رفتار واقعی هوش‌مصنوعی برای دانشمندان داده

انتشارات محمد رحیمی

۱۳ ژوئن ۲۰۲۳

فهرست مطالب

تشکر نامه

مایلم از مشارکت کنندگان زیر برای کمک‌های عالی و متنوعشان در این کتاب تشکر کنیم: پیتر هرشوک (اخلاق بودایی)، جان هکر رایت (اخلاق فضیلت)، ساموئل جی لوین و دانیل سینکلر (اخلاق یهودی)، کالین مارشال (اخلاق دئونتولوژیک)، جوی میلر و آندریا سالیوان کلارک (اخلاق بومی) و جان مورانگی (اخلاق آفریقایی). هدف ما ترسیم تصویری اخلاقی است که تا حد امکان منتوع و جذاب باشد. بدون کمک این عزیزان خردمند، توانستیم این کتاب را ایجاد کنیم!

توضحات لازم

اخلاق دئونتولوژیک یا اخلاق واجب‌گرایانه *(Deontological Ethics)* یک نظریهٔ اخلاقی است که بر ترکیب ویژگی‌های اخلاقی عمل و رعایت وظایف و اصول اخلاقی تمرکز دارد. این نظریه بر ایدهٔ آن تأکید می‌کند که برخی از اعمال به طور ذاتی صحیح یا نادرست هستند، بدون توجه به پیامدهای آن‌ها. در اخلاق واجب‌گرایانه، اخلاقیت یک عمل توسط نیت پشت آن و رعایت قوانین یا وظایف اخلاقی تعیین می‌شود. این نظریه بر مفاهیمی مانند عدالت، انصاف و احترام به حقوق فردی تأکید می‌کند. نظریه‌های اخلاقی واجب‌گرایانه شامل اخلاق کانتی و نظریهٔ فرمان الهی می‌شوند.

اخلاق فضیلت‌گرا *(Virtue Ethics)* اخلاق فضیلت‌گرا یک نظریهٔ اخلاقی است که بر توسعهٔ صفات فضیلت‌آمیز و اخلاقی در افراد تأکید دارد. این نظریه بر ایدهٔ اینکه بودن یک فرد با اخلاق خوب برای رفتار اخلاقی اساسی است تمرکز دارد. اخلاق فضیلت‌گرا کمتر بر قوانین یا وظایف خاص تأکید می‌کند و به جای آن بر تقویت و تمرین ویژگی‌های فضیلت‌آمیز مانند راستگویی، مهربانی، شجاعت و حکمت تأکید می‌کند. تمرکز بر توسعه و عمل به این فضیلت‌ها به منظور اتخاذ تصمیمات اخلاقی صحیح و زندگی فضیلت.

اخلاق فایده‌گرا *(Utilitarian Ethics)* اخلاق فایده‌گرا، همچنین به عنوان نتیجه‌گرایی شناخته می‌شود، یک نظریهٔ اخلاقی است که اخلاقیت یک عمل را بر اساس پیامدها یا نتایج آن تعیین می‌کند. این نظریه به نگرشی می‌پردازد که عملی صحیح، عملی است که به حداکثر شادی کلی یا خوبی برای بیشترین تعداد افراد منجر می‌شود. اخلاق فایده‌گرا بر اولویت دادن به حداکثر خوبی برای بیشترین تعداد و کاهش رنج یا آسیب تمرکز می‌کند. این نظریه بر محاسبه و ارزیابی نتایج برای تعیین مسیر اخلاقی عمل تأکید دارد.

فهرست همکاران و مشارکت کنندگان

همکاران:

پیتر سینگر
مرکز دانشگاهی برای ارزش‌های انسانی دانشگاه پرینستون

ییپ فای تسه
مرکز دانشگاهی برای ارزش‌های انسانی دانشگاه پرینستون

سامول جی لوین
مرکز حقوقی تورو

کولین مارشال
گروه فلسفه دانشگاه واشینگتن

جویی میلر
گروه فلسفه دانشگاه وست چستر

جان مورونگی
گروه فلسفه دانشگاه وست چستر

مشارکت کنندگان:

جان هکر رایت
گروه فلسفه دانشگاه گوئلف

دنیل سینکلر
دانشکده حقوق دانشگاه فوردهام

پیتر دی هرشوک
مرکز شرقی-غربی

آندریا سالیوان کلارک
گروه فلسفه دانشگاه ویندزور

فصل ۱

مقدمه: ماشین‌های اخلاقی

ماشین‌های اخلاقی

این کتاب، برای دانشمندان داده و افراد علاقه‌مندی است که در برخی جهات آن دچار تردید شده‌اند، یا به طور خلاصه، راه درست و غلط استفاده از دیتا را به افراد نشان دهد و از استفاده‌ی غیراخلاقی آن جلوگیری کند.

به نظر می‌رسد که اهمیت علم داده در زندگی روزمره بسیار کم است! این امر باعث می‌شود که مردم عادی حتی در درک کردن این فناوری قدرتمند، کاملا ناتوان باشند؛ چه رسد به شکل‌دهی یا اداره‌ی آن! از طرفی خیلی از دانشمندانی که در این زمینه مشغول فعالیت هستند، نه زمان کافی برای کسب معلومات اخلاقی را دارند و نه منابع کافی برای اینکه ذهن خود را در گیر اهمیت اخلاق در این زمینه کنند. در صورتی که این فناوری می‌تواند تأثیرات اخلاقی زیادی را بر جامعه وارد کند. این کتاب در جهت کاهش این کمبودها نوشته‌شده است تا چراغ راهی باشد برای کسانی که به اخلاق در این حوزه اهمیت می‌دهند. برای این ایده که «دانشمندان باید اخلاق را بیاموزند»، تفکراتی مانند «شما نمی‌توانید چیزی در مورد اخلاق به کسی بیاموزید، مردم آن را میسازند» وجود دارد. البته قسمتی از آن درست است، مردم یک جامعه، اخلاق را می‌سازند.

امروزه ما با یک پدیده‌ی بسیار قدرتمند و البته بسیار پر خطر به نام «علم داده» روبرو هستیم؛ بنابراین، باید اخلاقیات و ضوابط این حوزه به صورت گسترده آموزش داده شود.

برای اینکه این کتاب تا حد امکان مفید و دوستانه واقع شود، سعی کردیم مطالب را با لحنی ساده بیان کنیم. در این کتاب، ۷ مثال واقعی که استفاده نادرست از علم داده را نشان می‌دهند، بیان می‌کنیم. ما همچنین با چندین دانشمند برجسته‌ی اخلاق تماس گرفتیم تا در هر مورد نظراتشان

را بپرسیم. همچنین برای ارائه‌ی طیف وسیع رفتارها و اخلاقیات انسانی، از سه دیدگاه غرب نسبت به اخلاق، فراتر رفتیم، سه رویکرد غرب عبارتند از: نتیجه‌گرایی(فایده‌گرایی)، دین شناسی و رفتار با تقوا. رویکردهایی که به طور اضافی بررسی کردیم: بودایی، یهودی، بومی و آفریقایی. هر یک از این رفتارها و رویکردها، می‌توانند زاویه‌ی دید متنوعی را ارائه کنند که ممکن است به آن فکر نکرده باشیم. هدف ما این است که درک کاملی از هر رویکرد ارائه دهیم، یک جعبه ابزار کامل برای روبرویی با چالش‌های آینده.

همانطور که می‌دانیم، یک مشکل خاص، می‌تواند با زوایای دید متفاوت (رویکردهای اخلاقی متفاوت که اشاره کردیم)، به طور مختلف تحلیل و بررسی شود. توانایی تحلیل معضل از دیدگاه‌های مختلف، لازمه‌ی «تفکر انتقادی» است. امیدوارم این کتاب دیدگاه گسترده‌ای را در اختیار خواننده قرار دهد!

علم داده چیست؟

علم داده، اصولی است برای استخراج دیتاهای غیر بدیهی و الگوها از مجموعه دیتاهای بزرگ. از طرفی هوش‌مصنوعی را می‌توان هر گونه پردازش اطلاعات که کارکرد روانی را انجام می‌دهد، اطلاق کرد. مثلا پیش‌بینی، تداعی کردن، تخیل کردن، برنامه‌ریزی و به طور کلی، هر پردازشی که تا کنون موجودات زنده قادر به انجام آن بودند.

ماشین لرنینگ *(ML)*، زیرمجموعه‌ای از علم داده و بخش رو به رشدی از این زمینه است. بر خلاف *GOFAI*، ماشین لرنینگ *(ML)* شکلی از هوش‌مصنوعی است که از رویکردهای آماری برای یافتن الگوها در دنیا (که بهم ریخته است) استفاده می‌کند. در خیلی جهات، *(ML)* پاسخی برای شکست‌های زودهنگام هوش‌مصنوعی سمبلیک *(GOFAI)* در بیرون از فضای آزمایشگاهی بود، به دلیل اینکه *GOFAI* قادر به پردازش پیچیدگی دنیای واقعی نبود.

