیت‌های اندکی که هوش مصنوعی دارد قرار بگیریم، دقیقاً همین فرضیه‌های ترکیب

^{۱۵} The Anti-Defamation League (ADL) یک سازمان غیردولتی بین‌المللی است که مرکز آن در شهر نیویورک قرار دارد. اتحادیه ضد افترا که خود را مهم‌ترین سازمان حمایت از حقوق بشر در آمریکا قلمداد می‌کند، مهم‌ترین وظیفه‌ی خود را «مقابله با یهودستیزی و حمایت از حقوق مدنی شهروندان و هر نوع نژادپرستی و مقابله با هر نوع نفرت پراکنی» معرفی می‌کند. - م.

^{۱۶} Defense Language Aptitude Battery (D-Lab) آزمایشی که وزارت دفاع ایالات متحده‌ی آمریکا به منظور سنجش توانایی افراد برای یادگیری زبان‌های خارجی انجام می‌دهد تا بتوانند تعیین کند که چه کسی می‌تواند به عنوان زبان‌شناس نظامی تربیت شود. - م.

^{۱۷} Online Hate Index (OHI)

^{۱۸} hate speech

^{۱۹} World Wide Web

^{۲۰} Tim Berners-Lee

^{۲۱} Pierre de Fermat

شده است مبنی بر این که «تکنولوژی پیش می‌آورد = تکنولوژی حل می‌کند»^{۲۲}، همراه با این باور که تکنولوژی امروزی به صورت ویژه‌ای سرشار از خطرات اخلاقی تازه است.

خود مفهوم «نفرت‌پراکنی» بحث‌برانگیز است. برخی بر این باورند تلاش برای از بین بردن کاربردهای خاصی از زبان، چه توسط قانون چه توسط کسانی که مدیریت پلتفرم‌های آنلاین^{۲۳} را بر عهده دارند، در راستای رسیدن به اهدافی مانند از بین بردن تبعیض، ضروری‌اند. برخی نگران این مسئله‌اند که چنین چیزی خطری در برابر آزادی بیان به شمار می‌رود و بازتاب نوعی از سانسورشدگی است. بحث‌هایی هم بر سر این مسئله وجود دارد که آنچه از نظر برخی «نفرت‌پراکنی» به شمار می‌رود، از نظر دیگران صرفاً یک شوخی کنایه‌دار تلقی می‌شود و به دشواری می‌توانیم خارج از متن، چیزی را به عنوان «نفرت‌پراکنی» طبقه‌بندی کنیم. افزون‌براین، نگرانی‌هایی هم در این راستا وجود دارند که با توجه به حجم وسیع مطالبی که به صورت آنلاین پُست می‌شوند، هر گونه نظارت بر «نفرت‌پراکنی»ها، تکه‌تکه و نامنظم خواهد بود، همچنین ممکن است برخی از افراد یا گروه‌ها به صورت ناخواسته طعمه‌ی این نفرت‌پراکنی‌ها قرار بگیرند.

اگر می‌توانستیم نفرت‌پراکنی‌ها را به صورت خودکار تشخیص دهیم، احتمالاً به طبقه‌بندی و انسجام‌بخشی آن‌ها کمک می‌کرد. شاخص تنفر آنلاین می‌تواند با استفاده از پردازش سریع‌تر داده‌ها نسبت به یک انسان و به درستی و بدون تعصبات اعمال کردن سیاست‌های مربوطه، به این مسائل پاسخ دهد (البته این مسئله بستگی دارد به این که چگونه برنامه‌ریزی شده باشد. اگر به شیوه‌ای برنامه‌ریزی شده باشد که متعصب باشد، با بازدهی بسیار بالایی متعصب خواهد بود).

بنابراین مشکل اینجاست. اگر از این ایده هیجان‌زده شویم که هوش مصنوعی می‌تواند «نفرت‌پراکنی‌ها» را تشخیص دهد و آن‌ها را طبقه‌بندی و حذف کند، اما هم چنان از این واقعیت که جهان آنلاین دارد همه‌ی ما را به ترول‌های^{۲۴} سرشار از نفرت و وظیفه‌شناس تبدیل می‌کند، در هراس باشیم، می‌توانیم از این تکنولوژی با توجه به این مسئله استفاده کنیم که مشکلات دیگر مربوط به نفرت‌پراکنی تا حدی در این روش نادیده گرفته خواهند شد. بنابراین این امر می‌تواند این بحث را که به سختی می‌توانیم آن را به «نفرت‌پراکنی در برابر آزادی بیان» خلاصه کنیم در جهت خاصی قرار دهد. به عبارت دیگر، می‌تواند به ما کمک کند تا به ارزش‌هایمان شکل دهیم. می‌تواند به راحتی به تغییر

^{۲۲} اشاره به این که مشکلات حاصل از تکنولوژی را خود تکنولوژی حل می‌کند. -م.

^{۲۳} online platforms

^{۲۴} troll

شیوه‌های برقراری ارتباط آنلاین کمک کند، چون این نگرانی وجود دارد که بات‌های نفرت‌پراکنی، مردم را از پلت‌فرم مدنظرشان بیرون اندازد.

مزایای هوش مصنوعی شامل این موارد می‌شود: توانایی ویژه‌ی آن در به اشتراک گذاشتن داده‌ها برای رسیدن به دیدگاه جامعی از چیزها، قابلیت آن در کمک به محدود کردن تعصبات انسانی و سرعت و کارآمدی آن در انجام کارها. هوش مصنوعی می‌تواند ظرفیت انسانی را در تمام این موارد افزایش دهد. اما تمام این مزایا باید در برابر ارزش‌های دیگر ما سنجیده شوند. بدون این کار، قدرت هوش مصنوعی ممکن است ما را از خود بی‌خود کند و به آن اجازه دهیم تا مسئولیت تعیین نحوه‌ی اندیشیدن ما درباره‌ی برخی از مهم‌ترین ارزش‌ها و فعالیت‌هایمان را بر عهده بگیرد.

منبع: فلسفیدن