

بسم الله الرحمن الرحيم

عصر
فضای
مجازی

گزارش شماره ۸۰

شهر یور ۱۴۰۰



مرکز ملی فضای مجازی
پژوهشگاه فضای مجازی

گزارش مؤسسه AI Now در سال ۲۰۱۸

محتوای انتشار یافته در این اثر
الزاماً بیانگر دیدگاه مرکز ملی فضای مجازی نیست

تهیه شده در پژوهشگاه فضای مجازی
(گروه مطالعات اخلاقی فضای مجازی)

تهیه کننده: بهزاد خداقلی زاده
ناظر علمی: جواد درویش، محمدسجاد زارع
محمد مهدی نصرهرندی

حقوق مادی و معنوی این اثر متعلق به مرکز ملی فضای
مجازی است و استفاده از آن با ذکر منبع مجاز می باشد.

نشانی: تهران، میدان آرژانتین، خیابان بیهقی، نش
خیابان ۱۶ غربی، پلاک ۲۰
تلفن: ۰۲۱-۸۶۱۵۱۰۶۱
کد پستی: ۱۵۱۵۶۷۴۳۱۱

فهرست

۵ سخن نخست
۹ چکیده
۱۳ مقدمه

بخش اول

درباره‌ی موسسه‌ی AI Now ۲۱

بخش دوم

مشکلات تشدیدشده ۴۱

بخش سوم

راه حل‌های سال ۲۰۱۸ ۷۵

بخش چهارم

چه چیزی برای بعد نیاز است؟ ۹۷

جمع بندی ۱۲۳

منابع ۱۲۹

سخن نخست



فضای مجازی با شتاب شگرف و رو به تزایدی که در حال بسط و گسترش است تمام ساحات اجتماعی، اقتصادی، سیاسی و فرهنگی زندگی بشر را درنوردیده و هر روز بخش بزرگی از زندگی واقعی را در خود فرو برده و حیات متفاوت و جدیدی به آن می‌دهد. لذا به نظر می‌رسد دو نگاه کلان به فضای مجازی وجود دارد: نگاه اول که بالاخص در ابتدای رشد و تکوین فضای مجازی مسلط شده بود، آن را همچون ابزاری کنار سایر ابزارهای بشری تصویر می‌کرد که تنها طریقت داشت. اما نگاه دوم، در نتیجه رشد تحولات خیره‌کننده فضای مجازی و سایه گسترتری آن در حوزه‌ها و شئون بشر در یک دهه اخیر آن را چون سکویی می‌داند که بسیار فراتر از شأن ابزاری حیات انسان‌ها را سامان جدیدی داده و ادعای تمدن نوینی را دارد. رویکردی که از قضا از چشمان بصیر رهبر انقلاب نیز دور نمانده و انتظاری تمدنی از فضای مجازی در ایران را مطالبه داشته‌اند.

در همین راستا گزارش‌های عصر فضای مجازی تلاش می‌کند تا فهم سازمان‌ها و دستگاه‌های مرتبط با حوزه فضای مجازی را ارتقاء بخشیده و آن‌ها را برای مواجهه فعال و خردمندانه با تحولات این عرصه مهیا سازد.

سید ابوالحسن فیروزآبادی
دبیر شورای عالی و رئیس مرکز ملی فضای مجازی

چکیده



با گسترش دامنه‌ی نفوذ فناوری هوش مصنوعی در حوزه‌های متعدد اجتماعی، مسائل بسیاری برای عامه‌ی مردم و گروه‌های خاص به وجود آمده است. استفاده از سیستم‌های هوش مصنوعی در فناوری‌های نظارتی، بکارگیری سیستم‌های تصمیم‌گیری خودکار در دستگاه‌های اجرایی و آزمایش سیستم‌های هوش مصنوعی در فضاهای کنترل نشده، از جمله مسائلی هستند که موسسه‌ی AI Now به آن‌ها پرداخته است. در این گزارش سعی می‌شود، بر اساس آسیب‌شناسی مسائل مذکور، راه‌حلهایی برای آن‌ها ارائه شود. این موسسه که توسط زنان هدایت می‌شود، به نقد و بررسی وضعیت کار در شرکت‌های هوش مصنوعی و آسیب‌هایی که این فضاها برای گروه‌های به حاشیه رانده شده و زنان دارند، می‌پردازد.

واژگان کلیدی: هوش مصنوعی، تبعات اجتماعی، سیستم‌های تصمیم‌گیری خودکار، جانبداری، بی‌طرفی.

مقدمه



چالش‌های اجتماعی هوش مصنوعی در سال ۲۰۱۸

به رغم وجود دورنمایی [نامیدکننده] از نابرابری فزاینده، پوپولیسیم سیاسی و رسوایی‌های صنعتی، سال گذشته شاهد ادغام سیستم‌های قدرتمند هوش مصنوعی در نهادهای اجتماعی اصلی بودیم. تا به حال، جنبش‌های قابل توجهی از داخل و خارج شرکت‌های فناوری وجود داشته‌است که خواهان پاسخ‌گویی و عدالت بیشتر [در این موضوع] هستند. گزارش AI Now در سال ۲۰۱۸ بر روی این موضوعات تمرکز می‌کند و به بررسی نقش تشکلهای، تنظیم مقررات و شکاف میان اصول اخلاقی هوش مصنوعی و پاسخ‌گویی هدفمند می‌پردازد. بطور خلاصه، سالی دراماتیک در حوزه‌ی هوش مصنوعی بود. قابل توجه‌ترین داستان سال این می‌توانست باشد که به طور معمول در هر سالی، Cambridge Analytica، با استفاده از داده‌های شبکه‌های اجتماعی و تبلیغات الگوریتمی هدفمند، به دنبال دستکاری انتخابات ملی در ایالات متحده و بریتانیا است. اما در سال ۲۰۱۸ این تنها یکی از رسوایی‌های متعدد بود. مثلاً چندین فاجعه در فیس‌بوک اتفاق افتاد؛ از جمله رخنه‌ی گسترده‌ی اطلاعاتی در ماه سپتامبر،

طرح دعوی در سطوح مختلف قانونی بخاطر تبعیض، اتهام به ترویج نسل‌کشی اقلیت‌ها در میانمار، خطر نقض بالقوه قانون مسکن‌سازی عادلانه و میزبانی از تعداد عظیمی از حساب‌های جعلی روسی. در این سال، مدیران این شرکت بارها برای شهادت دادن به دادگاه احضار شدند. هم‌چنین مارک زاکربرگ^۱ در ماه آوریل به سنای آمریکا و در ماه مه به پارلمان اروپا احضار شد. زاکربرگ به هنگام حضورش در کنگره برای شهادت، بیش از ۳۰ بار از فناوری‌های هوش مصنوعی به عنوان درمان همه‌ی مسائل شرکت یاد کرد؛ به خصوص در مسائلی پیچیده مانند سانسور، عدالت و بی‌طرفی محتوا.

اما فیس بوک تنها شرکتی نبود که در بحران بود. در ماه مارس گزارشی منتشر شد مبنی بر این که گوگل در حال ساخت سیستم‌های هوش مصنوعی برای برنامه‌ی نظارت پهبادی وزارت دفاع ایالات متحده‌ی آمریکا، یعنی پروژه‌ی Maven است. این خبر باعث شد که یک موج بی‌سابقه از جانب تشکل‌های کارگری و کارمندی حوزه‌ی فناوری به وجود بیاید و مخالفت خود را در این باره نشان بدهند. در ماه ژوئن، زمانی که دولت ترامپ سیاست جداسازی خانواده را مطرح کرد که کودکان مهاجر را به اجبار از والدینشان جدا می‌کرد، کارمندان آمازون، Salesforce و مایکروسافت همگی از شرکت‌های خود خواستند تا قراردادهای‌شان را با سازمان مهاجرت و گمرک آمریکا (ICE) به پایان برسانند. کم‌تر از یک ماه بعد معلوم شد که ICE الگوریتم ارزیابی ریسک خود را به شکلی اصلاح کرده بود که تنها می‌توانست یک نتیجه را تولید کند: سیستم برای ۱۰۰ درصد مهاجران بازداشت‌شده، حکم «زندانی» را پیشنهاد می‌کرد.

1. Mark Zuckerberg

در طول سال ۲۰۱۸، سیستم‌های هوش مصنوعی همچنان بر روی جمعیت‌های زنده و در حوزه‌های پرخطر آزمایش می‌شدند که عواقب جدی‌ای را نیز در پی داشت. در ماه مارس، خودروهای خودگردان چند راننده و عابر پیاده را کشتند. پس از آن، در ماه می، یک سیستم تشخیص صدا در بریتانیا که برای تشخیص کلاهبرداری در حوزه‌ی مهاجرت طراحی شده بود، به اشتباه موجب لغو شدن هزاران ویزا و اخراج افراد از خاک بریتانیا شد. اسناد منتشر شده در ماه جولای نشان داد که ماشین‌واتسون که متعلق به شرکت IBM است، برای درمان سرطان توصیه‌های خطرناک و نادرست پیشنهاد می‌داد. همچنین یک پژوهش در ماه سپتامبر نشان داد که IBM با همکاری اداره پلیس شهر نیویورک (NYPD) سعی دارد تا با استفاده از تصاویر گرفته شده با دوربین‌های پلیس از هزاران تن از مردم در خیابان‌های نیویورک که بدون اجازه و خبرداشتن آن‌ها اتفاق افتاده است، سامانه‌ای برای تشخیص قومیت بسازد تا جستجوی چهره‌ی افراد را بر اساس قومیت ممکن سازد.

اتفاقاتی که ذکر شد، تنها نمونه‌هایی از مجموعه‌ای غیرعادی از حوادث سال ۲۰۱۸ است. در واکنش [به این اتفاقات] موجی از انتقادات گسترده، به همراه تقاضا برای پاسخگویی بیشتر از جانب شرکت‌های فناوری و سیستم‌هایی که توسط آن‌ها ساخته شده است، به وجود آمده است. در نتیجه، برخی شرکت‌ها فراخوان‌هایی عمومی برای دولت ایالات متحده‌ی آمریکا تنظیم کرده‌اند تا برای فناوری‌هایی نظیر تشخیص چهره مقرراتی وضع کند. دیگر شرکت‌ها [اسنادی ذیل عنوان] اصول اخلاقی هوش مصنوعی منتشر کرده‌اند و برای ایجاد

راه‌حل‌های فنی برای مسئله‌ی جانبداری و تبعیض در سیستم‌های هوش مصنوعی، تلاش‌هایشان را افزایش داده‌اند. اما بسیاری از این رویکردهای اخلاقی و فنی، نگاهی محدود نسبت به مسائل دارند؛ این رویکردها نه با بافت تاریخی یا اجتماعی دست و پنجه نرم می‌کنند، و نه برای پاسخگویی به عموم، نظارت عمومی و دخالت دادن رویه‌های عادلانه، سازوکاری را ارائه می‌دهند. تقریباً ناممکن است که عموم مردم، این [گزاره] را که «به مسائل جاری [هوش مصنوعی] پرداخته شده است»، را تصدیق کنند.

همانطور که بسیاری از محققان اشاره کرده‌اند، فرهنگ رازداری صنعتی و قانونی که بر توسعه هوش مصنوعی نیز چیره است، یک مانع مهم پاسخ‌گویی این فناوری است. فرهنگ صنعت است که بسیاری از فناوری‌های هوش مصنوعی را مانند «جعبه‌ی سیاه» خلق می‌کند. بسیاری از اجزای اصلی سازنده‌ای که برای درک سیستم‌های هوش مصنوعی و حصول اطمینان از پاسخ‌گویی آن‌ها لازم است - از داده‌های آموزش‌دهنده و مدل‌های داده و کدهای دستوردهنده‌ی کارکرد الگوریتمی گرفته تا اجرای دستورها و نرم‌افزار و تصمیماتی که طراحی و توسعه را هدایت می‌کنند - به ندرت برای بازنگری در دسترس هستند و بخاطر قوانین رازداری شرکت پنهان می‌شوند.

هم‌چنین شکاف پاسخ‌گویی فعلی ناشی از انگیزه‌هایی است که تحقیقات فنی هوش مصنوعی را با سرعت سرسام‌آوری جلو می‌برد. تشویق کردن به «نوآوری»، چاپ سریع مقالات و اضافه کردن یک ایده‌ی تازه به حوزه فنی، باعث شده تا آهنگی پرشتاب در حوزه‌ی هوش مصنوعی و به طور کلی در رشته‌های فنی شروع به نواختن

بکند. این آهنگ به قیمت مواجهه با سوالاتی تجربی درباره‌ی بافتار و کاربرد [سیستم‌های هوش مصنوعی] و درگیری عمیق با نگرانی‌های اخلاقی نواخته می‌شود. به طور مشابه، شرکت‌های فناوری با [سیاست] تشویق به «راه‌اندازی و تکثیر» سیستم‌هایشان هدایت می‌شوند، در حالی که فرض می‌کنند پرسش‌های پیچیده اجتماعی و سیاسی توسط سازمان‌های سیاست‌گذار و حقوقی، مدیریت و پاسخ داده می‌شوند؛ این امر موجب می‌شود که توسعه دهندگان و ادارات فروش [شرکت‌های هوش مصنوعی] از مسئولیت توجه به زیان‌های بالقوه [محصولاتشان] آزاد باشند. فرهنگ «سریع حرکت کن و همه چیز را درهم‌بشکن!» برای تضمین پاسخ‌گویی عمومی هدفمند یا مشارکت جمعیت‌های آسیب‌پذیر انگیزه کمی باقی می‌گذارد. این مسئله همان‌قدر که استفاده‌ی بی‌امان از هوش مصنوعی در حوزه‌های حساس اجتماعی و سیاسی خطرانی را برای جوامع محروم و آسیب‌دیده به وجود می‌آورد، به طور خاص مشکل‌زاست.

چالش ایجاد مدیریتی بهتر و پاسخ‌گویی گسترده‌تر برای هوش مصنوعی، هنگامی که این سیستم‌ها در تاروپود دولت و نهادهای عمومی تنیده شده‌اند، با مسائل خاصی مواجه می‌شود. فقدان شفافیت، آگاهی، مشارکت معنادار، پاسخ‌گویی و نظارت، موانع ساختاری جدی‌ای را برای رویه‌های عادلانه و جبران خسارت ناشی از تصمیمات ناعادلانه و تبعیض‌آمیز ایجاد می‌کند.

از آن‌جا که سیستم‌های هوش مصنوعی به صورت گسترده‌تر در نهادهای اداره‌کننده‌ی زندگی روزمره نقش دارند، در گزارش امسال تعدادی از مسائل اضطراری که با آن‌ها مواجه هستیم را ارزیابی

می‌کنیم. در این گزارش بر روی بزرگترین بازیگران صنعت تمرکز می‌کنیم؛ چرا که تعداد شرکت‌هایی که قائل به تولید سیستم‌های هوش مصنوعی در مقیاس بالا هستند، بسیار کم است، در حالی که قدرت و دسترسی آن‌ها در ابعاد جهانی است. ما گستره‌ای از واکنش‌ها که در حال حاضر از جانب صنعت، دولت‌ها، پژوهشگران، فعالان و جامعه مدنی [نسبت به مسائل هوش مصنوعی ارائه شده است را] در مقیاس بزرگ ارزیابی می‌کنیم. مجموعه‌ای از رویکردهای اساسی را پیشنهاد و ده توصیه دقیق را ارائه می‌کنیم. در نهایت، جدیدترین استراتژی‌های تحقیق و سیاست‌گذاری را به اشتراک می‌گذاریم که می‌تواند به پاسخگویی بیشتر و درک غنی‌تر از سیستم‌های هوش مصنوعی در زمینه‌ی اجتماعی گسترده‌تر کمک کند.

بخش اول

درباره‌ی موسسه‌ی AINow



درباره‌ی موسسه‌ی AI Now

موسسه‌ی AI Now، یک موسسه‌ی تحقیقاتی میان‌رشته‌ای در دانشگاه نیویورک است که فعالیت‌های خود را به بررسی تبعات اجتماعی فناوری‌های هوش مصنوعی اختصاص داده است. این موسسه اولین مرکز تحقیقاتی دانشگاهی است که به طور خاص بر اهمیت اجتماعی هوش مصنوعی تمرکز کرده است. موسسه‌ی AI Now، که توسط کیت کرافورد^۱ و مردیت وایت‌تاکر^۲ بنیان گذاشته شده و هدایت می‌شود، یکی از معدود موسسه‌های هوش مصنوعی است که توسط زنان هدایت می‌شود.

موسسه‌ی AI Now در ائتلاف گسترده‌ای از ذینفعان^۳، از جمله پژوهشگران دانشگاهی، صنعت، جامعه مدنی، سیاست‌گذاران و جمعیت‌های آسیب‌دیده کار می‌کند تا مسائلی را شناسایی کرده و پاسخ بدهد که با رواج‌یافتن سریع هوش مصنوعی در سرتاسر حوزه‌های اجتماعی رشد یافته است. موسسه‌ی AI Now، پژوهش‌هایی میان‌رشته‌ای را به وجود می‌آورد تا این اطمینان را به وجود بیاورد که سیستم‌های هوش مصنوعی به جوامع و بافتارهایی که باید به آن‌ها خدمت‌رسانی کنند، پاسخ‌گو هستند و به شیوه‌هایی که کار گرفته

1. Kate Crawford
2. Meredith Whittaker
3. Stakeholders

می‌شوند که عدالت^۱ و برابری^۲ را تقویت می‌کنند. دستور کار پژوهشی فعلی موسسه بر روی چهار حوزه اصلی تمرکز دارد: جانبداری^۳ و شمول^۴، حقوق و آزادی‌ها، کار و خودکارسازی و زیرساخت‌های ایمنی و حیاتی. آخرین مکتوبات ما عبارتند از:

- شکایت کردن از الگوریتم‌ها^۵، گزارشی مهم از ارزیابی پرونده‌های اخیر دادگاهی درباره‌ی استفاده دولت از الگوریتم.
- کالبدشناسی یک سیستم هوش مصنوعی^۶، یک نقشه‌ی بزرگ مقیاس که با همکاری آزمایشگاه SHARE تولید شده است که به بررسی نیروی کار انسانی، اطلاعات و منابع جهانی مورد نیاز برای تولید یک دستگاه Amazon Echo، می‌پردازد.
- گزارش ارزیابی تاثیر الگوریتمی^۷، که به جمعیت‌های آسیب‌دیده و ذینفعان کمک می‌کند که استفاده از هوش مصنوعی و تصمیم‌گیری الگوریتمی در سازمان‌های عمومی را ارزیابی کنند.
- جعبه‌ابزار سیاست پاسخ‌گویی الگوریتمی^۸، که برای افرادی مناسب است که علاقه‌مند به دانستن استفاده‌های دولتی از سیستم‌های الگوریتمی هستند.

ما همچنین میزبان کارگاه‌های تخصصی و رویدادهای عمومی در گستره وسیعی از موضوعات هستیم. کارگاه ما در مورد **مهاجرت، داده و اتوماسیون در دوره‌ی ترامپ**^۹، با همکاری مرکز عدالت برنان^{۱۰} و مرکز حریم خصوصی و تکنولوژی دانشگاه قانون جورج تاون^{۱۱}، بر موضوع استفاده‌ی دولت ترامپ از گردآوری داده، تحلیل پیشگويانه^{۱۲} و

1. Justice
2. Equity
3. Bias
4. Inclusion
5. Litigating Algorithms
6. Anatomy of an AI System

7. Algorithmic Impact Assessment (AIA) Report
8. Algorithmic Accountability Policy Toolkit
9. Immigration, Data, and Automation in the Trump Era
10. Brennan Center for Justice
11. Center for Privacy and Technology at Georgetown Law
12. predictive analytics

یادگیری ماشین^۱ به منظور هدف قرار دادن جوامع مهاجر توجه داشت. **گروه کاری پیدایش داده^۲**، متخصصانی از صنعت و دانشگاه را فرا می‌خواند تا به بررسی جنبه‌های مهندسی تولید و نگهداری مجموعه داده‌ها بپردازند. هم‌چنین میزگرد ما درباره **یادگیری ماشینی، نابرابری و جانبداری^۳** در برلین، با همکاری آکادمی رابرت بوش^۴، محققان و سیاستگذاران را از سراسر اروپا گرد هم جمع می‌کند که به مسائل جانبداری، تبعیض و بی‌طرفی^۵ در فناوری یادگیری ماشین و فناوری‌های مرتبط بپردازند.

سمپوزیوم عمومی سالانه‌ی ما، رهبران حوزه‌ی دانشگاهی، صنعت، دولت و جمعیت‌های مدنی را فرا می‌خواند تا بزرگ‌ترین چالش‌هایی که با آن‌ها روبرو هستیم را، هم‌زمان که هوش مصنوعی در زندگی روزمره‌ی ما وارد می‌شود، مورد بررسی قرار دهیم. سمپوزیوم موسسه‌ی AI Now در سال ۲۰۱۸، به تلاقی اخلاق، شکل‌سازی و پاسخ‌گویی هوش مصنوعی پرداخت و رویدادهای بسیار مهم سال گذشته را بررسی کرد. در این برنامه که رایگان و برای عموم آزاد بود، بیش از ۱۰۰۰ نفر ثبت‌نام کردند. اطلاعات بیشتر در سایت www.ainowinstitute.org در دسترس است.

توصیه‌ها

۱- دولت‌ها باید با گسترش اختیارات نهادهایی که در حوزه‌هایی خاص ورود می‌کنند، هوش مصنوعی را [به نحوی] سامان‌دهی کنند که این فناوری‌ها با توجه به حوزه‌ی خدمت‌رسانی‌شان،

1. Machine learning
2. The Data Genesis Working Group
3. Machine Learning, Inequality and Bias

4. Robert Bosch Academy
5. Fairness

نظارت^۱، بررسی دقیق^۲ و کنترل شوند. به کارگیری سیستم‌های هوش مصنوعی، بدون وجود مدیریت، نظارت یا نظام‌های پاسخگویی کافی، به سرعت در حال رشد است. حوزه‌هایی مانند سلامت، آموزش، عدالت کیفری و رفاه، همگی تاریخچه، چارچوب‌های مقرراتی و خطرات خاص خود را دارند. با این همه، یک استاندارد عمومی هوش مصنوعی به سختی با الزامات تخصصی یک حوزه‌ی خاص که برای تنظیم دقیق مقررات لازم است، تطابق پیدا می‌کند. ما به رویکرد حوزه‌خاص^۳ نیاز داریم که فناوری را اولویت قرار ندهد، بلکه بر کاربردهای آن فناوری در حوزه‌ای خاص تمرکز کند. مثال‌های مفید از رویکردهای حوزه‌خاص عبارتند از اداره‌ی هوانوردی فدرال ایالات متحده و اداره‌ی ملی ایمنی حمل و نقل جاده‌ای.

۲- برای حفاظت از مردم، فناوری تشخیص چهره^۴ و تشخیص تمایلات^۵ به تنظیم مقررات سختگیرانه نیاز دارد. چنین مقرراتی باید شامل قوانین ملی باشد که به نظارت قوی، حذف محدودیت‌ها و شفافیت عمومی نیاز دارد. جوامع باید این حق را داشته باشند که استفاده از این فناوری‌ها در حوزه‌های عمومی و خصوصی را نپذیرند. با توجه به خطرات سرکوب‌گرانه و مداوم نظارت شدید و گسترده، صرفِ آگاهی عمومی نسبت به استفاده از آن‌ها کافی نیست و باید حد بالایی از رضایت هر فرد [جهت بکارگیری این فناوری‌ها] وجود داشته باشد. تشخیص تمایلات شایسته‌ی توجه خاص‌تری است. تشخیص تمایلات، زبرده‌ای از تشخیص چهره است که بر اساس تصاویر و ویدئوی چهره، چیزهایی مانند شخصیت،

1. oversee
2. audit
3. sector-specific

4. Facial recognition
5. affect recognition

احساسات درونی، سلامت ذهنی و «مشارکت فرد شاغل در کار» را تشخیص می‌دهد. این ادعاها توسط شواهد علمی محکم پشتیبانی نمی‌شوند و به شکل‌های غیراخلاقی و غیرمسئولانه‌ای در حال استفاده شدن هستند که ما را به یاد شبه‌علمی مانند جرمه‌خوانی^۱ و سیماشناسی^۲ می‌اندازد. ارتباط یافتن فناوری تشخیص تمایلات با مسائلی مانند استخدام، دسترسی به بیمه، آموزش و امنیت عمومی، در هر دو سطح شخصی و عمومی، خطرات عمیقی را می‌آفریند.

۳- صنعت هوش مصنوعی نیاز مبرمی به رویکردهایی جدید نسبت به حکمرانی دارد. همانطور که این گزارش نشان می‌دهد، ساختارهای حکمرانی داخلی در اغلب شرکت‌های فناوری قادر به تضمین پاسخ‌گویی سیستم‌های هوش مصنوعی نیستند. مقررات دولتی مولفه‌ی مهمی است اما شرکت‌های پیشرو در صنعت هوش مصنوعی نیز نیاز به ساختارهای درونی پاسخگویی دارند که فراتر از دستورالعمل‌های اخلاقی برود. این ساختارها باید مسائلی مانند حضور نمایندگان کارمندان در هیئت مدیره، وجود هیئت‌های مشاوره‌ی اخلاقی و اجرای مستقل اقدامات مربوط به کنترل و شفافیت را شامل بشود. متخصصان شخص ثالث^۳ باید توانایی بررسی دقیق سیستم‌های هوش مصنوعی کلیدی و افشاگری درباره‌ی آن‌ها داشته باشند و شرکت‌ها باید این را تضمین کنند که از «صدر تا ذیل» زیرساخت‌های هوش مصنوعی آن‌ها از جمله کاربرد و استفاده‌ی نهایی آن‌ها قابل بررسی است.

۴- شرکت‌های هوش مصنوعی باید از رازداری تجاری^۴ و دیگر دعاوی

حقوقی‌ای چشم‌پوشی کنند که بر سر راه پاسخگویی در حوزه‌ی عمومی قرار می‌گیرد. فروشندگان و توسعه‌دهندگان که هوش مصنوعی و سیستم‌های تصمیم‌گیری خودکار^۱ را برای استفاده‌های دولتی می‌سازند باید با چشم‌پوشی از رازداری تجاری و سایر دعوای حقوقی که مانع از بررسی و فهم دقیق نرم‌افزار آن‌ها می‌شود، موافقت نکنند. قوانین رازداری شرکت‌ها مانعی بر سر راه رویه‌های عادلانه قانونی^۲ هستند. این قوانین باعث می‌شوند که سیستم‌هایی مبهم و غیرپاسخگو به وجود بیاید که ارزیابی‌های جانب‌داری آن‌ها یا اعتراض به تصمیمات آن‌ها را مشکل می‌کند. هر کسی که این فناوری‌ها را برای استفاده در حوزه‌ی عمومی تدارک می‌کند، باید از فروشندگان بخواهد که آن‌ها پیش از هرگونه توافقی باید از این دعوای [قانونی] چشم‌پوشی کند.

۵- شرکت‌های فناوری باید امنیت منتقدان خیرخواه، تشکل‌های کارمندی و سوت‌زن‌های اخلاقی^۳ را تامین کنند. سازماندهی و مقاومت تشکل‌های کارمندی شرکت‌های فناوری به عنوان نیرویی به نفع پاسخگویی و تصمیم‌سازی اخلاقی در شرکت‌های فناوری ظهور کرده است. شرکت‌های فن‌آوری باید از توانایی کارکنان‌شان در افشاگری و تصمیم‌سازی اخلاقی در پروژه‌هایشان، حمایت کنند. این حمایت‌ها باید شامل سیاست‌های روشنی باشد که با منتقدان خیرخواه همکاری و از آن‌ها محافظت کند و این را تضمین کند که کارکنان حق دارند از آن‌چه بر روی آن کار می‌کنند باخبر باشند و این اطمینان را بدهد که کارکنان توانایی برهیز کردن از انجام پروژه‌ها را بدون تلافی و مجازات دارند.

1. automated decision systems
2. due process
3. ethical whistleblowers

همان‌گونه که از سوت‌زن‌هایی که نگران منافع عمومی هستند باید محافظت شود، کارکنانی که نگرانی‌های اخلاقی را افزایش می‌دهند نیز باید مورد حمایت قرار بگیرند.

۶- موسسات حمایت از مصرف‌کننده باید قوانین «صداقت در تبلیغات» را برای محصولات و خدمات هوش مصنوعی اعمال کنند. تبلیغات غلوآمیز درباره‌ی هوش مصنوعی به شدت در حال رشد است و منجر به ایجاد شکافی فزاینده میان وعده‌های تجاری و عملکرد واقعی محصول می‌شود. وجود این شکاف‌ها که موجب بالارفتن احتمال خطر برای افراد و مشتریان تجاری می‌شود، اغلب عواقب ناگواری را برای آن‌ها در پی دارد. مانند بسیاری از محصولات و خدمات دیگری که پتانسیل آسیب جدی یا بهره‌برداری از جوامع را دارند، فروشندگان سیستم‌های هوش مصنوعی نیز باید استانداردهای سفت و سختی را برای آن‌چه وعده می‌دهند، بپذیرند؛ مخصوصاً زمانی که شواهد علمی کافی‌ای برای وعده‌هایشان وجود نداشته باشد و عواقب بلندمدت آن‌ها نامعلوم باشد.

۷- شرکت‌های فن‌آوری باید از «مدل خط لوله» فراتر بروند و باید متعهد شوند که پاسخگوی مسئله‌ی اعمال محرومیت و تبعیض در محیط‌های کاری خود باشند. شرکت‌های فناوری و حوزه هوش مصنوعی به عنوان یک کل بر روی «مدل خط لوله» متمرکز شده‌اند و به دنبال آموزش و استخدام کارمندان متنوعی هستند. این امر اگرچه مهم است، اما به نظر می‌رسد که این مدل اتفاقاتی را نادیده می‌گیرد که به هنگام استخدام در محیط‌های کاری رخ می‌دهد؛ اتفاقاتی که در آن براساس جنسیت، نژاد یا

معلولیت، افراد را دچار محرومیت و آزار کرده یا به طور سیستماتیکی آن‌ها را کم‌ارزش تلقی می‌کنند. شرکت‌ها باید مسایل ژرف‌تری را که در محیط‌های کار بروز می‌دهد مورد بررسی قرار دهند و هم‌چنین باید رابطه‌ی بین فرهنگ‌های پردشده^۱ در محیط‌های کاری و محصولات ساخته‌شده‌ای که می‌تواند ابزارهایی را تولید کند که تداوم‌بخشِ تعصب و تبعیض باشند را بررسی کنند. این تغییر تمرکز باید با اقداماتی عملی همراه باشد؛ از جمله تعهد برای پایان دادن به نابرابری فرصت‌ها و دستمزد و وجود معیارهای شفافیت درباره استخدام.

۸- بی‌طرفی، پاسخگویی و شفافیت در هوش مصنوعی نیاز به گزارشی دقیق از «زنجیره‌ی کامل تامین^۲» دارد. برای معنادار بودن پاسخگویی یک سیستم هوش مصنوعی، نیاز است که اجزای سازنده‌ی یک سیستم هوش مصنوعی و زنجیره کامل تامین آن را دنبال کنیم؛ یعنی شرح دقیق منابع و کاربرد داده‌ی آموزش‌دهنده‌ی ماشین، داده‌ی تست، مدل‌ها، رابط‌های برنامه کاربردی (APIs)^۳ و دیگر اجزای زیرساختی که در طول چرخه عمر محصول دخیل هستند. این شرح را شرح «زنجیره کامل تامین» سیستم‌های هوش مصنوعی می‌نامیم و این شرطی ضروری برای شکل مسئولانه‌تری از بررسی دقیق این سیستم‌هاست. زنجیره کامل تامین موجب فهمیدن هزینه‌کرد واقعی زیست‌محیطی و نیروی کار در سیستم‌های هوش مصنوعی می‌شود. این زنجیره، فهم ما نسبت به مصرف انرژی، استفاده از نیروی کار کشورهای در حال توسعه برای ایجاد محتوا و ایجاد داده‌های آموزش‌دهنده

1. exclusionary cultures
2. full stack supply chain
3. application program interfaces

و اتکا به کلیک‌کارها^۱ برای توسعه و نگهداری سیستم‌های هوش مصنوعی را بیشتر می‌کند.

۹- به منظور اقامه دعوی حقوقی، سازمان‌دهی تشکل‌های کارگری، مشارکت اجتماعی درباره‌ی مسائل پاسخگویی هوش مصنوعی، بودجه و حمایت بیشتر مورد نیاز است. بیشتر افرادی که در معرض آسیب سیستم‌های هوش مصنوعی قرار دارند، اغلب قادر به مقابله با نتایج این سیستم‌ها نیستند. به منظور ایجاد سازوکارهای قوی جهت جبران قانونی خسارات و مشارکت مدنی در این مسائل به حمایت بیشتری نیاز داریم. این حمایت‌ها باید شامل حمایت از فعالان عمومی‌ای باشد که نماینده‌ی آن دسته از افرادی هستند که بخاطر تصمیم‌سازی الگوریتمی، از خدمات اجتماعی محروم شده‌اند؛ همچنین این حمایت‌ها باید شامل جوامع مدنی و تشکل‌های شغلی بشود که از گروه‌هایی پشتیبانی می‌کنند که در معرض خطر از دست دادن شغل و بهره‌برداری سیستم‌های هوش مصنوعی قرار دارند؛ و در نهایت این حمایت‌ها باید شامل زیرساخت‌های جامعه‌محوری باشد که مشارکت عمومی را میسر می‌سازد.

۱۰- برنامه‌های تحصیلی دانشگاهی مرتبط با هوش مصنوعی باید فراتر از علوم کامپیوتر و رشته‌های مهندسی توسعه پیدا کنند. هوش مصنوعی به عنوان یک رشته میان‌رشته‌ای آغاز به کار کرد، اما در طول این دهه‌ها به یک رشته فنی تقلیل یافته است. با افزایش کاربرد سیستم‌های هوش مصنوعی در حوزه‌های اجتماعی، این حوزه باید جهت‌گیری تحصیلی و پژوهشی خود را

نیز وسعت دهد. این نکته به معنی توجه داشتن به شکل‌های مختلف تخصص در رشته‌های علوم اجتماعی و علوم انسانی [ابر موضوع هوش مصنوعی] است. تلاش‌های صنعت هوش مصنوعی که واقعا آرزوی پرداختن به تبعات اجتماعی را دارد، نمی‌تواند تنها در حوزه‌ی علوم و مهندسی کامپیوتر باقی بماند؛ حوزه‌ای که اساتید و دانشجویان در آن، برای تحقیق در دنیای اجتماعی آموزش ندیده‌اند. وسعت‌بخشی به نگاه پژوهشی در زمینه‌ی هوش مصنوعی، توجه عمیق‌تری را به زمینه‌های اجتماعی در پی خواهد داشت.