الگوریتم‌های *ML*، با لایه‌های موازی اطلاعاتی که ارائه می‌شوند، آموزش داده می‌شوند و می‌توانند به روش‌هایی بیاموزند که نظارت نشده و نسبتا مرموز هستند! خیلی شبیه عملکرد مغز ما (از یک روش یا تابع استفاده می‌کند و آن را بر روی مجموعه‌ای از دیتا اعمال می‌کند). مانند تابعی که ایمیل‌های به درد نخور (هرزنامه) را شناسایی می‌کند؛ این تابع بر روی مجموعه‌ای از ایمیل‌ها اعمال می‌شود یا مشخص شود که کدام ایمیل به درد نخور است.

ویژگی‌های هرزنامه‌ها و غیر هرزنامه‌ها قبلا توسط انسان‌هایی که تفاوت را می‌دانند، برای الگوریتم برچسب گذاری می‌شود. از طرف دیگر، یادگیری بدون نظارت، شامل هیچ برچسب‌زنی‌ای نمی‌شود و ما نمی‌دانیم که دنبال چه فاکتورهایی هستیم! این الگوریتم در ابتدا مجموعه‌ای دیتا دریافت می‌کند و بررسی می‌کند که کدام ویژگی‌ها مرتبط هستند. برای مثال، یک الگوریتم بدون نظارت، ممکن است که به تصاویر متعددی از سگ نگاه کند و تعیین کند چه ویژگی‌هایی جوهره‌ی «سگ بودن» را به وجود می‌آورد. زمانی هم که با یک تصویر جدید روبرو می‌شود، می‌تواند تصمیم بگیرد که سگ است یا خیر.

امروزه ابزارهای علم داده خیلی کاربرپسندتر شدند و تازه‌واردان و حتی افرادی که آموزش کمی دارند، به راحتی می‌توانند وارد این زمینه شوند. این به این معنی است که هیچ وقت انجام کار با نتایج بد در این زمینه، به این آسانی نبوده است! بنابراین عواقب پروژه‌هایی بد، باید توسط کسانی که وظیفه‌ی طراحی یا اجرای آن را دارند، پیش‌بینی شود.

همانطور که کِلِهِر *(Kelleher)* توضیح می‌دهد: «دیتا یا داده»، عنصری است که از دنیای واقعی انتزاع شده است و «اطلاعات»، داده‌هایی هستند که سازماندهی شدند تا مفید واقع شوند و «دانش» درک دقیق اطلاعاتی هست که داده‌ها به ما می‌دهند. اما با ارزش‌تر از همه، خرد است؛ که زمانی رخ می‌دهد که دانش را برای هدف خوب به کار ببریم. هدف ما این است که به خوانندگان خود کمک کنیم تا این خرد را توسعه دهند؛ که فکر می‌کنیم در قلب اخلاق علم داده قرار دارد.

بنابراین، اخلاق فقط بخشی از انجام خوب علم داده است. این یعنی، یک مشکل در دنیای واقعی، بسیار فراتر از جنبه‌های فنی آن است و البته اینکه یک سیستم چگونه قرار است زندگی افراد را تحت تاثیر قرار دهد نیز، اهمیت دارد!

موارد مطالعاتی

مورد۱ – مورد اول اخلاق تحقیق و روش علمی

مورد مطالعاتی اول، خواننده را با مفاهیمی مانند تکثیرپذیری، دقت و اعتبار آشنا می‌کند. بسیاری از این بحث‌ها بر اساس تلاش‌های اخیر در روانشناسی و همچنین علوم اجتماعی و پزشکی استوار شده است تا به واقعیتی که قسمت قابل توجهی از نتایج منتشر شده قابل تکثیر یا اعتبارسنجی نیستند، پاسخ دهند.

این مورد، سوءرفتار تحقیقاتی در آزمایشگاه غذایی کورنل *(Cornell Food and Brand Lab)* به وسیله‌ی برایان وانسینک *(Brian Wansink)* را شرح می‌دهد. مشخص شد که او برای نتایج از چندین روش غیر علمی و البته غیر اخلاقی استفاده کرده است. روش‌هایی از جمله: *cherry picking* (علنی کردن نتایج دلخواه)، روش *HAEKing* (فرضیه سازی پس از مشخص شدن نتایج تجربی) و روش *p-hacking* (دستکاری داده‌ها برای به دست‌آوردن یک نتیجه آماری معنی دار).

آقایان «سینگر» و «فای تسه» تفسیری بر رفتار «وانیسنک» از دیدگاه فایده‌گرایی ارائه می‌دهند. این دو بر اهمیت راست بودن نتایج علمی که دیگران به آن تکیه می‌کنند، تأکید دارند. کسانی که این وظیفه را به عهده گرفته‌اند تا شواهد علمی و تجربی‌ای را که دیگران از آن استفاده می‌کنند، ارائه دهند، درواقع بار سنگینی را بر دوش دارند. آن‌ها باید این کار با به بهترین نحو ممکن انجام دهند.

مورد۲ – مدل‌های ماشین در دادگاه

الگوریتم‌های *ML*، در چندین پرونده‌ی جنایی مورد استفاده قرار گرفت و البته اشکال اخلاقی را نیز به بار آورد. حتی بهترین مدل‌های تائید شده نیز در زمینه‌های اجتماعی مختلف نیز عملکر متفاوتی دارند. حتی مدل‌های عالی نیز که توسط انسان استفاده می‌شوند، می‌توانند عواقب ناخواسته‌ای را شامل: «تبعیض»، «تعصب» و «سوءاستفاده‌ی عمدی» به بار بیاورند.

استفاده از مدل *Markov chain Monte Carlo (MCMC)* منجر به معضل اخلاقی می‌شود. زیرا این مدل نمی‌تواند به طور کامل تکرار شود و درنتیجه شواهد تولید شده توسط آن نیز قابل تکرار نیست.

این مسائل از طریق یک مطالعه درباره الگوریتم‌های ترکیب *DNA* و نقش آن‌ها در محاکمه

نادرست «اورال نیکولاس هیلاری» *(Oral Nicholas Hillary)* در قتل یک پسر جوان در پاتسدام، نیویورک، توضیح داده می‌شوند. در این مورد، تحقیقات پلیس و محاکمه شامل **عوامل قویی از تعصب شخصی و نژادی** علیه هیلاری بود، که یک مربی محبوب و موفق با ریشه‌های آفریقایی-کارائیبی بود. این موجب شد تفسیر بسیار تعصب‌آمیزی از شواهد *DNA* برای متهم، ساخته شود. بازرسی شد که شواهد ناقص و غیرقابل اعتماد بوده و به درستی توسط دادگاه از پرونده حذف شده است، که در نتیجه به هیلاری تبرئه شد.

«ساموئل جِی لووین» از دیدگاه اخلاق یهودی، استفاده از مدل‌های یادگیری ماشینی در سامانه عدالت کیفری را مورد بررسی قرار می‌دهد و به بررسی تنش‌ها (در واقع تناقض) بین جبر و اراده آزاد در اخلاق یهودی پرداخته است.

ما همه عاملان اخلاقی هستیم و مسئول انتخاب‌های خودمان هستیم. اما اگر اعمال ما پیش از این تعیین شده باشند، آیا ما در حقیقت دیگران را برای تصمیماتی که نگرفته‌اند، قضاوت نمی‌کنیم؟ طوری دیگر بیان می‌کنم: اگر اعمال ما از پیش تعیین شده باشند، این به معنای آن است که پیش‌تر مشخص شده‌اند و نشان می‌دهد که ما کنترلی بر روی آن‌ها نداریم. جمله‌ای که شما ارائه داده‌اید، سؤالی را مطرح می‌کند که آیا عادلانه است که افراد را برای تصمیماتی که آن‌ها جز گزینه‌هایی که می‌توانستند انتخاب کنند، قضاوت کنیم؟ به عبارت دیگر، اگر انتخاب‌های یک شخص از پیش تعیین شده باشد و او اراده آزادی نداشته باشد، آیا منصفانه است که او را برای آن تصمیمات مسئول دانسته و قضاوت کنیم؟

این را می‌توان در استفاده گسترده از مدل‌های ماشینی برای پیش‌بینی میزان تکرار جرم و احکام و تصمیم‌گیری برای دریافت وثیقه برای متهمان مشاهده کرد. بسیاری از قوانین جزایی ما مبتنی بر ایده‌هایی در مورد اختیار است که از یهودیت گرفته شده و از طریق ادیان ابراهیمی منتشر شده است. تنش بین جبرگرایی و اراده آزاد در بسیاری از تصمیم گیری‌ها در سیستم عدالت کیفری ما نفوذ می‌کند. در همین حال، کسانی که به‌جای عدالت به دنبال قدرت هستند، می‌توانند از فناوری‌های علمی به گونه‌ای سوء استفاده کنند که از مرزهای اخلاقی خارج شود.