چکیده‌ی اجرایی

در دل تمام رسوایی‌های سیل‌واری که در سال ۲۰۱۸ در حوزه‌ی هوش مصنوعی اتفاق افتاد، پرسش‌هایی درباره‌ی مسئله‌ی پاسخگویی وجود دارد: زمانی که سیستم‌های هوش مصنوعی به ما آسیب می‌رسانند، مسئولیت با چه کسی است؟ چگونه این آسیب‌ها را می‌فهمیم و چگونه آن‌ها را درمان می‌کنیم؟ نقاط مداخله کجا هستند و چه تحقیقات و مقررات اضافی مورد نیاز است تا اطمینان حاصل شود که این مداخلات موثر هستند؟ در حال حاضر پاسخ‌های کم و ناکافی‌ای به این پرسش‌ها وجود دارد و چارچوب‌هایی که در حال حاضر فناوری هوش مصنوعی را مدیریت می‌کنند، قادر به تضمین پاسخگویی نیستند. همانطور که فراگیری، پیچیدگی و ابعاد [استفاده از] این سیستم‌ها در حال رشد است، فقدان پاسخ‌گویی و نبود نظارت

معنادار - از جمله پاسداشت اساسی از مسئولیت پذیری، تعهد قانونی و رویه‌های عادلانه - به نگرانی اضطراری و فزاینده‌ای تبدیل شده است. براساس گزارش‌های ما در سال‌های ۲۰۱۶ و ۲۰۱۷، گزارش AI Now در سال ۲۰۱۸ با این مشکل مرکزی (پاسخگویی) دست و پنجه نرم می‌کند و به مسایل کلیدی زیر می‌پردازد:

۱- شکاف در حال گسترش مسئولیت‌پذیری در هوش مصنوعی، که به قیمت آسیب دیدن حداکثری عده‌ای، به نفع کسانی است که این سیستم‌ها را می‌سازند و به کار می‌گیرند؛

۲- استفاده از هوش مصنوعی برای حداکثری کردن و تقویت نظارت، به ویژه توامان با مسئله‌ی تشخیص چهره و تشخیص تمایلات که موجب افزایش بالقوه‌ی کنترل متمرکز و سرکوب می‌شود؛

۳- افزایش استفاده دولتی از سیستم‌های تصمیم‌گیری خودکار که بدون ساختار پاسخگویی، افراد و جوامع را به طور مستقیم تحت‌تاثیر قرار می‌دهد؛

۴- شکل‌های بدون مقررات و بدون کنترل آزمایش‌های هوش مصنوعی بر روی جمعیت‌های انسانی؛

۵- محدودیت راه‌حل‌های فناورانه برای مسائل بی‌طرفی، تعصب و تبعیض.

در هر کدام از این مسائل، چالش‌های نوظهور و تحقیقات جدید را شناسایی کرده و در مورد توسعه، به‌کارگیری و تنظیم مقررات برای هوش مصنوعی توصیه‌هایی ارائه می‌کنیم. ما مسیرهایی عملی که از تحقیقات الهام گرفته شده را پیشنهاد می‌دهیم تا سیاستگذاران، عموم مردم و متخصصان فناوری بهتر بتوانند خطرات را فهمیده و

کاهش دهند. با توجه به این که موقعیت مکانی موسسه AI Now و تخصص منطقه‌ای آن در ایالات متحده متمرکز است، این گزارش در درجه‌ی اول بر بافتار آمریکا که در آن چندی از بزرگ‌ترین شرکت‌های هوش مصنوعی مستقر هستند، متمرکز خواهد شد.

شکاف پاسخ‌گویی هوش مصنوعی رو به گسترش است:

رسوایی‌های فناوری در سال ۲۰۱۸ نشان داده است که فاصله‌ی میان کسانی که هوش مصنوعی را توسعه می‌دهند و از آن سود می‌برند و آن‌هایی که از عواقب اثرات منفی آن به احتمال زیاد رنج می‌برند، [نه تنها] کم نشده، [بلکه] در حال بیشتر شدن است. دلایل مختلفی برای این امر وجود دارد: از جمله فقدان مقررات دولتی، تمرکز بیش از حد روی حوزه‌ی مهندسی هوش مصنوعی، ساختارهای مدیریتی ناکافی در شرکت‌های فناوری، عدم تقارن قدرت بین شرکت‌ها و مردمی که [شرکت‌ها] به آن‌ها خدمت می‌کنند و وجود یک جدایش فرهنگی شدید میان مهندسانی که مسئولیت تحقیقات فنی دارند و جمعیت‌های بسیار متنوعی که سیستم‌های هوش مصنوعی در میان آن‌ها مستقر شده‌اند. این شکاف‌ها نگرانی فزاینده‌ای درباره‌ی مسائل جانبداری، تبعیض، رویه‌های عادلانه‌ی قانونی، تعهد قانونی و مسئولیت‌پذیری ایجاد می‌کنند. این گزارش بر نیاز مبرم به تحقیقات و مقررات سفت و سخت در حوزه‌های خاصی که هوش مصنوعی در آن‌ها دخالت دارد، تاکید می‌کند.

هوش مصنوعی نظارت گسترده را تقویت می‌کند:

نقش هوش مصنوعی در نظارت گسترده در ایالات متحده، چین و بسیاری از کشورهای دیگر در سرتاسر جهان، به شدت افزایش یافته است. این اتفاق در استفاده‌ی رو به رشد از شبکه‌های حسگر، ردیابی [افراد در] شبکه‌های اجتماعی، تشخیص چهره و تشخیص تمایلات مشاهده می‌شود. این توسعه نه تنها حریم خصوصی افراد را تهدید کرده، بلکه خودکارسازی نظارت را نیز تسریع می‌کند و در نتیجه فراگیری و دسترسی به آن را نیز افزایش می‌دهد. این اتفاق، خطرات جدیدی را به ما عرضه می‌کند و بسیاری از نگرانی‌های دیرینه را بزرگ می‌کند. استفاده از فناوری تشخیص تمایلات که بر اساس شبه علم‌هایی است که دروغ بودن ادعاهایشان ثابت شده است نیز در حال افزایش است. فناوری تشخیص تمایلات تلاش می‌کند که با تحلیل دقیق چهره، به احساسات درونی پی ببرد. این فناوری با ادعاهایی ساختگی درباره‌ی خلق و خو، سلامت روانی، سطح تعهد و اغلب بدون آگاهی مردم، برای اهداف تبعیض‌آمیز و غیراخلاقی مورد استفاده قرار می‌گیرد. فناوری تشخیص چهره نیز خطرات خاص خود را به همراه دارد و از عدالت کیفری گرفته تا آموزش و اشتغال را تحت تاثیر اقدامات تبعیض‌آمیز ممکن قرار می‌دهد و خطراتی را برای حقوق بشر و آزادی‌های مدنی در کشورهای مختلف به وجود می‌آورد.

دولت‌ها استفاده از سیستم‌های تصمیم‌گیری خودکار را به سرعت و بدون محافظت کافی از حقوق مدنی گسترش می‌دهند:

در سراسر جهان، سازمان‌های دولتی تحت عناوینی مانند بهره‌وری و صرفه‌جویی در هزینه، در حال فراهم کردن و به کارگیری سیستم‌های تصمیم‌گیری خودکار (ADS) هستند. با این حال، بسیاری از این سیستم‌ها هنوز تست نشده و برای انجام وظایف خود به شکل بدی طراحی شده‌اند و در نتیجه، برخلاف قانون اساسی و به صورت نامشروعی به نقض حقوق فردی منجر می‌شوند. بدتر این‌که، زمانی که این سیستم‌ها اشتباه می‌کنند و تصمیمات نادرست می‌گیرند، در اغلب موارد زیرسوال بردن این تصمیمات، مقابله با آن‌ها و اصلاح‌شان دشوار یا غیرممکن است. برخی از نهادها در تلاشند تا سازوکارهایی برای شفافیت، توجه به رویه‌های عادلانه‌ی قانونی و سایر حقوق اساسی فراهم کنند، اما رازداری تجاری و چنین قوانینی بررسی دقیق و تست کافی این سیستم‌ها را تهدید به جلوگیری می‌کنند. با توجه به تلاش‌های نهادهای کنش‌گر اجتماعی و اقامه‌ی قانونی شکایات که اخیراً اتفاق افتاده است، ما راه‌هایی را برای پاسخگویی سیستم‌های تصمیم‌گیری خودکار طراحی کرده‌ایم.

آزمایش‌های بی‌حد و حصر سیستم‌های هوش مصنوعی روی جمعیت‌های انسانی و در شرایط کنترل نشده:
دره سیلیکون^۱ با طرز فکر «سریع حرکت کن و همه چیز را درهم بشکن!»

معروف شده است؛ جایی که در آن شرکت‌ها به سمت آزمایش کردن با تکنولوژی‌های جدید، به سرعت و بدون توجه کافی به تاثیر شکست این آزمایش‌ها و اینکه چه کسی متحمل خطرات آن می‌شود، هل داده می‌شوند. در سال گذشته، شاهد تعداد رو به افزایشی از آزمایش‌هایی بودیم که سیستم‌های هوش مصنوعی را در شرایط کنترل نشده، بدون پروتکل‌های مناسب جهت اطلاع [مردم]، رضایت آن‌ها یا پاسخگویی به آن‌ها به کار گرفته‌اند. چنین آزمایش‌هایی به دلیل عدم وجود پیامدهایی که شکست آن‌ها را نشان بدهد، ادامه دارند. زمانی که آسیبی رخ دهد، اغلب مشخص نیست که مسئولیت با چه کسی است. تحقیق و تخصیص مسئولیت و تعهد قانونی متناسب، اولویتی اضطراری باقی می‌ماند.

محدودیت راه‌حل‌های فناورانه برای مسائلی چون بی‌طرفی، جانبداری و تبعیض:

تلاش‌های جدیدی به منظور طراحی مدل‌هایی ریاضی برای بی‌طرفانه ساختن سیستم‌ها، انجام شده‌است؛ به طوری که وقتی ماشین‌ها خروجی‌ها را محاسبه می‌کنند هدفشان اجتناب از تبعیض باشد. با این حال، بدون چارچوبی که تاریخ و بافتارهای اجتماعی و سیاسی را توضیح بدهد، این فرمول‌های ریاضی برای بی‌طرفی، تقریباً بدون شک عوامل کلیدی دخیل را از دست می‌دهد، و در مواجهه با مشکلات عمیق‌تر به شیوه‌ای به آن‌ها پردازد که در نهایت به افزایش آسیب یا نادیده گرفتن عدالت منجر شود. به منظور اطمینان از اینکه ما قادر به پرداختن به موضوعات اصلی

هستیم و برای اطمینان از این که توجه ما از تساوی به سوی عدالت خواهد بود، وسعت بخشی به چشم اندازها و گسترش دامنه‌ی تحقیقات، رو به سوی مسئله‌ی بی‌طرفی و جانبداری در هوش مصنوعی و فراتر رفتن از نگاه صرفاً ریاضیاتی حیاتی است.

حرکت به سوی اصول اخلاقی:

در سال جاری، انتشار اصول و دستورالعمل‌های متعدد اخلاقی برای ساخت و به‌کارگیری فناوری‌های هوش مصنوعی مشاهده شد که بسیاری از آن‌ها در واکنش به نگرانی‌های فزاینده در مورد تبعات اجتماعی هوش مصنوعی بودند. اما همانطور که مطالعات نشان می‌دهند، اگر توسعه‌ی نرم‌افزاری مستقیماً به ساختارهای پاسخگویی در ارتباط نباشند، این نوع تعهدات اخلاقی تاثیر چندانی بر عملکرد توسعه نرم‌افزاری نخواهند داشت. علاوه بر این، این کدها و دستورالعمل‌ها هیچ‌گونه سازوکار تحمیل قانونی، نظارت یا داشتن عواقب انحراف از تعهدات، در خودشان ندارند. در صورتی که کدهای اخلاقی، به شکلی صادقانه در فرایندهای توسعه هوش مصنوعی اعمال شوند و توسط سازوکارهای قابل اجرا و مسئولیت‌پذیرانه‌ای که به نفع جامعه پاسخگو هستند، پشتیبانی شوند، تنها در این صورت می‌توانند کمک کنند که شکاف پاسخگویی هوش مصنوعی کم‌تر بشود. گزارش زیر این موضوعات را با جزئیات بیشتر توسعه می‌دهد و درباره‌ی آخرین پژوهش‌های آکادمیک تامل می‌کند و هفت استراتژی را برای حرکت رو به جلو تشریح می‌کند:

۱- توسعه تحقیقات بی‌طرفانه‌ی هوش مصنوعی، به سوی مسائل

معطوف به عدالت و فراتر رفتن از توجه به برابری ریاضیاتی و بی‌طرفی آماری

۲- مطالعه و یافتن مجموعه کامل زیرساخت‌های مورد نیاز برای ایجاد هوش مصنوعی، از جمله بررسی مو به موی زنجیره‌های تامین مواد

۳- شرح دقیق شکل‌های مختلف کار مورد نیاز برای ساخت و نگهداری سیستم‌های هوش مصنوعی

۴- تعهد به میان‌رشته‌ای‌تر شدن هوش مصنوعی

۵- تحلیل [نقش] نژاد، جنسیت و قدرت در هوش مصنوعی

۶- توسعه مداخلات جدید در سیاست‌های حاکمیتی و اقامه دعوی استراتژیک

۷- ایجاد ائتلاف میان پژوهشگران، جامعه مدنی و تشکل‌های حوزه‌ی فناوری

این رویکردها برای پی‌ریزی دوباره‌ی حوزه‌ی هوش مصنوعی طراحی شده است و به عدم تعادل قدرتی رو به رشد می‌پردازد که در حال حاضر به نفع افرادی است که به قیمت آسیب دیدن بسیار محتمل برخی از جمعیت‌ها، سیستم‌های هوش مصنوعی را توسعه داده و از آن‌ها بهره می‌برند.

بخش دوم

مشکلات تشدید شده



به هنگام شناسایی مهم‌ترین تبعات اجتماعی هوش مصنوعی در سال جاری، با نگاهی دقیق‌تر، به نقش هوش مصنوعی در نظارت گسترده در کشورهای مختلف در سراسر جهان و تبعاتی که برای حقوق و آزادی‌های افراد دارد می‌رسیم. به طور خاص در این گزارش، به استفاده‌ی رو به افزایش از فناوری تشخیص چهره و زیررده‌ی از این فناوری که با عنوان تشخیص تمایلات شناخته‌شده توجه می‌کنیم و درخواست‌های رو به رشد [مردم] به منظور تنظیم مقررات را ارزیابی می‌کنیم. در مرحله‌ی بعدی، یافته‌های خود را درباره‌ی استفاده دولت از سیستم‌های تصمیم‌گیری خودکار به اشتراک می‌گذاریم. به این می‌پردازیم که وقتی که این سیستم‌ها توسط قوانین تجاری رازداری و دیگر قوانینی که از بررسی دقیق آن‌ها جلوگیری می‌کنند، محافظت می‌شوند، چه سوالاتی درباره‌ی مسائلی مانند بی‌طرفی، شفافیت و فرآیندهای قانونی به وجود می‌آید؟ در نهایت، به شیوه‌های آزمایش سیستم‌های هوش مصنوعی در شرایط کنترل نشده که بر روی جمعیت‌های انسانی آزمایش می‌شوند، نگاهی می‌اندازیم. تحلیل می‌کنیم که چه کسی بیش‌ترین بهره را می‌برد و چه کسی در

معرض آسیب قرار دارد.

۱-۲- هوش مصنوعی، نظارت گسترده را تقویت می کند

در سال جاری، شاهد آن بودیم که فناوری هوش مصنوعی نظارت در ابعاد گسترده را تقویت کرده است؛ این اتفاق به واسطه‌ی تکنیک‌هایی انجام شده که ویدیو، صوت، تصاویر و محتوای شبکه‌های اجتماعی کل مردم را تجزیه و تحلیل کرده و افراد و گروه‌هایی را مورد هدف قرار می‌دهند. محققان و فعالان، مدت‌ها در مورد خطرات جمع‌آوری انبوه داده و نظارت هشدار داده‌اند و طبق یافته‌های آنان، هوش مصنوعی خطر را در سه حوزه افزایش می‌دهد:

خودکارسازی^۱، مقیاس تحلیل، و قابلیت پیش‌بینی. مشخصاً، سیستم‌های هوش مصنوعی باعث می‌شوند که قابلیت‌های نظارتی سیستم‌ها، بسیار فراتر از محدودیت‌های بررسی انسانی و تجزیه و تحلیل کد نویسی دستی، خودکارسازی بشوند. بنابراین آن‌ها می‌توانند این قابلیت‌ها را در دسترس تعداد کمی از افراد، متمرکز کنند. همچنین این سیستم‌ها [قابلیت] تحلیل و دنباله‌یابی^۲ حجم عظیمی از داده را به صورت نمایی افزایش می‌دهند تا بتوانند که ارتباطی‌ها و نتیجه‌گیری‌هایی که پیش از این سیستم‌ها، ناممکن یا سخت بود را انجام بدهند. در نهایت، این سیستم‌ها قابلیت‌های پیش‌بینی جدیدی برای تصمیم‌گیری درباره شخصیت افراد و خطر آفرین بودن آن‌ها فراهم می‌کنند که امکان کنترل دانه‌ای جمعیت را افزایش می‌دهند. چندین نمونه از نظارت‌های نگران‌کننده‌ای که توسط هوش مصنوعی ممکن شده است را، چین در سال جاری از خود به نمایش

1. Automation
2. Tracking

گذاشته است، و آگاه بودن ما از آن به خاطر آن است که خود دولت آشکارا آن‌ها را تایید می‌کند. با این حال، مهم است که توجه کنیم که بسیاری از زیرساخت‌های مشابه، در حال حاضر در ایالات متحده و جاهای دیگر وجود دارند که اغلب توسط شرکت‌های خصوصی‌ای که بازاریابی آن‌ها استفاده از این موارد را مهم می‌شمارد، تولید و تقویت می‌شوند. در آمریکا، همانگونه که بررسی خواهیم کرد، استفاده از چنین ابزارهایی توسط ضابطین قانون و دولت به ندرت در معرض موشکافی عمومی قرار می‌گیرد و چیزهای زیادی وجود دارد که ما از آن‌ها بی‌خبریم. این زیرساخت‌ها و قابلیت‌ها در آمریکا بدون آشکارسازی و نظارت عمومی، و بسته به انگیزه‌های تجاری و اراده‌ی سیاسی، می‌توانند به راحتی در جهت اهداف نظارتی قرار بگیرند. در کشور چین، تکنولوژی نظارت خودکار که توسط ارتش و دولت پشتیبانی می‌شود برای کنترل کردن بخش عظیمی از مردم به کار گرفته می‌شود که اغلب جمعیت‌های به حاشیه رانده‌شده را مورد هدف قرار می‌دهد. گزارش‌ها حاکی از نصب ابزارهای تشخیص چهره در مرز هنگ‌کنگ-شنژن، استفاده از انبوهی از ربات‌هایی که به شکل کبوتر هستند در پنج استان کشور و سیستم نظارت بر اعتبار اجتماعی است که هر کدام از آن‌ها نشان می‌دهد که چگونه سیستم‌های نظارتی که توسط هوش مصنوعی تقویت شده‌اند را می‌توان به عنوان ابزاری برای کنترل اجتماعی از راه دور استفاده کرد. بنا بر گزارش‌ها، سرکوب‌گرانه‌ترین استفاده از این سیستم‌ها در منطقه خودمختار سین جیان در حال رخ دادن است؛ به طوری که نشریه‌ی *The Economist* آن را با عنوان «دولت پلیسی که مانند

هیچ‌جای دیگر نیست» توصیف می‌کند. نظارت در این منطقه که متعلق به اقلیت قومی اویغور است، فراگیر بوده و [بدین ترتیب] گستره‌ی متنوعی از ایستگاه‌های بازرسی فیزیکی، دوربین‌های نظارتی، جاسوس افزارها، دستگاه‌های Wi-Fi و دستگاه‌های جمع‌آوری اطلاعات بیومتریک که گاهی اوقات به طور مخفیانه استفاده می‌شود، مورد نیاز است. ابزارهای یادگیری ماشین، این زنجیره‌ی داده‌ها را با هم ادغام می‌کند تا فهرست‌های مفصلی از مظنونینی که باید بازداشت شوند، تولید کند تا این افراد در اردوگاه‌هایی که توسط دولت برای منضبط کردن این گروه‌ها ساخته شده، دوره‌ی بازآموزی را سپری کنند. برآوردها از تعداد افرادی که در این اردوگاه‌ها دستگیر شده‌اند، از صدها هزار تا یک میلیون نفر متغیر است.

این زیرساخت‌ها فقط منحصر به چین نیستند. ونزوئلا استفاده از کارت شناسایی هوشمند جدیدی به نام «carnet de patria» را اعلان عمومی کرد. این کارت، با ادغام پایگاه داده‌های دولتی که با برنامه‌های اجتماعی [دولت] در ارتباط است، می‌تواند دولت را قادر به نظارت بر امور مالی، سابقه پزشکی و فعالیت‌های انتخاباتی شهروندان کند. در ایالات متحده نیز شاهد تلاش‌های مشابهی بوده‌ایم. پنتاگون از پژوهش درباره‌ی نظارت بر شبکه‌های اجتماعی با استفاده از هوش مصنوعی، حمایت مالی کرده است تا بتواند رفتار مردم در ابعاد بالا را پیش‌بینی کند. هم‌چنین اداره مهاجرت و گمرک ایالات متحده (ICE) در حال حاضر از یک سیستم مدیریت که توسط Palantir توسعه پیدا کرده استفاده می‌کند؛ هم‌چنین این اداره از خدمات شبکه‌ای Amazon در عملیات اخراج مهاجران بهره می‌برد.

این سیستم اطلاعات عمومی شده را با اطلاعات خریداری شده از دلان اطلاعات خصوصی ادغام می‌کند تا برای مهاجران پرونده درست کرده و به اداره برای ثبت مشخصات، پی‌گیری و اخراج افراد کمک کند. این مثال‌ها نشان می‌دهند که سیستم‌های هوش مصنوعی، چگونه یکپارچه شدن فن‌آوری‌های نظارتی و مدل‌های داده‌محور کنترل اجتماعی را افزایش می‌دهد و چگونه قدرت چنین اطلاعاتی را تقویت می‌کند، خطرات سو استفاده را بزرگ می‌کند؟ این مثال‌ها سوالات اضطراری و مهمی را مطرح می‌کنند که چگونه حقوق و آزادی‌های اساسی محافظت خواهد شد؟

علم‌غزنده‌ی تشخیص تمایلات و تاریخ خطرناک آن

در حال حاضر ظهور خطرات جدیدی که از سیستم‌های تشخیص چهره که مقرراتی برای آن‌ها وضع نشده است، قابل مشاهده است. این سیستم‌ها تشخیص و بازشناسی چهره‌های افراد در تصاویر یا ویدیو را تسهیل می‌کنند و می‌توانند با ابزارهای دیگر برای انجام شکل‌های پیچیده‌تر نظارت ترکیب شوند تا در اعمالی مانند لبخوانی خودکار که توانایی مشاهده و تفسیر صحبت کردن از راه دور را عرضه می‌کند، استفاده شوند.

در میان تکنیک‌های نظارت و ردیابی که هوش مصنوعی آن‌ها را ممکن ساخته، فناوری تشخیص چهره در حوزه‌ی آزادی‌های مدنی نگرانی‌ها را افزایش می‌دهد. از آنجا که ویژگی‌های چهره، شکلی کاملاً شخصی از هویت بیومتریک است که تغییر دادن آن بسیار دشوار است، از بین بردن آن و یا «صرف نظر کردن» از عملکرد آن دشوار

است. برخلاف دیگر ابزارهای ردیابی، سیستم‌های تشخیص چهره‌ی فعلی، از هوش مصنوعی برای چیزی بیش از تشخیص چهره‌ی ساده بهره می‌جویند. پس از تشخیص، یک چهره می‌تواند با اشکال دیگر سوابق شخصی و داده‌های قابل‌شناسایی مانند اعتبار بانکی، طبقه‌ی اجتماعی و یا سابقه جنایی ارتباط پیدا کند.

تشخیص تمایلات، زیر مجموعه‌ای از تشخیص چهره است، که هدف آن تفسیر چهره است تا به طور خودکار حالات احساسی درونی یا حتی مقاصد پنهان را شناسایی کند. این رویکرد نوعی پیش‌بینی آب و هوای عاطفی را نوید می‌دهد: تحلیل صدها هزار تصویر چهره، شناسایی «خرده حالت‌ها»، و نگاشت این حالت‌ها به «احساساتی واقعی» تشخیص تمایلات، سنت کهن سیماشناسی را بازفعال می‌کند - شبه‌علمی که ادعا می‌کند ویژگی‌های صورت می‌تواند جنبه‌های درونی سیرت و شخصیت ما را آشکار کند. پس از زمان باستان، علاقه علمی به سیماشناسی در قرن نوزدهم به شدت افزایش یافت؛ هنگامی که [این شبه‌علم] به روشی اصلی برای اشکال علمی نژادپرستی و تبعیض تبدیل شد. اگر چه سیماشناسی پس از ارتباط آن با علم نژاد نازی از چشم‌ها افتاد، اما محققان [همچنان] از ظهور ایده‌های سیماشناسانه در کاربردهای فناوری تشخیص تمایلات نگران هستند. این ایده که سیستم‌های هوش مصنوعی ممکن است قادر به بیان احساسات واقعی یک دانش‌آموز، مشتری یا مظنون جنایی هستند یا این که می‌توانند شخصیت درونی افراد را تشخیص دهند، هم برای شرکت‌ها و هم دولت‌ها جذاب است، حتی اگر توجیه علمی این ادعاها بسیار سوال‌برانگیز باشد و تاریخچه‌ی اهداف

تبعیض آمیز آن‌ها به خوبی مستند شده باشد.

تشخیص تمایلات نشان می‌دهد که چگونه می‌توان به راحتی از سیستم‌های یادگیری ماشین برای تشدید کردن شکل‌های طبقه‌بندی و تبعیض استفاده کرد؛ حتی زمانی که پایه‌های اصلی این نظریه‌ها در میان روانشناسان بحث‌برانگیز باقی بماند. یکی از دانشمندانی که با تشخیص تمایلات بوسیله‌ی هوش مصنوعی همبستگی خاصی دارد، پل اکمن^۱ است، که به نظر او احساسات را می‌توان در مجموعه کوچکی از مقولات پایه مانند عصبانیت، تنفر، ترس، شادی، غم و تعجب دسته‌بندی کرد. بر اساس نظرات اکمن، مطالعه‌ی چهره، خوانشی عینی از حالات واقعی درونی ایجاد می‌کند که پنجره‌ای به سوی روح است. اساس اعتقاد او این ایده است که احساسات، فارغ از در نظر گرفتن بافتارهای فرهنگی، ثابت و همگانی، یکسان در میان افراد و قابل مشاهده در مکانیسم‌های بیولوژیکی هستند. اما آثار اکمن عمیقاً توسط روانشناسان، انسان‌شناسان، و پژوهشگران دیگر مورد انتقاد قرار گرفته است. لیزا فلدمن^۲ که یک روانشناس است به همراه همکارانش استدلال کرده که فهمیدن احساسات [انسانی] از نظرگاه این مقولات سفت و سخت و عوامل فیزیولوژیکی ساده‌سازی شده، دیگر منطقی نیست. با این حال، پژوهشگران هوش مصنوعی کار اکمن را درست می‌پندارند و از آن به عنوان پایه‌ای برای خودکارسازی شناسایی احساسات استفاده کرده‌اند.

عوامل بافتاری، اجتماعی و فرهنگی، نسبت به آنچه که اکمن و افراد مانند او باور داشتند، نقش گسترده‌تری را در بیان احساسی ایفا می‌کنند. در سایه‌ی این فهم علمی جدید از احساسات، هر گونه نگاهت

ساده‌انگاره‌ای که توسط هوش مصنوعی از حالات چهره به مقولات اساسی احساسات انجام می‌شود، احتمالاً اشتباهات یک پارادایم علمی قدیمی را بازتولید می‌کند. همچنین نگرانی‌های اخلاقی را ایجاد می‌کند درباره‌ی این که داور شخصیت و احساسات «واقعی» کسی در خارج از او قرار می‌گیرد، و درباره‌ی سوء استفاده بالقوه از قدرت که می‌تواند براساس این ادعاهای معیوب توجیه شود. جیمی متزل^۱ که یک روان‌پزشک است، مثال هشدار دهنده‌ی تازه‌ای را مستندسازی می‌کند: میان افراد سیاه‌پوستی که در دهه‌ی ۱۹۶۰ میلادی دارای شیزوفرنی بودند و کسانی که از جنبش آزادی مدنی حمایت می‌کردند، الگوسازی شده است.

تشخیص تمایلات در کنار تشخیص چهره در مقیاس بزرگ، پتانسیل افزایش چنین سو استفاده‌های سیاسی از مشخصه‌های روانشناختی را دارد. در حوزه آموزش، برخی از دانشگاه‌های آمریکا متوجه شده‌اند که از نرم‌افزارهای تحلیل تمایلات بر روی دانش‌آموزان استفاده می‌شود. دانشگاه سنت توماس در مینه‌سوتا، به استفاده از سیستمی پی برده است که با استفاده از یک دوربین وبکم و پشتیبانی سیستم تشخیص چهره مایکروسافت و دستگاه‌های تشخیص تمایلات، بر دانش‌آموزان در کلاس نظارت می‌کند. این سیستم حالت احساسی دانش‌آموزان را پیش‌بینی می‌کند. معلم می‌تواند کلیتی از حالات احساسی دانش‌آموزان را مشاهده کند؛ معلم بر اساس قضاوتی که سیستم انجام می‌دهد، می‌تواند تدریس خود را به گونه‌ای تغییر دهد که «مشارکت دانش‌آموزان را تضمین کند». این مساله سوالات جدی را در چند سطح مطرح می‌کند: اگر این سیستم با مدلی ساده‌انگاره

1. Jamie Metz

از احساسات نتواند حالات پیچیده تری را درک کند، چه اتفاقی می‌افتد؟ چگونه یک دانش‌آموز می‌تواند با تصمیمی که توسط این سیستم گرفته می‌شود مقابله کند؟ اگر دانش‌آموزان مختلف [توسط سیستم] «شاد» دیده شوند، اما دیگران «عصبانی» باشند، چه اتفاقی خواهد افتاد؟ - چگونه باید معلم درس را دوباره به جریان بپردازد؟ تبعات مربوط به حریم خصوصی چنین سیستمی چیست؟ خصوصاً این که در هنگام اجرای برنامه آزمایشی، هیچ مدرکی وجود ندارد که دانش‌آموزان از بکارگیری این سیستم مطلع باشند.

خارج از فضای کلاس و مدرسه، دستیارهای شخصی مثل Alexa و Siri را نیز مشاهده می‌کنیم که به دنبال فهمیدن لحن‌های احساسی گفتار انسان هستند؛ حتی شرکت‌ها تا آنجا پیش رفته‌اند که از روش‌های انحصاری بازاریابی که بر مبنای تشخیص احساسات، سلامت ذهنی و سلامت فیزیکی افراد است، استفاده می‌کنند.

شرکت Affectiva که دارای سیستم احساسات‌سنج مجهز به هوش مصنوعی است، در حال حاضر این وعده را می‌دهد که می‌تواند با کنترل «احساسات، حالات شناختی و واکنش‌های راننده و سرنشینان که از چهره و صدای آن‌ها حاصل می‌شود» رانندگی ایمن‌تری را تضمین کند. با این وجود شواهد کمی وجود دارد از این که آیا این سیستم‌ها در میان افراد، بافتارها و فرهنگ‌های متنوع به درستی کار می‌کنند یا این که آیا تضمینی وجود دارد که از نگرانی درباره‌ی مسائلی مانند حریم خصوصی، جانبداری یا تبعیض - در حین کار این سیستم‌ها - کاسته شود. علاوه بر این، همانطور که در نوشتجات گسترده‌ی حوزه‌ی جانبداری و بی‌طرفی مشاهده شده است، طبقه‌بندی

ماهیت احساسی انسان‌ها نه تنها اثرات مستقیم بر زندگی آن‌ها دارد بلکه داده‌هایی را برای آموزش و تاثیر بر دیگر سیستم‌های هوش مصنوعی فراهم می‌کند. این مساله خطرات استفاده از فناوری تشخیص تمایلات را بالا می‌برد، و تاکید بیشتری را به وجود می‌آورد که چرا این فناوری باید به طور جدی مورد بررسی قرار گرفته و کاربرد آن به شدت محدود شود.

تشخیص چهره نگرانی‌های مربوط به حقوق مدنی را افزایش می‌دهد

نگرانی‌هایی در حال تشدید شدن است نسبت به این که تشخیص چهره تبعیض نژادی و جانبداری‌های دیگر در نظام عدالت کیفری را افزایش می‌دهد. در اوایل امسال، اتحادیه آزادی‌های مدنی آمریکا (ACLU)^۱ افشا کرد که هم اداره پلیس اورلاندو و هم اداره کلانتری حوزه‌ی واشینگتن در حال استفاده از سیستم Rekognition شرکت آمازون هستند که می‌تواند «بی‌درنگ از میان ده‌ها میلیون عکس، تشخیص چهره را انجام دهد» و «تا ۱۰۰ چهره را در تصاویر پرجمعیت و چالش‌برانگیز» تشخیص دهد. در حوزه‌ی واشینگتن، آمازون به طور خاص با اداره‌ی کلانتری کار می‌کرد تا نرم‌افزاری را برای موبایل بسازد که بتواند چهره‌ها را اسکن کرده و آن‌ها را با [اطلاعات] پایگاه داده‌ای که حداقل ۳۰۰،۰۰۰ عکس از تبهکاران را در خود دارد، مقایسه کند. یکی از نمایندگان آمازون اخیراً در طی گفتگویی اظهار داشت که در حال بررسی کاربردهایی [از سیستم‌های خود] هستند که در آن، شبکه دوربین‌های نظارتی اورلاندو می‌توانند توامان با تکنولوژی تشخیص چهره استفاده شوند تا «فرد تحت تعقیب»

1. American Civil Liberties Union

را در هر جایی از شهر که ممکن است باشد، پیدا کنند . علاوه بر نگرانی‌های مربوط به حریم خصوصی و نظارت گسترده، استفاده سازمان‌های مجری قانون از فناوری تشخیص چهره نیز با نگرانی‌هایی درباره‌ی جانبداری‌های نژادی تلاقی پیدا کرده است. محققان ACLU و دانشگاه کالیفرنیا در برکلی سیستم Rekognition شرکت آمازون را با مقایسه‌ی تصاویر اعضای کنگره‌ی آمریکا با پایگاه داده‌ای که حاوی ۲۵۰۰۰ تصویر از افراد بازداشت‌شده بود، آزمایش کردند. نتایج این آزمایش، درصد قابل توجهی از بی‌دقتی این دستگاه را نشان می‌داد: سیستم Rekognition شرکت آمازون ۲۸ عضو کنگره را به اشتباه، به عنوان افرادی شناسایی کرد که در پایگاه اطلاعاتی افراد بازداشت‌شده بودند. علاوه بر این، به طور غیرمتناسبی تشخیص‌های مثبت و کاذب (یعنی تشخیص‌هایی که افراد را به عنوان تبهکار نشان می‌دهند اما در واقع این‌گونه نیست) در میان اعضای غیرسفیدپوست کنگره با نرخ خطای نزدیک به ۴۰٪ در مقایسه با نرخ خطای ۵٪ برای اعضای سفیدپوست، اتفاق داده است . چنین نتایجی سلسله‌ای از یافته‌ها را منعکس می‌کند که نشان‌دهنده‌ی آن است که تکنولوژی تشخیص چهره به طور متوسط، در تشخیص افراد سفیدپوست و مردان نسبت به افراد سیاه‌پوست و زنان بهتر عمل می‌کند . آمازون در واکنش به ACLU، اذعان کرده که «نتایج سیستم Rekognition، در صورت استفاده از پایگاه داده‌ای از چهره‌ها که نمایشگر مناسبی [از مجرمان] نیست، به طور قابل توجهی منحرف شود. » « با توجه به جانبداری‌های نژادی و تاریخی ژرفی که در نظام عدالت کیفری وجود داشته، بعید است که

پایگاه‌های داده‌ای که در اختیار مجریان قانون قرار دارند، «نمایشگر مناسبی [از مجرمان]» باشد. علی‌رغم این مشکلات جدی، فشارهای مداوم از جانب فعالان حقوق مدنی و تظاهرات کارمندان آمازون در اعتراض به سو استفاده‌ی بالقوه از این فناوری‌ها، اندرو جاسی^۱ مدیر عامل خدمات شبکه‌ای آمازون، اخیراً به کارکنان گفت: «ما از این که Rekognition برای مشتریان ما -در همه‌ی ابعاد و همه‌ی صنایع، در داخل یا خارج از سازمان‌های مجری قانون- ارزش افزوده ایجاد کرده است، احساس واقعاً خوبی داریم و احساس قدرت می‌کنیم.» تنها آمازون نیست که فناوری‌های تشخیص چهره را به روش‌های غیرمسئولانه پیاده می‌کند. روزنامه‌نگاران تحقیقی اخیراً فاش کردند که IBM و اداره پلیس شهر نیویورک (NYPD) برای توسعه‌ی چنین سیستمی همکاری می‌کنند که یکی از بخش‌های سفارشی آن [توسط پلیس] «جستجوی قومیت» است، که [این بخش از نرم‌افزار] با هزاران ساعت از فیلم‌های نظارتی NYPD آموزش داده شده است. استفاده از نرم‌افزار تشخیص چهره در بخش خصوصی نیز گسترش یافته است. خرده فروشان و مراکز تجاری اصلی هم اکنون برای شناسایی قاپ‌زن‌ها، کنترل جمعیت، و حتی «شناسایی مشتریان ناراضی»، شروع به استفاده از سیستم‌های تشخیص چهره‌ی مجهز به قابلیت‌های «تشخیص تمایلات» کرده‌اند.

این نگرانی‌ها به دلیل فقدان قوانین و مقررات تشدید شده‌اند. در حال حاضر هیچ قانون فدرالی که به دنبال ارائه استانداردها، محدودیت‌ها، الزامات یا دستورالعمل‌هایی مربوط به توسعه یا بکارگیری فناوری تشخیص چهره باشد، وجود ندارد. در حقیقت، اغلب قوانین فدرال

1. Andrew Jassy

موجود به دنبال ترویج بکارگیری فناوری تشخیص چهره به منظور نظارت، اجرای قوانین مهاجرت، استخدام و سیستم‌های ورود و خروج محلی هستند. قوانین ما از تکه‌های مختلفی درست شده است و هیچ یک از آن تکه‌ها به طور خاص به مسئله‌ی تشخیص چهره نپرداخته‌اند. یکی از این قوانین که در سال ۲۰۰۸ در ایلنویز تصویب شده، قانون محرمانگی اطلاعات بیومتریک است، که مقررات سخت و سختی را در رابطه با جمع‌آوری اطلاعات بیومتریک وضع می‌کند. با این حال، با توجه به این که در سال ۲۰۰۸ فناوری تشخیص چهره به طور گسترده در دسترس نبوده، این قانون از این فناوری صحبتی به عمل نمی‌آورد. اما تفسیر منطقی که از آن شده مبتنی بر آن است که بسیاری از الزامات نظیر گرفتن رضایت، باید اعمال شود. اخیراً، چندین شهرداری و یک سیستم ترانزیت محلی حکمی را تصویب کرده‌اند که به دنبال ایجاد شفافیت و نظارت بیشتر بر جمع‌آوری داده و بکارگیری الزاماتی در رابطه با استفاده از فناوری‌های نظارتی است، که شامل فناوری تشخیص چهره نیز خواهد شد. مخالفت با استفاده از ابزارهای تشخیص چهره در سازمان‌های دولتی در حال گسترش است. پس از آن که ACLU از اسنادی که نشان‌دهنده‌ی بکارگیری سیستم Rekognition در سازمان‌های مجری قانون بود، پرده‌برداری کرد، در اوایل سال جاری، موسسه‌ی AI Now به ACLU و بیش از ۳۰ نهاد تحقیقاتی و فعال اجتماعی دیگر پیوست و از آمازون خواست تا از فروش نرم‌افزار تشخیص چهره به سازمان‌های دولتی دست بردارند. اعضای کنگره نیز در حال فشار آوردن بر آمازون برای ارائه اطلاعات بیشتر هستند.

برخی دیگر پا را فراتر گذاشتند و خواستار ممنوعیت کامل [استفاده از این سیستم‌ها] شدند. پژوهشگرانی مانند وودرو هارتزوغ^۱ و ایوان سلینگر^۲ استدلال می‌کنند که فناوری تشخیص چهره «بزاری برای ظلم و ستم است که به شکل تمام‌عیاری برای دولت‌ها مناسب است تا نظارت اقتدارطلبانه و ماشین ناقض حریم خصوصی خود را به نمایش بگذارند» که پیش از اعمال آن به اکوسیستم دیجیتال معاصر ما، دوراندیشی و سخت‌کوشی بسیار زیادی را می‌طلبید.

در نقد پروژه‌ی مطالعاتی دانشگاه استنفورد درباره‌ی gaydar که مدعی شده بود که شبکه عصبی عمیق gaydar، در پیش‌بینی تمایلات جنسی از روی تصویر چهره، نسبت به انسان دقیق‌تر است؛ فرانک پاسکال^۳ نوشت که «برخی از برنامه‌های تحقیقاتی علمی به بهترین شکل پی‌گیری نمی‌شوند - و این ممکن است یکی از آن‌ها باشد.» کید کراکفورد^۴، مدیر تکنولوژی برنامه آزادی در ACLU ماساچوست، به نفع ممنوعیت [استفاده از این فناوری‌ها] مطلبی نوشت و بیان کرد که «فناوری‌های هوش مصنوعی نظیر سیستم‌های تشخیص چهره اساساً توازن قدرت بین مردم و دولت را به هم می‌زنند ... برخی از فناوری‌ها برای این توازن قدرت، چنان خطرناک هستند که نباید پذیرفته شود.» برد اسمیت^۵، رئیس مایکروسافت، خواستار تنظیم مقررات دولتی برای سیستم‌های تشخیص چهره شده است؛ ریک اسمیت^۶ نیز که مدیر عامل فناوری‌های مجری قانون شرکت آکسون^۷ است، اخیراً اظهار داشت که «آستانه دقت» دستگاه‌های تشخیص چهره، «در حدی که در تصمیم‌گیری‌های عملیاتی به آن نیازمندیم، نیست.»

1. Woodrow Hartzog
2. Evan Selinger
3. Frank Pasquale
4. Kade Crockford

5. Brad Smith
6. Rick Smith
7. Axon

اتفاقات سال جاری، نیاز مبرم به مقررات سخت‌گیرانه‌تر بر تشخیص چهره و تشخیص تمایلات را قویاً مورد تأکید قرار می‌دهد. چنین مقرراتی باید استفاده‌ی بخش خصوصی و دولتی را به شدت محدود کرده و تضمین بدهد که جوامع تحت‌تأثیر این فناوری‌ها، قاضی نهایی این امر هستند که از آن‌ها استفاده بکنند یا خیر. این مسئله، به طور خاص در موقعیتی که در آن حقوق و آزادی‌های اساسی در معرض خطر هستند، مهم است و به نظارت سفت و سخت، بررسی دقیق و شفافیت نیاز دارد. ارتباط میان پایگاه‌های داده‌ی خصوصی و دولتی نباید مجاز باشد. در حال حاضر و با توجه به شواهد موجود، سیاست‌گذاران نباید برای استفاده از این سیستم‌ها در حوزه‌ی عمومی سرمایه‌گذاری کنند و آن‌ها را گسترش دهند.

۲-۲- خطرات استفاده از سیستم‌های تصمیم‌گیری خودکار در دولت

در طی سال گذشته شاهد افزایش قابل توجه بکارگیری سیستم‌های تصمیم‌گیری خودکار (ADS) در بخش‌های دولتی از جمله نظام عدالت کیفری، رفاه کودکان، آموزش و مهاجرت بوده‌ایم. این سیستم‌ها اغلب ذیل این نظریه به کار گرفته می‌شوند که بازدهی دولت و وضعیت هزینه‌ها را بهبود می‌بخشند و به دنبال کمک به فرآیند تصمیم‌گیری و سیاست‌گذاری بوده و روش‌های متعدد تصمیم‌گیری را کنار می‌گذارد. با این حال، به دلیل این که مالکیت مدل‌های تصمیم‌گیری غالباً با بخش خصوصی است و سیستم‌ها اغلب قبل از بکارگیری آزمایش نشده‌اند، بسیاری از فعالان اجتماعی نگرانی‌های قابل توجهی در مورد عدم وجود رویه‌های عادلانه‌ی قانونی، عدم

پاسخگویی، فقدان مشارکت اجتماعی و عدم بررسی دقیق در این سیستمها مطرح کرده‌اند. تمی دابز^۱ از جمله افراد آسیب‌دیده بود. او در سال ۲۰۰۸ به آرکانزاس رفت و در برنامه‌ی حمایت از معلولین نام‌نویسی کرد تا به او که فلج مغزی بود، کمک [پزشکی] کنند. بر اساس این برنامه، دولت پرستار واجد شرایطی را جهت ارزیابی تمی فرستاد تا تعداد ساعاتی که تمی [به مراقبت] نیاز داشت را تعیین کند. از آنجا که تمی بیشتر ساعات بیداری خود را بر صندلی چرخدار سپری کرده و دستانش دچار کوفتگی بود، در ارزیابی اولیه، ۵۶ ساعت مراقبت خانگی در هفته به او اختصاص یافته است. این اتفاق ادامه داشت تا این‌که در سال ۲۰۱۶ ارزیاب دولت با سیستم تصمیم‌گیری الگوریتمی جدید بر روی لپ‌تاپ خود وارد خانه‌ی تمی شد. با بکارگیری الگوریتمی اختصاصی، این سیستم تعداد ساعات اختصاص داده‌شده را محاسبه کرد. بدون هیچ توضیح و یا فرصت برای اظهار نظر، بحث یا ارزیابی مجدد، این سیستم، ۳۲ ساعت در هفته را به تمی اختصاص داد؛ یک افت عظیم و ناگهانی که تمی هیچ فرصتی برای آماده شدن برای آن نداشت و کیفیت زندگی اش را به شدت کاهش داد.

وضعیت تمی یک استثنا نبود. بر اساس گفته‌های کوین دی‌لیبان^۲ که وکیل قانونی ایالت آرکانزاس است، صدها معلول دیگر نیز متحمل کاهش ساعات نگهداری شده‌اند. با این حال، آن‌ها هیچ فرصت قابل توجهی برای مقابله با این تخصیص ساعات نداشتند. وکلای قانونی متعاقباً از ایالت آرکانزاس شکایت کردند و در نهایت به این نتیجه رسید که برنامه‌ی جدید تخصیص الگوریتمی نادرست و

1. Tammy Dobbs
2. Kevin De Liban

مغایر با قانون اساسی است. با این وجود، تا آن زمان زندگی افراد معلول متحمل آسیب‌های بسیاری شده بود.

موارد معلولیت در آرکانزاس نمونه‌ای عینی از خطرات فراوانی را نشان می‌دهد که به هنگام استفاده‌ی دولت‌ها از ADS در تصمیماتی که تأثیرات مستقیم بر جمعیت آسیب‌پذیر دارند، اتفاق می‌افتد. در حالی که ارزیاب‌ها هم ممکن است [در تصمیماتشان] از جانبداری یا منطق ناقص رنج ببرند، اما تأثیر [منفی] تصمیمات مورد به مورد آن‌ها در هیچ کجا به اندازه یا مقیاسی که یک سیستم ADS معیوب می‌تواند کل جمعیت را تحت تأثیر قرار بدهد، نمی‌رسد.

افزایش روزافزون چنین سیستم‌هایی در زمانی اتفاق می‌افتد که بر اساس گزارش‌های پایگاه اطلاعاتی جهانی نابرابری درآمدی^۱، ایالات متحده بالاترین میزان نابرابری درآمدی در تمام کشورهای غربی را داراست. علاوه بر این، داده‌های بانک مرکزی ایالات متحده آمریکا نشان می‌دهد که نابرابری ثروت به رشد خود ادامه می‌دهد؛ نابرابری ثروت نژادی در ۵۰ سال گذشته سه برابر شده و سیاست‌های تنظیم‌شده‌ی کنونی نیز در جهت وخیم‌تر شدن این مشکلات هستند. فقط در سال ۲۰۱۸، در آمریکا شاهد قطع شدن بودجه‌ی برنامه‌های اجتماعی بوده‌ایم که به نفع فقیرترین شهروندان بود؛ در کنار این موضوع، بودجه‌ی پیشنهادی فدرال که موجب کاهش قابل توجه مسکن ارزان قیمت می‌شود، اجرای الزامات کار طاقت‌فرسا برای دریافت بیمه‌ی بهداشت مستمندان و پیشنهاد قطع کمک‌های غذایی به افراد کم‌درآمد و معلولین را شاهد بوده‌ایم.

در بافتار چنین سیاست‌هایی، سازمان‌ها برای کاهش هزینه‌ها تحت فشار

زیادی قرار دارند و بسیاری از آن‌ها به ADS به عنوان وسیله‌ای جهت خودکارسازی تصمیمات دشواری نگاه می‌کنند که تاثیرات عینی بر نیازمندان دارد. به این ترتیب، بسیاری از سیستم‌های ADS اغلب با هدف انجام کار بیشتر با داشته‌های کمتر و در جهت سیاست‌های ریاضتی و هزینه‌کاه، بکار گرفته می‌شوند. آن‌ها اغلب برای رسیدن به این اهداف طراحی و پیگیربندی شده‌اند؛ اثربخشی نهایی آن‌ها براساس توانایی‌شان در کاهش هزینه‌ها ارزیابی می‌شود، که اغلب به ضرر جمعیت‌هایی تمام می‌شود که این ابزارها قرار است به آن‌ها خدمت‌رسانی کنند. همانطور که ویرجینیا ایبنکس استدلال می‌کند «چیزی که به نظر در حال تلاش برای کاهش موانع و حذف جانبداری انسانی است، اغلب اثر معکوس دارد و صدها هزار نفر را از دریافت خدماتی که شایسته آن هستند، محروم می‌کند.» هنگامی که این مشکلات بوجود می‌آیند، درمان آن‌ها اغلب دشوار است. تعداد کمی از سیستم‌های تصمیم‌گیری خودکار به گونه‌ای طراحی یا پیاده‌سازی شده‌اند که به افراد آسیب‌دیده اجازه اعتراض به تصمیمات درست یا نادرست، روشن‌سازی یا اصلاح آن‌ها را می‌دهند. علاوه بر این، غالباً آزادی عمل و توانایی مدیران، کارمندان و دیگرانی که آموزش دیده‌اند جهت دخالت در تصمیم یک سیستم یا لغو آن، به طور قابل توجهی محدود یا حذف می‌شود. کارمندان بخاطر ناتوانی در تغییر دادن تصمیمات نامنعطفی که توسط سیستم‌های خودکار گرفته شده است، به واسطه‌هایی صرف و [غیر موثر] تبدیل می‌شوند.

با این‌که در گذشته کارمندان و خدمتگذاران عمومی مسئولیت چنین

تصمیماتی را داشته‌اند، اما [در حال حاضر] بسیاری از سیستم‌های تصمیم‌گیری خودکار از فروشندگان خصوصی خریداری می‌شود و غالباً بدون آزمایش، بازبینی یا بررسی دقیقی که تناسب آن دستگاه با یک حوزه‌ی خاص را تضمین کند، به کار گرفته می‌شوند. این سیستم‌ها به طور معمول بدون هیچ شکل شفاف‌ی از نظارت یا پاسخگویی ساخته می‌شوند. این امر کشف خروجی‌های مشکل‌زای این سیستم‌ها را دشوار می‌سازد، مخصوصاً این‌که خطاها و شواهد حاکی از تبعیض، به صورت جمعی آشکار می‌شوند و تنها در الگوهایی میان اتفاقات فردی متعدد قابل شناسایی هستند. شناسایی چنین مشکلاتی مستلزم نظارت و کنترل است. همچنین این شناسایی مستلزم دسترسی به داده‌هایی است که اغلب نه در دسترس فعالان اجتماعی و عموم مردم است و نه توسط سازمان‌های دولتی نظارت می‌شود. به عنوان مثال، اتحادیه‌ی معلمان هیوستون از حوزه‌ی آموزش و پرورش مستقل هیوستون، بخاطر تهیه کردن یک سیستم تصمیم‌گیری خودکار شخص ثالث که با استفاده از اطلاعات امتحانات دانش‌آموزان اقدام به تصمیم‌گیری درباره‌ی اشتغال معلمان می‌کرد، شکایت کرد؛ این تصمیم‌گیری‌ها شامل آن بود که چه کسی ترفیع بگیرد و چه کسی به کارش خاتمه دهد. علی‌رغم این‌که حوزه‌ی آموزش و پرورش مستقل هیوستون به تمام داده‌های اصلی دسترسی داشت، مشخص شد که هیچ‌کس در این حوزه - حتی یک کارمند- نمی‌تواند تصمیمات اتخاذ شده توسط سیستم را تبیین یا حتی بازتولید کند. به آموزگاران‌ی که به دنبال اعتراض به این تصمیمات بودند، گفته شد که به این سیستم «جعبه سیاه» صرفاً اطمینان

کنند و نمی‌توانند آن را زیرسوال ببرند. حتی هنگامی که معلمان دادخواستی را مبنی بر نقض قانون اساسی، حقوق مدنی و نقض قانون کار اقامه کردند، فروشندگان ADS در برابر هر گونه دسترسی دادن به چگونگی کارکرد سیستم خود ایستادند. در نتیجه، قاضی حکم داد که استفاده از این سیستم ADS در مواردی که با کار کارمندان دولت تلاقی پیدا می‌کند، می‌تواند با رویه‌های عادلانه‌ای که توسط قانون اساسی حفاظت می‌شود، اختلاف پیدا کند؛ به خصوص هنگامی که رازداری تجاری جلوی قابلیت کارکنان برای فهمیدن چگونگی اتخاذ تصمیمات را می‌گیرد. با موافقت حوزه‌ی آموزش و پرورش مستقل هیوستون با کنار گذاشتن سیستم ADS شخص ثالث، این پرونده متعاقباً حل شد. در سال ۲۰۱۳ به طور مشابه، بخش ایالتی لس‌آنجلس به منظور ارزیابی آسیب‌های قریب‌الوقوع کودکان و پیش‌بینی احتمال بازگشت یک خانواده به شرایط مدنظر نظام رفاه کودکان در ۱۲ الی ۱۸ ماه، یک سیستم تصمیم‌گیری خودکار را به کار برد. این بخش ایالتی سیستم را بررسی و بازبینی نکرد و یا ثمربخشی استفاده از این تحلیل‌های پیشگویانه را برای سلامت و رفاه کودکان ارزیابی نکرد. تنها پس از مرگ یک کودک بود که رهبری بخش ایالتی به بررسی [سیستم] پرداخت، که [این بررسی] سوالاتی جدی در رابطه با اعتبار آن سیستم‌ها بوجود آورد. این بررسی این نکته را به طور خاص متذکر شد که این سیستم نتوانسته بود تصویر جامعی از یک خانواده‌ی نوعی را ارائه دهد، "بلکه در عوض، بدون وزن‌دهی به نکات ظریف، تنها بر روی چند نکته‌ی برجسته تمرکز کرده بود. ویرجینیا اینکس^۱ مشکلات مشابهی را در بررسی یک سیستم تصمیم‌گیری

1. Virginia Eubanks

خودکار یافت که توسط همان فروشنده خصوصی برای استفاده در بخش ایالتی الگینی^۱ توسعه داده شده بود. این سیستم نتایج جانبدارانه‌ای ایجاد کرد، زیرا به طور قابل توجه کودکان فقیری که به جمعیت طبقه کارگر و به ویژه جوامع رنگین پوست متعلق بودند را تحت تاثیر قرار می‌داد؛ در واقع، والدین و کودکان فقیر را در معرض تحقیقات مکرر قرار می‌داد.

حتی با وجود مسئله‌ی جانبداری و خطرات بالقوه در حوزه‌های آسیب‌پذیر، این سیستم‌ها با سرعت سرسام‌آوری به کار گرفته می‌شوند. وزارت توسعه‌ی اجتماعی نیوزیلند از به کارگیری یک سیستم تصمیم‌گیری خودکار پیش‌بینی‌کننده برای شناسایی کودکان در معرض خطر بدرفتاری حمایت کرده است؛ در حالی که «نگرانی‌های قابل توجه اخلاقی» درباره‌ی این سیستم‌ها تصدیق شده است. آن‌ها از این [سیستم] دفاع می‌کنند، بر این مبنا که منفعت آن که شامل بازپیکربندی مسئله‌ی رفاه کودکان به عنوان یک مسئله‌ی آماری است، «احتمالاً» به آسیب‌های بالقوه‌اش می‌چربد.

این مثال‌ها نه تنها نیاز به شفافیت، نظارت و پاسخگویی بیشتر را در بکارگیری، توسعه و پیاده‌سازی ADS مورد تاکید قرار می‌دهد، بلکه نیاز به بررسی محدودیت‌های کلی این سیستم‌ها و عوامل اقتصادی و سیاسی موثر در بکارگیری این سیستم‌ها را نیز برجسته قلمداد می‌کند. ویرجینیا ایبنکس، که استفاده از ADS را در مسئله‌ی رفاه کودکان مورد بررسی قرار داده، این موضوع و تعدادی از مطالعات موردی را بررسی کرده است تا نشان دهد چگونه ADS اغلب در جهت اجتناب از مشکلات وسیع ساختاری و نظام‌مند اجتماعی استفاده می‌شود؛

مشکلاتی که رسیدگی معنادار به آن‌ها، اغلب فراتر از ظرفیت ادارات بی‌پول است .

سیستم‌های خودکار دیگر هم‌چنان به عنوان راه مبارزه‌ای با مشکلات موجود در ساختارهای دولتی پیشنهاد می‌شوند. سال‌ها است که فعالان عدالت کیفری و پژوهشگران برای حذف وثیقه [از نظام کیفری]، که به طور نامتناسبی براساس نژاد و وضعیت اجتماعی-اقتصادی، به افراد آسیب می‌رساند، تلاش می‌کنند؛ با این حال، [گرفتن این وثیقه‌ها] منجر به افزایش امنیت عمومی نشده است . در واکنش به این موضوع، نیوجرسی و کالیفرنیا به تازگی قوانینی را برای پاسخ دادن به این نگرانی‌ها تصویب کردند. با این حال، به جای اینکه به سادگی به گرفتن وثیقه پایان دهند، آن را با یک سیستم «ارزیابی پیش از محاکمه»^۱ طراحی کردند که «امتیاز ریسک» تولیدشده توسط آن تعیین می‌کند که آیا فرد باید آزاد شود یا در زندان منتظر محاکمه بماند .

تغییر از سیاست‌هایی مانند دریافت وثیقه به سیستم‌های خودکار و سیستم‌های ارزیابی ریسک امتیازی اتفاقی نسبتاً جدید است و حتی بدون وجود پژوهش‌های قابل توجهی که امکان تقویت تبعیض در نظام عدالت کیفری را بررسی کنند، هنوز در حال ادامه یافتن هستند. با این وجود، چند شاخص اولیه‌ای وجود دارد که نگرانی‌ها را افزایش می‌دهد. قانون نیوجرسی در سال ۲۰۱۷ به اجرا درآمد و با این که دولت با کاهشی در آمار افراد پیش‌محاکمه‌شونده مواجه شده، اما فعالان اجتماعی ابراز نگرانی کرده‌اند که نابرابری‌های نژادی در سیستم ارزیابی ریسک پابرجاست . به طور مشابه، هنگامی

1. pretrial assessment system

که قوانین کالیفرنیا در اوایل سال جاری تصویب شد، بسیاری از حامیان عدالت کیفری که خواهان پایان دریافت وثیقه بودند و [حتی] نسخه اولیه‌ی این لایحه را حمایت کردند، به دلیل الزامات [غلط] ارزیابی ریسک، با نسخه نهایی آن مخالفت کردند.

سیاست‌های آموزشی نیز از سیستم‌های تصمیم‌گیری خودکار تاثیر می‌پذیرند. یکی از اساتید دانشگاه کالج لندن از جمله کسانی است که به نفع جایگزینی [آزمون‌های] هوش مصنوعی به جای آزمون‌های معمول استدلال کرده است؛ او همچنین پیشنهاد کرده که سیستم AIAssess که متعلق به آزمایشگاه دانش UCL است، می‌توانند «در ارزیابی دانش و فهم فرزندانمان قابل اعتماد باشد» و می‌تواند در جهت جایگزینی یا تقویت آزمون‌های سنتی عمل کند. با این حال، مجموعه‌ی رو به رشدی از تحقیقات وجود دارد که نشان می‌دهد سیستم‌های امتیازدهی خودکار مانند بسیاری از انواع دیگر هوش مصنوعی، ممکن است جانبداری علیه گروه‌های زبانی و نژادی خاص را به گونه‌ای کدگذاری کند که الگوهای محروم‌سازی آن زبان‌ها و اقوام را بازتولید کند. تصمیمات غیرعادلانه‌ای که براساس امتیازدهی خودکار به دانش‌آموزانی اختصاص می‌یابد که متعلق به گروه‌هایی هستند که به صورت نظام‌مند و از لحاظ تاریخی محروم هستند، به احتمال زیاد پیامدهای عمیقی بر زندگی [آن] کودکان خواهد داشت و به تشدید شدن نابرابری در دسترسی به فرصت‌های شغلی و منابع خواهد انجامید.

تبعات سیستم‌های ADS آموزشی از موضوعاتی مانند تکلیف مدرسه و حتی حمل و نقل هم فراتر می‌رود. شهر بوستون در سال جاری

پس از دو تلاش ناموفق برای حل مسئله‌ی تخصیص برابر مدارس با استفاده از سیستم‌های خودکار، در کانون توجه قرار گرفت. در مرحله‌ی اول، اداره‌ی مدارس، یک الگوریتم تخصیص جغرافیایی برای مدارس به کار برد که هدف آن فراهم کردن دسترسی به مدارس باکیفیت‌تر و نزدیک‌تر به خانه برای دانش‌آموزان بود. هدف اصلی مسئولان شهر، افزایش یکپارچگی نژادی و جغرافیایی در حوزه‌ی مدارس بود؛ اما گزارش ارزیابی نشان داد که [این سیستم] دقیقاً خلاف آن [هدف] را انجام داده است: با این حال که [این سیستم] رفت و آمدهای روزانه‌ی دانش‌آموز را کم کرده، اما در نهایت یکپارچگی مدارس را [نیز] کاهش داده است. محققین اشاره کردند که این امر به نوبه خود، به این دلیل بود که برای سیستم غیرممکن بود که هدف مورد نظر خود را با توجه به تاریخ و زمینه‌ای که در آن مورد استفاده قرار گرفته بود، برآورده کند. پیش از این اتفاق، توزیع جغرافیایی کیفیت مدارس در بوستون غیرمنصفانه بود و نابرابری‌های نژادی پیش‌موجود در این مدارس بود که در بوجود آمدن پیچیدگی‌هایی که فائق آمدن بر آن‌ها برای یک الگوریتم ناممکن بود. پس از این اتفاق، اداره‌ی مدارس بوستون دوباره تلاش کرد تا از یک سیستم الگوریتمی برای بهبود نابرابری استفاده کند؛ به طوری که این بار [این سیستم را] به منظور پیکربندی مجدد زمان شروع مدارس طراحی کرد؛ این کار با هدف شروع دیرتر مدارس دبیرستان و شروع زودتر مدارس راهنمایی انجام شد. این کار، بر اساس تشخیص ساعت زیستی دانش‌آموزان در سنین مختلف، به منظور تلاش در جهت بهبود سلامت و عملکرد دانش‌آموزان و برای بهینه‌سازی استفاده از

اتوبوس‌های مدرسه جهت صرفه‌جویی انجام شد. همچنین مقصود دیگر این برنامه، افزایش یکسان‌نگری نژادی بود؛ چرا که دانش‌آموزان رنگین‌پوست عمدتاً با زمان‌های نامناسبی برای شروع و اتوبوس‌سواری‌های طولانی مواجه هستند. مسئولان این شهر سیستم تصمیم‌گیری خودکاری را توسعه دادند که برای این اهداف بهینه شده بود. با این حال، به دلیل واکنش عمومی شدید پیاده‌سازی نشد و در نهایت منجر به برکناری سرپرست اداره‌ی مدارس شد.

در این مورد، فرآیند طراحی به اندازه کافی نیازهای خانواده‌ها را تشخیص نداد، یا آن‌ها را در تعریف و بازبینی اهداف سیستم در نظر نگرفت. با اعمال سیستم پیشنهادی، والدین دانش‌آموزان راهنمایی و دبیرستان، به دلیل تغییر زمان شروع و پایان مدارس، باید زمان‌بندی‌های خود را تغییر اساسی می‌دادند که منجر به فشار آمدن به افرادی می‌شد که انعطاف این تغییرات را نداشتند. انجمن ملی پیشرفت رنگین‌پوستان (NAACP) و کمیته حقوق مدنی و عدالت اقتصادی و کلاً^۲ با این طرح مخالفت کردند؛ چرا که اداره‌ی مدارس در ارزیابی والدین رنگین‌پوست و کم‌درآمدی که معمولاً شغلشان انعطاف‌پذیری زمانی ندارد و آن‌هایی که قدرت مالی برای هزینه‌های اضافی نگهداری کودکان را ندارند، موفق نبودند.

این تلاش‌های ناموفق دو مساله مهم را نشان می‌دهند که سیاست‌گذاران باید هنگام ارزیابی کاربرد این سیستم‌ها در نظر بگیرند. اول، مشکلات ساختاری و نظام‌مند ما ادامه خواهند یافت و اگر قبل از طراحی و پیاده‌سازی، این سیستم‌ها مورد بررسی قرار نگیرند، احتمالاً منافع بالقوه این سیستم‌ها [توسط آن مشکلات]

تضعیف خواهند شد. دوم، مشارکت اجتماعی قدرتمند و هدفمند باید در فرآیند هدف‌گذاری یک سیستم گنجانده شود.

در چارچوب تحقیقاتی ارزیابی تاثیر الگوریتم که متعلق به موسسه‌ی AI Now است، مشارکت اجتماعی -هم در مرحله طراحی و هم پیش از پیاده‌سازی، هنگام و پس از آن- بخش مهمی از فرآیندهای پاسخگویی ADS است. هنگامی که جوامع آسیب‌دیده این فرصت را داشته باشند که استفاده از سیستم‌هایی که قابل قبول نیستند را ارزیابی کرده و نپذیرند، و پیش از آن که [نتایج] نقص‌های اساسی سیستم به جای گذاشته شود، امکان اعتراض به آن‌ها وجود داشته باشد، در این صورت اعتبار و مشروعیت [استفاده از] این سیستم به طور گسترده‌ای بهبود می‌یابد. چنین مشارکتی در جهت خدمت به جوامع و نهادهای دولتی است: اگر والدین رنگین‌پوست و کم‌درآمد در بوستون به طور معنی‌داری در ارزیابی سیستم آغاز زمان مدارس مشارکت می‌کردند، ممکن بود مسائل آن‌ها در طراحی سیستم، حفظ زمان و منابع شهری و فراهم کردن یک مدل مورد نیاز برای نظارت لحاظ شود.

مهم‌تر از همه، هنگامی که این سیستم‌ها مانند «جعبه‌ی سیاه» هستند، پاسخگویی درباره‌ی استفاده‌ی دولتی از آن‌ها ممکن نیست. زمانی که فروشندگان واسطه بر روی [قانون] رازداری تجاری پافشاری می‌کنند تا [جزئیات] سیستم‌های‌شان را مبهم نگه دارند، هر راهی برای اصلاح یا درخواست تجدید نظر بسیار دشوار می‌شود. به همین دلیل است که فروشندگان باید از رازداری تجاری و سایر ادعاهای قانونی که مانع توانایی فهم، بررسی دقیق یا آزمایش سیستم‌های‌شان

در مورد مسائلی نظیر جانبداری، خطا و غیره می‌شوند، چشم‌پوشی کنند. برای اشخاص دولتی و کسانی که اثرات این سیستم‌ها را مطالعه می‌کنند مهم است که بفهمند که چرا [یک ماشین، این] توصیه‌های خودکار را ارائه می‌دهد و چرا باید به اعتبار آن‌ها اعتماد کنند. حتی مهم‌تر آن است که کسانی که زندگی شان تحت تاثیر منفی این سیستم‌ها قرار می‌گیرند، باید بدانند که قادر به اعتراض و تجدید نظر نسبت به تصمیمات نامطلوب آن سیستم‌ها هستند. باشند. دولت‌ها باید محتاط باشند: با این که سیستم‌های تصمیم‌گیری خودکار ممکن است نویدبخش صرفه‌جویی در هزینه و بازدهی بالا باشند، این دولت‌ها هستند - و نه فروشندگان واسطه - که در نهایت مسئول به نتیجه نرسیدن آن‌ها خواهند بود. بدون شفافیت، پاسخگویی و نظارت کافی، این سیستم‌ها [تنها] در حال عرضه و تقویت روش‌های ناعادلانه و خودسرانه در تصمیم‌گیری‌ها و سیاست‌گذاری‌های مهم دولت هستند. می‌کنند.