«کالین مارشال» اثبات‌های تولیدشده توسط مدل‌های ماشین را از دیدگاه اخلاق «دانتولوژیک» بررسی می‌کند، که به مدت طولانی نگران شناسایی و از بین بردن اشکالاتی از نوع ناعادلانه‌گرایی در تصمیم‌گیری اخلاقی بوده است که برخی افراد را نسبت به دیگران در مزیت قرار می‌دهد. تصمیمات اخلاقی باید آزمون همگانی را پشت سر بگذارند: اگر یک اقدام همگانی نباشد، به این معناست که همه انجام‌دهندگان در یک موقعیت مشابه، به احتمال زیاد یک انتخاب مشابه را انجام خواهند داد. این باید همه دانشمندان داده را تشویق کند که به تأثیرات مدل‌هایشان از دیدگاه افرادی که تحت تأثیر قرار می‌گیرند نگاهی بیندازند. این نیازمند این است که عاملان اخلاقی از دیدگاه خود، گاهی اوقات فایده‌گرایی، خارج شوند. اگر همه به این شیوه عمل کنند، سیستمی که ما می‌خواهیم در آن به طور غیرعادلانه توسط یک تحلیلگر جزئی یا یک الگوریتم تعصبی اتهام شویم، چگونه خواهد بود؟ سیستم‌هایی که شامل ناعدالتی غیرمشروع هستند، از نظر اخلاقی غیرمجاز هستند و باید استفاده نشوند.

مورد۳ - رسانه‌های ساختگی و خشونت سیاسی

در این مورد، دو مثال را بررسی می‌کنیم:

۱- مثال اول در مورد سخنرانی رئیس جمهور «گابن علی بونگو» *(Gabonese Ali Bongo)* در شب سال نو سال ۲۰۱۹ است. این ویدئو برای فرونشاندن ترس‌ها در مورد بیماری اخیر «بونگو» طراحی شده بود. اما زمانی که این ویدئو به عنوان یک deepfake *(deepfake)* شناخته شده، تنش‌های سیاسی را برانگیخت. سربازان گارد جمهوری خواه، کودتای نافرجامی را در لیبرویل راه انداختند؛ به این دلیل که «بونگو» دیگر در رأس کار نیست و نمی‌توان به حزب حاکم اعتماد کرد. کودتا با خشونت سرکوب شد و منجر به کشته شدن دو سرباز و بازداشت خیلی‌ها شد. ولی بعدا

مشخص شد که ویدئو کاملا واقعی بوده! در ویدئو به نظر می‌رسد که اثرات *(deepfake)* روی «بونگو» مشاهده می‌شود، ولی درواقع تاثیرات بعد از عمل باعث این موضوع شده بود! چشم‌های «بونگو» به طور غیر طبیعی در ویدئو مشاهده می‌شود که گمان *(deepfake)* را می‌رساند.

۲- مثال دوم به «فیک‌های کم عمق» *(shallow fakes)* علیه زنان در هند، به ویژه روزنامه نگاران و سیاستمدارانی که از حزب حاکم انتقاد می کنند، می پردازد. این رسانه‌های دستکاری شده شامل سایت‌های حراج جعلی هستند که مدعی «فروش» زنان هستند و آن‌ها را در شرایط تحقیرآمیز جنسی «پورنوگرافی» به تصویر می‌کشند. خبرنگاران زن در هند متوجه شده‌اند که پورنوگرافی تقلبی می‌تواند باعث ویرانی روحی‌شان شود و زندگی حرفه‌ای آن‌ها را به پایان برساند فقط به این دلیل که در دسترس عموم قرار گرفته‌اند! و نه به این دلیل که کسی باور کند این عکس‌ها و ویدئوها "واقعی" هستند.

در واقع، رسانه‌های مصنوعی در زمینه‌های بیشتری وارد زندگی ما می شوند. محتوایی که می‌خوانیم به‌طور فزاینده‌ای توسط هوش‌مصنوعی تولید می‌شود! پدیده‌ای که اخیراً حتی در مجلات علمی با داوری مشابه نیز دیده می‌شود. یعنی هوش‌مصنوعی مطلب علمی تولید می‌کند!

معنای زندگی در دنیای رسانه‌های دستکاری شده چیست؟ دنیایی که دیگر نمی توان حقیقت را به طور قابل اعتماد تعیین کرد و روی آن توافق کرد، یا حتی **در دنیایی که حقیقت دیگر اهمیتی ندارد؟** الگوریتم‌های هوش‌مصنوعی ممکن است به ما در شناسایی و حذف رسانه‌های مصنوعی کمک کنند، اما نمی‌توانند این مشکلات عمیق‌تر را برطرف کنند.

«سینگر» و «فای تسه» به مشکلات ناشی از رسانه‌های مصنوعی (شبکه‌های اجتماعی دستکاری‌شده) از دریچه اخلاق فایده گرایانه نگاه می کنند. آن‌ها بر اهمیت حقیقت تأکید می کنند که (حقیقت) به معنای استفاده از روش علمی و شواهد تجربی معتبر برای تصمیم گیری است. در غیر این صورت، فقط اعتمادمان را نسبت به نهادهای اصلی بیشتر از دست می‌دهیم. اعتماد و نهادهای قوی (اصلی) که رفاه افراد جامعه ما را ارتقا می دهند، با ارزش گذاری ما ساخته می شوند. پورنوگرافی چه به

صورت کم عمق و چه به صورت عمیق، به طور ویژه سلامتی را تخریب می‌کند و توسط اخلاق «فایده گرایانه» رد می‌شود. این (پورنوگرافی عمیق و کم عمق) در خدمت تقویت این عقیده که: "زنان وسیله‌ای برای سرگرمی دیگران هستند" است. این امر می‌تواند آسیب‌هایی از جمله: ارعاب (ایجاد رعب و وحشت برای زنان)، ظلم و ستم و وادار کردن زنان به انجام کارهایی خلاف قوانین زندگی عادی، داشته‌باشد.

«مورانگی» از دیدگاه اخلاق «اوبونتو» می‌نویسد. [به طور کلی، «اخلاق اوبونتو به عنوان مجموعه‌ای از ارزش‌ها تعریف می‌شود که از میان آن‌ها می‌توان به روابط متقابل، خیر مشترک، روابط مسالمت‌آمیز، تأکید بر کرامت انسانی، و ارزش زندگی انسانی و نیز اجماع، مدارا، و احترام متقابل اشاره کرد».] او تشویق می‌کند که داده‌شناسان نقش خود را به عنوان معماران جهانی که در آن زندگی می‌کنیم درک کنند و بر پیامدهای ساخت و ساز جهان خود تأمل کنند. او خاطرنشان می‌کند که علم داده در حال حاضر به عنوان یک تلاش خنثی و غیرسیاسی تدریس می‌شود، اما اثرات آن بر مردم آفریقا چیزی جز خنثی است. در اخلاق بومی آفریقایی، رفاه جامعه هم‌زمان، اخلاقی و سیاسی است و شامل هر دو «فرد» و «جامعه» می‌شود. هوش‌مصنوعی هر دو را با تجاوز به (فرهنگ) جوامع در آفریقا و سراسر جهان، و با تضعیف حس مشترک، نظم اجتماعی بومی و ارزش‌های اجتماعی که اساس زندگی‌های اصیل و اخلاقی را تشکیل می دهند، تضعیف می‌کند.

«میلر» و «سالیوان-کلارک» از اخلاق بومی برای بحث در مورد راه‌های مختلف استفاده از داده‌ها برای دستکاری، اجبار، کنترل، سرکوب و سلب حق رای گروه‌های خاص استفاده می‌کنند. افراد بومی اغلب هدف داده‌هایی با هدف مشخص، از این طریق بوده‌اند. این امر، منجر به رشد «جنبش حاکمیت داده‌های بومی» شده است که استقلال و کنترل بر داده‌هایشان و نحوه‌ی استفاده از آن‌ها را به خودشان برمی‌گرداند.

مورد۴ - بیومتریک و تشخصی چهره

در این مورد، به بررسی مسائل اخلاقی ناشی از استفاده از «بیومتریک» به عنوان نوعی کلیدشناسایی می‌پردازیم. در دنیایی که اطلاعات مانند گنج است، بسیار مهم است تا افراد ابزار معتبری برای احراز هویت و جلوگیری از دسترسی به دیتاهای خصوصی خود داشته‌باشند. کلیدهای (فرم‌های) بیومتریک، خیلی از این مشکلات را حل می‌کنند؛ آن‌ها کلیدهایی کامل و قابل اعتماد هستند و از سایر اَشکال شناسه‌ها (پسورد، شماره‌ی تلفن، ایمیل، الگوها و ...)، امن‌تر هستند. هرچند، زمانی که بیومتریک‌ها توسط مقامات برای نظارت و پزشکی قانونی استفاده شوند، می‌توانند مشکلات اخلاقی ایجاد کنند. در این مورد، ما به استفاده‌ی نادرست «پلیس سواره‌ی سلطنتی کانادا *(RCMP)*» اشاره می‌کنیم، که توسط کمیسیونر حریم خصوصی کانادا؛ به عنوان نقص قوانین حریم خصوصی شناخته‌شد.

افراد *(RCMP)* از سیستم تولید شده توسط شرکتی به نام *(Clearview AI)* برای جستجوی مظنونان و یافتن کودکانِ قربانی استثمار جنسی آنلاین، استفاده کرده بود. عکس‌های استفاده شده توسط *Clearview*، از شبکه‌های اجتماعی و سایر سایت‌های اینترنتی گرفته شده بود و اسماً «عمومی» بود. بنابراین *RCMP* استدلال کرد که استفاده آن‌ها از این عکس‌ها قوانین حریم خصوصی را نقض نمی‌کند. ولی با این حال باید توجه داشت که دیتای گرفته‌شده از سایت‌های عمومی نیز باید با رضایت کابر آن دیتا باشد؛ وگرنه منجر به نقض مشکلات حریم خصوصی می‌شود.