۳-۲- آزمایش بر روی جامعه: چه کسی این بار را به دوش می‌کشد؟

طی ده سال گذشته، تمرکز و سرمایه‌گذاری بر تحقیق و توسعه‌ی فنی هوش مصنوعی شتاب گرفته است. اما تلاش برای تضمین اینکه این سیستم‌ها ایمن و غیر تبعیض‌آمیز هستند، منابع و توجه یکسانی را جلب نکرده است. در حال حاضر، روش‌های کمی برای اندازه‌گیری، اعتباربخشی^۱ و نظارت بر اثرات سیستم‌های هوش مصنوعی که در شرایط غیرکنترل‌شده اعمال می‌شوند، وجود

1. validating

دارد. سیستم‌های هوش مصنوعی که وظیفه تصمیم‌گیری‌های مهم را دارند، اغلب با نظارت کم یا چارچوب نظارتی غیرشفافی، بر روی جمعیت‌های زنده آزمایش می‌شوند.

برای مثال، در مارس ۲۰۱۸، یک ماشین خودران شرکت اوبر، در حومه شهر فینیکس در حال حرکت بود که نتوانست یک زن را «ببیند» و با او برخورد کرد و او را کشت. در ماه مارس گذشته، تسلا تایید کرد که دومین راننده نیز در ماشینی که تکنولوژی راننده‌ی خودکار روی آن نصب بوده، در یک تصادف کشته شده است. هیچ یک از دو شرکت متحمل عواقب جدی نشدند، و فقط در مورد اوبر، شخصی که مسئول مراقبت از ماشین خودران بود، در نهایت سرزنش شد. با وجود این اشتباهات مرگبار، Waymo که متعلق به شرکت Alphabet است، اخیراً از طرحی با عنوان «پروژه‌ی سواری زود هنگام» در شهر فینیکس رونمایی کرده است. شهروندان می‌توانند در این طرح ثبت‌نام کرده که سوژه‌ی آزمایش Waymo بشوند و سوار ماشین‌های خودران بشوند.

بسیاری ادعا می‌کنند که تلفات کم ماشین‌های خودران باید در بافتار اکوسیستم موجود فهم شود که در آن تلفات زیادی، بدون وجود هوش مصنوعی نیز رخ می‌دهد. با این حال، خطراتی که از برهم‌کنش انسان و ماشین ایجاد می‌شوند، [تنها] بخاطر این که نظام‌های مقرراتی و مسئولیتی، با انسان و ماشین متفاوت برخورد می‌کنند، در سطح نظارتی یا پاسخگویی کاملاً مجزایی قرار نمی‌گیرند. انگیزه‌های قوی برای تنظیم مقررات قانونی و حقوقی در این حوزه و بسیاری از حوزه‌های دیگر هوش مصنوعی وجود دارد.

برای مثال، این واقعیت که شهر فینیکس محل آزمایشی برای Waymo و Uber است، تصادفی نیست.

در اوایل سال جاری، آریزونا، رسماً آنچه را که دولت در سال ۲۰۱۵ اجازه داده بود را اعلام کرد: رانندگی ماشین‌های کاملاً خودران، بدون راننده در معابر عمومی مجاز است. این سیاست بدون هیچ چارچوب مقرراتی وضع شده است؛ چارچوب مقرراتی که از الزامات مقابله با مسایل پیچیده‌ای به حساب می‌آید که درباره‌ی پاسخ‌گویی و الزامات قانونی بوجود می‌آید. به گفته فینیکس تایمز: «آریزونا موافقت کرده که کنار برود و شاهد چگونگی رشد این فناوری باشد. اگر مشکلی هم پیش بیاید، هنوز برنامه‌ای برای آن وجود ندارد.» این شکاف پاسخگویی مقرراتی به وضوح در مورد حادثه‌ی اوبر که ظاهراً به علت ترکیبی از سودجویی شرکت و حواس‌پرتی راننده‌ی پشتیبان بوده، قابل‌مشاهده است.

در حالی که شاید ماشین‌های خودران واضح‌ترین خطر غیرنظامی هوش مصنوعی به امنیت انسانی باشند، اما حوزه‌های دیگر هوش مصنوعی نیز نگرانی‌هایی جدی را مطرح می‌کنند. به عنوان مثال، «واتسون برای آنکولوژی» که متعلق به شرکت IBM است، در حال حاضر در بیمارستان‌های سرتاسر جهان مورد آزمایش قرار می‌گیرد که در تشخیص بیماری و مراقبت‌های بالینی کمک می‌کند. در حال حاضر به شکل فزاینده‌ای، اثربخشی آن و وعده‌های تجاری IBM مورد سوال قرار می‌گیرند. خبرنگاران تحقیقی به مدارک محرمانه‌ای دست یافتند که تصویرگر طرحی مشکل‌زا از سیستم IBM، از جمله توصیه به «روش‌های درمانی غیرایمن و نادرست سرطان» است. در حالی که

این سیستم هنوز در مرحله آزمایشی خود بود، نگرانی‌های جدی در مورد انگیزه‌های محرک گسترش این فن‌آوری، فقدان اعتبار پزشکی و نبود تحقیقات هم‌تادآوری‌شده را نشان داد. کرد.

چنین رویدادهایی سرعت گسترش هوش مصنوعی را در بخش سلامت و درمان کم نکرده‌اند. اخیراً، اداره غذا و داروی آمریکا (FDA) تصمیمی بحث‌برانگیز در این باره گرفت که Apple Watch جدید را، که در آن الکتروکاردیوگرام (EKG) به کار رفته و می‌تواند کاربر را از ریتم نامنظم قلبش مطلع سازد، به عنوان محصولی ایمن برای مصرف کنندگان اعلام کند. حال، نگرانی‌هایی وجود دارد مبنی بر این که تلاش FDA برای حرکت هم‌سرعت با نوآوری، می‌تواند با نگرانی‌های مربوط به حریم خصوصی و امنیت داده در تلافی باشد. به طور مشابه، تصمیم DeepMind Health برای استفاده از فناوری Streams، که ابزاری طراحی‌شده توسط گوگل، به منظور پشتیبانی از تصمیمات پرستاران و کارورزان بهداشتی است، موجب نگرانی درباره‌ی محقق نشدن وعده‌های DeepMind در مورد به اشتراک نگذاشتن داده‌های بیمار بشود. بشوند.

غالباً کودکان و جوانان سوژه‌ی چنین آزمایش‌هایی هستند. در اوایل سال جاری مشخص شد که پیرسون^۱، یکی از فروشندگان اصلی سیستم‌های آموزشی هوش مصنوعی، «تیمه‌دات اجتماعی-روانشناختی» را به یکی از نرم‌افزارهای آموزشی خود اضافه کرد تا ببیند که چه پاسخ‌هایی از ۹۰۰۰ دانش‌آموز می‌گیرد. آن‌ها این کار را بدون رضایت یا آگاهی دانش‌آموزان، والدین یا معلمان انجام دادند. سپس شرکت پی‌گیری کرد که آیا دانش‌آموزانی که از طریق

1. Pearson

نرم‌افزار آموزشی [عنوان] «فکر رشدیافته» را دریافت کرده‌اند، مسائل بیشتری را نسبت به دانشجویانی که این کار را انجام نداده‌اند، حل کرده‌اند یا نه. این آزمایش روان‌شناختی بر روی افراد ناآگاه، به ویژه کم سن و سالی که در نظام آموزشی [درس می‌خوانند]، نگرانی‌هایی را مرتبط با اخلاق و حریم خصوصی افزایش می‌دهد. همچنین، این مسئله اثرات فزاینده شرکت‌های خصوصی در دامنه‌های ظاهراً عمومی و فقدان شفافیت و رویه‌های عادلانه که همراه شیوه‌های کنونی استقرار و یکپارچه‌سازی هوش مصنوعی را باشند را برجسته می‌کند. حال، نه تنها شاهد نمونه‌هایی از آسیب‌های واقعی هستیم که می‌توانند از سیستم‌های جانبدارانه و نادقیق هوش مصنوعی ناشی شوند، بلکه شواهدی از تمایل صنعت هوش مصنوعی برای اعمال نسخه‌های جدید آزمایشی بر روی جمعیت‌های انسانی وجود دارد. اخیراً که آمازون با تاکتیک‌های تبلیغاتی فریبنده، برای کسب درآمد از ازدواج و ثبت تولد بچه مورد نقد واقع شد، این‌گونه واکنش نشان داد: «ما دائماً در حال آزمایش هستیم.» وقتی که تسلط بازار و سود نسبت به ایمنی، شفافیت و اعتماد ارزش بیشتری دارد، این اتفاقات [بخشی از] الگویی تکرارشدنی هستند. بدون چارچوب معناداری از پاسخگویی و همچنین ساختارهای مقرراتی مستحکم، اندازه و مقیاس این نوع آزمایش‌های کنترل‌نشده گسترش خواهد یافت و خطرات بالقوه رشد خواهد کرد.

بخش سوم

راه‌حل‌های سال ۲۰۱۸



۳-۱- محدودیت های اصلاحات فناورانه

طی سال گذشته، درباره‌ی این که سیستم های هوش مصنوعی جانبداری را توسعه و تداوم می‌بخشند و روش های محاسباتی ذاتاً بی‌طرف و عینی نیستند، شاهد افزایش اتفاق نظر بودیم. این بازشناسی به کمک پروژه‌ی «عدالت، پاسخگویی و شفافیت» در «انجمن یادگیری ماشین» و در پی سلسله‌ای از اتفاقات رخ داد؛ از جمله، شواهدی از جانبداری در ارزیابی الگوریتمی مخاطرات پیش از محاکمه. همچنین پروژه‌ی عدالت، پاسخگویی و شفافیت در انجمن یادگیری ماشین نیز به این فهم کمک کرد. این انجمن در هسته‌ی بدنه‌ی تحقیقات آکادمیک نوظهور در زمینه جانبداری و بی‌طرفی هوش مصنوعی قرار دارد که با روش‌هایی معطوف به حل این مسائل، چشم‌اندازی به سوی ماهیت‌شان ایجاد می‌کند.

به منظور یافتن «بی‌طرفی الگوریتمی»، تعاریف متعددی از بی‌طرفی، به همراه استراتژی‌هایی برای دست یافتن به آن، عمدتاً توسط جامعه مهندسی در چند سال اخیر پیشنهاد شده است. این پروژه توسعه الگوریتم‌هایی جدید و تکنیک‌هایی آماری که هدفشان

تشخیص و کاهش جانبداری است را توصیه می‌کند. موفقیت این تکنیک‌ها عموماً در مقایسه با تعریف‌های محاسباتی دیگر از بی‌طرفی و براساس مجموعه‌ای ریاضیاتی از نتایج سنجیده می‌شود. با این حال، مشکلاتی که نهایتاً این روش‌ها به دنبال درمان آن‌ها هستند، ریشه‌های عمیق اجتماعی و تاریخی دارند. در ادامه، بررسی مختصری از برخی از رویکردهای مهم درباره‌ی فهمیدن و تعریف مسائل مرتبط با جانبداری و بی‌طرفی الگوریتمی آمده است.

• آسیب‌های تخصصی، اثراتی از سیستم‌های هوش مصنوعی را توصیف می‌کند که خدمات، منابع یا فرصت‌ها را از بعضی افراد به طور غیر منصفانه دریغ می‌دارد. چنین صدماتی توجه کسانی را به خود جلب کرده که اهتمام به ایجاد تمهیدات فنی‌ای دارند که بی‌طرفی سیستم‌های هوش مصنوعی را تضمین می‌کند؛ تا حدی به این دلیل که (از لحاظ نظری) امکان کمی‌سازی و درمان این آسیب‌ها وجود دارد. با این حال، توجه ما نسبت به اصلاح سیستم‌هایی که آسیب‌های بازنمایانه^۱ را تقویت و بازتولید می‌کنند، کمتر شده است: آسیب‌های حاصل‌شده از سیستم‌هایی که رفتارهای کلیشه‌ای آسیب‌زننده را بازتولید و تقویت می‌کنند و اغلب به روش‌هایی عمل می‌کنند که مفروضات بکار رفته برای توجیه تبعیض و نابرابری را انعکاس می‌دهند. در سخنرانی افتتاحیه کنفرانس پردازش عصبی اطلاعات (NeurIPS) در سال ۲۰۱۷، کیت کرافورد، موسس AI Now روشی را توصیف کرد که در آن، الگوهای تاریخی تبعیض و طبقه‌سازی که اغلب بخاطر تفاوت‌های دریافت‌شده، بازنمایی آسیب‌زننده‌ای از مردم می‌سازند، در فرضیات

1. representational harms

و داده‌های ورودی سیستم‌های هوش مصنوعی منعکس می‌شوند و اغلب منجر به آسیب‌های تخصیصی^۱ می‌شوند. این دیدگاه مستلزم آن است که مسئله را چیزی بیشتر از وجود جانبداری در یک الگوریتم یا یک مجموعه‌ی داده ببینیم. ما باید به «نقش هوش مصنوعی در بازنمایی آسیب‌زننده به هویت انسان» و روشی که چنین بازنمایی‌های مضر را شکل می‌دهد توجه کنیم؛ هم‌چنین باید به این بازنمایی‌ها که فهم اجتماعی و فرهنگی ما از خود و دیگری را می‌سازد، توجه کرد. کنیم.

• استراتژی‌های مشاهده‌تی بی‌طرفی، با زیر ذره‌بین قرار دادن مجموعه‌ای از داده‌ها (خواه داده‌های استفاده‌شده برای آموزش مدل‌های هوش مصنوعی باشد، خواه داده‌های پردازش‌شده‌ی ورودی توسط همین مدل‌ها) و اعمال روش‌هایی مبادرت به تشخیص و کاهش جانبداری می‌کند. این روش‌ها به منظور تشخیص آن است که آیا این داده‌ها براساس ویژگی‌هایی از قبیل نژاد، جنسیت، یا طبقه‌ی اجتماعی-اقتصادی، جانبداری علیه افراد یا گروه‌ها رمزگذاری می‌کند یا نه. این ویژگی‌ها معمولاً ویژگی‌هایی حفاظت‌شده یا حساس به حساب می‌آیند. اکثر رویکردهای مشاهده‌تی بی‌طرفی می‌توانند شکل‌هایی از فرآیندهای ضدطبقه‌بندی، توازن طبقه‌بندی یا کالیبراسیون باشند که توسط سام کوربت-دیویس^۲ و شاراد گوئل^۳ مطرح شده‌است. استراتژی‌های مشاهده‌تی بی‌طرفی به طرز فزاینده‌ای به واسطه‌ی تلاش‌های جامعه برای مقابله با محدودیت‌های پروژه‌ی بی‌طرفی مهندسی بوجود آمده و روزه‌های ورود به دیگر رشته‌ها را فراهم کرده است.

- استراتژی‌های ضد طبقاتی‌سازی زمانی یک مدل یادگیری ماشین را بی‌طرف می‌دانند که این مدل وابسته به ویژگی‌های حفاظت‌شده‌ی موجود در مجموعه داده نباشد. به عنوان مثال، این استراتژی، ارزیابی ریسک پیش از محاکمه دو متهمی که از نظر نژاد یا جنسیت متفاوت بوده اما از نظر دیگر اطلاعات شخصی شان برابر هستند را در صورتی «بی‌طرفانه» به حساب می‌آورد که [این مدل] به آن‌ها ریسک یکسانی نسبت بدهد. این استراتژی اغلب به حذف همه ویژگی‌های محافظت‌شده و تمام ویژگی‌های مرتبط با آن‌ها از مجموعه‌ی داده‌هایی ختم می‌شود که برای آموزش مدل استفاده می‌شود. می‌کند.
- توازن طبقاتی زمانی یک مدل را بی‌طرف می‌داند که عملکرد پیش‌بینی‌گری آن برای [همه‌ی] گروه‌هایی که با ویژگی‌های محافظت‌شده‌شان تعریف می‌شوند، برابر باشد. به عنوان مثال، توازن طبقاتی تضمین می‌کند که برای دریافت یک وام، اگر همه‌ی افراد خوش حساب باشند، درصد افرادی که یک الگوریتم، آن‌ها را برای دریافت وام لایق نمی‌داند، میان جمعیت‌های سیاه‌پوست و سفیدپوست یکسان باشد. در عمل، این استراتژی به منظور یکی شدن جمعیت‌های خاص با دیگران، اغلب منجر به کاهش «دقت» نسبت به آن جمعیت‌های خاص می‌شود.
- استراتژی‌های کالبراسیون، زمانی که یک سیستم هوش مصنوعی تصمیم یا پیش‌بینی‌ای ایجاد کرده، کم‌تر به اطلاعات و بیشتر به نتیجه توجه می‌کنند. این استراتژی‌ها برای حصول اطمینان از این که نتایج به ویژگی‌های حفاظت‌شده وابسته نیستند، عمل

می‌کنند. به عنوان مثال، در مورد اشاره‌شده درباره‌ی ارزیابی ریسک پیش از محاکمه، با اعمال استراتژی کالیبراسیون می‌توان اطمینان حاصل کرد که در میان گروهی از متهمان با امتیاز ریسک مشابه، احتمال این که کسی واقعا ارتکاب مجدد جرم را انجام بدهد، در میان همه‌ی افراد با ویژگی‌های محافظت‌شده‌ی غیر یکسان، یکسان خواهد بود.

محققان بسیاری محدودیت‌های این رویکردها را شناسایی کرده‌اند. درباره‌ی استراتژی ضدطبقاتی‌سازی، برخی استدلال می‌کنند که موارد مهمی وجود دارد که در آن‌ها ویژگی‌های محافظت‌شده - مانند نژاد یا جنسیت - باید به منظور حصول اطمینان از تصمیمات بی‌طرفانه، در داده‌های مورد استفاده برای آموزش هوش مصنوعی لحاظ شوند. به عنوان مثال، کوربت-دیویس و گوئل درباره‌ی اهمیت دخالت دادن جنسیت در ارزیابی ریسک پیش از محاکمه بحث می‌کنند. در حالی که زنان در بسیاری از حوزه‌های قضایی کم‌تر از مردان به ارتکاب مجدد جرم دچار می‌شوند، ارزیابی‌های ریسک که [به ظاهر] نسبت به جنسیت خنثی هستند، در مورد تکرار جرم زنان دچار بزرگنمایی می‌شوند «که می‌تواند منجر به تصمیمات غیرضروری و ناپه‌نچار قضایی شود». در نتیجه، [در حال حاضر] برخی از حوزه‌های قضایی [در واقع] از ابزارهای ارزیابی ریسکی که نسبت به جنسیت حساس هستند، استفاده می‌کنند. این مثال‌ها در برابر این دیدگاه رایج که حذف کافی اطلاعات از مجموعه داده‌ها در نهایت منجر به «غیرجانبدارانه شدن» یک سیستم هوش مصنوعی می‌شود، می‌ایستند. از آنجا که تقریبا همیشه همبستگی بین متغیرها در مجموعه‌ای

از داده‌ها وجود دارد، حذف این متغیرهای محافظت‌شده می‌تواند اطلاعات بسیار کمی را باقی بگذارد و بنابراین باعث می‌شود که عملکرد پیش‌بینی‌گری، بدون توانایی اندازه‌گیری آسیب‌های بالقوه، ضعیف شود.

دوم، برخی استدلال کرده‌اند که معیارهای مختلف بی‌طرفی ریاضیاتی به صورت دو به دو، مانع‌الجمع هستند. از این رو، به جز در موارد بسیار محدود، عموماً غیر ممکن است که همزمان هم استراتژی کالیبراسیون و هم انواع توازن طبقاتی را ارضا کرد ساخت. این «امکان‌ناپذیری در نتایج» نشان می‌دهند که چگونه هر استراتژی بی‌طرفی‌ای، در مورد آنچه که بی‌طرفانه است یا نه، مفروضات پنهان مختلفی را در نظر می‌گیرد. در نهایت، این یافته‌ها [درباره‌ی مفروضات مذکور] به پیچیده‌شدن بحث‌های سیاست‌گذاری که به حل مسایل جانبداری با [استفاده از] ابزارهای ریاضیاتی توجه دارند، منجر می‌شود. چیزی که این یافته‌ها روشن می‌کنند این است که حل مسائل پیچیده سیاست‌گذاری که مرتبط با جانبداری و تبعیض هستند، از طریق بکارگیری یک یا چند معیار ریاضیاتی بی‌طرفانه، دور از موفقیت به نظر می‌رسد. این بدان معنا نیست که چنین معیارهایی مفید نیستند: معیارهای مشاهده‌تی، ممکن است به فهمیدن این موضوع کمک کنند که آیا مجموعه داده‌ها و سیستم‌های هوش مصنوعی با مفاهیم مختلفی از بی‌طرفی و جانبداری کار می‌کنند یا نه، که متعاقباً به آگاهان ما از بحثی غنی‌تر درباره‌ی اهدافی کمک می‌کند که امیدواریم به هنگام بکارگیری هوش مصنوعی در بافتارهای پیچیده‌ی اجتماعی به آن‌ها برسیم.

گسترش روش‌های مشاهدتی بی‌طرفی نگرانی‌هایی در مورد امکان ایجاد حس اطمینان کاذب را گسترش می‌دهد. در حالی که محققان اغلب احساسات کم‌رنجی درباره‌ی محدودیت‌های ابزار خود دارند، دیگرانی که آن‌ها را پیاده‌سازی می‌کنند ممکن است زمانی که به دنبال اصلاحات سریع هستند، چنین محدودیت‌هایی را نادیده بگیرند. این ایده که به هنگام «درمان شدن» سیستم‌های هوش مصنوعی با چنین روش‌هایی آن‌ها عاری از جانبداری هستند و برای استفاده در حوزه‌های حساس ایمن هستند، می‌تواند حس امنیت کاذبی که مخاطره‌انگیز است را ایجاد کند - امنیتی که به شدت متکی به تعاریف ریاضیاتی از بی‌طرفی بوده و به بافت عمیق‌تر اجتماعی و تاریخی بی‌توجه است. فرانک پاسکال که محقق حوزه‌ی حقوق است، هشدار می‌دهد: «الگوریتم‌ها به تنهایی نمی‌توانند الگوریتم‌های دیگری را به شکل معناداری پاسخگو نگه دارند.»

در حالی که توجه بیشتر به مشکلات بی‌طرفی و جانبداری در هوش مصنوعی یک پیشرفت مثبت محسوب می‌شود، برخی نسبت به «ریاضیاتی‌سازی اصول اخلاقی» ابزار نگرانی کرده‌اند. همانطور که شیرا میچل^۱ استدلال کرده است:

«ما به عنوان صاحب‌نظران علم آمار در عرصه سیاست‌گذاری، باید نسبت به مخاطرات جایگزین شدن سیاست‌گذاری با گفتمان تخصصی هوشیار باشیم. به طور کلی، هر نوع مداخله آماری در گفتگو، منجر به افزایش ورودی‌های فنی می‌شود، مگر اینکه این مداخله به گفتگویی بین متخصصان فنی، کاهش پیدا کند. آیا ما به خاطر قدرت آمار را می‌خواهیم؟ یا اینکه قدرت را صرفاً با ابزارهای جدیدی که در جهت

به حاشیه راندن دغدغه های سیاستی غیر کمی بکار می‌روند، فراهم می‌کنیم؟ می‌یابد.»

چنین نگرانی‌هایی جدید نیستند. اثر در حال چاپ هاتچینسون^۱ و میچل، پنجاه سال تلاش به منظور ساختن تعاریف کمی از بی‌طرفی در حوزه‌های مختلف را بررسی می‌کند. کار آن‌ها دوره‌ای بین سال‌های ۱۹۶۴ و ۱۹۷۳ را یادآور می‌شود که در آن، محققان برای ارزیابی‌های تحصیلی روی تعریف بی‌طرفی تمرکز کرده بودند؛ تکرار این بحث‌ها به نوعی در مباحثات کنونی بی‌طرفی در هوش مصنوعی را دیده می‌شود. تلاش آن‌ها پس از اینکه نتوانستند در مورد «راه‌حل‌های تکنیکی فراگیر در مسائلی که بی‌طرفی در آن دخیل است» به توافق برسند، متوقف شد. این سوابق بر آن چه پروژه‌ی «عدالت، پاسخگویی و شفافیت» در «انجمن یادگیری ماشین» در حال یافتن بود، تاکید دارد: بدون «ارتباط تنگاتنگ با تاثیرات دنیای واقعی»، ارزش افزوده‌ی معیارهای جدید بی‌طرفی و الگوریتم‌های موجود در انجمن یادگیری ماشین [تنها] می‌تواند حداقلی باشد. به منظور رسیدن به تحقیقات هدفمند در موضوعاتی مانند بی‌طرفی و جانبداری الگوریتمی، ما باید مداوماً تخصص‌ها و نگرش‌های جمعیت‌های بیرون از رشته‌های مهندسی را در کنار مهندسان قرار دهیم.

رهیافت‌هایی گسترده‌تر

دوب و همکاران به تعریفی از جانبداری نزدیک شده‌اند که در آثار اولیه‌ی حوزه‌ی طراحی حساس به ارزش (VSD) به منظور ارائه‌ی نگرش گسترده‌تری از مسئله‌ی بی‌طرفی پیشنهاد شده شده بود.

1. Hutchinson
2. Dobbe

طراحی حساس به ارزش که در دهه نود میلادی توسط باتیا فریدمن^۱ و هلن نیسن باوم^۲ نظریه پردازی شد، ادعا می‌کند که جانبداری در سیستم‌های کامپیوتری پیش از وجود آن سیستم‌ها، وجود دارد. این جانبداری در داده‌هایی که وظیفه‌ی آموزش سیستم را دارند منعکس شده و در فرضیاتی که در طول ساخت یک سیستم کامپیوتری در نظر گرفته می‌شوند، ریشه دارد. این جانبداری، به علت وجود حلقه‌های بازخوردی و ناسازگاری میان سیستم و بافت پویای اجتماعی و فرهنگی، به هنگام کارکرد سیستم‌ها آشکار می‌شود. رویکرد طراحی حساس به ارزش، راهی است برای داشتن چشم‌اندازی وسیع‌تر به این مسائل؛ به طوری که بر منافع و دیدگاه‌های ذی‌نفعان مستقیم و غیر مستقیم، در طول فرآیند طراحی تاکید می‌کند.

رویکرد دیگر «تحلیل سیستم‌های اجتماعی» نام دارد که اولین بار توسط کیت کرافورد و رایان کالو در مجله‌ی Nature بسط داده شد. این روش، با تحلیل کالبدسنجانه‌ی یک سیستم تکنیکی، شیوه‌های کمی و کیفی پژوهش را ترکیب می‌کند و در عین حال، فناوری مدنظر را هنگام استفاده در محیط‌های اجتماعی نیز مطالعه می‌کند. این رویکرد پیشنهاد می‌کند که در هر مرحله، اعم از مفهوم‌سازی اولیه، طراحی، پیاده‌سازی و تنظیم مقررات یک فناوری، با تاثیرات اجتماعی [متناظر با آن] در سرتاسر چرخه حیات دست و پنجه نرم کنیم. همچنین شاهد افزایش تمرکز بر روی بررسی خاستگاه و روش ساخت داده‌های استفاده‌شده جهت آموزش به سیستم‌های هوش مصنوعی هستیم. این داده‌ها «جهان‌بینی» سیستم‌های هوش مصنوعی را شکل می‌دهند؛ دانستن چگونگی ایجاد این داده‌ها و

آنچه که بازنمایی می‌کنند، برای درک محدودیت سیستم‌هایی که توسط آن داده‌ها آموزش می‌بینند، امری ضروری است. به عنوان راه‌حلی اولیه برای این مسئله، گروهی از محققان به رهبری تیمنیت گبرو^۱ پیشنهاد شکلی استاندارد از مستندسازی با عنوان «datasheets for datasets» را دادند که مجموعه داده‌های استفاده‌شده جهت آموزش به سیستم‌های هوش مصنوعی را مورد مطالعه قرار می‌دهد. مقاله‌ای تکمیلی در حال حاضر نگاهی به استانداردسازی خاستگاه داده برای مدل‌های هوش مصنوعی دارد. این رویکردها به متخصصین هوش مصنوعی و کسانی که نظارت بر کاربردهای هوش مصنوعی در بافتاری خاص و ارزیابی آن‌ها را بر عهده دارند، اجازه می‌دهد تا بفهمند که آیا داده‌هایی که یک مدل مشخص را شکل می‌دهد متناسب، بازنماینده‌ی واقعیت یا دارای مشکلات قانونی یا اخلاقی هستند یا نه.

پیشرفت رویکردهای بی‌طرفی و ضدجانبداری، نشانه‌های محکمی هستند که حوزه هوش مصنوعی قبول کرده که این نگرانی‌ها واقعی هستند. با این حال، تا زمانی که نگرش‌های وسیع‌تری در نظر گرفته نشوند، محدودیت مدل‌های کوتاه‌نگرانه‌ی ریاضیاتی همچنان در جهت سست کردن این رویکردها تداوم خواهند داشت. رهیافت‌های مرتبط با بی‌طرفی و جانبداری، باید هم آسیب‌های تخصصی و هم آسیب‌های بازنمایانه را در نظر داشته باشند، و رهیافت‌هایی که درباره‌ی تعاریف بی‌طرفی و جانبداری بحث می‌کنند باید افراد و جوامعی که بیش‌ترین تاثیرپذیری را دارند، بازشناسی کرده و به آن‌ها فرصت اظهار نظر بدهند. هر فرمولاسیونی از بی‌طرفی که جمعیت‌های

آسیب‌دیده و بافت نهادی که یک سیستم در آن بکار گرفته می‌شود را نادیده بگیرد، بسیار محدود است.

۲-۳- کاربردهای صنعتی: جعبه‌ابزارها

امسال شاهد بودیم که چندین شرکت فن‌آوری تعاریف، معیارها و ابزارهای بی‌طرفی را عملیاتی کردند. در سال گذشته، چهار شرکت بزرگ هوش مصنوعی، ابزارهای کاهش جانبداری را عرضه کردند. آی بی ام جعبه‌ابزار متن‌باز «هوش مصنوعی ۳۶۰» را منتشر کرد که ۹ الگوریتم مختلف و چندین معیار بی‌طرفی دیگر را شامل می‌شد که توسط محققان پروژه‌ی «عدالت، پاسخگویی و شفافیت» در «انجمن یادگیری ماشین» توسعه داده شده بود. این جعبه‌ابزار به منظور ترکیب شدن با شبکه‌ی توسعه نرم‌افزاری از مراحل اولیه تا پیش‌پردازش اطلاعات، تا خود فرآیند آموزش در نظر گرفته شده بود که با استفاده از مدل‌های ریاضیاتی خاصی، استراتژی‌های کاهش جانبداری را پیاده‌سازی می‌کردند. می‌کنند.

«گروه تحقیقاتی مردم و هوش مصنوعی» گوگل (PAIR) ابزار متن‌باز «What-If» را منتشر کرد که به محققان اجازه می‌دهد تا اثرات استراتژی‌ها و معیارهای مختلف کاهش جانبداری را پیش‌بینی کنند؛ همچنین ابزاری به نام «Facets» نیز تصمیم‌گیری درباره‌ی این‌که کدام معیار بی‌طرفی مورد استفاده قرار گیرد را پشتیبانی می‌کند. می‌دهد. فیس فیس بوک ساخت و آزمایش، ابزاری به نام «جریان بی‌طرفی» را که برای استفاده‌ی داخلی مهندسين فیس بوک است را رسماً اعلام کرد؛ این ابزار از تعداد زیادی الگوریتم مشابه

تشکیل شده تا به ما در شناسایی جانبداری در مدل‌های یادگیری ماشین کمک کند. حتی شرکت مشاوره‌ی Accenture، ابزارهایی نرم‌افزاری را توسعه داده تا به مشتری‌ها کمک کند که جانبداری در الگوریتم‌ها را فهمیده و به طور کامل آن‌ها را حذف کنند.

مجموعه‌ی استانداردهای صنعتی نیز در پاسخ به درخواست‌های بخش عمومی و صنعتی و به منظور تضمین پاسخگویی، اقداماتی مرتبط با مسئله‌ی بی‌طرفی را بر عهده گرفته‌اند. موسسه مهندسان برق و الکترونیک (IEEE) به تازگی از گواهینامه‌ی اخلاق برای سیستم‌های خودکار و هوشمند رونمایی کرده است، به امید ایجاد «نشانه‌هایی» که بتواند به بخش گسترده‌تری از عموم مردم گواهی دهد که یک سیستم هوش مصنوعی شفاف، پاسخ‌گو و بی‌طرف است. در حالی که این اقدام به تازگی اتفاق افتاده است و در حالی که IEEE روش‌های زیربنایی این گواهی‌نامه را منتشر نکرده است، به سختی می‌توان دید که با وجود پیچیدگی این مسایل، چگونه در تمام بافتارها و تمام سیستم‌های هوش مصنوعی اتکا بر یک گواهی استاندارد می‌تواند ممکن - یا نهایتاً قابل اعتماد- باشد، به نحوی که این اطمینان را بدهد که این سیستم‌ها به شیوه‌هایی ایمن و اخلاقی استفاده می‌شوند. می‌آید.

هم در بکارگیری پرسرعت روش‌های آکادمیک بی‌طرفی در صنعت و هم در صدور پرشتاب گواهی‌نامه‌ها، اشتیاقی را به «حل» و «حذف» مشکلات مربوط به جانبداری و بی‌طرفی با استفاده از رهیافت‌ها و مهارت‌هایی می‌بینیم که از ضرورت تغییرات مهم ساختاری اجتناب می‌کنند و در بررسی عوامل پیچیده اجتماعی و تاریخی دخیل، ناموفق

هستند. ترکیب «اعتبار آکادمیک» اصلاحات فنی بی‌طرفی با چیزی مثل گواهینامه‌های یادشده، به خطر ابزاری شدن بی‌طرفی منجر می‌شود؛ به طوری که به صنعت اجازه می‌دهد بگوید که این مشکلات را حل کرده و ممکن است توجهات را از بررسی آسیب‌های ادامه‌دار منحرف کند. این اتفاق، همچنین شرکت‌ها را از مسئولیت بررسی اشکال پیچیده و پرهزینه‌ی بررسی دقیق و اصلاح [سیستم‌ها] معاف می‌کند. به جای اتکا به راه‌حل‌های سریع، ابزارها و گواهینامه‌ها، مسئله‌ی جانبداری و بی‌طرفی نیاز به ملاحظات عمیق‌تر و چارچوب‌های پاسخگویی مستحکم‌تری دارد.

۳-۳- چرا اخلاق کافی نیست؟

یک توصیه سطح بالا در گزارش سال ۲۰۱۷ موسسه‌ی AI Now این بود که «کدهای اخلاقی که حوزه‌ی هوش مصنوعی را هدایت می‌کنند، باید با سازوکارهای مستحکم نظارت و پاسخگویی همراه باشند.» در حالی که شاهد هجوم پر شتابی در بکارگیری چنین کدهایی بوده‌ایم، که در بسیاری از موارد به عنوان ابزاری برای پاسخدهی به بحث‌های پیرامون طراحی و پیاده‌سازی سیستم‌های هوش مصنوعی به شمار می‌آیند، [اما] نظارت و پاسخگویی قوی‌ای را نسبت به این تعهدات اخلاقی ندیده‌ایم.

پس از آن که مشخص شد که گوگل در حال همکاری با پنتاگون بر سر پروژه‌ی Maven (توسعه‌ی سیستم‌های هوش مصنوعی برای نظارت پهبادی) است، بحث‌ها درباره‌ی نقش هوش مصنوعی در سامانه‌های جنگ‌افزاری به شدت افزایش یافت. پروژه‌ی Maven، اعتراض

قابل توجهی را در بین کارکنان گوگل ایجاد کرد و این کارکنان از رؤسای این شرکت درخواست کردند که به هنگام پایان قرارداد فعلی، مشارکتشان در این برنامه را پایان بدهند. در پاسخ، سوندار پیچای، مدیر عامل گوگل، مجموعه‌ای از هفت «اصل راهنما» را به طور عمومی منتشر کرد که برای اطمینان از اینکه کار این شرکت در زمینه هوش مصنوعی مسئولیت اجتماعی را بر عهده خواهد داشت، طراحی شده بود. این اصول اخلاقی شامل تعهد به «سودبخشی اجتماعی»، و «اجتناب از ایجاد یا تقویت جانبداری ناعادلانه» است. این اصول همچنین شامل بخشی با عنوان «کاربردهای هوش مصنوعی که ما به دنبالشان نیستیم» بود که این کاربردها شامل «جنگ‌افزارها و فن‌آوری‌های دیگری است که هدف اصلی یا کاربرد آنها، آسیب رساندن مستقیم یا غیر مستقیم به مردم است» - واکنشی مستقیم به تصمیم شرکت که دیگر قرارداد خود با وزارت دفاع را تمدید نکند. اما برای عموم مردم واضح نیست که چه کسی بر اجرای این اصول نظارت می‌کند و همچنین هیچ هیات [نظارت بر] امور اخلاقی [توسط شرکت‌ها] معرفی نشده است.

در انتشار کدها و دستورالعمل‌های اخلاقی، گوگل تنها نبود. سایر شرکت‌ها، از جمله مایکروسافت، فیس بوک و آکسون که سازنده‌ی دوربین‌های پلیس است، هیات‌ها، مشاوران و تیم‌های [نظارت بر امور] اخلاقی را گرد هم آوردند. دادند. بعلاوه، نهادهای عضویتی مهندسی چندی از کدهای اخلاقی خود را بروز رسانی کردند. انجمن IEEE آئین‌نامه اخلاقی خود را بازبینی کرد تا چالش‌های هوش مصنوعی و سیستم‌های خودکار را در آن منعکس سازد؛ محققان

انجمن ماشین‌آلات محاسباتی (ACM) درخواست کردند تا ساختار فرایندهای هم‌تادآوری مقالات، تجدید بشود تا مولفان مقالات مهندسی ملزم شوند که استفاده‌های احتمالی مضرّ پروژه‌های‌شان را در نظر بگیرند. دانشگاه‌هایی مانند دانشگاه هاروارد، دانشگاه نیویورک، استنفورد و MIT دوره‌های جدیدی را در زمینه اصول اخلاقی هوش مصنوعی و توسعه اخلاقی آن ارائه داده‌اند، که هدف از آن‌ها شناسایی مسائل و در نظر داشتن پیامدهای نوآوری فناورانه پیش از پیاده‌سازی در مقیاس بالا است. دانشگاه مونترال جریان گسترده‌ای را به منظور تنظیم بیانیه‌ای درباره‌ی توسعه مسئولانه‌ی هوش مصنوعی به راه انداخت که شامل نشست‌های تخصصی و سخنرانی‌های عمومی - به منظور مشارکت شهروندان - بود.

چنین تحولاتی دلگرم‌کننده هستند و لازم به ذکر است که آن‌هایی که در قلب توسعه هوش مصنوعی فعالیت می‌کنند، اعلام کرده‌اند که اخلاق را جدی می‌گیرند. [آندوین] اصول اولیه‌ی اخلاقی کمک کرده که زبانی مشترک رواج پیدا کند تا با آن زبان درباره‌ی مسائل اجتماعی و سیاسی بحث و گفتگو کنیم. این اصول اولیه برای توسعه دهندگان، کارکنان شرکت‌ها و سایر ذینفعان، مجموعه‌ای از گزاره‌های ارزش^۱ یا اهداف سطح بالایی را فراهم می‌کنند که می‌توانند بعداً مورد قضاوت قرار گیرند. اصول اولیه‌ی اخلاقی، آموزشی نیز هستند و معمولاً سطح آگاهی از خطرات خاص هوش مصنوعی را - چه در یک نهاد مشخص، چه در حوزه‌های عمومی‌تر - بالا می‌برند.

با این حال، توسعه سیستم‌های هوش مصنوعی بی‌طرف و برابرنگر، نیازمند چیزی بیش از زبانی اخلاقی است که شاید با حسن‌نیت

تنظیم شده باشد. ما در رویکرد فعلی به اخلاق، شاهد دو نوع از مشکلات هستیم. اولی مربوط به اجرا و پاسخگویی است. رهیافت‌های اخلاقی صنعت به طور ضمنی خواستار این هستند که وقتی شرکت‌ها می‌گویند که رویه‌های خود را به شیوه‌های اخلاقی پیش خواهند برد، مردم به سادگی بپذیرند که آن‌ها بر سر حرفشان می‌مانند. در حالی که مردم می‌توانند تصمیم یک شرکت را با اصول اخلاقی‌اش مقایسه کنند، یک رهیافت اخلاقی باعث نمی‌شود که نسبت به آن تصمیم دید کافی داشته باشیم، یا باعث نمی‌شود که قدرت بازگرداندن آن تصمیم یا هدایت کردن آن را داشته باشیم. لوسی ساچمن^۱ که متخصصی پیشرو در زمینه‌ی تعامل انسان با کامپیوتر است، در تحلیل خود از اصول هوش مصنوعی گوگل، استدلال می‌کند که بدون «تهدادهای لازم جهت مذاکره، فرجام‌خواهی، و جبران خسارات» اصول اخلاقی مبهمی مانند «شیطان صفت نباشید» یا «کار درست را انجام بدهید»، پوچ هستند. این شکل «به ما اعتماد کنید» از خودحاکمیتی شرکت‌های بزرگ، پتانسیل کنار گذاشتن یا پیشی گرفتن از شکل‌های جامع و الزام‌آور مقررات دولتی را نیز دارد. بن واگنر^۲ از دانشگاه اقتصاد و تجارت وین استدلال می‌کند: «در صورت ناتوانی یا بی‌تمایلی به ارائه راهکارهای مناسب مقرراتی، اخلاق به عنوان گزینه‌ی «آسان» یا «نرمی» تلقی می‌شود که می‌تواند به ساختار کمک کرده و به اصول اولیه‌ی اخلاقی خودنظارت‌گر موجود معنابخشی کند.»^۳ به عبارت دیگر، کدهای اخلاقی ممکن است، انتقادات را با اذعان به اینکه مشکلات همیشه وجود دارند، منحرف کنند؛ بدون این که

1. Lucy Suchman
2. Ben Wagner

اختیار تنظیم مقررات یا تغییر روش توسعه‌ی فناوری یا استفاده از آن را به نهاد دیگری واگذار کنند. این حقیقت که مدیر سابق عملیات‌های فیس بوک ادعا می‌کند که «ما نمی‌توانیم به فیس بوک اعتماد کنیم که خودش بر خودش نظارت کند»، باید در هنگام ارزیابی کدهای اخلاقی در صنعت مد نظر قرار گیرد.

نوع دوم از مشکلات، مربوط به فرضیات عمیق‌تر و جهان‌بینی طراحان کدهای اخلاقی در فناوری است. گرین^۱ و همکاران در واکنش به رشد تعداد دستورالعمل‌های اخلاقی شرکت‌ها، مرور انتقادی و نظام‌مندی از «بیانیه‌های ماموریت و چشم‌انداز^۲ هوش مصنوعی» که توجه زیادی را به خود جلب کرده، به عمل آورده‌اند. یکی از یافته‌های آن‌ها این بود که این بیانیه‌ها می‌خواهند جهان‌بینی موجب‌گرایی فناوریانه^۳ را اتخاذ کنند؛ جهان‌بینی‌ای که در آن عاملیت و تصمیم‌گیری اخلاقی به متخصصان محول شده‌است؛ «دایره‌ای محدود از کسانی که می‌توانند دغدغه‌های اخلاقی مربوط به هوش مصنوعی یا یادگیری ماشین» را از طرف بقیه ما مورد قضاوت قرار دهند. این بیانیه‌ها بدون اذعان به خطرات خاص هوش مصنوعی برای جمعیت‌های محروم، اغلب حاکی از آن هستند که تکنولوژی هم منافع و هم خطرات بزرگی را برای تمام بشریت وعده می‌دهد. به جای پرسیدن سوالات اخلاقی و سیاسی اساسی در مورد اینکه آیا باید سیستم‌های هوش مصنوعی ساخته شوند یا نه، این اسناد به طور ضمنی پیشرفت فن‌آوری را اجتناب‌ناپذیر می‌دانند و خواستار ابداع سیستم‌های بهتر نیز هستند.

مطالعه تجربی درباره‌ی استفاده از این کدها تازه آغاز شده است، اما نتایج

اولیه امیدوارکننده نیستند. یکی از مطالعات اخیر نشان داد که «آموزش دادن به مهندسان به منظور اینکه کدهای اخلاقی ACM را در تصمیم‌گیری‌های‌شان لحاظ کنند، در مقایسه با حالتی که گروه نظارتی بر آن‌ها نظارت می‌کرد، هیچ اثری نداشته است.» با این حال، این پژوهشگران دریافتند که گزارش‌های رسانه‌ای یا گزارش‌های تاریخی درباره‌ی بحث‌های اخلاقی در مهندسی، مانند رسوایی دیزل‌گیت، ممکن است اثرات قابل تامل‌تری داشته باشند. شاید مهم‌ترین مدرک مربوط به محدودیت این کدهای اخلاقی نوظهور، چگونگی عمل‌کرد شرکت‌ها پس از تنظیم این کدها است. در بین کاربردهایی که گوگل، بر اساس بخشی از اصول هوش مصنوعی خود، قول می‌دهد که آن‌ها را پیگیری نکند، «فناوری‌هایی هستند که هدفشان به طور گسترده‌ای مغایر با اصول پذیرفته‌شده حقوق بین‌الملل و حقوق بشر است. باشد.» این ادعا در اوایل سال جاری پس از اینکه روزنامه‌نگاران تحقیقی فاش کردند که گوگل بدون هیچ سر و صدایی، نسخه سانسور شده از موتور جستجوی خود را با نام Dragonfly برای بازار چین توسعه داده است، آزموده شد. سازمان‌ها این پروژه را با عنوان تخطی از قوانین حقوق بشر و همین‌طور تخطی از اصول هوش مصنوعی گوگل، محکوم کردند. کارمندان گوگل نیز علیه این اقدام اعتراض کردند. تا هنگام نوشتن این گزارش، پروژه‌ی مذکور هنوز لغو نشده است و توسعه مداوم آن نیز، با وجود تعهد واضحی که در اصول هوش مصنوعی گوگل آمده، توضیح داده نشده است؛ با این حال، مدیرعامل گوگل از این پروژه به عنوان پروژه‌ای «آزمایشی» دفاع کرده است. است.

بدین ترتیب، نیاز آشکاری به پاسخگویی و نظارت در صنعت وجود دارد، و تاکنون حرکت به سمت اصول اخلاقی این نیاز را برآورده نکرده است. این اتفاق احتمالاً بخاطر انگیزه‌های بازارمحوری است که در کنار پیاده‌سازی‌های صنعت‌محور کار می‌کنند: پس از آن‌که توئیتر و فیس بوک فعالیت خود برای مبارزه با اطلاعات نادرست و افزایش هزینه جهت حفظ امنیت و حریم خصوصی را اعلام کردند، افتی شدید در قیمت سهام آن‌ها رخ داد.

این مثال‌ها توجیهی برای دنبال نکردن دستورالعمل‌های اخلاقی‌تر نیست، بلکه پیشنهاد می‌کند که ما باید نسبت به این‌که شرکت‌ها به طور داوطلبانه برای پیاده‌سازی تصمیمات اخلاقی اقدام کنند و ما به آن‌ها اطمینان کنیم، محتاط باشیم؛ چرا که بسیاری از انگیزه‌های حاکم بر این شرکت‌های بزرگ سهامی عام، اقدام اخلاقی را تخطئه می‌کنند. به منظور اینکه این مکانیزم‌ها با صورت معناداری از پاسخگویی خدمات‌رسانی کنند، لازم است که نظارت خارجی و شفافیت اعمال شود تا این اطمینان حاصل شود که علاوه بر تربیت هنجارها و ارزش‌های اخلاقی در تخصص مهندسی و شرکت‌های فناوری، از خارج نیز سامانه‌ای کنترلی وجود دارد.

بخش چهارم

چه چیزی برای بعد نیاز است؟



چه چیزی برای بعد نیاز است؟

زمانی که ما گزارش AI Now در سال ۲۰۱۶ را منتشر کردیم، فرمول‌های بی‌طرفی، جعبه‌ابزارهای غیرجانبداری و دستورالعمل‌های اخلاقی برای هوش مصنوعی بسیار کم بودند. این واقعیت که امروزه آن‌ها اموری عادی تلقی می‌شوند، نشان می‌دهد که این حوزه تا چه حد پیشرفت کرده‌است. با این حال باید کارهای بیشتری انجام شود. در ادامه هفت استراتژی برای پیشرفت آتی در این موضوعات را طرح می‌کنیم.

۱-۴- از بی‌طرفی به سوی عدالت

در هر بحثی در مورد جانبداری و بی‌طرفی باید به مسائل قدرت و سلسله‌مراتب آن توجه داشته باشد؛ به این که چه کسی در جایگاه تولید و سود بردن از این سیستم‌ها است؛ به این که ارزش‌های چه کسانی در این سیستم‌ها جاسازی می‌شود؛ چه کسی «توابع هدف» این سیستم‌ها را تنظیم می‌کند؛ و این سیستم‌ها در چه بافتاری قرار است کار کنند. با یادآوری مجدد فراخوان انجمن ماشین‌آلات محاسباتی (ACM) مبنی بر توجه به «تبعات منفی» به عنوان الزامی

برای هم‌تادآوری مقالات مهندسی، باید به روش‌هایی که هوش مصنوعی می‌تواند به عنوان ابزاری برای بهره‌برداری و نظارت استفاده شود، توجه بیشتری کرد. همچنین باید مراقب باشیم که مسائل سیاسی را در چهارچوب دغدغه‌های مهندسی بازسازی نکنیم. هنگامی که این مسائل با «روش‌های تعمیر» مهندسی چهارچوب‌بندی می‌شوند، راه‌حل‌های از بین بردن جانبداری به ندرت اجازه می‌دهند که سوالاتی درباره‌ی مناسب بودن یا اثربخشی یک سیستم هوش مصنوعی پرسیده شود؛ یا به ندرت اجازه می‌دهند که درباره‌ی بافت نهادی‌ای که در نهایت، سیستم «تعمیرشده» در آن اعمال خواهد شد، پرس‌وجو بشود. به عنوان مثال، الگوریتم پیش‌بینی‌کننده‌ای که جانبداری آن از بین رفته است و به دقت پیش‌بینی می‌کند که جرم در کجا رخ خواهد داد، توسط ضابطین قانون برای آزار و سرکوب جوامع رنگین‌پوست مورد استفاده قرار می‌گیرد؛ در نتیجه این سیستم هنوز اساساً یک سیستم جانبدارانه است. در این راستا، تعریف ما از «بی‌طرفی» باید زمینه‌های ساختاری، تاریخی و سیاسی‌ای را که در آن سیستم الگوریتمی استفاده می‌شود، فراگیرد. علاوه بر این، بی‌طرفی واژه‌ای است که می‌تواند به راحتی برای فریب دادن افراد مورد استفاده قرار بگیرد؛ بنابراین پرسش‌های مهمی از قبیل «بی‌طرف برای چه کسی؟ و در چه زمینه‌ای؟» همیشه باید پرسیده شود. به عنوان مثال، ساختن یک سیستم تشخیص چهره که عملکرد آن نسبت به سیاه‌پوستان و سفیدپوستان یکسان است، ممکن است نوعی پیشرفت فنی در زمینه‌ی برابری به شمار بیاید، اما اگر این فناوری به شکل نامتناسبی روی جوامع رنگین‌پوست و کم‌درآمد

استفاده شود، آیا واقعا «بی طرفانه» است؟ به همین دلیل است که اگر تعاریف بی طرفی صرفا در حوزه مهندسی باشند، با محدودیتی دشوار مواجه هستند؛ به طور خلاصه، «برابری عدالت نیست.»»

۲-۴- تفکر زیرساختی

به منظور داشتن فهمی بهتر و دنبال کردن پیچیدگی سیستم‌های هوش مصنوعی، باید از فن‌آوری و گزاره‌گویی‌های مربوط به آن فراتر برویم تا بافتار وسیعتری را توضیح بدهیم که هم هوش مصنوعی آن را شکل می‌دهد و هم توسط نیروهای اجتماعی و مادی شکل می‌گیرد. همانطور که ادواردز^۱ و همکاران استدلال می‌کنند: «هنگامی که با زیرساخت‌ها سر و کار داریم، باید به همه‌ی انواع شکل‌های سازمانی، حرفه‌ها و نهادهایی که با فناوری جدید همراه می‌شوند یا امکان ساخته شدن آن را فراهم می‌کنند یا مسیر توسعه‌ی آن را تغییر می‌دهند، توجه کنیم» انجام این کار نیازمند رهیافت‌های روش‌شناختی تجربی و نظریه‌سازی شده‌ی تک‌سیستم‌هایی که برای ابعاد محلی یا محدود جداسازی شده‌ی تک‌سیستم‌هایی که برای ابعاد محلی یا جهانی استفاده می‌شود، گسترش بیابد. درک پیچیدگی‌های هوش مصنوعی نیازمند بررسی راه‌هایی است که در آن فن‌آوری‌ها با روابط اجتماعی، وابستگی‌های مادی و اهداف سیاسی درهم‌تنیدگی دارد. در پروژه‌ی «کالبد شناسی یک سیستم هوش مصنوعی»، که حاصل آن، مقاله‌ای در سال ۲۰۱۸ و نقشه‌ای بزرگ مقیاس بود، بنیانگذار موسسه‌ی AI Now، کیت کرافورد و پروفیسور ولادان جولر^۲، دستگاه Amazon Echo را در نظر گرفته و همه‌ی شکل‌های منابع محیطی و

نیروی کار مورد نیاز جهت توسعه، تولید، نگهداری و در نهایت از بین بردن این شی ساده و شیک را تحلیل کردند. زمانی که از Alexa بخواهید که آهنگ مورد علاقه‌تان را پخش کند، شما زنجیره‌ی وسیع به‌هم‌پیوسته‌ای از فرآیندهای استخراجی^۱ را به راه انداخته‌اید. این زنجیره شامل این موارد می‌شود: استخراج لیتیوم از معادن بولیوی، ایجاد مجموعه داده‌های آموزش‌دهنده‌ی ماشین در مقیاس بزرگ توسط کلیک‌کارها در جنوب شرقی آسیا، کشتی‌های کانتینربر و لجستیک بین‌المللی، استخراج داده و تحلیل آن از طریق سرویس صوتی الکسا (AVS) در سراسر مراکز پراکنده‌ی داده. این فرآیند در قبرستان تمام دستگاه‌های هوش مصنوعی که استفاده‌های مصرفی دارند، به پایان می‌رسد: در توده‌های زباله در غنا، پاکستان و چین. پروژه «کالبدشناسی یک سیستم هوش مصنوعی» به رهیافت‌هایی که می‌توان در مواجهه با تبعات جهانی هوش مصنوعی اتخاذ کرد اشاره می‌کند و طبیعت چند لایه‌ی برکشیدن ارزش‌ها [از یک فناوری] و بهره‌برداری از جهان در حال توسعه به نفع جهان توسعه‌یافته را نشان می‌دهد. این کار به روشن کردن نقاط تاریک‌تری که به ندرت در تحلیل سیستم‌های هوش مصنوعی لحاظ می‌شوند، کمک می‌کند. به طور خاص، تحلیل زیرساختی هوش مصنوعی نشان می‌دهد که جعبه سیاه‌هایی درون جعبه‌های سیاه وجود دارد: نه تنها در سطح الگوریتمی، بلکه در سطح قوانین رازداری تجاری، شیوه‌های کار و زنجیره‌های تامین غیرقابل ردیابی مواد معدنی نایابی که برای ساخت دستگاه‌های هوش مصنوعی استفاده می‌شوند. این جعبه‌های سیاه، نه تنها اثرات مادی سیستم‌های هوش مصنوعی را مخفی می‌کنند،

1. extractive

بلکه جلوی دیده شدن کار عظیمی که مربوط به حفظ و نگهداری این سیستم‌ها است را نیز می‌گیرد. می‌شود. همانطور که نیک سیور^۱ می‌گوید: «اگر نمی‌توانید انسانی را در این حلقه ببینید، فقط باید به سراغ یک حلقه بزرگ‌تر بروید.»

تنها با دنبال کردن این لایه‌های اجتماعی-تکنیکی می‌توانیم آنچه را که «زنجیره تامین کامل» هوش مصنوعی می‌نامیم درک کنیم؛ این زنجیره‌ی تامین شامل مولفه‌های انسانی و غیر انسانی است که سیستم‌های هوش مصنوعی را در مقیاس جهانی می‌سازند. زیرساخت‌های اجتماعی-تکنیکی بسیاری برای اطلاعات وجود دارد که برای ترتیب دادن کارکردهای هوش مصنوعی لازم هستند. این زیرساخت‌ها شامل این موارد هستند: داده‌های آموزش دهنده، داده‌های تست، APIها، مراکز داده، شبکه‌های فیبر نوری، کابل‌های زیردریایی، مصرف انرژی، نیروی کار لازم جهت تعدیل محتوا و تکیه‌ی مداوم بر کلیک‌کارها به منظور توسعه و نگهداری سیستم‌های هوش مصنوعی. ما بدون فهمیدن این زنجیره‌ی تامین کامل، نه می‌توانیم اثرات محیطی و تبعات شغلی این ابزارها را که راحتی روزمره را برای ما به ثمر رسانده‌اند، ببینیم؛ و نه می‌توانیم به طور هدفمندی از بی‌طرفی، پاسخگویی و شفافیت در سیستم‌های هوش مصنوعی حمایت کنیم.

۳-۴- لحاظ کردن کار پنهان در سیستم‌های هوش مصنوعی

یکی دیگر از حوزه‌های نوظهور پژوهش که انتظار کارساز بودن بیشتری از آن می‌رود، روی نیروهای کاری تمرکز می‌کند که به ساخت، نگهداری و آزمایش سیستم‌های هوش مصنوعی کمک

می‌کنند، اما در عین حال کمتر به آن‌ها حقوق داده می‌شود و به عنوان کسانی که کاری را از پیش می‌برند شناخته نمی‌شوند. این کار انسانی پنهان، شکل‌های بسیاری را به خود می‌گیرد: از کار زنجیره تامین و کلیک‌کاری انبوه‌سپاری‌شده^۱ گرفته تا مشاغل سنتی خدماتی صنعت. کار پنهان در تمام مراحل شبکه‌ی زنجیره‌ای هوش مصنوعی وجود دارد؛ از تولید و انتقال مواد معدنی خامی که برای ایجاد زیرساخت اصلی سیستم‌های هوش مصنوعی مورد نیاز است، گرفته تا فراهم کردن کار انسانی ناپیدایی که معمولاً مقوم ادعای «جادویی بودن» هوش مصنوعی، به هنگام به کارگیری این سیستم‌ها در محصولات و خدمات هستند. لیلی ایرانی که متخصص علوم ارتباطات است، به این کار پنهان با عنوان «خودکارسازی با سوخت انسانی» اشاره می‌کند. می‌نامد. پژوهش‌های او متوجه تجربیات کلیک‌کارها یا «ریزکارگرهای^۲» است که وظایف دیجیتالی تکرارشونده‌ای را انجام می‌دهند؛ مانند برچسب زدن به داده‌های آموزش‌دهنده و بازبینی محتوای علامت‌گذاری شده.

در حالی که وجود این نیروی کار برای این که سیستم‌های هوش مصنوعی «کار» کنند ضروری است، اما پرداخت‌ها برای این کارها معمولاً ناکافی است. مطالعه سال ۲۰۱۸ سازمان بین‌المللی کار (ILO) ۳۵۰۰ ریزکارگر از ۷۵ کشور را مورد بررسی قرار داد که به آن‌ها به صورت روزمره بر روی پلتفرم‌های همگانی ریزکاری^۳ مانند Clickworker و Mechanical Turk، Crowdfunder، Microworker کار پیشنهاد می‌شود. این گزارش نشان داد که تعداد قابل توجهی از این افراد - با وجود اینکه ۵۷ درصد پاسخ‌دهندگان مدارج تحصیلی

1. crowdsourced
2. microworkers
3. microtask

پیشرفته در رشته‌های علم و فن‌آوری داشته‌اند- کمتر از حداقل دست‌مزد محلی خود دریافت می‌کنند. می‌کردند. به طور مشابه، آن‌هایی که کار تعدیل محتوا را انجام می‌دهند و مطالب مشکل‌دار بارگذاری شده بر روی شبکه‌های اجتماعی را کنترل می‌کنند، به رغم کار حیاتی و هیجانی آن‌ها، دست‌مزد ناکافی دریافت می‌کنند. این مساله در جامعه‌ی پژوهش‌های مهندسی هوش مصنوعی فراموش نشده است؛ این جوامع پرداختن به نقش حیاتی و کم‌تر توجه‌شده‌ی این نوع کار را آغاز کرده‌اند و به مسئولیت خود برای دخالت در این مسئله توجه دارند. سیلبرمن^۱ و دیگران در مورد این موضوع بحث می‌کنند که چگونه محققانی که در حال انجام مطالعات هوش مصنوعی هستند، به طور فزاینده‌ای به نیروی کار انبوه‌سپاری شده وابسته هستند. آن‌ها نوشته‌اند که بین سال‌های ۲۰۰۸ تا ۲۰۱۶، تکرار واژه «انبوه‌سپاری کردن» در مقالات علمی، از ۱۰۰۰ به بیش از ۲۰۰۰۰ رسید. با در نظر گرفتن ریزکارگرهای آنلاینی که قوانین فعلی کار برای آن‌ها مقرراتی تنظیم نکرده است، از پژوهشگران خواسته می‌شود در آنچه که «رفتار اخلاقی» در جامعه پژوهش هوش مصنوعی محسوب می‌شود، تجدید نظر کنند. سیلبرمن و همکاران به نفع این موارد استدلال می‌کنند: این که با کلیک کارها به عنوان همکار رفتار شود، این که حداقل حقوق تصویب‌شده در موقعیت جغرافیایی فرد پرداخت شود و این که به هیئت بازبینی نیاز است.

اقدام به بررسی کار پنهان انسانی، از سنخ تحقیقات فمینیستی است. به عنوان مثال، مفهوم «کار نامرئی»^۲ نشأت گرفته از پژوهش‌هایی

است که مرتبط با کارهای مراقبتی زنان است و نیز نشأت گرفته از تحقیقاتی درباره‌ی استانداردهایی سازمانی است که بر «کار عاطفی» [زنان] تکیه دارد؛ به ویژه در حوزه‌هایی «زنانه شده» مانند پرستاری و خدمات پرواز. محققان دریافتند که فعالیت‌های رایج انجام‌شده توسط کارکنان زن، مانند آرام کردن بیماران مضطرب یا مدیریت مشتریان سرکش، به رغم ضروری بودن آن‌ها، به عنوان کار به رسمیت شناخته نمی‌شود و این کارها [توسط کارفرما] جبران نمی‌شود. میراث فمینیستی برای فهم خصوصیات این نوع کار پنهان مفید است؛ کاری که در عین ضروری بودن، اغلب خارج از روایت هوش مصنوعی قرار می‌گیرد، به ندرت به حساب می‌آید یا به ندرت جبران می‌شود.

آسترا تیلور^۱ در مقاله‌اش با عنوان «معمای خودکارسازی» عبارت «fauxtomation» را پیشنهاد می‌کند تا توجه‌ها را به شکاف میان لفاظی بازاری هوش مصنوعی بعنوان محصول یا خدمتی یکپارچه و واقعیت درهم و برهم اتوماسیون که بر این نیروی کار ناشناخته تکیه دارد، جلب کند. تیلور هشدار می‌دهد: «اتوماسیون، همانطور که بُعدی فناورانه دارد، کارکردی ایدئولوژیکی نیز دارد.» در این راستا، او روایت‌های زیادی که مربوط به آینده کار هستند را نقد می‌کند؛ روایت‌هایی که افق نزدیکی را فرض می‌کند که در آن ماشین‌ها جایگزین کارگران می‌شوند. بنظر او چنین ادعاهایی کارکرد تضعیف‌شاغلین را دارد: کارگران چه اهرم فشاری برای افزایش دست‌مزد و مزایای بیشتر در مواجهه با اتوماسیون پیش رو دارند؟ در سال ۲۰۱۶ در مک‌دونالد، هنگامی که انگیزه‌هایی برای جایگزینی کیوسک‌های خودکار

1. Astra Taylor

به جای صندوق‌داران به وجود آمد، باعث ایجاد جنبش فراگیر «مبارزه برای ۱۵ دلار» شد. اد رنسی^۱، مدیر عامل سابق مک‌دونالد، استدلال کرد که کارگرانی که برای حقوق بهتر می‌جنگیدند، در نهایت به وضع بدی دچار می‌شدند، زیرا تقاضای آن‌ها برای دست‌مزد، شرکت را وادار می‌کرد که [کار] آن‌ها را خودکار کرده و آن‌ها از حذف کنند. پس از دو سال بررسی ادعاهای رنسی می‌بینیم که این ادعاها تماماً درست نیست. چه اتوماسیون باشد چه نباشد، شاغلین و کارگران هنوز مورد نیاز هستند: مثلاً پس از این‌که مک‌دونالد به فروشگاه اصلی شیکاگو چند کیوسک اضافه کرد، تعداد کارمندان افزایش پیدا کرد.

یکپارچه‌سازی اتوماسیون و هوش مصنوعی در محیط کار نه تنها قصد خودکارسازی وظایف شاغلین را دارد، بلکه به مدیریت، نظارت و ارزیابی خود کارگران نیز توجه دارد. پژوهش فرهنگ‌شناسی رانندگان شرکت اوبر که در سال ۲۰۱۸ توسط الکس روزنبلات^۲ انجام شده، جزئیات این را توضیح می‌دهد که چگونه تحت تاثیر خواست یک سیستم مجهز به هوش مصنوعی متمرکز، بی‌ثباتی و عدم قطعیت برای زندگی یک نفر به وجود می‌آید. منطق الگوریتمی که برنامه‌ی هم‌سفری آنلاین را مدیریت می‌کند، می‌تواند به راحتی رانندگان را از کار بازدارد و منجر به دستمزدهایی غیرقابل اطمینان و هزینه‌هایی غیر منتظره بشود و افراد را به ساعات کاری طولانی‌تر سوق دهد و موجب ناامن شدن شرایط رانندگی بشود. این پلتفرم‌ها کارکنان را از هم جدا می‌کنند و فعالیت‌های هماهنگ و شکل‌سازی کارگری را دشوار می‌سازند. این پلتفرم‌ها همچنین در ایجاد عدم تقارن

اطلاعاتی قابل توجه میان کارگران و شرکتهایی که از نظر داده غنی هستند و قصد جذب کردن ارزش [افزوده] از کارگران را دارند، کارکرد دارد. با این وجود، در سال ۲۰۱۸ شاهد افزایش مخالفت چنین کارگرانی بوده‌ایم. برخی از نمونه‌های برجسته این اعتراضات کارگری، اعتراضات پیکسوارانی است که در کنار کارمندان صنایع فست‌فود بریتانیا اعتصاب کردند و همچنین اعتراضات رانندگان تاکسی است که خواستار امنیت شغلی بودند.

شاغلین پیمانی سیلیکون ولی که در شرکت‌های بزرگ فناوری در حوزه‌هایی مانند امنیت، غذا و مشاغل خدماتی فعالیت می‌کنند نیز اقدام به تشکیل‌سازی کرده‌اند و در پی دستمزدی کافی برای امرار معاش و حمایت‌های دیگر هستند. این افراد در زمره‌ی هزاران شاغلی هستند که در کنار هم‌تایان تمام‌وقت خود در حوزه‌ی فناوری کار می‌کنند، اما در طبقه‌ی کارکنان پیمانی مستقل قرار داده می‌شوند. بخاطر چنین نام‌گذاری‌ای، اغلب دستمزد کمی به آن‌ها پرداخت می‌شود و مزایا و حمایت‌های کمی برای آن‌ها فراهم می‌شود. اگرچه آن‌ها نسبت بزرگی از نیروهای کار حوزه‌ی فناوری را تشکیل می‌دهند و کارهای مهمی را انجام می‌دهند، اما به ندرت به عنوان کارکنان رسمی به حساب می‌آیند. مثلاً امسال تعداد کارکنان پیمانی گوگل، برای اولین بار در تاریخچه‌ی شرکت، از تعداد کارمندان اصلی آن پیشی گرفت. این موج فزاینده‌ی مخالفت‌ها، تنشی اجتماعی را نمایان می‌کند که در عمل پنهان‌سازی و کم‌توجهی به انواع مهم کار وجود دارد.