«داودزول» و «گلتز» اظهار داشتند که تصمیم هوش‌مصنوعی *Clearview* برای به کارگیری فناوری تشخیص چهره در جنگ اوکراین، هم قوانین بین‌المللی درگیری‌های مسلحانه و هم ارزش‌های بشردوستانه را نقض می‌کند. ما استدلال می‌کنیم که سیستم‌های داده‌ای که پتانسیل هدف قرار دادن غیرنظامیان یا نقض قوانین درگیری مسلحانه را دارند، غیراخلاقی هستند و استفاده از آن‌ها باید ممنوع شود.

«میلر» و «سالیوان کلارک» داده‌های بیومتریک را از منظر ارزش‌های بومی تجزیه و تحلیل می

کنند. داده‌های بیومتریک به خودی خود غیراخلاقی و مضر نیستند، اما ممکن است به روش‌هایی غیراخلاقی و آسیب رسان مورد استفاده قرار گیرند. این امر، به خوبی توسط افراد بومی درک می‌شود، زیرا اغلب تجربه کرده‌اند که از داده‌ها، علیه آن‌ها استفاده شده است. برای بررسی حریم خصوصی، نباید یک دیدگاه و زاویه‌ی دید را برای همه تعمیم داد، بلکه برای هر قوم یا گروه، باید ارزش‌ها و هنجارهای آن‌ها و حتی ارزش‌های بومی را نیز دخیل کرد. بررسی این موضوع به صورت تک بعدی، کار درستی نیست. می‌توانیم از ارزش‌های فطری برای این سؤال استفاده کنیم که آیا استفاده از داده‌های بیومتریک، تعادل و هماهنگی در روابط را مختل می‌کند؟ آیا استفاده از داده‌های بیومتریک در دادرسی کیفری، باعث ایجاد حس اعتماد بیش‌ازحد به گناهکاری متهم می‌شود؟ و باعث می‌شود که از فروتنی که یک ارزش فطری است، دلسرد شود؟ راه درست این است که برای استفاده از دیتای مردم، از آن‌ها اجازه گرفته شود و البته حاکمیت و خودمختاری برای داده‌هایشان را به مردم برگردانده شود.

مورد۵ - تعدیل محتوا: سخنرانی خطرناک و پاکسازی قومی در میانمار

این مطالعه موردی به بررسی استفاده از هوش‌مصنوعی در تعدیل محتوا می پردازد یعنی اینکه چگونه باید محتوا و دیتا در فضای مجازی کنترل شود. مثالی هم از سخنرانی ضد «روهینگیا» (شهری در میانمار) در *Facebook* که پاکسازی قومی آن‌ها توسط نیروی دولتی در میانمار را در پی داشت. این موضوع، بحث‌های زیادی در مورد اینکه چه محتوایی باید در پلتفرم‌های شبکه‌های اجتماعی ممنوع شود ایجاد کرده‌است.

الگوریتم‌های *ML*، باید در کنار نیروی انسانی کار کنند تا مؤثر باشند. در عین حال، تعدیل و کنترل محتوای گذاشته‌شده توسط انسان، کاری سخت و خطرناک است؛ زیرا می‌تواند شکایت صاحبان محتوا را در پی داشته‌باشد.

شرکت‌های شبکه‌های اجتماعی، در حال توسعه‌ی دستورالعمل‌های تعدیل محتوا هستند و البته هیچ‌وقت معلوم نیست که چه محتوایی باید ممنوع شود! محدودیت‌هایی که توسط هوش‌مصنوعی شناسایی و حذف می‌شود، می‌تواند تاثیر دلخراشی در «آزادی بیان» داشته‌باشد. برای کسانی که دیتایشان حذف و کنترل شده، اغلب اطلاعات کمی در اختیار می‌گذارند و برای کسانی که آزادیشان محدود شده، هیچ توسلی وجود ندارد.

در عین حال، برای کسانی که از طریق تهدید، آزار و اذیت، پورنوگرافی جعلی، رادیکالیسم یا رادیکال‌سازی (هواداری از تغییرات ریشه‌ای در جامعه، تغییرات بنیادی و ریشه‌ای) یا کلاهبرداری توسط محتوا در شبکه‌های اجتماعی آسیب دیده‌اند، اغلب راه‌حل‌های کمی در برابر پلتفرم‌های شبکه‌های اجتماعی قرار دارد.

«هرشاک»، محتوای مدیریت را از دیدگاه اخلاق بودایی تحلیل می‌کند. Facebook مسئولیت اخلاقی دارد که تعقیب منافع تجاری خود را به گونه‌ای انجام دهد که آسیب نرساند، و وقتی که اجازه داد تا سخنان بد و نفرت‌انگیز در برابر «روهینگیا» در پلتفرم خود گسترش یابد، این مسئولیت‌ها را نادیده گرفت. «هرشاک»، اشاره به ابهام مرزهای اخلاقی توسط پلتفرم‌هایی مانند *Facebook* دارد. این کار باعث پخش مسئولیت و آسیب در بین گروه‌ها، افراد و عوامل متنوع می‌شود.

در تصمیم‌گیری درباره نحوه‌ی مدیریت محتوا در آینده، ارزش‌های «بودایی» به ما یاد می‌دهند که از سوءاستفاده، داستان‌سازی و شایعه، غیبت، تهمت، دروغ و نفرت که در شبکه‌های اجتماعی بسیار شایع هستند، پرهیز می‌کنند. این ویژگی‌های اخلاقی و رفتاری در بودیسم (آئین بودایی) ارزشمند است: «شفقت، مهربانی، متانت، و شادی در اقبال دیگران» - که به وضوح وجود ندارند و ما می‌توانیم در تمام تعاملات خود با دیگران، از جمله در شبکه‌های اجتماعی، آن‌ها را پرورش دهیم.

«هکر-رایت» از دیدگاه اخلاق فضیلت‌محور به سخنان نفرت‌انگیز و بد در مدیریت محتوا نگاه می‌کند. چه نوع فضایلی باید از طریق پاسخ‌های ما به مدیریت محتوا ترویج یابند یا کنار گذاشته شوند؟ ما هرگز نمی‌توانیم همه محتوای مضر را از شبکه‌های اجتماعی حذف کنیم، و این به تنهایی باعث ارتقای یک جامعه نمی‌شود. از این گذشته، رسانه‌ها دقیقاً به این دلیل مؤثر هستند که ما به آن‌ها

اجازه می دهیم دیدگاه‌های قبلیِ ما را تقویت کنند. به این ترتیب، همه ما خواسته یا ناخواسته، در دستکاری رسانه‌های اجتماعی شرکت می‌کنیم (منظور این است که در رسانه‌ها، خواسته یا ناخواسته، فعالیت می‌کنیم). ما می‌توانیم با پرورش فضیلت‌های مهمی مانند شجاعت اخلاقی، تفکر انتقادی و تمایل به ارتقای جامعه (از جمله کسانی که با ما مخالف هستند) با این کار مقابله کنیم. به این ترتیب، می‌توانیم به نوعی خرد عملی دست یابیم که از پرورش عادتی و صد البته آگاهانه‌یِ فضایلی که ارسطو معتقد بود منجر به رفاه است، ناشی می‌شود.

«میلر» و «سالیوان کلارک» به تعدیل محتوا از دریچه اخلاق فطری و ذاتی نگاه می‌کنند که همه چیز را به هم مرتبط می بیند. *Facebook* درک نکرد که محتوای موجود در پلتفرم آن‌ها چگونه روی «روهینگیا» تأثیر می‌گذارد. الگوریتم‌های آن‌ها، سخنرانی‌های بسیار جذاب را در اولویت قرار می‌داد، حتی زمانی که نفرت و دشمنی را ترویج می‌داد و خشونت علیه یک گروه آسیب‌پذیر را تقویت می‌کرد! این اقدامات باعث ایجاد نوعی ناهماهنگی و عدم تعادل می‌شود که اغلب باعث آسیب می‌شود. *Facebook* همچنین نتوانست ارزش مهم فروتنی در تفکر را پرورش دهد. کسانی که سیستم‌های *ML* را طراحی و اجرا می‌کنند، موظفند از محدودیت‌های الگوریتم‌های خود و همچنین پتانسیل آن‌ها برای سوءاستفاده، آگاه باشند. «میلر» و «سالیوان-کلارک» به نکته‌ی مهمی اشاره می‌کنند که چندین مشارکت‌کننده به آن اذعان دارند، این که: «کلمات قدرت دارند». الگوریتم‌هایی که گفتار خاصی را برای دیگران تبلیغ می‌کنند یا باعث ایجاد ابهام می‌شوند نیز قدرت دارند، و باید با فروتنی و در نظر گرفتن رفاه دیگران از آن‌ها استفاده کرد.