در خبرهای اخیر که شرایط کارگران انبارهای آمازون و رانندگان تحویل

بسته‌ی پرایم^۱ را شرح می‌دهند، هزینه‌های جسمانی، عاطفی و مالی برخورد با شاغلین به مثابه‌ی «قطعه‌ای کد» و کاستن از ارزش کار و بهزیستی آنان برجسته شده است. کارگران انبارهای آمازون در اروپا به تازگی دست به اعتصاب زده و به شرایط سخت [کار] اعتراض کردند. بنابر گفته‌های یکی از اعتصاب‌کنندگان «شما کارتان در شرکت را با سلامتی آغاز می‌کنید و به هنگام رها کردن آن، انسانی شکسته می‌شوید» و نیز بسیاری از کارگران بخاطر شرایط محیط کار به عمل جراحی نیاز دارند.

شناسایی همه‌ی نیروهای کاری که برای «کار کردن هوش مصنوعی» لازم هستند، می‌تواند به ما کمک کند تا تبعات توسعه و استفاده از آن را درک کنیم. هم‌چنین تحقیق در این حوزه‌ها به ما کمک می‌کند تا تمرکزی که بر قریحه‌ی مهندسی در روایت‌های مربوط به آفرینش هوش مصنوعی وجود دارد را مورد بازنگری قرار بدهیم و این را بفهمیم که مهارت‌های مهندسی تنها بخشی از یک تلاش بسیار بزرگ‌تر را تشکیل می‌دهند. این تحقیقات ما را قادر می‌سازد تا سیاست‌های متعدد شغلی را مورد سوال قرار بدهیم؛ سیاست‌هایی مانند فشار آوردن به کارگران تا مهارت‌های برنامه‌نویسی یا علم داده را، به عنوان روشی برای مطمئن شدن از این که آن‌ها به رسمیت شناخته خواهند شد و حقوق دریافت خواهند کرد، کسب کنند. این تحقیقات هم‌چنین به ما کمک می‌کنند تا تشخیص دهیم چه کسی احتمالاً سود می‌کند، و چه کسی در کنار تولید و پیاده‌سازی شبکه‌های زنجیره‌ای هوش مصنوعی، به احتمال زیاد آسیب خواهد دید.

۴-۴- میان‌رشته‌ای بودن بیشتر

پژوهشگران و توسعه دهندگان هوش مصنوعی در حال ساخت فن‌آوری‌هایی هستند که در زمینه‌های گسترده‌ای مانند قانون، جامعه‌شناسی و پزشکی، تبعات مهمی برای جمعیت‌های گوناگون دارند. با این حال، بخش زیادی از این توسعه، از تجربه و تخصص این گروه‌ها بسیار جدا شده است. این امر باعث شده که خواستی برای گسترش ترکیب دانشی افرادی که در طراحی، توسعه و نقد هوش مصنوعی درگیر هستند، بوجود بیاید؛ این گسترش باید از تخصص صرفاً مهندسی بگذرد کند. از آن زمان به بعد، در این مسیر شاهد تحولاتی بوده‌ایم. به تازگی، دانشگاه MIT، طرح‌هایی را درباره‌ی تاسیس یک کالج جدید علوم محاسباتی اعلام کرده که هدف از آن «پیشبرد فعالیت‌های پیشگامانه در رابطه با استفاده‌ی اخلاقی از هوش مصنوعی و تاثیر اجتماعی آن» از طریق تقویت آموزش یکپارچه‌ی میان‌رشته‌ای و «آموزش دو زبان آینده» است.

چنین ابتکاراتی حیاتی هستند: از آنجا که هوش مصنوعی عمیقاً در حال ریشه دواندن در حوزه‌هایی مانند بهداشت، عدالت کیفری، استخدام، مسکن و نظام‌های آموزشی است، تخصص در این حوزه‌ها -در صورتی که کار هوش مصنوعی را پیش‌بینی بدانیم- ضروری است. در یکپارچه‌سازی این دیدگاه‌ها، این نکته مهم است که آن‌ها صرفاً «زبان‌هایی» نیستند که باید به دانشمندان و مهندسين کامپیوتری که به دنبال گسترش کار خود در زمینه‌های جدید هستند، آموخته شود. در عوض، علوم اجتماعی و علوم انسانی باید در دانش بنیادین حوزه هوش مصنوعی و در مسیر آینده به عنوان کمک‌کننده حضور

داشته باشند و ما را قادر بسازند تا روش‌های جدید تحلیل و مداخله روش‌شناختی را تقویت کنیم.

۴-۵- نژاد، جنسیت و قدرت در هوش مصنوعی

در سال جاری، اقدامات سیاسی بسیاری حول مسائل تبعیض، آزار و نابرابری در حوزه‌ی صنایع فن‌آوری، به ویژه در حوزه هوش مصنوعی به وجود آمد. شد. این نگرانی فزاینده، تعدادی از مسائل مرتبط را به هم متصل می‌کند؛ از جانبداری در سیستم‌های هوش مصنوعی گرفته تا عدم توفیق در لحاظ کردن تنوع [فرهنگی] و تلاش برای شمول [گروه‌های به حاشیه رانده شده] در صنعت و دانشگاه و همچنین فعالیت‌های مردم عادی برای مقابله با آزار جنسی و سو استفاده از قدرت در محل‌های کاری و کلاس‌های درسی.

پس از هم‌صدا شدن با جنبش گسترده‌ی #MeToo، شاهد مسائلی مربوط به تنوع و شمول در حوزه‌ی هوش مصنوعی بودیم که در دستور کار عمومی را قرار گرفت:

- پس از کنفرانس سیستم‌های پردازش اطلاعات عصبی در سال ۲۰۱۷، اعضای انجمن‌های هوش مصنوعی و یادگیری ماشینی درباره مشکلات آزار و تبعیض در محیط‌های کنفرانس شروع به ابراز نگرانی کردند که منجر به برجسته شدن جنبشی با هشتگ #ProtestNIPS شد؛ جنبشی که قصد آن، برجسته کردن نمونه‌هایی از ستمی بودن فضای جامعه و لزوم پرداختن به آن بود. در کنار چیزهای دیگر، این جنبش باعث ایجاد تغییر در سرنام^۱ این کنفرانس شد، که بخاطر مفاهیم جنسیتی و تاریخی آن، به صورت

طولانی مدت موضوعی حساسیت‌زا به حساب می‌آید. این کنفرانس که قبلاً «NIPS» نامیده می‌شد، اکنون با نام NeurIPS برگزار می‌شود.

- ما همچنین شاهد تاکید مجدد بر دستورالعمل‌های اولیه‌ای درباره‌ی مسئله‌ی شمول در هوش مصنوعی بودیم که برای ساخت پلتفرم‌ها اختصاص یافته بود؛ مانند دستورالعمل‌های سیاه‌پوستان در هوش مصنوعی^۱، زنان در یادگیری ماشینی^۲، لاتین‌ها در هوش مصنوعی^۳، و هم‌جنس‌گرایی در هوش مصنوعی^۴.

- در سرتاسر صنعت هوش مصنوعی، شاهد جنبش‌های کارگری در حال رشدی بودیم که با این مسائل ارتباط داشت. به طور خاص، «اعتصاب سراسری گوگل» دستورالعملی را در پیش گرفت که طی آن تصدیق شد که نژاد، طبقه‌ی اجتماعی و جنسیت با شکل‌هایی از تبعیض جنسیتی درهم‌تنیده است. این اعتصابات آشکارا هسته‌ی نیازهای کارکنان موقتی و خرده‌فروشان را نشانه می‌رفت؛ افرادی که فاقد امنیت شغلی و مزایای کارکنان برخوردارتر هستند. بودند. این تلاش‌ها منجر به تغییرات ساختاری- به ویژه، پایان حکمیت اجباری در چند شرکت بزرگ هوش مصنوعی درباره‌ی ادعاهایی که مربوط به آزار و اذیت جنسی بود- شده است.

- در عرصه‌های دیگر، هیات مدیره‌ی برخی شرکت‌ها، پیشنهاد سهام‌داران‌شان را که هدف آن‌ها ایجاد فضای کاری تبعیض‌آمیز بوده است، نادیده گرفته‌اند یا در غیر این صورت از پرداختن به پیشنهادها خودداری کرده‌اند. در سال ۲۰۱۶، اپل پیشنهادی را رد کرد که [طبق آن] گوناگونی در [افراد] هیات‌مدیره و مدیریت ارشد الزامی

1. Black in AI
2. Women in Machine Learning
3. Latinx in AI

4. Queer in AI

می‌شد. در میان این اقدامات، طرفداران گوناگونی^۱ در حوزه‌ی هوش مصنوعی، به دنبال رسیدگی به مسئله‌ی جنسیت، آزار نژادی و اجحاف در جامعه فناوری، و دیگر انواع بی‌عدالتی و سو استفاده از قدرت هستند. اما هنوز نبردی دشوار در پیش است: در حالی که توجه زیادی به مسائل جانبداری در سیستم‌های هوش مصنوعی وجود دارد، اما باید شاهد تحقیقات بسیاری در زمینه‌هایی که متمرکز بر عدالت و گوناگونی در حوزه‌ی هوش مصنوعی هستند، باشیم. برآورد اخیرى که توسط مجله‌ی Wired و Element AI انجام شده، نشان داد که تنها ۱۲ درصد از محققانی که به سه کنفرانس برتر یادگیری ماشین در سال ۲۰۱۷ شرکت کرده‌اند، زن هستند. این شکاف جنسیتی در شرکت‌های بزرگ فن‌آوری مانند فیس بوک و گوگل نیز وجود دارد؛ وب سایت‌های این شرکت‌ها نشان می‌دهند که [به ترتیب] تنها ۱۵ درصد و ۱۰ درصد کارکنان پژوهشی هوش مصنوعی آن‌ها زن هستند. و هیچ اطلاعات موثقی درباره‌ی تنوع نژادی در این حوزه، یا نرخ تداوم شغلی رنگین‌پوستان وجود ندارد. به طور کلی، شواهد محدود نشان می‌دهند که هوش مصنوعی به عنوان یک حوزه، حتی از علوم کامپیوتر به مثابه‌ی یک کل که در نقطه‌ی حوضی تاریخی خود قرار دارد، کم‌تر دارای گوناگونی است. زنان تنها ۱۸٪ از افراد فعال در رشته‌های علوم کامپیوتر در آمریکا را تشکیل می‌دهند؛ افولی که از ۳۷٪ در سال ۱۹۸۴ اتفاق افتاد.

این اتفاق وقتی دراماتیک‌تر می‌شود که در مقایسه با حوزه‌ی علم، فناوری، مهندسی و ریاضیات قرار بگیرد که تنوع جنسیت در آن‌ها بهبود قابل توجهی را نشان داده است. با این حال، این مشکلات جدید

نیستند: تحقیق Wired و Element AI با مطالعه‌ای که درباره‌ی حوزه هوش مصنوعی توسط متخصصین IEEE در سال ۱۹۹۲ منتشر شد، تفاوت چندانی ندارد؛ مطالعه‌ی IEEE نشان می‌دهد که تنها ۱۳ درصد از نویسندگان مقالات منتشر شده در ژورنال در طی چهار سال، زن بوده‌اند. هم‌چنین در دهه ۱۹۸۰، در گزارشی با عنوان «موانع برابری در دانشگاه: زنان در دانشکده علوم کامپیوتر MIT» دانشجویان فارغ‌التحصیل زن در آزمایشگاه‌های علوم کامپیوتر و هوش مصنوعی MIT تجربیات خود را از محیط‌های سمی کارشان به طور کامل ثبت کردند. زمان آن رسیده‌است که ارتباط بین تبعیض و آزار در جامعه هوش مصنوعی و تعصب در محصولات فنی که توسط این جامعه تولید شده است را مورد توجه قرار دهیم. محققان در مطالعات علم و فن‌آوری مدت‌ها است که بررسی کرده‌اند که ارزش‌ها و باور کسانی که فناوری‌ها را خلق می‌کنند، این فناوری‌ها را شکل می‌دهند. بلندنگری به منظور شناسایی این ارتباط ما را مطمئن می‌کند که بهتر می‌توان به مسائلی پرداخت که با رشد سریع هوش مصنوعی در حوزه‌های اجتماعی به وجود آمده‌اند. همان‌طور که یک پژوهشگر هوش مصنوعی می‌گوید: «تعصب تنها در مجموعه داده‌های ما نیست، بلکه در کنفرانس‌ها و اجتماعات ما نیز وجود دارد.»

یک مثال جدید وجود این ارتباطات را ترسیم می‌کند و نشان می‌دهد که چگونه اعمال تبعیض‌آمیز موجود در فرهنگی که یک سیستم هوش مصنوعی را تولید می‌کند می‌تواند در خود سیستم منعکس و تقویت بشود. آمازون اخیراً یک سیستم آزمایشی هوش مصنوعی را برای کمک به رتبه‌بندی داوطلبان شغلی توسعه داده است.

آمازون به این امید که افراد واجد شرایط را به شکل موثری شناسایی کند، این سیستم را با داده‌هایی آموزش داد که سوابق اولویت‌های استخدامی شرکت را نمایش می‌دادند. اما این سیستم آن‌طور که انتظار می‌رفت کار نکرد: براساس سوابق استخدام شرکت، این سیستم جانبداری آشکاری را ضد داوطلبان زن نشان داد، و رزومه‌های داوطلبانی که در دو کالج دخترانه درس خوانده بودند را در درجه‌ی پایین‌تری قرار داد، و حتی رزومه‌هایی که حاوی کلمه «زن» بود را دچار عواقبی کرد. بعد از کشف این جانبداری، شرکت تلاش کرد تا این سیستم را اصلاح کند، و تلاش کرد که به منظور عادلانه‌تر رفتار کردن در این شرایط این الگوریتم را تعدیل کند. این کار عملی نشد و این پروژه در نهایت کنار گذاشته شد. تبعیض جنسیتی به شکل عمیقی در این سیستم ریشه دوانده بود که نمی‌شد با استفاده از روش‌های غیر جانبدارانه کردن که معمولاً در حوزه‌ی هوش مصنوعی به کار گرفته می‌شود، ریشه‌کن شود.

محققانی مانند سافیا نابل^۱ و مار هیکس^۲، متوجه شده‌اند که خط واسطی میان الگوهای سابقه‌دار تبعیض و آزار در هوش مصنوعی و روش‌هایی که فن‌آوری هوش مصنوعی می‌تواند در به حاشیه راندن افراد و نابرابری اجتماعی دخالت کند و آن‌را تقویت کند. وجود دارد. الگوهای تبعیض فرهنگی اغلب در سیستم‌های هوش مصنوعی به شیوه‌های پیچیده و هدفمندی جاسازی می‌شوند، و ما باید بفهمیم که چگونه این تأثیرات توسط جوامع مختلف احساس می‌شوند. این فضا، فضایی است که مدت‌ها نادیده گرفته شده‌است و تحقیقات در آن به شدت مورد نیاز است. موسسه‌ی AI Now در حال

انتشار گزارشی اختصاصی در مورد این مسائل است و ما یک پروژه تحقیقاتی چند ساله داریم که به بررسی این چالش‌ها اختصاص دارد.

۶-۴- اقامه‌ی دعوی راهبردی و مداخلات سیاستی

امسال شاهد افزایش چالش‌های دادگاهی مرتبط با استفاده از سیستم‌های خودکار بودیم؛ به خصوص زمانی که ادارات دولتی از آن‌ها در تصمیماتی که بر حقوق فردی تأثیر می‌گذارند، استفاده کردند. در گزارش اخیری که موسسه‌ی AI Now با نام «شکایت از الگوریتم‌ها» منتشر کرده است، پنج مطالعه موردی اخیر از دادرسی‌های مربوط به استفاده از سیستم‌های خودکار را مستندسازی کرده‌ایم: پرونده‌ی مزایای بیمه‌ی بهداشت مستمندان؛ پرونده‌ی مزایای معلولیت؛ پرونده‌ی ارزشیابی اشتغال معلمان دولتی؛ پرونده‌ی ارزیابی خطرات جرایم نوجوانان و نهایتاً پرونده‌ی آنالیز DNA جنایتکاران. یافته‌های این مطالعات، چند مسیر جدید را پیش روی ما به وجود آورد. اول آن که این موارد، شواهدی عینی را فراهم کردند مبنی بر این که دولت‌ها به طور معمول سیستم‌های تصمیم‌گیری خودکار (ADS) را به عنوان راهی برای «کاهش هزینه‌ها» یا ساده کردن کار بکار می‌گیرند. با این حال، دولت‌ها در ارزیابی این که چگونه این سیستم‌ها ممکن است به طور نامتناسبی به جمعیت‌هایی که قرار است به آن‌ها خدمت‌رسانی کنند آسیب برسانند، ناموفق است؛ به ویژه آن جمعیت‌هایی که بیش‌ترین آسیب را می‌بینند و کم‌ترین چاره یا حتی دانشی درباره‌ی این که این سیستم‌ها بر زندگی‌شان تأثیر عمیقی می‌گذارند. در بسیاری از موارد، حتی یک کارمند دولتی

هم وجود نداشت که بتواند تصمیم خودکار را تبیین کند، اشتباهات را درست کند یا نتایج تصمیم را بررسی دقیق کند. به واسطه‌ی مجموعه‌ای از توافقات فروشنده و پیمان‌کار، تقریباً تمام راه‌های پی‌بردن به تأثیرات این سیستم‌ها یا اعتراض به آن‌ها با حمایت‌هایی قانونی همچون قانون محرمانگی تجاری محافظت شد.

دوم اینکه، تعداد کمی از ادارات دولتی، به هنگام عوض کردن ساختار تصمیم‌گیری انسان‌محور به ساختار الگوریتم‌محور، اقداماتی واقعی برای تضمین عدالت و حفاظت از رویه‌های عادلانه انجام داده‌اند. مکانیزم‌های بررسی دقیق، درخواست تجدید نظر و پاسخگویی از طراحی سیستم‌های خودکار به طور کامل غایب بوده است. خوشبختانه، با دادخواهی قضایی موفق توسط وکلای اتحادیه آزادی‌های مدنی آمریکا (ACLU) از آیداهو، موسسه‌ی کمک‌های قانونی آرکانزاس، اداره‌ی معلمان هوستون، انجمن کمک حقوقی نیویورک و فعالان مختلف عمومی توانستند پیروزی‌های موکلین خود را تضمین کنند و توانستند براساس قانون اساسی و رویه‌های سازمانی عادلانه با این موارد غیرقانونی مبارزه کنند.

کتابچه‌ای که برای چگونگی شکایت‌کردن از الگوریتم‌ها تنظیم شده، هنوز در حال نگارش است، اما گزارش ما از چند استراتژی مفید برای تقویت راه‌حل‌های این مسئله رونمایی کرد. اول این که استدلال‌هایی که بر پایه رویه‌های عادلانه‌ی آیین‌نامه‌ای هستند، با وجود اکثریت عظیمی از قاضی‌ها که به نفع حمایت‌های قانون اساسی و حقوق مدنی افراد رای می‌دهند، چالش‌هایی جدی برای ادعاهای محرمانگی تجاری فروشندگان خصوصی به وجود آورده است.

دوم این که عدم آگاهی دادن به افراد و جوامع تحت تاثیر مهم است: اداراتی که از تعامل با گروه‌های اجتماعی‌ای که به استفاده از این سیستم‌ها ارتباط پیدا می‌کنند، غفلت کرده بودند، اغلب [در دادگاه‌ها] اینطور قضاوت شده‌اند که فرصت مناسبی برای آگاهی عمومی و دریافت نظرات فراهم نکرده‌اند؛ به این معنی که پیاده‌سازی سیستم‌های هوش مصنوعی توسط آن‌ها به شکلی بالقوه مغایر با قانون اساسی بود. سوم این که هم‌کاری میان رشته‌ای زمانی که برای تعیین عملکرد اشتباه سیستم تلاش می‌کنیم، مهم است؛ به خصوص زمانی که مدارک را به قضات ارسال می‌کنیم. در مواردی که وکلا با کارشناسان علوم مهندسی و علوم اجتماعی کار می‌کردند، قضات توانستند درباره‌ی فریب‌های علمی در این سیستم‌ها و همچنین پیامدها و آسیب‌های اجتماعی آن‌ها چیزهایی یاد بگیرند. با نگاهی به آینده، پیش‌بینی می‌کنیم که پرونده‌های حقوقی آتی درباره‌ی اقامه‌ی دعوی علیه الگوریتم‌ها درس‌های بیشتری برای ما داشته باشند. بررسی این پرونده‌ها، فهم بیشتری نسبت به این سیستم‌ها و پاسخگویی در مان‌گرانه‌ای برای آن‌ها ایجاد می‌کند؛ حتی در شرایطی که ادارات دولتی تلاش کرده‌اند که مالکیت یا نظارت روی آن‌ها را کتمان کنند. با ترکیب کردن ابزارهایی مانند «چارچوب ارزیابی تاثیر الگوریتمی» که AI Now تنظیم کرده و نظام‌های مقرراتی نظارتی قاطع، می‌توانیم به شناسایی، سنسجش و در صورت لزوم دخالت در اقداماتی که از هوش مصنوعی و سیستم‌های خودکار استفاده‌های آسیب‌رسان می‌کنند، مبادرت بورزیم. در هر حال، به منظور ادامه دادن به پیشرفت اخیر، وکلا و فعالان اجتماعی که نماینده

افراد حاضر در این دادخواهی‌ها هستند، به سرمایه‌گذاری و حمایت بیشتر و همچنین شبکه‌ای از متخصصان این حوزه نیازمند هستند که بتوانند مشاوره درباره‌ی استراتژی و بررسی دقیق سیستم‌ها به وکلا و فعالان اجتماعی کمک کنند.

۳-۴- پژوهش و تشکل‌سازی: یک ائتلاف جدید

گسترش سریع استفاده از هوش مصنوعی و سیستم‌های مرتبط در زندگی روزمره، یک نگرانی درباره‌ی آینده نیست. همه‌چیز همین جاست؛ هیچ نشانه‌ای از کاهش سرعت نیست. با درک این موضوع، و براساس سنت‌های قدیمی فعالیت اجتماعی و تشکیلاتی به منظور تقاضا برای تغییرات ساختاری به منظور پاسخگویی بیشتر، مجموعه‌ای از استراتژی‌ها طراحی شده است.

فعالیت اجتماعی توسط افراد متخصص در زمینه‌ی فناوری اتفاق جدیدی نیست. در اوایل دهه ۱۹۸۰، سازمان «متخصصان کامپیوتر در جستجوی مسئولیت اجتماعی» برای مخالفت با استفاده از کامپیوترها در جنگ تشکیل شد. اخیراً در سال ۲۰۱۶، پیمان «دوباره نه» هزاران کارگر را در بخش‌های مختلف فن‌آوری سازمان داده است که تعهدی را امضا کنند که طبق آن ساخت پایگاه‌های داده و مدیریت مجموعه‌های داده به قصد هدف قرار دادن اقلیت‌های مذهبی یا تسهیل کردن اخراج دسته‌جمعی [مهاجران] انجام نشود. در حالی که سازمان‌دهی و فعالیت اجتماعی در سال ۲۰۱۸ از یک سنت قدیمی الگو می‌گیرد، ابعاد آن در حوزه‌ی فن‌آوری اتفاقی جدید است. کارکنان فناوری در مخالفت با تصمیمات تکنیکی و تجاری

کارفرمایان خود، در حال پیوستن به جمع محققان و سازمان‌های جامعه مدنی هستند.

کارمندان گوگل در سال ۲۰۱۸ به صورت عمومی کار تشکیلاتی خود را در مخالفت با پروژه Maven آغاز کردند؛ پروژه‌ای که طی آن، پنتاگون به دنبال استفاده از قابلیت‌های ماشین بینایی گوگل در پروژه‌ی نظارت پهبادی وزارت دفاع آمریکا بود. محققان و سازمان‌های حقوق بشر به این مساله ورود کردند و در ماه ژوئن، گوگل رسماً اعلام کرد که این پروژه را رها خواهد کرد. در شرکت‌های آمازون، Salesforce و میکروسافت، کارمندان به منظور پایان بخشیدن به قرارداد با اداره‌ی مهاجرت و گمرک آمریکا (ICE)، نامه‌ای را امضا کردند که از سوی سازمان‌های حمایت از مهاجران حمایت شد. کارمندان شرکت Amazon، در واکنش به افشا کردن قراردادهای آمازون توسط ACLU در زمینه‌ی فروش سیستم‌های تشخیص چهره به دتسگاه‌های اجرایی، طوماری را برای پایان دادن به فروش این سیستم‌ها امضا کردند. به دنبال مسئله‌ی Maven، کارکنان گوگل مجدداً علیه پروژه Dragonfly برخاستند؛ نسخه‌ای از موتور جستجوی گوگل که برای بازار چین طراحی شده بود و سانسور و نظارت دولتی را ممکن می‌ساخت. در واکنش به گزارش رسانه‌ها در فاش کردن این اقدامات محرمانه، کارکنان خواستار نظارت و پاسخگویی اخلاقی شدند و بیش از ۷۰۰ نفر از آن‌ها از سازمان عفو بین‌الملل خواستند که این پروژه را لغو کنند و در این راستا نام و امضای خود را نیز پای نامه‌ای ثبت کردند.

مهم‌ترین برهه در اوایل نوامبر بود؛ هنگامی که ۲۰۰۰۰ شاغل گوگل

در اقدامی که «اعتصاب برای تغییری واقعی» نام گرفت در سراسر دنیا اعتصاب کردند. این اعتصاب، گوگل را شرکتی توصیف کرد که در آن «سو استفاده از قدرت، نژاد پرستی سیستماتیک و تصمیم‌گیری غیرپاسخگویانه امری عادی به حساب می‌آید.» سازمان‌دهندگان این اعتراضات از مدیران گوگل پنج خواسته داشتند: از جمله پایان دادن به نابرابری در دستمزد و فرصت‌ها، حذف داوری اجباری در موارد آزار و تبعیض جنسیتی و اضافه کردن نمایندگی کارکنان به هیات‌مدیره. یک هفته پس از اعتصاب، گوگل به بخش کوچکی از این درخواست‌ها پاسخ داد و توافق کرد که به داوری اجباری در موارد آزار جنسی خاتمه دهد (اما به طور قابل توجهی تبعیض را نادیده گرفت).^{۱۰} این جنبش به سرعت در سراسر صنعت، از جمله فیس بوک، eBay، Square، و Airbnb تکرار شد.

با پیوستن کارکنان به جمع پژوهشگران و گروه‌های جامعه مدنی، این موج جدید تشکل‌های کارگری باعث ایجاد گوناگونی و گشودگی بیشتر در حوزه تحقیقات هوش مصنوعی می‌شود. است. این جنبش‌ها دیدگاه‌های گوناگونی که متعلق به طبقات، حوزه‌ها و رشته‌های مختلف هستند را دخیل می‌کنند تا اطمینان حاصل شود که این دیدگاه‌ها قابلیت فهمیدن هزینه‌های واقعی فعالیت‌های شرکت، از جمله تاثیر سیستم‌های ساخته‌شده توسط آن‌ها را دارند. کارکنان گوگل که در اعتصاب شرکت کردند، ائتلاف خود را در سراسر طبقه‌ها و بخش‌ها گسترش دادند و بر نقش کارگران قراردادی در خواسته‌هایشان تاکید کردند و خودشان را در جنبش پیش‌رونده‌ای می‌دیدند که «نه تنها حوزه‌ی فناوری را شامل می‌شد، بلکه در

سراسر کشور، معلمان، کارگران فست فود و دیگرانی را در بر می‌گرفت که از قدرت خود برای ایجاد تغییرات واقعی استفاده می‌کردند. « موج اخیر فعالیت‌های اجتماعی در شرکت‌های فن‌آوری تا حد زیادی توسط سوت‌زن‌ها هدایت شده است، که اطلاعات مربوط به پروژه‌های محرمانه را برای روزنامه‌نگاران افشا ساخته‌اند. این افشاگری‌ها به آموزش مردمی که از چنین دسترسی‌هایی محروم بوده‌اند، کمک کرده است و به محققین و فعالان کمک کرده تحلیل‌های بیشتری را ارائه دهند. سوت‌زنی با ایجاد زمینه‌ی مشترکی از حقیقت، ائتلاف‌های گسترده‌ای ساخته که این جنبش‌ها را شاخص می‌کند. نقش مهم سوت‌زنی اخلاقی در طول سال گذشته هم اهمیت اجتماعی آن و هم نبود حمایت از کسانی که چنین افشاگری‌هایی را انجام می‌دهند را برجسته کرده است.

ائتلاف گسترده رهبران تشکل‌های کارگری، محققان و جامعه مدنی نقش فزاینده‌ای در فشار آوردن به حوزه‌ی فناوری به منظور پاسخگویی ایفا می‌کند. بسیاری از کارمندان مهندسی قدرت چانه‌زنی خود را دارند و در موقعیت منحصر به فردی برای خواستن تغییر از کارفرمایان خود قرار دارند. استفاده از این قدرت در جهت پاسخگویی بیشتر، مدلی امیدوارکننده از تشکل‌سازی کارکنان درباره‌ی مسائلی است که مربوط منافع عمومی هستند؛ به خصوص با توجه به فقدان فعلی مقررات دولتی، نظارت خارجی و دیگر اهرم‌های هدفمند که قادر به بازبینی و هدایت تصمیم‌گیری‌های شرکت‌های فن‌آوری هستند.

جمع بندی



در سال جاری شاهد این بودیم که سیستم‌های هوش مصنوعی به سرعت در حوزه‌های اجتماعی بیشتری وارد شدند و شمار بسیاری از مردم را در معرض خطر قرار دادند. در حالی که تکنیک‌های هوش مصنوعی هنوز وعده وعیده‌های قابل توجهی می‌دهند، بکارگیری برق‌آسای این سیستم‌ها بدون ارزیابی، پاسخگویی و نظارت مناسب می‌تواند خطراتی جدی ایجاد کند. ما نیاز مبرم داریم که در هر حوزه‌ای به تنظیم مقررات درباره‌ی سیستم‌های هوش مصنوعی بپردازیم؛ این تنظیم مقررات باید توجه خاصی به فناوری تشخیص چهره و تشخیص تمایلات داشته باشد و باید با تحقیقات دقیقی به آن‌ها رسید. اما مقررات تنها در صورتی موثر هستند که موانع قانونی و فناوری‌های که از بررسی دقیق این سیستم‌ها، فهمیدن نتایج آن‌ها و مداخله در آن‌ها جلوگیری می‌کنند، حذف شوند. در سال ۲۰۱۶، ما در اولین گزارش AI Now توصیه کردیم که قانون کلاهبرداری و سوء استفاده کامپیوتری (CFAA) و قانون کی‌رایت هزاره‌ی دیجیتال (DMCA) نباید در جهت محدود کردن تحقیقات مرتبط با پاسخگویی و بررسی دقیق هوش مصنوعی استفاده شود. امسال، فراتر می‌رویم: شرکت‌های

هوش مصنوعی باید از رازداری تجاری و سایر ادعاهای قانونی که از پاسخگویی الگوریتمی در حوزه‌ی عمومی جلوگیری می‌کنند، چشم‌پوشی کنند. دولت‌ها و موسسات عمومی باید توان فهمیدن و تبیین کردن این را داشته باشند که چگونه و چرا تصمیمات [الگوریتمی] اتخاذ می‌شوند، به خصوص زمانی که این تصمیمات، دسترسی مردم به مراقبت‌های بهداشتی، مسکن و اشتغال را تعیین می‌کند. خیلی وقت است که دیگر سوال ما این نیست که آیا در [استفاده از] سیستم‌های هوش مصنوعی آسیب و جانبداری وجود دارد یا خیر. این بحث حل شده است: شواهد در سال گذشته از حد شک فراتر رفته‌اند. مرحله بعدی پرداختن به این آسیب‌ها است. این مسئله با توجه به ابعاد پیاده‌سازی این سیستم‌ها و با توجه به شیوه‌ی کارکردشان در متمرکز کردن قدرت و نظارت در دستان عده‌ای قلیل و توزیع نابرابر هزینه‌ها و مزایایی که با این متمرکز کردن همراه است، مسئله‌ای ضروری و مبرم است. به منظور دنبال کردن فرآیند توسعه و پیاده‌سازی یک محصول در طول چرخه‌ی عمرش و لحاظ کردن هزینه‌های زیست‌محیطی و هزینه‌های مربوط به نیروی کار، ما به تحلیل‌های عمیق تری از «زنجیره تامین کامل» سیستم‌های هوش مصنوعی نیاز داریم.

علاوه بر این، مدت‌هاست انتظار می‌رود که شرکت‌های فن‌آوری مستقیماً به فرهنگ [اشتباه] محروم‌سازی و تبعیض در محل کار توجه کنند. نبود گوناگونی و وجود روش‌های جاری آزار و اذیت، محرومیت و نابرابری دست‌مزد، نه تنها برای کارمندان این شرکت‌ها آسیب‌زننده است بلکه بر محصولات هوش مصنوعی که این شرکت‌ها عرضه

می‌کنند نیز تاثیر می‌گذارد؛ چرا که ابزارهایی را تولید می‌کنند که تعصب و تبعیض را تداوم می‌بخشند .

ساختار حاضری که در آن توسعه و گسترش هوش مصنوعی رخ می‌دهد، در ضدیت معناداری با این مسایل ضروری قرار دارد. کسانی که در موقعیت منفعت‌اندوژی قرار دارند، این انگیزه را دارند که توسعه و بکارگیری این سیستم‌ها را تسریع می‌بخشند؛ این کار بدون زمان گذاشتن برای تشکیل تیم‌های مختلف، ایجاد حفاظت‌های ایمنی یا تست تاثیرات متمایز این سیستم‌ها انجام می‌شود. آن‌هایی که بیش از همه در معرض آسیب این سیستم‌ها قرار دارند، عموماً فاقد ابزارهای مالی هستند و دسترسی به مکانیزم‌های پاسخگویی که اجازه جبران خسارت و یا درخواست‌های تجدید نظر قانونی را می‌دهد، ندارند . به همین دلیل است که ما به نفع تامین بودجه‌ی بیشتر برای اقامه دعوی عمومی، سازمان‌دهی تشکل‌های کارگری و مشارکت اجتماعی استدلال می‌کنیم؛ چرا که سیستم‌های الگوریتمی و هوش مصنوعی، توازن قدرت را در بسیاری از نهادها و محیط‌های کاری تغییر داده‌اند.

ضروری است که توازن قدرت به نفع مردم تغییر کند. این اتفاق نیازمند تغییرات ساختاری مهمی است که از تمرکز صرف بر سیستم‌های فنی فراتر می‌رود؛ تغییرات ساختاری مانند اشتیاق به تغییر دادن مفروضات عملیاتی استاندارد که بازیگران صنعت مدرن هوش مصنوعی را مدیریت می‌کنند. تمرکز فعلی بر اصلاحات تکنیکی یکایک سیستم‌ها باید [به نحوی] گسترش پیدا کند که رشته‌های اجتماعی، گذشته‌های جوامع و استراتژی‌هایی را در بر بگیرد که ظرفیت

فهم عمیق‌تر از بافت‌های مختلف اجتماعی را فراهم کند؛ بافت‌هایی که توسعه و استفاده از سیستم‌های هوش مصنوعی را شکل می‌دهند. همانطور که بیشتر دانشگاه‌ها توجه خود را به مطالعه مفاهیم اجتماعی هوش مصنوعی معطوف می‌کنند، علم و مهندسی کامپیوتر دیگر نمی‌توانند کانون بی‌چون و چرای مسئله باشند؛ بلکه باید با رشته‌های اجتماعی و انسانی، سازمان‌های جامعه مدنی و جوامع آسیب‌دیده به شکل مساوی هم‌کاری کنند.

خوشبختانه، دیده شدن ائتلاف‌های جدید میان پژوهشگران، فعالان، وکلا، کارگران فن‌آوری و سازمان‌های جامعه مدنی در حال شکل‌گیری است تا از نظارت، پاسخگویی و کنترل مداوم سیستم‌های هوش مصنوعی حمایت کنند. برای آن‌که این ارتباطات مهم رشد کنند، حمایت‌های بیشتری نیاز است؛ مانند تعهدی از سوی شرکت‌های فن‌آوری برای محافظت از منتقدان خیرخواهی که نمی‌خواهند در قراردادهای نظامی یا انتظامی کار کنند و تعهد به حمایت از کارکنانی که در فعالیتهای تشکیلاتی و سوت‌زنی درگیر هستند. در سال گذشته بسیاری از دشوارترین چالش‌ها برای پاسخگویی و عدالت به وجود آمد؛ چرا که سیستم‌های هوش مصنوعی به اعماق دنیای اجتماعی راه یافتند. با این حال، برهه‌های فوق‌العاده‌ای مانند نزاع‌های عمومی مهم و فرم‌های امیدوارکننده‌ای از اعتراضات وجود داشت که ممکن است در نهایت مسیرهایی برای تغییراتی پیامدگرایانه و مثبت را روشن کند.

منابع



۱. همانطور که پیشگامان حوزه‌ی هوش مصنوعی، استوارت راسل و پیتر نورویگ اشاره کردند، تاریخچه‌ی هوش مصنوعی تعریف واضحی از آن را ارائه نداده است؛ اما مشاهده می‌شود که این فناوری [به شکل‌های مختلفی به چهار هدف ممکن اهمیت می‌دهد: «سامانه‌هایی که مانند انسان فکر می‌کنند، سامانه‌هایی مانند انسان عمل می‌کنند، سامانه‌هایی که عقلانی می‌اندیشند، سامانه‌هایی که عقلانی عمل می‌کنند.» ما در این گزارش از واژه‌ی هوش مصنوعی برای اشاره به اجتماعی گسترده از فناوری‌هایی که همه‌شان به آرایه‌ای از داده‌ها و زیرساخت‌های محاسباتی وابسته هستند، استفاده می‌کنیم. این اجتماع از فناوری‌ها، فناوری‌هایی مانند تشخیص گفتار، ترجمه‌ی زبان، تشخیص تصویر، پیش‌بینی‌ها و تعیین‌ها -وظایفی که در کنار چهار هدفی که راسل و نورویگ مشخص می‌کنند، به صورت سنتی بر عهده‌ی انسان بوده است- را پوشش می‌دهد. با این‌که هوش مصنوعی [پدیده‌ای] جدید نیست، پیشرفت‌های اخیر در توانایی جمع‌آوری داده و ذخیره‌ی مقادیر زیادی از آن‌ها به همراه توسعه‌ی قدرت محاسباتی، در کنار فشار برای تجاری‌سازی هوش مصنوعی و استفاده از آن در زمینه‌های اصلی اجتماعی، منجر به پیشرفت‌های غیرمنتظره‌ای در حوزه‌ی هوش مصنوعی شده است. ببینید:

Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, (Englewood Cliffs, NJ: Prentice Hall, 1995), 2

2. Carole Cadwalladr and Emma Graham-Harrison, "Revealed: 50 Million Facebook Profiles Harvested for Cambridge Analytica in Major Data Breach," *The Guardian*, March 2018, 17, <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>.

3. Guy Rosen, "Security Update," *Facebook Newsroom*, September 2018, 28, <https://newsroom.fb.com/news/09/2018/security-update/>

4. Josh Eidelson, "Facebook Tools Are Used to Screen Out Older Job Seekers,"

Lawsuit Claims,” Bloomberg, May 2018 ,29, <https://www.bloomberg.com/news/articles/29-05-2018/facebook-tools-are-used-to-screen-out-older-job-seekers-lawsuit-claims>.

5. Bloomberg Editorial Board, “Think the U.S. Has a Facebook Problem? Look to Asia,” Bloomberg, October 2017 ,22, <https://www.bloomberg.com/opinion/articles/22-10-2017/facebook-has-a-bigger-problem-than-washington> .

6. Andrew Liptak, “The US Government Alleges Facebook Enabled Housing Ad Discrimination,” The Verge , August 2018 ,19, <https://www.theverge.com/17757108/19/8/2018/us-department-of-housing-and-urban-development-facebook-complaint-race-gender-discrimination> .

7. Elizabeth Weise, “Russian Fake Accounts Showed Posts to 126 Million Facebook Users,” USA TODAY , October 2017 ,30, <https://www.usatoday.com/story/tech/30/10/2017/russian-fake-accounts-showed-posts-126-million-facebook-users/815342001/> .

8. Hamza Shaban, Craig Timberg, and Elizabeth Dwoskin, “Facebook, Google and Twitter Testified on Capitol Hill. Here’s What They Said,” Washington Post , October 2017 ,31, <https://www.washingtonpost.com/news/the-switch/wp/31/10/2017/facebook-google-and-twitter-are-set-to-testify-on-capitol-hill-heres-what-to-expect/> ;

Casey Newton, “Mark Zuckerberg’s Appearance before European Parliament Yields an Empty Spectacle,” The Verge , May 2018 ,22, <https://www.theverge.com/17381250/22/5/2018/mark-zuckerberg-european-parliament-facebook> .

9. Drew Harwell, "AI will solve Facebook's most vexing problems, Mark Zuckerberg says. Just don't ask when or how," Washington Post , April 2018 ,11, <https://www.washingtonpost.com/news/the-switch/wp/11/04/2018/ai-will-solve-facebooks-most-vexing-problems-mark-zuckerberg-says-just-dont-ask-when-or-how/> ...

10. Kate Conger and Dell Cameron, "Google Is Helping the Pentagon Build AI for Drones," Gizmodo , March 6, 2018, <https://gizmodo.com/google-is-helping-the-pentagon-build-ai-for-drones-1823464533> .

11. Rick Paulas, "A New Kind of Labor Movement in Silicon Valley," The Atlantic , September 4, 2018, <https://www.theatlantic.com/technology/archive/2018/09/tech-labor-movement/567808/>

12. Hamza Shaban, "Amazon Employees Demand Company Cut Ties with ICE," Washington Post , June

22,2018, <https://www.washingtonpost.com/news/the-switch/wp/2018/06/22/amazon-employees-demand-company-cut-ties-with-ice/>

Jacob Kastrenakes, "Salesforce Employees Ask CEO to 'Re-Examine' Contract with Border Protection Agency," The Verge , June 25, 2018, <https://www.theverge.com/2018/6/25/17504154/salesforce-employee-letter-border-protection-ice-immigration-cbp> ;

Colin Lecher, "The Employee Letter Denouncing Microsoft's ICE Contract Now Has over 300 Signatures," The Verge , June 21, 2018, <https://www.theverge.com/2018/6/21/17488328/microsoft-ice-employees-signatures-protest> .

13. Nikhil Sonnad, "US Border Agents Hacked Their "Risk Assessment" Sys

tem to Recommend Detention

100% of the Time,” Quartz , June 26, 2018, <https://qz.com/1314749/us-border-agents-hacked-their-risk-assessment-system-to-recommend-immigrant-detention-every-time/> .

14. Daisuke Wakabayashi, “Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam,” The New York Times , July 30, 2018, <https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html> .

15. Nikhil Sonnad, “A Flawed Algorithm Led the UK to Deport Thousands of Students,” Quartz , May 3, 2018, <https://qz.com/1268231/a-toeic-test-led-the-uk-to-deport-thousands-of-students/> .

16. Casey Ross and Ike Swetlitz, “IBM’s Watson Recommended ‘unsafe and Incorrect’ Cancer Treatments,” STAT , July 25, 2018, <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/> .

17. George Joseph and Kenneth Lipp, “IBM Used NYPD Surveillance Footage to Develop Technology That Lets Police Search by Skin Color,” The Intercept , September 6, 2018, <https://theintercept.com/2018/09/06/nypd-surveillance-camera-skin-tone-search/> .

۱۸. برای مشاهده‌ی گاه‌شمار رخدادهای سال ۲۰۱۸

Kate Crawford and Meredith Whittaker, “AI in 2018: A Year in Review,” Medium , October 14, 2018, <https://medium.com/@AINowInstitute/ai-in-2018-a-year-in-review-8b161ead2b4e>.

19. Jon Evans, “The Techlash,” TechCrunch , June 17, 2018, <https://techcrunch.com/2018/06/17/the-techlash/>

20. "Microsoft Calls for Facial Recognition Technology Rules given 'Potential for Abuse,'" The Guardian , July 14, 2018, <https://www.theguardian.com/technology/2018/jul/14/microsoft-facial-recognition-technology-rules-potential-for-abuse> .
21. Natalie Ram, "Innovating Criminal Justice," Northwestern University Law Review 112, no. 4 (February 1, 2018): 659–724, <https://scholarlycommons.law.northwestern.edu/nulr/vol112/iss4/2> ;
Rebecca Wexler, "Life, Liberty, and Trade Secrets," Stanford Law Review 70, no. 5 (May 2018): 1343–1429, <https://www.stanfordlawreview.org/print/article/life-liberty-and-trade-secrets/> ;
Danielle Keats Citron and Frank A. Pasquale, "The Scored Society: Due Process for Automated Predictions," Washington Law Review 89, no. 1 (2014): 1–33, <https://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1318/89WLR0001.pdf> .
22. Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, (Cambridge: Harvard University Press, 2015).
23. D. Sculley, Jasper Snoek, Alex Wiltschko, and Ali Rahimi, "Winner's Curse? On Pace, Progress and Empirical Rigor," 6th International Conference on Learning Representations (ICLR), (Vancouver, 2018), <https://openreview.net/pdf?id=rjWF0Fywf> .
24. Kate Crawford, "The Test We Can—and Should—Run on Facebook," The Atlantic , July 2, 2014, <https://www.theatlantic.com/technology/archive/2014/07/the-test-we-canand-shouldrun-on-facebook/373819/> ;
Molly Jackman and Lauri Kanerva, "Evolving the IRB: Building Robust Re

view for Industry Research,” Washington & Lee Law Review Online 72, no. 8 (June 14, 2016): 442–457;

Zoltan Boka, “Facebook’s Research Ethics Board Needs to Stay Far Away from Facebook,” Wired , June 23, 2016, <https://www.wired.com/2016/06/facebook-research-ethics-board-needs-stay-far-away-facebook/>

25. “Sandvig v. Sessions — Challenge to CFAA Prohibition on Uncovering Racial Discrimination Online,” September 12, 2017, American Civil Liberties Union , <https://www.aclu.org/cases/sandvig-v-sessions-challenge-cfaa-prohibition-uncovering-racial-discrimination-online> .

26. Simone Browne, Dark Matters: On the Surveillance of Blackness (Durham: Duke University Press, 2015);

Alvaro M. Bedoya, “What the FBI’s Surveillance of Martin Luther King Tells Us About the Modern Spy Era,” Slate , January 18, 2016, <https://slate.com/technology/2016/01/what-the-fbis-surveillance-of-martin-luther-king-says-about-modern-spying.html> ;

James Ball, Julian Borger, and Glenn Greenwald, “Revealed: How US and UK Spy Agencies Defeat Internet Privacy and Security,” The Guardian , September 6, 2013, <https://www.theguardian.com/world/2013/sep/05/nsa-gchq-encryption-codes-security> ;

Shoshana Zuboff, “Big Other: Surveillance Capitalism and the Prospects of an Information Civilization,” Journal of Information Technology 30, no. 1 (March 1, 2015): 75–89, <https://doi.org/10.1057/jit.2015.5> .

27. Alice Shen, “Facial Recognition Tech Comes to Hong Kong-Shenzhen Border,” South China Morning Post , July 24, 2018, <https://www.scmp.com/>

news/china/society/article/2156510/china-uses-facial-recognition-system-deter-tax-free-traders-hong .

28. Stephen Chen, "China's Robotic Spy Birds Take Surveillance to New Heights," South China Morning Post , June 24, 2018, <https://www.scmp.com/news/china/society/article/2152027/china-takes-surveillance-new-heights-flock-robotic-doves-do-they> .

29. David Z. Morris, "China Will Block Travel for Those With Bad 'Social Credit,'" Fortune , March 18, 2018, <http://fortune.com/2018/03/18/china-travel-ban-social-credit/> .

30. Nathan Vanderklippe, "Chinese Blacklist an Early Glimpse of Sweeping New Social-Credit Control," The Globe and Mail , January 3, 2018, <https://www.theglobeandmail.com/news/world/chinese-blacklist-an-early-glimpse-of-sweeping-newsocial-credit-control/article37493300/>

31. "China Has Turned Xinjiang into a Police State like No Other," The Economist , May 31, 2018, <https://www.economist.com/briefing/2018/05/31/china-has-turned-xinjiang-into-a-police-state-like-no-other> .

32. Emily Feng and Louise Lucas, "Inside China's Surveillance State," Financial Times , July 20, 2018, <https://www.ft.com/content/2182eebe-8a17-11e8-bf9e-8771d5404543>

33. Angus Berwick, "A New Venezuelan ID, Created with China's ZTE, Tracks Citizen Behavior," Reuters , November 14, 2018, <https://www.reuters.com/investigates/special-report/venezuela-zte/> .

34. Nafeez Ahmed, "Pentagon Wants to Predict Anti-Trump Protests Using Social Media Surveillance," Motherboard , October 30, 2018, https://motherboard.vice.com/en_us/article/2018/10/pentagon-wants-to-predict-anti-trump-protests-using-social-media-surveillance

erboard.vice.com/en_us/article/7x3g4x/pentagon-wants-to-predict-anti-trump-protests-using-social-media-surveillance .

35. Karen Hao, "Amazon Is the Invisible Backbone behind ICE's Immigration Crackdown," MIT Technology Review , October 22, 2018, <https://www.technologyreview.com/s/612335/amazon-is-the-invisible-backbone-behind-ices-immigration-crackdown/> .

36. "Who's behind Ice?" (Empower LLC, Mijente, The National Immigration Project, and the Immigrant Defense Project, October 23, 2018), https://mijente.net/wp-content/uploads/2018/10/WHO%E2%80%99S-BEHIND-ICE-The-Tech-and-Data-Companies-Fueling-Deportations_v3-.pdf .

37. Brendan Shillingford et al., "Large-Scale Visual Speech Recognition," arXiv preprint [Cs], arXiv:1807.05162, July 13, 2018.

38. "Machine Vision Algorithm Learns to Recognize Hidden Facial Expressions," MIT Technology Review , November 13, 2015, <https://www.technologyreview.com/s/543501/machine-vision-algorithm-learns-to-recognize-hidden-facial-expressions/> .

39. Richard T. Gray, *About Face: German Physiognomic Thought from Lavater to Auschwitz* (Detroit: Wayne State University Press, 2004); Sharrona Pearl, *About Faces: Physiognomy in Nineteenth-Century Britain* (Cambridge, MA: Harvard University Press, 2010).

40. *Nineteenth-Century Britain* (Cambridge, MA: Harvard University Press, 2010).

Blaise Aguera y Arcas, Margaret Mitchell, and Alexander Todorov, "Physiognomy's New Clothes," Medium, May 7, 2017, <https://medium.com/@>

blaisea/physiognomys-new-clothes-f2d4b59fdd6a .

41. Ruth Leys, "How Did Fear Become a Scientific Object and What Kind of Object Is It?" *Representations* 110, no. 1 (2010): 66–104, <https://doi.org/10.1525/rep.2010.110.1.66> .

روث لیز به برنامه‌ی پژوهشی اکمن انتقاداتی را وارد کرده است که در کتاب زیر می‌توانید این انتقادات را بخوانید:

Ruth Leys, *The Ascent of Affect: Genealogy and Critique* (Chicago: University of Chicago Press, 2017).

42. Alan J. Fridlund, *Human Facial Expression: An Evolutionary View* (San Diego: Academic Press, 1994).

43. Lisa Feldman Barrett, "Are Emotions Natural Kinds?" *Perspectives on Psychological Science* 1, no. 1 (March 2006): 28–58, <https://doi.org/10.1111/j.1745-6916.2006.00003.x> ;

Erika H. Siegel, Molly K. Sands, Wim Van den Noortgate, Paul Condon, Yale Chang, Jennifer Dy, Karen S. Quigley, and Lisa Feldman Barrett. 2018. "Emotion Fingerprints or Emotion Populations? A Meta-Analytic Investigation of Autonomic Features of Emotion Categories." *Psychological Bulletin* 144 (4): 343–93, <https://doi.org/10.1037/bul0000128>.

۴۴. برای مثال، علی‌رغم انتقادات دفتر پاسخگویی دولتی ایالات متحده‌ی آمریکا، اداره‌ی امنیت سرمایه‌گذاری کرده است؛ پروژه‌های SPOT حمل و نقل آمریکا بیش از یک میلیارد دلار در پروژه‌ی که قصد دارد تروریست‌های احتمالی را از طریق همین نشان‌گرهای رفتاری شناسایی کند.

"Aviation Security: TSA Should Limit Future Funding for Behavior Detection Activities" (Washington, DC: U.S. Government Accountability Office, November 13, 2013), <https://www.gao.gov/products/GAO-14-159>.

45. Jonathan Metz, *The Protest Psychosis: How Schizophrenia Became a Black Disease* (Boston: Beacon Press, 2009).
46. Mark Lieberman, "Sentiment Analysis Allows Instructors to Shape Course Content around Students' Emotions," *Inside Higher Education* , February 20, 2018, <https://www.insidehighered.com/digital-learning/article/2018/02/20/sentiment-analysis-allows-instructors-shape-course-content> .
47. Will Knight, "Emotional Intelligence Might Be a Virtual Assistant's Secret Weapon," *MIT Technology Review*, June 13, 2016, <https://www.technologyreview.com/s/601654/amazon-working-on-making-alexa-recognize-your-emotions>;
48. "Affectiva Automotive AI," *Affectiva* , accessed November 18, 2018, <http://go.affectiva.com/auto> .
49. Jeff Weiner, "ACLU: Amazon's Face-Recognition Software Matched Members of Congress with Mugshots," *Orlando Sentinel* , July 26, 2018, <https://www.orlandosentinel.com/news/politics/political-pulse/os-amazon-rekognition-face-matching-software-congress-20180726-story.html> ;
"Amazon Rekognition Announces Real-Time Face Recognition, Text in Image Recognition, and Improved Face Detection," *Amazon Web Services* , November 21, 2017, <https://aws.amazon.com/about-aws/whats-new/2017/11/amazon-rekognition-announces-real-timeface-recognition-text-in-image-recognition-and-improved-face-detection/> .
50. Chris Adzima, "Using Amazon Rekognition to Identify Persons of Interest for Law Enforcement," *Amazon Web Services* , June 15, 2017, <https://aws.amazon.com/blogs/ai/using-amazon-rekognition-to-identify-persons-of-interest-for-law-enforcement/> .

<https://aws.amazon.com/blogs/machine-learning/using-amazon-rekognition-to-identify-persons-of-interest-for-law-enforcement/> .

51. Ranju Das, "Image & Video Rekognition Based on AWS" (Amazon Web Services Summit, Seoul, 2018), <https://youtu.be/sUzuJc-xBEE?t=1889> .

52. Jacob Snow, "Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots," American Civil Liberties Union, July 26, 2018, <https://www.aclu.org/blog/privacy-technology/surveillance-technoloamazons-face-recognition-falsely-matched-28>.

53. Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," in Conference on Fairness, Accountability and Transparency (New York, 2018), 77-91, <http://proceedings.mlr.press/v81/buolamwini18a.html> .

54. Matt Wood, "Thoughts On Machine Learning Accuracy," AWS News Blog , July 27, 2018, <https://aws.amazon.com/blogs/aws/thoughts-on-machine-learning-accuracy/> .

55. Sidney Fussell, "Amazon Accidentally Makes Rock-Solid Case for Not Giving Its Face Recognition Tech to Police," Gizmodo , July 27, 2018, <https://gizmodo.com/amazon-accidentally-makes-rock-solid-case-for-not-givin-1827934703> .

56. Bryan Menegus, "Amazon Breaks Silence on Aiding Law Enforcement Following Employee Backlash," Gizmodo , August 11, 2018, <https://gizmodo.com/amazon-breaks-silence-on-aiding-law-enforcement-follo-wi-1830321057>

57. Joseph and Lipp, "IBM Used NYPD Surveillance Footage to Develop Tech

nology That Lets Police Search by Skin Color”

58. Jenna Bitar and Jay Stanley, “Are Stores You Shop at Secretly Using Face Recognition on You?,” American Civil Liberties Union , March 26, 2018, <https://www.aclu.org/blog/privacy-technology/surveillance-technologies/are-stores-you-shop-secretly-using-face> .

59. John Brandon, “Walmart Will Scan for Unhappy Shoppers Using Facial Recognition (Cue the Apocalypse),” VentureBeat , August 9, 2017, <https://venturebeat.com/2017/08/09/walmart-will-scan-for-unhappy-shoppers-using-facial-recognition-cue-the-apocalypse/> .

۶۰. قوانین زیر را ببینید:

AG and Legal Workforce Act, H.R. 6417, 115th Cong. (2017-2018), <https://www.congress.gov/115/bills/hr6417/BILLS-115hr6417ih.pdf> ;

Securing America’s Future Act of 2018, H.R. 4760, 115th Cong. (2017-2018), <https://www.congress.gov/115/bills/hr4760/BILLS-115hr4760ih.pdf>;

Strong Visa Integrity Secures America Act,” H.R. 2626, 115th Cong. (2017-2018), <https://www.congress.gov/115/bills/hr2626/BILLS-115hr2626ih.pdf> ;

H.R. Security and Immigration Reform Act of 2018, 6136, 115th Cong. (2017-2018), <https://www.congress.gov/115/bills/hr6136/BILLS-115hr-6136ih.pdf> .

61. Biometric Information Privacy Act, 740 ILCS 14/1 (Statutes current through the end of the 2018 Regular Session of the 100th General Assembly), <https://advance-lexis-com.proxy.library.nyu.edu/api/document?collection=statutes-legislation&id=urn:contentItem:5C66-0WY1-6YS3-D06V->

00000-00&context=1516831 .

62. Brian Hofer, "BART Board Approves Surveillance Ordinance, Lake Merritt Development," KTVU , September 13, 2018, <http://www.ktvu.com/news/bart-board-approves-surveillance-ordinance-lake-merritt-development>.

63. "Letter from Nationwide Coalition to Amazon CEO Jeff Bezos Regarding Rekognition," American Civil Liberties Union , June 18, 2018, <https://www.aclu.org/letter-nationwide-coalition-amazon-ceo-jeff-bezos-regarding-rekognition> .

64. Abrar Al-Heeti, "Congress Still Wants Answers from Amazon about Its Facial Recognition Tech," CNET , November 29, 2018, <https://www.cnet.com/news/congress-still-wants-answers-from-amazon-about-its-facial-recognition-tech/> .

65. Woodrow Hartzog and Evan Selinger, "Facial Recognition Is the Perfect Tool for Oppression," Medium , August 2, 2018, <https://medium.com/s/story/facial-recognition-is-the-perfect-tool-for-oppression-bc2a08f0fe66> .

66. Yilun Wang and Michael Kosinski, "Deep Neural Networks Are More Accurate than Humans at Detecting Sexual Orientation from Facial Images," Journal of Personality and Social Psychology 114, no. 2 (2018): 246–57, <http://dx.doi.org/10.1037/pspa0000098> .

67. Frank Pasquale, "When Machine Learning Is Facially Invalid," Communications of the ACM 61, no. 9 (August 2018): 25–27, <https://doi.org/10.1145/3241367> .

68. Kade Crockford, "Massachusetts Should Ban Face Recognition Technology," WBUR , August 1, 2018,

<http://www.wbur.org/cognoscenti/2018/08/01/kade-crockford-face-surveillance-technology-ban> .

همچنین ببینید:

Hartzog and Selinger, "Facial Recognition Is the Perfect Tool for Oppression."

69. Brad Smith, "Facial Recognition Technology: The Need for Public Regulation and Corporate Responsibility," Microsoft on the Issues , July 13, 2018, <https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/> ;

Sidney Fussell, "Axon CEO Says Face Recognition Isn't Accurate Enough for Body Cams Yet," Gizmodo , August 8, 2018, <https://gizmodo.com/axon-ceo-says-face-recognition-isnt-accurate-enough-for-1828205723> .

70. "Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems" (New York: AI Now Institute, September 2018), <https://ainowinstitute.org/litigatingalgorithms.pdf> ;

Dillon Reisman, Jason Schultz, Kate Crawford, and Meredith Whittaker, "Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability" (New York: AI Now Institute, April 2018), <https://ainowinstitute.org/aiareport2018.pdf> ;

Micah Altman, Alexandra Wood, and Effy Vayena, "A Harm-Reduction Framework for Algorithmic Fairness," IEEE Security Privacy 16, no. 3 (May 2018): 34–45, <https://doi.org/10.1109/MSP.2018.2701149>.

71. Lecher, "A Healthcare Algorithm Started Cutting Care, and No One Knew Why."

73. Carlos Gradín, "World Income Inequality Database (WIID3.4)," United Nations University , October 27, 2015, <https://www.wider.unu.edu/project/wiid-world-income-inequality-database> .

74. Serena Lei, "Nine Charts about Wealth Inequality in America," The Urban Institute, 2015, <http://urbn.is/wealthcharts> ;

Lisa J. Dettling, Joanne W. Hsu, Lindsay Jacobs, Kevin B. Moore, and Jeffrey P. Thompson, "Recent Trends in Wealth-Holding by Race and Ethnicity: Evidence from the Survey of Consumer Finances," (Washington: Board of Governors of the Federal Reserve System, September 27, 2017), <https://doi.org/10.17016/2380-7172.2083> .

75. Andrew Restuccia, Sarah Ferris, and Helena Bottemiller Evich, "Behind Trump's Plan to Target the Federal Safety Net," POLITICO , December 11, 2017, <http://politi.co/2AJfCzL> .

76. Brakkton Booker, "White House Budget Calls For Deep Cuts To HUD," NPR.org , February 13, 2018, <https://www.npr.org/2018/02/13/585255697/white-house-budget-calls-for-deep-cuts-to-hud>.

77. Alison Kodjak, "Federal Judge Blocks Medicaid Work Requirements In Kentucky," NPR.org , June 29, 2018, <https://www.npr.org/sections/health-shots/2018/06/29/624807533/federal-judge-blocks-medicaidwork-requirements-in-kentucky> ;

Rachel Garfield et al., "Implications of Work Requirements in Medicaid: What Does the Data Say?," The Henry J. Kaiser Family Foundation , June 12, 2018, <https://www.kff.org/medicaid/issue-brief/implications-of-work-re>

quirements-in-medicaid-what-does-the-data-say/ .

78. Stacy Dean, "President's Budget Would Shift Substantial Costs to States and Cut Food Assistance for Millions," Center on Budget and Policy Priorities, May 23, 2017, <https://www.cbpp.org/research/food-assistance/presidents-budget-would-shift-substantial-costs-to-states-and-cut-food> .

79. David Pegg and Niamh McIntyre, "Child Abuse Algorithms: From Science Fiction to Cost-Cutting Reality," The Guardian , September 16, 2018, <https://www.theguardian.com/society/2018/sep/16/child-abuse-algorithms-from-science-fiction-to-cost-cutting-reality> .

80. David Scharfenberg, "Computers Can Solve Your Problem. You May Not like the Answer," The Boston Globe , September 21, 2018, <https://apps.bostonglobe.com/ideas/graphics/2018/09/equity-machine> ;

Joe Flood, *The Fires: How a Computer Formula, Big Ideas, and the Best of Intentions Burned Down New York City—and Determined the Future of Cities* (New York: Riverhead Books, 2011).

81. Virginia Eubanks, "We Created Poverty. Algorithms Won't Make That Go Away," The Guardian , May 13, 2018, <https://www.theguardian.com/commentisfree/2018/may/13/we-created-poverty-algorithms-wont-make-that-go-away> ;

82. Leo Morales, "Federal Court Rules Against Idaho Department of Health and Welfare in Medicaid Class Action," ACLU of Idaho , March 30, 2016, <https://www.acluidaho.org/en/news/federal-court-rules-against-idaho-department-health-and-welfare-medicaid-class-action> .

۸۳. به مقاله‌ی موسسه‌ی AI Now با عنوان "Litigating Algorithms". نگاه کنید.

85. Michael Nash, "Examination of Using Structured Decision Making and Predictive Analytics in Assessing Safety and Risk in Child Welfare" (Los Angeles: County of Los Angeles Office of Child Protection, May 4, 2017), http://file.lacounty.gov/SDSInter/bos/bc/1023048_05.04.17OCPReportonRiskAssessmentTools_SDMandPredictiveAnalytics_.pdf .

86. Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (New York: St. Martin's Press, 2018).

87. "Vulnerable Children Predictive Modelling" (Wellington: New Zealand Ministry of Social Development), accessed November 18, 2018, <https://www.msd.govt.nz/about-msd-and-our-work/publications-resources/research/predictive-modelling/> .

88. Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (New York: St. Martin's Press, 2018).

89. Margaret Talbot, "The Case Against Cash Bail," The New Yorker , August 25, 2015, <https://www.newyorker.com/news/news-desk/the-case-against-cash-bail> .

90. Sam Levin, "Imprisoned by Algorithms: The Dark Side of California Ending Cash Bail," The Guardian , September 7, 2018, <https://www.theguardian.com/us-news/2018/sep/07/imprisoned-by-algorithms-the-dark-side-of-california-ending-cash-bail> ;

Maddie Hanna, "What Happened When New Jersey Stopped Relying on Cash Bail," The Philadelphia Inquirer , February 16, 2018, <http://www2>.

philly.com/philly/news/new_jersey/new-jersey-cash-bail-risk-assessment-20180216.html .

91. Colleen O'Dea, "Civil Rights Coalition Calls for End to Core Element of NJ Bail Reform," NJ Spotlight , July 31, 2018, <http://www.njspotlight.com/stories/18/07/30/civil-rights-groups-call-for-end-to-core-element-of-nj-bail-reform/> .

92. Jeremy B. White, "California Ended Cash Bail. Why Are So Many Reformers Unhappy About It?," POLITICO Magazine , August 29, 2018, <https://politi.co/2PPLnyz> .

93. Rose Luckin, "Towards Artificial Intelligence-Based Assessment Systems," Nature Human Behaviour 1 (March 1, 2017): 1–3, <https://doi.org/10.1038/s41562-016-0028> .

94. Brent Bridgeman, Catherine Trapani, and Yigal Attali, "Comparison of Human and Machine Scoring of Essays: Differences by Gender, Ethnicity, and Country," Applied Measurement in Education 25, no. 1 (January 2012): 27–40, <https://doi.org/10.1080/08957347.2012.635502> .

95. Nitin Madnani et al., "Building Better Open-Source Tools to Support Fairness in Automated Scoring," in Proceedings of the First ACL Workshop on Ethics in Natural Language Processing (Valencia: Association for Computational Linguistics, 2017), 41–52, <https://doi.org/10.18653/v1/W17-1605> .

96. Daniel T. O'Brien et al., "An Evaluation of Equity in the Boston Public Schools' Home-Based Assignment Policy" (Boston: Boston Area Research Initiative, July 2018), <https://news.northeastern.edu/wp-content/uploads/2018/07/BPSHBAP.pdf> .

۹۷. همان.

98. Scharfenberg, "Computers Can Solve Your Problem. You May Not like the Answer"

۹۹. همان.

100. Dillon Reisman, Jason Schultz, Kate Crawford, and Meredith Whittaker, "Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability" (New York: AI Now Institute, April 2018), <https://ainowinstitute.org/aiareport2018.pdf> .

۱۰۱. به مقاله‌ی موسسه‌ی AI Now با عنوان "Litigating Algorithms" نگاه کنید.

102. "Algorithmic Accountability Policy Toolkit" (New York: AI Now Institute, October 2018), <https://ainowinstitute.org/aap-toolkit.pdf> .

103. Reisman et al, "Algorithmic Impact Assessments."

104. Wakabayashi, "Self-Driving Uber Car Kills Pedestrian in Arizona."

105. Jack Stewart, "Tesla's Self-Driving Autopilot Involved in Another Deadly Crash," Wired , March 31, 2018, <https://www.wired.com/story/tesla-autopilot-self-driving-crash-california/> .

106. "Uber Autonomous-SUV Driver Streamed 'The Voice' Just before Deadly Arizona Crash, Report Says," Los Angeles Times , June 22, 2018, <https://www.latimes.com/business/autos/la-fi-hy-uber-self-driving-death-20180622-story.html> .

۱۰۷. ویدیوی زیر ببینید:

"Early Rider Program," Waymo , accessed November 16, 2018, <https://waymo.com/apply/> .

108. Aarian Marshall, "Wanna Save Lots of Lives? Put (Imperfect) Self-Drive

ing Cars on the Road, ASAP” Wired , November 7, 2017, <https://www.wired.com/story/self-driving-cars-rand-report/>.