مورد۶ – بدافزار ذهنی: الگوریتم‌ها و معماری انتخاب

در سال 2013، شرکت تجزیه و تحلیل داده «کمبریج آنالیتیکا» شروع به جمع‌آوری اطلاعات در *Facebook* برای ایجاد پروفایل‌های روانشناختی عمیق روی ده‌ها میلیون کاربر بدون، رضایت آن‌ها کرد. سپس این داده‌ها به بازاریابان فروخته شد، از جمله چندین کمپین سیاسی. این رسوایی منجر به ورشکستگی «کمبریج آنالیتیکا» و میلیاردها دلار جریمه برای *Facebook* شد! رسوایی «کمبریج آنالیتیکا» نشان داد که جمع‌آوری اطلاعات حساس روانشناختی از کاربران رسانه‌های اجتماعی و استفاده از این داده‌ها به روش‌هایی که آن‌ها را دستکاری می‌کنند (اغلب بر خلاف منافع مردم)، چقدر آسان است! این مورد، دقیقا یک مورد برجسته از چیزی است که ما «برافزار ذهنی» می‌نامیم. «بدافزار ذهنی» اغلب بر علیه کاربران به شیوه‌هایی استفاده می‌شود که نه صرفاً برای پیش‌بینی رفتار آن‌ها، بلکه برای تغییر رفتار آن‌ها طراحی شده‌اند (برای "تحت فشار دادن"، دستکاری و تغییر رفتارهای فردی و افکار عمومی).

بهره بردن از قدرت روانی الگوریتم‌ها برای افراد سیاسی آسان است. از زمان رسوایی «کمبریج آنالیتیکا»، انتقادات بیشتری مبنی بر: اینکه شرکت‌های شبکه‌های اجتماعی در مقابله با «لایک‌ها» و «فالوورها» نادرست و دستکاری شده و دیگر اشکال تعامل مصنوعی و غیرواقعی شکست خورده‌اند؛ و اینکه از این موضوع برای دستکاری در انتخابات و سرکوب استفاده شده است، وارد شده است.

«هرشاک» به تعدیل محتوا از دیدگاه اخلاق بودایی نگاه می‌کند. همیشه، همه‌ی شرکت‌های شبکه‌های اجتماعی، محتوا را برای کاربران خود فیلتر و تعدیل می‌کنند، و ما باید مراقب باشیم که آن‌ها چگونه این انتخاب‌ها را انجام می‌دهند، چه کسی مسئول این انتخاب است و چه ارزش‌هایی در اولویت هستند؟ معماری انتخاب دیجیتالی که ایجاد می‌کنیم باید رفاه فردی و اجتماعی را افزایش دهد. در حالی که نیاز به تقویت آزادی شخصی وجود دارد، ما باید آگاه باشیم که این پتانسیل این را نیز دارد که کاربران را در انتخاب‌های گذشته خود قفل کند و در نتیجه آزادی آن‌ها را محدودتر

کند. این ممکن است به سادگی منجر به این شود که توده‌های بشریت «زندگی‌هایی را داشته باشند که در آن هرگز لازم نیست از اشتباهات درس بگیریم یا در رفتار سازگارانه شرکت کنیم». در اخلاق بودایی، همه چیز به هم مرتبط است. زیرساخت‌های دیجیتالی‌ای که ما ایجاد می‌کنیم نه تنها بر انتخاب‌ها، رفتار و روابط اجتماعی ما تأثیر می‌گذارد، بلکه اساساً چیزی که هستیم را تغییر می‌دهد!

«هکر رایت» بحث خود را در مورد انتخاب‌ها، به وسیله‌ی رفتار با فضیلت ادامه می‌دهد. «ارسطو» فاعل نیکوکار را کسی توصیف می‌کند که به دنبال خیر است و از منکر دوری می‌کند. یک مامور پاکدامن به دنبال خیر خواهد بود، اما آن‌ها همچنان به سمت برخی از رذایل کشیده می‌شوند و با انتخاب درست مبارزه خواهند کرد. یک مامور رذیل نیز با جذب خود به سمت رذیله، بدی و تباهی مبارزه می‌کند، اما آن‌ها قدرت اراده کافی را برای مبارزه ندارند. پس از شکست خوردن در مبارزه بین فضیلت و رذیلت، ممکن است که احساس شرمندگی کنند. از سوی دیگر ، یک ایده‌آل غلط (که فکر می‌کنند خوب و درست است) را پذیرفته اند، و بنابراین آن‌ها با تسلیم شدن به رذیلت، فکر می‌کنند که رذیلت برای زندگی خوب است، ولی در اشتباه‌اند.

شبکه‌های اجتماعی و شرکت‌های بازی‌سازی طراحی شده‌اند تا از طریق تاکتیک‌های هوشمندانه‌ی دستکاری و غلبه بر قدرت اراده کاربران، همه را، به جز فضیلت‌ترین کاربران جذب کنند (فقط کاربران فضیل که درست‌کاراند، جذب نمی‌شوند). پرورش فضایل و تقویت اراده، می‌تواند ابزار موثری برای غلبه بر انبوه بدافزارهای ذهنی‌ای باشد که هر روزه با آن مواجه هستیم.

«مارشال» به بدافزارهای ذهنی از زاویه‌ی دئونتولوژیک (نگاه دینی) نگاه می‌کند. هرگونه تلاش برای تأثیرگذاری بر دیگران، ابتدا باید در جهت یک هدف اخلاقی باشد. در اخلاق دین شناسی، استفاده از دیگران به عنوان وسیله‌ای برای رسیدن به هدف ممنوع است. ما نمی‌توانیم برای تحقق منافع خود دیگران را زیر پا بگذاریم، کاری که بسیاری از شرکت‌ها و بازاریابان شبکه‌های اجتماعی انجام می‌دهند! دوم، هر نوع نفوذ اخلاقی و تأثیرگذاری باید مبتنی بر صداقت و گفتمان عقلانی و منطقی باشد. بدافزار ذهنی به دنبال جذب چیزی است که «کانمن» *(Kahneman)* آن را تفکر «**سیستم 1**» می‌نامد (تعریف سیستم 1: پاسخ‌های احساسی و خودکار (سریع و آسان) که در ابتدا به

اطلاعات جدید می‌دهیم). با این حال، هر تلاش برای تأثیرگذاری، حتما باید روش تفکر «**سیستم 2**» را نیز درگیر کند (تعریف سیستم 2: روش‌های تفکر آگاهانه، منطقی و مشورتی (آهسته و دشوار)). کسانی که الگوریتم‌ها را به صورت اخلاقی و برای تأثیرگذاری به کار می‌برند، باید در مورد نحوه عملکرد الگوریتم‌ها صادق و شفاف باشند. درنهایت، ما باید فرآیند مشورتی منطقی تصمیم‌گیری بر اساس اطلاعات خوب و داده‌های تجربی صحیح را نسبت به پاسخ‌های سریع و احساسی اولویت دهیم؛ چیزی که امروز دقیقا برعکس آن در شبکه‌های اجتماعی درحال انجام است!

مورد۷ – هوش‌مصنوعی و موجودات غیر انسان

انسان‌ها تنها موجودات زنده‌ای نیستند که سیستم‌های ML بر منافعشان تأثیر می‌گذارند (چه به صورت مثبت و چه به صورت منفی). در این فصل، «سینگر» و «تسه» تحقیقات خود را در مورد روش‌هایی که الگوریتم‌های هوش‌مصنوعی بر رفاه حیوانات تأثیر می‌گذارند ارائه می‌کنند.

اول، آن‌ها درباره‌ی تأثیرات مختلفی که نتایج موتورهای جستجو و الگوریتم‌های توصیهای می‌توانند بر نحوه تفکر ما در مورد حیوانات و در نتیجه نحوه برخورد ما با حیوانات تأثیر بگذارند، بحث می‌کنند. تعصب الگوریتمی در نتایج موتورهای جستجو و توصیه‌ی محتوا می‌تواند محتوا و تبلیغاتی را به ما ارائه دهد، که بر میزان تأثیر آن می‌افزاید. محصولات حیوانی‌ای که مصرف می‌کنیم در حالی که ظلم و آزار حیوانات در دنیای واقعی را پنهان می‌کنیم و کاربران را نسبت به این آسیب‌ها حساسیت زدایی می‌کنیم. مدل‌های زبانی، می‌توانند «بار نژادپرستی» زبان را تقویت کنند که حیوانات را تحقیر می‌کند. این تأثیر زیادی بر رفاه حیوانات دارد (منظور این است که فرضا صفت درنده برای ببر درست است، ولی درواقع یک صفت منفی به حساب می‌آید، مدل‌های زبانی ممکن است از این صفات استفاده کرده و ناخواسته محتوایی تولید کنند که گونه‌گرایانه و یا نژادپرستی را می‌رساند).