109. Lindsay Moore, “Autonomous Vehicles Continue to Drive Without Rules,” Phoenix New Times , November 9, 2017, <https://www.phoenixnewtimes.com/news/autonomous-vehicles-continue-to-drive-without-rules-9853743> .

۱۱۰. گزارش‌هایی گویند که راننده‌ی پشتیبان اوپرتنها کمی قبل از حادثه‌ی مرگبار آریزونا، صدا را به صورت زنده بارگذاری کرده است.

111. Ross and Swetlitz, “IBM’s Watson Recommended ‘unsafe and Incorrect’ Cancer Treatments.”

112. Angela Chen, “What the Apple Watch’s FDA Clearance Actually Means,” The Verge , September 13, 2018, <https://www.theverge.com/2018/9/13/17855006/apple-watch-series-4-ekg-fda-approved-vs-clearedmeaning-safe> .

113. Mark Vermette, “Apple Watch Approval Marks Shift In Device Development And Approvals,” Med Device Online , October 24, 2018, <https://www.meddeviceonline.com/doc/apple-watch-approval-marks-shift-in-device-development-and-approvals-0001> ;

Rajiv Leventhal, “4.4M Patient Records Breached in Q3 2018, Protenus Finds,” Healthcare Informatics Magazine , November 7, 2018, <https://www.healthcare-informatics.com/news-item/cybersecurity/44m-patient-records-breached-q3-2018-protenus-finds> .

114. Alex Hern, “Google ‘betrays Patient Trust’ with DeepMind Health Move,” The Guardian , November 14, 2018, <https://www.theguardian.com/>

technology/2018/nov/14/google-betrays-patient-trust-deepmind-health-care-move .

115. Benjamin Herold, "Pearson Tested 'Social-Psychological' Messages in Learning Software, With Mixed Results," Education Week , April 17, 2018, http://blogs.edweek.org/edweek/DigitalEducation/2018/04/pearson_growth_mindset_software.html?cmp=SOC-SHR-FB .

116. Rolfe Winkler and Laura Stevens, "New Parents Complain Amazon Baby-Registry Ads Are Deceptive," Wall Street Journal , November 28, 2018, <https://www.wsj.com/articles/new-parents-complain-amazon-ads-are-deceptive-1543417201> .

117. Julia Angwin, Jeff Larson, and Laura Kirchner, "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks," ProPublica , May 23, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> ;

Jeffrey Dastin, "Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women," Reuters , October 10, 2018, <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G> .

۱۱۸. مقاله و ویدئوی زیر را ببینید:

Shira Mitchell, "Mirror Mirror: Reflections on Quantitative Fairness," 2018, <https://shiraamitchell.github.io/fairness/> ;

Arvind Narayanan, Tutorial: 21 Fairness Definitions and Their Politics , accessed November 18, 2018, <https://www.youtube.com/watch?v=jlXluYdnyk> .

119. Solon Barocas, Kate Crawford, Aaron Shapiro, Hanna Wallach, 2017,

“The Problem with Bias: Allocative Versus Representational Harms in Machine Learning” (Measure, Model, Mix: Computer as Instrument: 9th Annual SIGCIS Conference, Philadelphia, 2017), SIGCIS Conference, <http://meetings.sigcis.org/uploads/6/3/6/8/6368912/program.pdf>.

۱۲۰. ویدئوی زیر را ببینید:

Kate Crawford, “The Trouble with Bias” (Conference on Neural Information Processing Systems, Long Beach, CA, 2017), https://www.youtube.com/watch?v=fMym_BKWQzk .

۱۲۱. همان.

122. Sam Corbett-Davies and Sharad Goel, “The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning,” arXiv preprint [CS] arXiv:1808.00023, July 31, 2018.

123. Solon Barocas and Moritz Hardt, “Fairness in Machine Learning” (Conference on Neural Information Processing Systems, Long Beach, CA, 2017), <https://mrtz.org/nips17/#/> ; Narayanan, 21 Fairness Definitions and Their Politics .

124. Tracy Jan, “Redlining Was Banned 50 Years Ago. It’s Still Hurting Minorities Today,” Washington Post , March 28, 2018, <https://www.washingtonpost.com/news/wonk/wp/2018/03/28/redlining-was-banned-50-years-ago-its-still-hurting-minorities-today/> .

125. Cynthia Dwork, Nicole Immorlica, Adam T. Kalai, Mark DM Leiserson, “Decoupled classifiers for group-fair and efficient machine learning”, Conference on Fairness, Accountability and Transparency (January 21, 2018), 119-133, <http://proceedings.mlr.press/v81/dwork18a/dwork18a.pdf> .

126. Jon Kleinberg, "Inherent Trade-Offs in Algorithmic Fairness," in Abstracts of the 2018 ACM International Conference on Measurement and Modeling of Computer Systems , SIGMETRICS '18 (New York: ACM, 2018), 40–40, <https://doi.org/10.1145/3219617.3219634> ;

Alexandra Chouldechova, "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments," arXiv preprint [Cs, Stat] arXiv:1610.07524, October 24, 2016.

127. Frank Pasquale, "Odd Numbers," Real Life , August 20, 2018, <https://reallifemag.com/odd-numbers/> .

128. Sebastian Benthall, "Critical Reflections on FAT* 2018: A Historical Idealist Perspective," Dataactive , April 11, 2018, <https://data-activism.net/2018/04/critical-reflections-on-fat-2018-a-historical-idealist-perspective/> .

129. Mitchell, "Mirror Mirror: Reflections on Quantitative Fairness."

130. Ben Hutchinson and Margaret Mitchell, "50 Years of Test (Un)Fairness: Lessons for Machine Learning" arXiv preprint [CS] , arXiv:1811.10104, November 25, 2018.

131. Roel Dobbe, Sarah Dean, Thomas Gilbert, and Nitin Kohli, "A Broader View on Bias in Automated Decision-Making: Reflecting on Epistemology and Dynamics," arXiv preprint [Cs, Math, Stat] , arXiv:1807.00553, July 2, 2018.

132. Batya Friedman and Helen Nissenbaum, "Discerning Bias in Computer Systems," in INTERACT '93 and CHI '93 Conference Companion on Human Factors in Computing Systems , (New York: ACM, 1993), 141–142, <https://>

doi.org/10.1145/259964.260152 .

133. Batya Friedman and Helen Nissenbaum, "Bias in Computer Systems," ACM Transactions on Information Systems 14, no. 3 (July 1996): 330–347, <https://doi.org/10.1145/230538.230561> .

134. Kate Crawford and Ryan Calo, "There Is a Blind Spot in AI Research," Nature 538, no. 7625 (October 20, 2016): 311, <https://doi.org/10.1038/538311a> .

۱۲۵. ویدیوی زیر را ببینید:

Meredith Whittaker, "Data Genesis: AI's Primordial Soup" (Eyeo Festival, Minneapolis, 2018), <https://vimeo.com/287094149> .

136. Timnit Gebru et al., "Datasheets for Datasets," arXiv preprint [Cs], arXiv:1803.09010, March 23, 2018.

137. Margaret Mitchell et al., "Model Cards for Model Reporting," arXiv preprint [Cs] arXiv:1810.03993, October 5, 2018.

138. Jevan A. Hutson, Jessie G. Taft, Solon Barocas and Karen Levy,, "Debiasing Desire: Addressing Bias & Discrimination on Intimate Platforms," Proceedings of the ACM On Human-Computer Interaction (CSCW) 2 (November 2018): 1–18, <https://doi.org/10.1145/3274342> .

139. Animesh Singh and Michael Hind, "AI Fairness 360: Attacking Bias from All Angles!," IBM Developer , September 19, 2018, <https://developer.ibm.com/blogs/2018/09/19/ai-fairness-360-attacking-bias-from-all-angles/> .

140. James Wexler, "The What-If Tool: Code-Free Probing of Machine Learning Models," Google AI Blog , September 11, 2018, <http://ai.googleblog.com/2018/09/the-what-if-tool-code-free-probing-of.html> ;

James Wexler, "Facets: An Open Source Visualization Tool for Machine Learning Training Data," Google AI Blog, June 17, 2017, <http://ai.googleblog.com/2017/07/facets-open-source-visualization-tool.html>.

141. Alekh Agarwal et al, "A Reductions Approach to Fair Classification," arXiv preprints [CS], arXiv:1803.02453, March 6, 2018.

142. Dave Gershgorn, "Facebook Says It Has a Tool to Detect Bias in Its Artificial Intelligence," Quartz, May 3, 2018, <https://qz.com/1268520/facebook-says-it-has-a-tool-to-detect-bias-in-its-artificial-intelligence/>.

143. Jeremy Kahn, "Accenture Unveils Tool to Help Companies Insure Their AI Is Fair," Bloomberg, June 13, 2018, <https://www.bloomberg.com/news/articles/2018-06-13/accenture-unveils-tool-to-help-companies-insure-their-ai-is-fair>.

144. "The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)," IEEE Standards Association, accessed November 19, 2018, <https://standards.ieee.org/industry-connections/ecpais.html>.

145. "TRUSTe Settles FTC Charges It Deceived Consumers Through Its Privacy Seal Program," Federal Trade Commission, November 17, 2014, <https://www.ftc.gov/news-events/press-releases/2014/11/truste-settles-ftc-charges-it-deceived-consumers-through-its>.

146. "AI Now 2017 Report."

147. Scott Shane, Cade Metz, and Daisuke Wakabayashi, "How a Pentagon Contract Became an Identity Crisis for Google," The New York Times, November 2, 2018, <https://www.nytimes.com/2018/05/30/technology/google-project-maven-pentagon.html>.

148. Sundar Pichai, "AI at Google: Our Principles," The Keyword , June 7, 2018, <https://www.blog.google/technology/ai/ai-principles/> .
149. Alan Boyle, "Microsoft Is Turning down Some Sales over AI Ethics, Top Researcher Eric Horvitz Says," GeekWire , April 10, 2018, <https://www.geek-wire.com/2018/microsoft-cutting-off-sales-ai-ethics-top-researcher-eric-horvitz-says/> ; Jordan Novet, "Facebook Forms Ethics Team to Prevent Bias in AI Software," CNBC , May 3, 2018, <https://www.cnbc.com/2018/05/03/facebook-ethics-team-prevents-bias-in-ai-software.html> ; "Axon AI and Policing Technology Ethics Board," Axon , accessed November 19, 2018, <https://www.axon.com/info/ai-ethics>
150. "Ethically Aligned Design Version 2: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems" (New York: IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2018), <https://ethicsinaction.ieee.org/> ;
- Brent Hecht, Lauren Wilcox, Jeffrey P. Bigham, Johannes Schöning, Ehsan Hoque, Jason Ernst, Yonatan Bisk, Luigi De Russis, Lana Yarosh, Bushra Anjum, Danish Contractor, and Cathy Wu. "It's Time to Do Something: Mitigating the Negative Impacts of Computing Through a Change to the Peer Review Process," ACM Future of Computing Blog , March 29, 2018, <https://acm-fca.org/2018/03/29/negativeimpacts/> .
151. Alina Tugend, "Colleges Grapple With Teaching the Technology and Ethics of A.I.," The New York Times , November 3, 2018, <https://www.ny-times.com/2018/11/02/education/learning/colleges-grapple-with-teaching-ai.html> .

152. “La Déclaration de Montréal pour un Développement Responsable de l’Intelligence Artificielle,” accessed November 19, 2018, <https://www.declarationmontreal-iaresponsable.com/> .

153. “Microsoft Professional Program for Artificial Intelligence,” Microsoft , accessed November 29, 2018, <https://academy.microsoft.com/en-us/professional-program/tracks/artificial-intelligence/> .

154. Lucy Suchman, “Corporate Accountability,” Robot Futures , June 11, 2018, <https://robotfutures.wordpress.com/2018/06/10/corporate-accountability/> .

155. Ben Wagner, “Ethics as Escape From Regulation: From Ethics-Washing to Ethics-Shopping?” In Being Profiling. Cogitas Ergo Sum , ed. Mireille Hildebrandt. (Amsterdam: Amsterdam University Press, forthcoming 2019), https://www.privacylab.at/wp-content/uploads/2018/07/Ben_Wagner_Ethics-as-an-Escape-from-Regulation_2018_BW9.pdf .

156. Sandy Parakilas, “We Can’t Trust Facebook to Regulate Itself,” The New York Times , January 20, 2018. <https://www.nytimes.com/2017/11/19/opinion/facebook-regulation-incentive.html> .

157. Dan M. Greene, Anna Laura Hoffman, and Luke Stark, “Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning” (Hawaii International Conference on System Sciences, Maui, forthcoming 2019), <http://dmgreene.net/wp-content/uploads/2018/09/Greene-Hoffman-Stark-Better-Nicer-Clearer-Fairer-HICSS-Final-Submission.pdf>.

159. Andrew McNamara, Justin Smith, and Emerson Murphy-Hill, "Does ACM's Code of Ethics Change Ethical Decision Making in Software Development?," in Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering , ESEC/FSE 2018 (New York: ACM, 2018), 729, <https://doi.org/10.1145/3236024.3264833> .
160. Pichai, "AI at Google: Our Principles."
161. Ryan Gallagher, "Google Plans to Launch Censored Search Engine in China, Leaked Documents Reveal," The Intercept , August 1, 2018, <https://theintercept.com/2018/08/01/google-china-search-engine-censorship/> .
162. Ronald Deibert, Rebecca Mackinnon, Xiao Qiang, and Lokman Tsui, "Open Letter to Google on Reported Plans to Launch a Censored Search Engine in China," Amnesty International , August 28, 2018, <https://www.amnesty.org/en/documents/document/?indexNumber=ASA17%2f9001%2f2018&language=en> ;
- Kate Conger and Daisuke Wakabayashi, "Google Employees Protest Secret Work on Censored Search Engine for China," The New York Times , September 10, 2018, <https://www.nytimes.com/2018/08/16/technology/google-employees-protest-search-censored-china.html>
163. Mark Bergen, "Google CEO Tells Staff China Plans Are 'Exploratory' After Backlash," Bloomberg August 17, 2018, <https://www.bloomberg.com/news/articles/2018-08-17/google-ceo-is-said-to-tell-staff-china-plans-are-exploratory> .
164. Matt Phillips, "Facebook's Stock Plunge Shatters Faith in Tech Compa

nies' Invincibility," The New York Times , October 17, 2018, <https://www.nytimes.com/2018/07/26/business/facebook-stock-earnings-call.html> ; Rupert Neate, "Twitter Stock Plunges 20% in Wake of 1m User Decline," The Guardian , July 27, 2018, <https://www.theguardian.com/technology/2018/jul/27/twitter-share-price-tumbles-after-it-loses-1m-users-in-three-months> .

165. For a more general description of justice as fairness, see: John Rawls, Justice as Fairness: A Restatement, ed. Erin I. Kelly (Cambridge, MA: Harvard University Press, 2001).

166. Brent Hecht et al, "It's Time to Do Something," <https://acm-fca.org/2018/03/29/negativeimpacts/>

167. Ben Green, "Fair' Risk Assessments: A Precarious Approach for Criminal Justice Reform" (5th Workshop on Fairness, Accountability, and Transparency in Machine Learning, Stockholm, 2018), <https://scholar.harvard.edu/files/bgreen/files/18-fatml.pdf> .

168. Ben Green, "Putting the J(ustice) in FAT," Berkman Klein Center , February 26, 2018, <https://medium.com/berkman-klein-center/putting-the-justice-in-fat-28da2b8eae6d> .

169. Kate Crawford, "Just An Engineer: The Politics of AI," (The Royal Society, London, July 2018), <https://www.youtube.com/watch?v=HPopJb5aDyA>.

170. Paul N. Edwards, Geoffrey C. Bowker, Steven J. Jackson, and Robin Williams. "Introduction: An Agenda for Infrastructure Studies," Journal of the Association for Information Systems 10, no. 5 (May 28, 2009): 364–74, <https://aisel.laisnet.org/jais/vol10/iss5/6> .

171. Susan Leigh Star, "The Ethnography of Infrastructure," *American Behavioral Scientist* 43, no. 3 (November 1, 1999): 377-91, <https://doi.org/10.1177/00027649921955326> ;
- Paul N. Edwards, Thomas J. Misa, and Philip Brey, "Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems," in *Modernity and Technology* (Cambridge, MA: MIT Press, 2003), 185-225;
- Shannon Mattern, "The Big Data of Ice, Rocks, Soils, and Sediments", *Places*, November 2017, <https://placesjournal.org/article/the-big-data-of-ice-rocks-soils-and-sediments/> ;
- Jean-Christophe Plantin, Carl Lagoze, Paul N. Edwards and Christian Sandvig. "Infrastructure studies meet platform studies in the age of Google and Facebook," *New Media & Society*, 20, no. 1 (2018): 293-310, <https://doi.org/10.1177/1461444816661553> .
172. Kate Crawford and Vladan Joler, "Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources," (AI Now Institute and Share Lab, September 7, 2018), <https://anatomyof.ai> .
173. Steven J. Jackson, "Rethinking Repair", in *Media Technologies: Essays on Communication, Materiality, and Society*, Tarleton Gillespie, Pablo J. Boczkowski and Kirsten A. Foot, eds. (Cambridge: MIT Press, 2014), 221-239.
174. Nick Seaver, "What Should An Anthropology of Algorithms Do?" *Cultural Anthropology*, 33 , No. 3 (2018): 375-385, <https://doi.org/10.14506/ca33.3.04> .

175. Li Yuan, "How Cheap Labor Drives China's A.I. Ambitions," The New York Times, November 27, 2018, <https://www.nytimes.com/2018/11/25/business/china-artificial-intelligence-labeling.html> ;

Mary L. Gray and Siddharth Suri, "The Humans Working Behind the AI Curtain," Harvard Business Review, January 9, 2017, <https://hbr.org/2017/01/the-humans-working-behind-the-ai-curtain> .

176. Lilly Irani, "The Hidden Faces of Automation," XRDS 23, no. 2 (December 2016): 34–37, <https://doi.org/10.1145/3014390> .

۱۷۷. همان.

178. Janine Berg et al., "Digital Labour Platforms and the Future of Work: Towards Decent Work in the Online World," (Geneva: International Labour Organization, September 20, 2018), http://www.ilo.org/global/publications/books/WCMS_645337/lang-en/index.htm .

179. Sarah T. Roberts, "Commercial Content Moderation: Digital Laborers' Dirty Work," in *The Intersectional Internet: Race, Sex, Class and Culture Online*, ed. Safiya Umoja Noble and Brendesha M. Tynes (New York: Peter Lang, 2016), 147–159.

180. M. S. Silberman et al., "Responsible Research with Crowds: Pay Crowdworkers at Least Minimum Wage," *Communications of the ACM* 61, no. 3 (February 2018): 39–41, <https://doi.org/10.1145/3180492> .

181. Arlene Kaplan Daniels, "Invisible Work," *Social Problems* 34, no. 5 (1987): 403–15, <https://doi.org/10.2307/800538> ;

Arlie Russell Hochschild, *The Managed Heart: Commercialization of Human Feeling* (Berkeley: University of California Press, 2012).

182. Astra Taylor, "The Automation Charade," Logic Magazine , October 2, 2018, <https://logicmag.io/05-the-automation-charade/> .

183. Jana Kasperkevic, "Ex-McDonald's CEO Suggests Replacing Employees with Robots amid Protests," The Guardian , May 25, 2016, <https://www.theguardian.com/us-news/2016/may/25/former-mcdonalds-ceo-threatens-replace-employees-robots> ;

Ed Rensi, "Thanks To 'Fight For \$15' Minimum Wage, McDonald's Unveils Job-Replacing Self-Service Kiosks Nationwide," Forbes , November 29, 2016, <https://www.forbes.com/sites/realspin/2016/11/29/thanks-to-fight-for-15-minimum-wage-mcdonalds-unveils-job-replacing-self-service-kiosks-nationwide/#5defa3eb4fbc>

184. Uliana Pavlova, "McDonald's Kiosks Mean More Staff at Chicago Flagship, Not Fewer," Bloomberg , August 8, 2018, <https://www.bloomberg.com/news/articles/2018-08-08/mcdonald-s-kiosks-mean-more-staff-at-chicago-flagship-not-fewer> .

185. Alex Rosenblat, Uberland: How Algorithms Are Rewriting the Rules of Work (Berkeley: University of California Press, 2018).

186. Ben Chapman, "Uber Eats and Deliveroo Riders Are Going on Strike This Week," The Independent , October 3, 2018, <https://www.independent.co.uk/news/business/news/uber-eats-deliveroo-strike-mcdonalds-wetherspoons-mcstrike-industrial-action-a8567286.html> ;

Carolyn Said, "Uber, Lyft Drivers Fear Getting Booted from Work," San Francisco Chronicle , October 14, 2018, <https://www.sfchronicle.com/business/article/Uber-Lyft-drivers-fear-getting-booted-from-work-13304052.php> .

187. Kate Conger, "Google and Facebook's Security Guards Are Fighting to Earn a Living Wage," Gizmodo , July 27, 2018, <https://gizmodo.com/google-and-facebooks-security-guards-are-fighting-to-earn-a-living-wage-1826104897> .

188. Mark Bergen and Josh Eidelson, "Inside Google's Shadow Workforce," Bloomberg , July 25, 2018, <https://www.bloomberg.com/news/articles/2018-07-25/inside-google-s-shadow-workforce> .

189. Thuy Ong, "Amazon Patents Wristbands That Track Warehouse Employees' Hands in Real Time," The Verge , February 1, 2018, <https://www.theverge.com/2018/2/1/16958918/amazon-patents-trackable-wristband-warehouse-employees> ;

Hayley Peterson, "Missing Wages, Grueling Shifts, and Bottles of Urine: The Disturbing Accounts of Amazon Delivery Drivers May Reveal the True Human Cost of 'Free' Shipping," Business Insider , September 11, 2018, <https://www.businessinsider.com/amazon-delivery-drivers-reveal-claims-of-disturbing-work-conditions-2018-8> .

190. Catie Keck, "Amazon Workers Across Europe Protest Black Friday, Citing Grueling Work Conditions," Gizmodo , November 23, 2018, <https://gizmodo.com/amazon-workers-across-europe-protest-black-friday-citi-1830622250> .

191. "AI Now 2017 Report."

192. L. Rafael Reif, "Letter to the MIT Community Regarding the MIT Stephen A. Schwarzman College of Computing," MIT News , October 15, 2018, <https://news.mit.edu/2018/letter-mit-community-regarding-mit-stephen-schwarzman-college-computing> ;

Steve Lohr, "M.I.T. Plans College for Artificial Intelligence, Backed by \$1 Billion," The New York Times , October 16, 2018, <https://www.nytimes.com/2018/10/15/technology/mit-college-artificial-intelligence.html> .

193. Lev Manovich, "Can We Think Without Categories?" Digital Culture & Society, 4, no. 1 (2018): 17-28.

194. AI Now Institute, "Gender, Race and Power: Outlining a New AI Research Agenda," Medium, November 15, 2018. <https://medium.com/@AINowInstitute/gender-race-and-power-5da81dc14b1b> .

195. Kristian Lum, "Statistics, we have a problem," Medium, December 13, 2017, <https://medium.com/@kristianlum/statistics-we-have-a-problem-304638dc5de5> .

196. Jennings Brown, " 'NIPS' AI Conference Changes Name Following Protests Over Gross Acronym," Gizmodo, November 19, 2018. <https://gizmodo.com/nips-ai-conference-changes-name-following-protests-ov-1830548185> ;

Corinna Cortes et al., "From the Board: Changing Our Acronym," NeurIPS November 16, 2018, <https://nips.cc/Conferences/2018/News>.

197. "NIPS Name Change," NeurIPS , October 17, 2018, <https://nips.cc/Conferences/2018/News>.

۱۹۸. مردیت وایت‌تاکر، یکی از موسسان موسسه‌ی AI Now یکی از سازمان‌دهندگان اعتصاب گوگل بود. همچنین این مقاله را نیز ببینید:

Caroline O'Donovan and Ryan Mac, "Google Engineers Are Organizing A Walkout To Protest The Company's Protection Of An Alleged Sexual Harasser", BuzzFeed, Oct. 30, 2018, <https://www.buzzfeednews.com/article/>

- carolineodonovan/googles-female-engineers-walkout-sexual-harassment .
199. Davey Alba and Caroline O'Donovan, "Square, Airbnb, And eBay Just Said They Would End Forced Arbitration For Sexual Harassment Claims," BuzzFeed, Nov. 15, 2018, <https://www.buzzfeednews.com/article/daveyalba/tech-companies-end-forced-arbitration-airbnb-ebay>.
200. Sharon Florentine, "Alphabet Dismisses Action on Diversity and Inclusion," CIO , June 15, 2018, <https://www.cio.com/article/3281867/it-industry/alphabet-dismisses-action-on-diversity-and-inclusion.html> ; Sara Ashley O'Brien, "Apple's Board Calls Diversity Proposal 'Unduly Burdensome and Not Necessary,'" CNNMoney , January 15, 2016, <https://money.cnn.com/2016/01/15/technology/apple-diversity/index.html> .
201. Tom Simonite, "AI Is the Future - But Where Are the Women?," Wired , August 17, 2018, <https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/> .
202. Jackie Snow, "'We're in a Diversity Crisis:' Cofounder of Black in AI on What's Poisoning Algorithms in Our Lives," MIT Technology Review , February 14, 2018, <https://www.technologyreview.com/s/610192/were-in-a-diversity-crisis-black-in-ais-founder-on-whats-poisoning-the-algorithms-in-our/> .
203. Paula A. Johnson, Sheila E. Widnall, and Frazier F. Benya, eds., Sexual Harassment of Women: Climate, Culture, and Consequences in Academic Sciences, Engineering, and Medicine (Washington, DC: The National Academies Press, 2018), <https://www.nap.edu/catalog/24994/sexual-harassment-of-women-climate-culture-and-consequences-in-academic> .

204. Caroline Clark Hayes, "Computer Science: The Incredible Shrinking Woman," in *Gender Codes: Why Women Are Leaving Computing* , ed. Thomas J Misa (Hoboken: Wiley, 50-25 ,(2010).
205. Dale Strok, "Women in AI," *IEEE Expert: Intelligent Systems and Their Applications* 7, no. 4 (August 22-7 :(1992, <https://doi.org/64.153460/10.1109> .
206. "Barriers to Equality in Academia: Women in Computer Science at M.I.T." (Cambridge, MA: The Laboratory of Computer Science and the Artificial Intelligence Laboratory at M.I.T, February 1983), <https://homes.cs.washington.edu/~lazowska/mit/Images/title.pdf> .
207. Langdon Winner, "Do Artifacts Have Politics?," *Daedalus* 109, no. 1 36-121 :(1980).
208. Stephen Merity, "Bias is not just in our datasets, it's in our conferences and community," *Smerity.com* , December 2017 ,11, https://smerity.com/articles/2017/bias_not_just_in_datasets.html .
209. Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women."
210. Safiya U. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, (New York: NYU Press, 2018);
211. Clemens Apprich, Wendy Hui Kyong Chun, Florian Cramer, and Hito Steyerl, *Pattern Discrimination* (Minneapolis: University of Minnesota Press, forthcoming 2019).
212. "Litigating Algorithms."
213. Reisman et al., "Algorithmic Impact Assessments."

214. "CPSR History," Computer Professionals for Social Responsibility , June 2005 ,1, <http://cpsr.org/about/history/> .

215. "The Never Again Pledge," accessed November 2018 ,29, <http://neveragain.tech/>

216. Scott Shane and Daisuke Wakabayashi, "'The Business of War:' Google Employees Protest Work for the Pentagon," The New York Times , November 2018 ,2, <https://www.nytimes.com/04/04/2018/technology/google-letter-ceo-pentagon-project.html> .

217. Lucy Suchman, Lilly Irani, and Peter Asaro, "Google's March to the Business of War Must Be Stopped," The Guardian , May 2018 ,16, <https://www.theguardian.com/commentisfree/2018/may/16/google-business-war-project-maven> ;

Mary Wareham, "Letter to Sergey Brin and Sundar Pichai," March 2018 ,13, https://www.stopkillerrobots.org/wp-content/uploads/04/2018/KRC_LtrGoogle_12March2018.pdf ;

Daisuke Wakabayashi and Scott Shane, "Google Will Not Renew Pentagon Contract That Upset Employees," The New York Times , November 2018 ,2, <https://www.nytimes.com/01/06/2018/technology/google-pentagon-project-maven.html> .

218. Shaban, "Amazon Employees Demand Company Cut Ties with ICE;" "Who's Behind Ice?;"

Caroline O'Donovan, "Employees Of Another Major Tech Company Are Petitioning Government Contracts," BuzzFeed News , June 2018 ,26, <https://www.buzzfeednews.com/article/carolineodonovan/salesforce-employees->

push-back-against

-company-contract ;

Sheera Frenkel, "Microsoft Employees Question C.E.O. Over Company's Contract With ICE," The New York Times , July 2018 ,27, <https://www.nytimes.com/2018/07/27/technology/microsoft-ice-immigration.html>;

Peter Kotecki, "Burning Man Protesters Raise Awareness of Palantir, AmazonICE Ties," Business Insider , August 2018 ,31, <https://www.businessinsider.com/burning-man-protestors-palantir-amazon-ice8-2018-219>.

219. Greg Sandoval, "Over 100 Amazon Employees Sign Letter Asking Jeff Bezos to Stop Selling Facial-Recognition Software to Police," Business Insider , June 2018 ,22, <https://www.businessinsider.com/over-100-amazon-employees-sign-letter-jeff-bezos-stop-selling-facial-recognition-software-police6-2018-> ;

Kade Crockford, "Over 150,000 People Tell Amazon: Stop Selling Facial Recognition Tech to Police," American Civil Liberties Union , June 2018 ,18, [https://www.aclu.org/blog/privacy-technology/surveillance-](https://www.aclu.org/blog/privacy-technology/surveillance-technologies/over-150000-people-tell-amazon-stop-selling-facial)

[technologies/over-150000-people-tell-amazon-stop-selling-facial](https://www.aclu.org/blog/privacy-technology/surveillance-technologies/over-150000-people-tell-amazon-stop-selling-facial) ;

Matt Cagle and Nicole Ozer, "Amazon Teams Up With Government to Deploy Dangerous New Facial Recognition Technology," ACLU Free Future , May 2018 ,22, <https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazon-teams-government-deploy-dangerous-new> .

220. Ryan Gallagher "Google China Prototype Links Searches to Phone Numbers" The Intercept , September 2018 ,14, <https://theintercept.com/14/09/2018/google-china-prototype-links-searches-to-phone->

numbers/.

221. Bryan Menegus, "Here's the Letter 1,400 Google Workers Sent Leadership in Protest of Censored Search Engine for China," Gizmodo , August 2018 ,16, <https://gizmodo.com/heres-the-letter-400-1-google-workers-sent-leadership-i1828393599-> ;

Google Employees Against Dragonfly, "We Are Google Employees. Google Must Drop Dragonfly," Medium, November 2018 ,27, <https://medium.com/@googlersagainstdragonfly/we-are-google-employees-google-must-drop-dragonfly4-c8a30c5e5eb> ;

Google Employees Against Dragonfly, "We Are Google Employees. Google Must Drop Dragonfly," Medium, November 2018 ,27, <https://medium.com/@googlersagainstdragonfly/we-are-google-employees-google-must-drop-dragonfly4-c8a30c5e5eb> ;

"Google Must Not Capitulate to China's Censorship Demands," Amnesty International , November 2018 ,28, <https://www.amnesty.org/en/latest/news/11/2018/google-must-not-capitulate-to-chinas-censorship-demands/> .

۲۲۲. مردیت وایت تاکر، یکی از موسسان موسسه‌ی AI Now، یکی از هشت سازمان‌دهنده‌ی اصلی اعتصاب گوگل بود

223. Google Walkout for Real Change, "Google Employees and Contractors Participate in 'Global Walkout for Real Change,'" Medium , November 2018,2, <https://medium.com/@GoogleWalkout/google-employees-and-contractors-participate-in-global-walkout-for-real-change389-c65517843> .

224. Richard Waters, "Google Ends Forced Arbitration for Sexual Harassment

Claims,” Financial Times , November 2018 ,8, <https://www.ft.com/content/ce3c11ec-e37e11-e-8a6e792428919-5cee> .

225. Kate Gibson, “Tech Signals End of Forced Arbitration for Sexual Misconduct Claims,” CBS MoneyWatch , November 2018 ,16, <https://www.cbsnews.com/news/tech-signals-end-of-forced-arbitration-for-sexual-misconduct-claims/> .

226. “AI Now 2017 Report.”

227. Google Walkout for Real Change, “#GoogleWalkout Update: Collective Action Works, but We Need to Keep Working,” Medium , November 2018 ,8, <https://medium.com/@GoogleWalkout/googlewalkout-update-collective-action-works-but-we-need-to-keep-working-b17f673ad513> ;

Noam Scheiber, “Google Workers Reject Silicon Valley Individualism in Walkout,” The New York Times , November 2018 ,7, <https://www.nytimes.com/06/11/2018/business/google-employee-walkout-labor.html>.

۲۲۸. برای نمونه، پروژه‌ی فوق سری گوگل با نام Dragonfly که برای سانسور نتایج جستجو در چین طراحی شده بود و شماره تلفن‌های شهروندان چینی را به گزارش جستجوی آن‌ها ارتباط می‌داد. هم‌چنین ببینید:

“We Are Google Employees. Google Must Drop Dragonfly,” Medium , November 2018 ,27, <https://medium.com/@googlersagainstdragonfly/we-are-google-employees-google-must-drop-dragonfly4-c8a30c5e5eb> .

229. Cade Metz, “Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent,” The New York Times , October 2017 ,22, <https://www.nytimes.com/22/10/2017/technology/artificial-intelligence-experts-salaries.html> .

230. “Recommendations” in “AI Now 2016 Report,” (New York: AI Now .

Institute, 2016), https://ainowinstitute.org/AI_Now_2016_Report.pdf

231. Crawford and Joler, "Anatomy of an AI System," <https://anatomyof.ai> .

۲۳۲. به عنوان یک نمونه از میان چندین نمونه نگاه کنید به:

Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women,"

<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G> .

233. Lecher, "A Healthcare Algorithm Started Cutting Care, and No One Knew Why."

234. Sonia Katyal, "Private Accountability in the Age of the Algorithm," *UCLA Law Review* 66 (forthcoming 2019), <https://www.uclalawreview.org/private-accountability-age-algorithm/> .



مرکز ملی فضای مجازی
پروژه نگاه فضای مجازی

csri.majazi.ir