دوم، آن‌ها در مورد استفاده از هوش‌مصنوعی در مزارع و کارخانه‌ها بحث می کنند. مدل‌های

ML در صنعت مزرعه‌ی کارخانه‌ای، برای جمع‌آوری اطلاعات در مورد حیوانات پرورشی، به منظور به دست‌آوردن سود حداکثری، استفاده می‌شوند. بیماری و مرگ و میر چقدر سود را به حداکثر می رساند؟ چه مقدار باید حیوانات تغذیه شوند تا رشد را، با پایین نگه داشتن هزینه‌ها متعادل کند؟ آن‌ها همچنین این موضوع مهم را مطرح می‌کنند که: چگونه رفتار حیوانات و حالات ذهنیشان را شناسایی و تفسیر می‌کنیم، زمانی که از دریچه چشم‌انداز خودمان، انسانی، نگاه می‌کنیم؛ ولی به راحتی حقوق‌شان را زیر پا می‌گذاریم. خروج هوش‌مصنوعی از ذهنیت انسانی و اتخاذ مجموعه‌ای از ارزش‌ها و دیدگاه‌های غیرانسانی به چه معناست؟ رفاه آینده حیوانات به نحوه حل این مسائل اخلاقی بستگی دارد.

«سینکلر» یک دیدگاه اخلاقی یهودی در مورد وظیفه رفتار با حیوانات به روشی درست و اخلاقی ارائه می‌دهد. در حالی که انسان‌ها نسبت به سایر موجودات برتری دارند، اولین مردم گیاهخوار بودند و بعدها که فاسد شدند، اجازه یافتند گوشت بخورند. مفهوم جلوگیری از ظلم به حیوانات عمیقاً در اخلاق یهودی گنجانده شده است، از جمله اجازه دادن به حیوانات کار برای استراحت در شب‌ها و لذت بردن از اوقات فراغت خود. «مارشال» درباره‌ی وضعیت اخلاقی حیوانات در اخلاق دئونتولوژیک بحث می‌کند. همه نسخه‌های اخلاق افضل (اخلاق وظیفه‌شناس، علماالاخلاق) اهمیت حقوق حیوانات را به رسمیت می‌شناسند، گرچه در مورد اهمیت حقوق حیوانات، موارد متفاوت و استثنا هم وجود دارد. اگر چنین است، پس استفاده از حیوانات به عنوان ابزاری صِرف برای اهداف خود از نظر اخلاقی غیرمجاز خواهد بود. او همچنین به این نکته مهم اشاره می‌کند که عدم احترام کافی به ادعاهای اخلاقی و حقوق حیوانات می‌تواند باعث شود که به طور کلی به ادعاهای اخلاقی و حقوق دیگران نیز احترام نگذاریم. باید از بی‌تفاوتی نسبت به رنج و احساس دیگران به شدت اجتناب شود (هرکس با هر احساسی).

«مورانگی» تفسیری درباره حقوق اخلاقی حیوانات از منظر اخلاق آفریقایی ارائه می‌دهد. او نقشی را که استعمارزدایی در اخلاق هوش‌مصنوعی بازی می‌کند، بررسی می‌کند و اینکه آیا می‌توانیم اخلاق هوش‌مصنوعی را طوری توسعه دهیم که به دنبال درک ماهیت اشتراکی «ما» باشد که

قلب رفتار «اوبونتو» را شرح می‌دهد؟ معماران این فناوری‌ها اغلب نمی‌توانند «خود را فرزندان هوش‌مصنوعی یا مادران و پدران هوش‌مصنوعی ببینند». درعوض، آن‌ها باید تشویق شوند تا به این فکر کنند که یک عامل اخلاقی به چه معناست، و چه چیزی به معنای رفاه است. این فضایی را برای یک هوش‌مصنوعی رهایی‌بخش به جای ظالمانه باز می‌کند (هوش‌مصنوعی‌ای که رفاه حیوانات را نیز ارتقا می‌دهد؛ زیرا اگر به اندازه‌ی کافی به اینکه چه کسی هستیم و چیستیم فکر نکنیم، نمی‌توانیم حیوانات را به عنوان موجوداتی که از حقوق برخوردار هستند، تصور کنیم).

همه‌ی مفسران این کتاب راهی به جلو برای دانشمندان داده ارائه می‌دهند تا در طراحی و استفاده از داده‌ها و سیستم‌های هوش‌مصنوعی اخلاق را رعایت کنند. در واقع، درگیر شدن با نظرات مشارکت‌کنندگان مطمئناً نوعی خرد را تقویت می‌کند که «کلهر» از آن حمایت کرده است، و این امر کمک زیادی به حرکت در دنیای الگوریتم‌ها و هوش‌مصنوعی می‌کند. اهمیت داده‌ها در این عصر قابل انکار نیست! این امر قدرت بزرگی را در دست دانشمندان داده قرار می‌دهد و همانطور که ضرب المثل قدیمی می گوید: "هرکه بامش بیش، برفش بیشتر". ما واقعاً امیدواریم که این کتاب ابزارهای ارزشمندی را ارائه دهد که به همه‌ی ما در انجام این مسئولیت بزرگ با عقل، شفقت و خرد کمک کند.

فصل ۲

مقدمه ای بر رویکردهای اخلاقی در علم داده

> دانش علم فیزیک مرا به خاطر ناآگاهی از اخلاق، تسلی نمی‌دهد، اما علم اخلاق همیشه مرا به خاطر ناآگاهی از علم فیزیکی تسلی می‌دهد. (منظور نویسنده، تأکید مهم بودن علم اخلاق و ارزش‌های انسانی است)
>
> *Blaise Pascal 1624-1624*

مقدمه

فناوری‌های یادگیری ماشین، در حال نفوذ به زندگی مردم عادی در سراسر جهان هستند. کاربران این فناوری‌ها، خواسته یا ناخواسته در زندگی اصولی دارند که رویکردهای آن، در میان فلفسه‌های غربی ارائه نشده. بنابراین، ما چندین رویکرد اخلاقی غیر غربی را در کتاب آورده‌ایم.

این‌ها برای طراحان ارزش دانستن دارد، هم برای اینکه بتوانند کاوش اخلاقی خود را عمیق‌تر کنند و هم به این ترتیب که بتوانند بهتر درک کنند که چگونه فن آوری‌هایشان تفسیر، اتخاذ، استفاده و تنظیم می‌شود. ما خوش‌شانس بوده‌ایم که تفسیرهایی از دانشمندان برجسته در زمینه‌های اخلاق دئونتولوژیک، اخلاق نتیجه‌گرا (فایده گرا)، و اخلاق فضیلت و فطری، و همچنین از اخلاق اوبونتو، اخلاق بودایی، اخلاق یهودی، و اخلاق بومی و ذاتی دریافت کرده‌ایم. ما امیدواریم که این به خواننده دید وسیع‌تری بدهد تا درباره‌ی فناوری‌های یادگیری ماشین از دیدگاه‌های مختلف فکر کند و بفهمد که چگونه آن‌ها توسط جوامع سراسر جهان پذیرفته می‌شوند و چگونه عمل می‌کنند. هر یک از این

رویکردهای اخلاقی در زیر به اختصار آورده شده است.

رفتار نتیجه گرایی و فایده‌گرایی

توسط پیتر سینگر و ییپ فای تسه

نتیجه‌گرایی خانواده‌ای از نظریه‌ها است که بر این عقیده هستند که درست یا نادرست بودن یک عمل بستگی به پیامدهای آن دارد یا به عبارت دیگر، وضعیتی که اعمال باعث ایجاد آن می‌شود.

فایده‌گرایی، در شکل کلاسیک خود، نظریه نتیجه گرایی است که منحصراً بر درد و لذت، یا شادی و بدبختی، به عنوان تنها پیامدهای اخلاقی مرتبط برای تعیین چگونگی ارزیابی پیامدهای اعمال تمرکز می‌کند. در اینجا تأکید بر این نکته حائز اهمیت است که فایده‌گرایی تنها در مورد ارزیابی درستی یا نادرستی اعمال نیست، بلکه در مورد ارزیابی خوب و بد حالت‌های امور است، که بی طرفانه در نظر گرفته می‌شوند. به طور خاص، فایده‌گرایان معتقدند که همه‌ی موجودات ذی‌شعور (آن‌هایی که می‌توانند درد و لذت را تجربه کنند) باید در نظر گرفته‌شوند و به علایق مشابه آن‌ها باید وزن مشابهی داده شود. در کنار هم، فایده‌گرایی، این دیدگاه است که یک عمل نه تنها باید منفعت برساند، بلکه از نظر اخلاقی نیز لازم است که بیشترین مازاد خالص ممکن را از شادی نسبت به بدبختی (یا لذت بر درد) به همراه داشته باشد. و هر عملی که بر خلاف این اصل باشد، ممنوع و غیرمجاز است.

اعتراضات رایج به سودگرایی

یک اعتراض رایج به سودگرایی این است که ما را به انجام اعمال آشکاراً غیراخلاقی هدایت می‌کند! «داستایوفسکی» در «برادران کارامازوف»، «ایوان» را به چالش می کشدکه یک نوزاد را تا سرحد مرگ شکنجه کند تا برای همه‌ی بشریت خوشبختی بیاورد. چالش «ایوان» به یک اعتراض معروف به سودگرایی تبدیل شده است. بیان ساختار اعتراض «داستایوفسکی» به طور رسمی این موضوع را بهتر نشان می‌دهد:

فرض ۱. اگر فایده‌گرایی درست بود، به درستی به ما می گفت که کدام اعمال درست و کدام نادرست است.

فرض ۲. فایده‌گرایی به ما می گوید که اگر شکنجه‌ی یک کودک بی گناه تا حد مرگ عواقب بهتری نسبت به هر عمل دیگری به همراه داشته باشد، آنگاه شکنجه یک کودک بی گناه تا حد مرگ کار درستی خواهد بود.

فرض ۳. شکنجه یک کودک بی گناه تا حد مرگ همیشه اشتباه است. نتیجه: فایده‌گرایی نادرست است.

بسیاری از ایرادات به فایده‌گرایی نیز به همین ترتیب مطرح می شوند: یک جراح به این فکر می‌کند که آیا مخفیانه اطمینان حاصل کند که یک عمل شکست می خورد؛ تا بیمار بمیرد و سپس از اعضای بدن او برای نجات جان چهار بیمار در انتظار اهدای اعضای ضروری استفاده شود. چنین نمونه‌هایی منعکس کننده‌ی دانش ما از نحوه عملکرد جهان نیستند. «ایوان» توضیح نداد که چگونه شکنجه‌ی کودک باعث شادی پایدار برای دیگران می‌شود. مثال پیوند عضو در نظر نمی‌گیرد که اگر کاری که

جراح انجام داده مشخص شود، ممکن است منجر به عواقبی شود که بسیار بیشتر از مزایای مورد نظر است (ممکن است افراد نسبت به پزشکان بی‌اعتماد شوند). چگونه جراح می‌تواند کاملاً مطمئن باشد که او گرفتار نخواهد شد؟ این فرض که شکنجه یک کودک بی گناه همیشه اشتباه است، متکی به ذات و فطرت انسانی دارد. بنابراین وقتی با نمونه‌های عجیب و خیالی سروکار داریم، فرض 3 مشکوک است و نمی‌توان به آن به عنوان مبنایی برای رد فایده‌گرایی اعتماد کرد.

ایراد اصلی دیگر این است که اندازه‌گیری درد و لذت، یا شادی و غم است. سودگرایان سه پاسخ اصلی به این اعتراض دارند. اولاً، این مشکلی محدود به فایده‌گرایی نیست. هر نظریه‌ی اخلاقی‌ای که مقداری به رفاه اهمیت می‌دهد از دشواری اندازه‌گیری رفاه افرادی که تحت تأثیر اعمال هستند نیز رنج می برد؛ و البته نظریه‌ی اخلاقی‌ای که تمام این ملاحظات رفاهی را نادیده می‌گیرد بسیار غیرقابل‌قبول خواهد بود.

ثانیاً، اگرچه اندازه‌گیری دقیق درد و لذت دشوار است، ترجیحات افراد و تا حدی حیوانات را می توان مشاهده، آزمایش و رتبه بندی کرد تا اولویت‌های آن‌ها آشکار مشخص. در برخی از مطالعات، روانشناسان با پرداخت هزینه به آزمایش‌شوندگان، سطوح خاصی از درد یا تحمل را در آن‌ها می‌سنجند. این موارد، اگرچه آن چیزی نیست که فایده‌گرایان کلاسیک آن را خیر می‌دانند، با این وجود، معیارهای مفیدی هستند که به ما ایده‌ای درباره‌ی درد و لذت می‌دهند. مدل دیگری که از موارد آشکار استفاده می‌کند، سال زندگی تعدیل‌شده با کیفیت یا *(QALY)*، حول این ایده است که یک سال زندگی با عملکرد یا سلامت مختل، به اندازه یک سال در سلامت عادی، خوب نیست. برای مثال، محققان از مردم می‌خواهند که خود را با آسیب‌های مختلف در سلامت تصور کنند (گاهی اوقات خود درد)، و سپس از آن‌ها می‌پرسند که حاضرید چند سال از زندگی خود را رها کنید تا این اختلال درمان شود؟ این روش اکنون در سطح جهانی توسط اقتصاددانان سلامت، محققان پزشکی و سیاست‌گذاران استفاده می‌شود. در نهایت، در اکثر موارد، عمل درست حتی بدون اندازه‌گیری واضح است. به عنوان مثال، پزشکی که ترتیب درد بیماران را در اولویت قرار می‌دهد، می‌تواند به وضوح ببیند که یک بیمار سوختگی، شدیدتر از فردی که از سرماخوردگی رنج می برد،

درد دارد و در معرض خسر مرگ بسیار بالاتری است؛ بنابراین، باید بیمار سوختگی را در اولیت قرار داد. یا مثلا اگر فردی از شما بپرسد که نزدیک‌ترین رستوران گیاه‌خواران کجاست؟ شما به احتمال خیلی زیاد با ارائه‌ی اطلاعات درست، او را راهنمایی می‌کنید، تا اینکه اصلا جواب ندهید یا پاسخ اشتباه بدهید!

اگرچه مواردی هم وجود دارند که پس از تجزیه و تحلیل هم شفاف نیستند؛ ولی با این جود می‌توان تصمیمات معقولی گرفت. نکته‌ی مهمی که در اینجا باید مورد توجه قرار گیرد، این است که نه تنها می توان بخش قابل توجهی از تصمیمات تحت فایده‌گرایی را بدون اندازه‌گیری لذت و درد اتخاذ کرد، بلکه آنچه در این دنیا در خطر است نیز معمولاً می‌تواند بدون اندازه‌گیری مستقیم درد و لذت تعیین شود. فقر جهانی (که باعث گرسنگی، تشنگی، بیماری‌ها و ... می‌شوند)، کشاورزی کارخانه‌ای و بیماری‌های همه‌گیر نمونه‌های مناسبی از مسائلی هستند که بدون شک، رنج عظیمی را برای تعداد زیادی از افراد به بار می‌آورند.

توصیه‌هایی برای به کارگیری صحیح اصول سودمندی

گسترده تر و طولانی تر فکر کنید

ما با «جان استوارت میل»، یک فایده‌گرای اولیه، موافقیم که باید «مفید بودن را به‌عنوان اصل نهایی در همه مسائل اخلاقی در نظر بگیریم. اما باید در فراگیرترین معنای آن فایده باشد». منظور ما از "فراگیرترین" این است که همه پیامدهای مرتبط، صرف نظر از زمان، فاصله فیزیکی، خویشاوندی و سایر ویژگی‌های اخلاقی نامربوط مانند جنسیت، نژاد، و عضویت در گونه باید در نظر گرفته شوند.

مسلماً، زمان یکی از بحث برانگیزترین آن‌هاست که از نظر اخلاقی نامربوط اعلام می‌شود. تخفیف زمان اغلب در زمینه‌های اقتصاد و یادگیری ماشین آموزش داده‌می‌شود و به کار می‌رود، ولی تصورات

آن‌ها در مورد ترجیحات زمانی با ایده‌های فایده‌گرایی متفاوت است. در اقتصاد، کاهش زمان، برای دریافت لذت و خوشی در زمان کمتر مد نظر است؛ یعنی ما مایل هستیم که لذت و خوشی را در زمان کمتری به دست بیاوریم تا اینکه بخواهیم برای آن صبر کنیم. در یادگیری ماشین به ویژه یادگیری تقویتی، "ضریب تخفیف" (γ)، متغیری است که تعیین می‌کند که عامل، تمایل به اهداف و پاداش‌های زودهنگام دارد یا دیرهنگام (اهمیت را برای پاداش‌های فوری یا آینده تعیین می‌کند). اگر مقدار (γ) نزدیک به 1 باشد، عامل به پاداش‌های آینده، بیشتر اهمیت می‌دهد و در نتیجه تمایل دارد تا مسیری را که باعث رسیدن به هدف در آینده می‌شود، دنبال کند. به عبارتی، عامل تمایل دارد پاداش‌های آینده را بیشتر به صورت بلندمدت مد نظر قرار دهد. از سوی دیگر، اگر مقدار (γ) نزدیک به 0 باشد، عامل بیشتر روی پاداش‌های فوری تمرکز می‌کند و تمایل دارد که از پاداش‌های فوری بهره‌برداری کند. به عبارتی، عامل در تصمیم‌گیری خود بیشتر به جوانب کوتاه‌مدت توجه می‌کند و پاداش‌های آینده، اهمیت نمی‌دهد. مثلا می‌توانیم بگوئیم به دلیل اینکه در آینده فلان بازار هدف وجود نخواهد داشت، (γ) را نزدیک به 0 د نظر می‌گیریم تا در کوتاه مدت، به نتیجه‌ی دلخواه برسیم، عکس این قضیه هم صادق است. به عنوان مثال، شکنجه در 100 سال به همان اندازه بد است که شکنجه‌ای اکنون به همان اندازه درد داشته‌باشد، اما اگر قطعیت کمتری داشته باشد (یعنی ممکن باشد که شکنجه انجام نشود)، ممکن است به همین دلیل آن را کاهش دهیم (یعنی شکنجه‌ی چیزی را می‌پذیریم که مارا شکنجه نکند یا آن موردی که قطعیت کمتری دارد).

بیایید سعی کنیم این اصول را در هوش‌مصنوعی و علم داده اعمال کنیم. برای مثال، در تصمیم‌گیری برای راه‌اندازی یک محصول، نه تنها باید تاثیری که ممکن است بر روی کاربران آن داشته باشد، بلکه باید در نظر داشت که چگونه جامعه وسیع‌تر افراد (در برخی موارد، حتی حیوانات) چه در کوتاه مدت و چه در بلند مدت ممکن است تحت تاثیر قرار گیرند. سؤالاتی از این قبیل باید پرسیده شود: آیا این محصول سوگیری‌ها، فرهنگ، ایدئولوژی‌ها، فضیلت‌ها یا سایر ارزش‌ها را در جامعه جذب و در نتیجه آن را تقویت می‌کند؟ آیا این محصول، یک صنعت بسیار ارزشمند را از بین می برد یا باعث به تاخیر انداختن یا جلوگیری از حذف یک صنعت غیراخلاقی می‌شود؟

از ارزش‌های مورد انتظار برای تصمیم گیری استفاده کنید

استفاده از تئوری ارزش مورد انتظار در تصمیم‌گیری، در تئوری تصمیم گیری، اقتصاد و علم داده، اساسی است (یعنی قبل از تصمیم‌گیری بسنجیم ببینم که دنبال چه چیزی هستیم و بر اساس آن تصمیم‌گیری بکنیم). اما باید در مورد نظریه‌های اخلاقی، به‌ویژه به حداکثر رساندن فاکتورهای اخلاقی مانند فایده‌گرایی نیز اعمال شود. مثال جراح در بخش قبل نشان می‌دهد که چرا سناریوهایی با ریسک بالا و کم احتمال، اهمیت دارند. مهم نیست که جراح چقدر با دقت سعی کرد عمل او را مخفی نگه دارد، او نتوانست به طور منطقی به این نتیجه برسد که احتمال افشای راز صفر است. با توجه به اثرات مشخص کشف شدن راز (اگر کشف می‌شد، مردم نسبت به پزشکان اعتمادشان را از دست می‌دادند)، جراح باید به این نتیجه برسد که انجام چنین عملی اشتباه است.

در حالی که محاسبه ارزش مورد انتظار اغلب ساده است (انجام آن عمل، چندین انسان را نجات می‌داد)، به دلیل سوگیری‌های شناختی انسان (مثلا اینکه شما بیمار من رو به خاطر اهدای عضو، به عمد به قتل رساندید!)، اغلب به درستی استفاده نمی‌شود یا حتی اصلاً اعمال نمی‌شود. «غفلت احتمالی» یک سوگیری شناختی است که افراد نسبت به «عدم قطعیت‌ها» نشان می‌دهند، به‌ویژه «احتمالات کوچک»، که تمایل دارند یا به طور کامل از آن‌ها غفلت کنند، یا تا حد زیادی (اغراق) آن را بزرگ کنند. یک مطالعه با دریافت اینکه مردم برای کاهش خطرات "رویدادهای نادر و پر تاثیر" یا ارزش خیلی زیاد یا بسیار پایین قائل هستند؛ (غفلت احتمالی) را تائید کرد. ما نیازی به جستجوی شواهدی مبنی بر غفلت جمعی از «رویدادهای نادر و پرتأثیر» نداریم! اگر قانون اجباری بستن کمربند در خودرو برداشته‌شود، به نظر شما چند نفر حاضرند تا کمربندشان ببینندند؟ (این خود نشان دهنده‌ی این است که مردم از سوگیری شناختی غفلت احتمالی استفاده می‌کنند!) این قضیه اصلا هم جالب نیست! زیرا «رویدادهایی با احتمال کم و تأثیر زیاد» اغلب دارای ارزش‌های

مورد انتظار بزرگ، اعم از منفی یا مثبت هستند، این دام در تفکر انسان نگران‌کننده است! این نشان می‌دهد که انسان اغلب به «ارزش‌های مورد انتظار» حتی فکر هم نمی‌کند! چه برسد که بخواهد آن را هنگام تصمیم‌گیری به کار ببرد!

سوگیری دیگری که ممکن است بر توانایی افراد در برآورد مقادیر مورد انتظار تأثیر بگذارد، «غفلت از محدوده» است. مطالعات نشان داده است که افراد ارزش‌گذاری خود را در تناسب با مقیاس یک مسئله تنظیم نمی‌کنند. به عنوان مثال، یک مطالعه از سه گروه از افراد در مورد تمایل آن‌ها به پرداخت هزینه برای نجات 2000 یا 20000 یا 200000 پرنده از غرق شدن در استخرهای نفتی بدون سرپوش پرسیده شد. میانگین‌های مربوطه 80، 78 و 88 دلار و میانگین پاسخ‌ها همگی 25 دلار بود. اگر ارزش‌گذاری افراد از برخی نتایج به‌درستی مقیاس‌پذیر نباشد، ارزش‌های مورد انتظار نیز نخواهد بود (یعنی اینجا باید هرکس با توجه به دارایی خود مبلغی را اعلام می‌کرد، ولی همه‌ی آن‌ها پاسخی نزدیک به 25 دلار داده‌بودند).

در انتخاب پروژه‌های خیریه، پروژه‌های (موارد) موثر را انتخاب کنید

از آنجایی که مردم معمولاً به جای تحقیق در مورد اثربخشی خیریه، بر اساس انگیزه و احساسات به خیریه می‌پردازند، اغلب از خیریه‌ها و اهداف بی‌اثر حمایت می‌کنند. ولی در عوض چیزهایی نذیر: نوع دوستی مؤثر، یک جنبش جهانی اخیر، بر اهمیت رفتار نوع دوستانه مؤثر، چه در قالب کمک‌های مالی و چه در قالب زمان مهم هستند!

چه خوب است که همین اصل (سراغ کارهایی برویم که اثربخشی بالا دارند) را در هوش‌مصنوعی پیاده‌سازی کنیم و به اهداف مهم‌تر، اولویت بالاتری بدهیم.

اخلاق دئونتولوژیک

نوشته‌ی کالین مارشال

رویکردهای «دئونتولوژیک» به اخلاق، بر مجموعه‌ای از ایده‌های مرتبط متمرکز است: احترام، استقلال، حقوق، و امتناع از رفتار با انسان‌ها (و شاید سایر موجودات) به‌گونه‌ای که گویی آن‌ها صرفاً چیزها یا ابزارهایی برای رسیدن به اهداف دیگر هستند. یک تصویر کلاسیک از رویکرد دئونتولوژیک شامل سناریوی زیر است: یک پزشک را تصور کنید که پنج بیمار دارد و هر یک از بیماران، نیاز فوری به اهدای عضو دارند. یک فرد قابل اعتماد و سالم وارد مطب دکتر می‌شود؛ دکتر می‌تواند فرد سالم را بکشد و اعضای بدن او را برای نجات پنج بیمار برداشت کند. حتی اگر پزشک بتواند این کار را بدون تشخیص انجام دهد، بسیاری از مردم قضاوت می‌کنند که نباید این کار را انجام دهند. این قضاوت به راحتی در اصطلاحات دئونتولوژیک، به این صورت بیان می‌شود: عدم احترام از طرف پزشک، به عنوان نقض حقوق فرد سالم، یا به عنوان دکتری که از فرد سالم به عنوان یک چیز صِرف (یک ظرف اندام) استفاده می‌کند.

رویکرد دئونتولوژیک اغلب با رویکردهای «نتیجه‌گرایانه» در تضاد است، که هر عملی را که بهترین نتیجه را به همراه داشته‌باشد توصیه می‌کند؛ یا مثلا اگر در مثلا قبل جزئیات به درستی تکمیل شوند رویکر «فایده‌گرا» پیشنهاد می‌دهد که فرد سالم را برای آن پنج بیماری قربانی کنیم. زیرا در این رویکرد نتیجه‌ای که حاصل می‌شود، این است که پنج انسان به زندگی برگشتند و فقط یک انسان کشته‌شد. ولی رویکرد «دئونتولوژیک»، از حق انسان سالم دفاع می‌کند. با این حال، در عمل، احکام رویکردهای اخلاقی «دئونتولوژیک» و «نتیجه‌گرایانه» غالباً منطبق هستند. به هر حال، در هر نسخه واقع بینانه‌ای از پرونده دکتر، هیچ تضمینی وجود ندارد که قتل مخفی بماند.

توجه به این موضوع، نتیجه‌گرایی، فاکتورگیری (کنارگذاری) ریسک‌های بزرگی را توصیه می‌کند، مانند کاهش اعتماد به متخصصان پزشکی (که در نتیجه افراد بیمار به دنبال کمک لازم نمی‌گردند)

و تأثیر روان‌شناختی مخرب احتمالی بر پزشک (که گناه و آسیب‌های روحی ممکن است آینده آن‌ها را مختل کند). در نتیجه چنین ملاحظاتی، بسیاری از نتیجه‌گرایان معتقدند که اگر مردم عموماً از منظر دئونتولوژیک به تصمیم‌گیری بپردازند، بهترین پیامدها تضمین می‌شود. به همین دلیل، می‌توانیم انتظار داشته‌باشیم که بسیاری از ارزیابی‌های «دئونتولوژیک» با ارزیابی‌های «نتیجه‌گرا» (و سایر موارد) همخوانی داشته‌باشند، حتی اگر رویکردهای مختلف بر عوامل متفاوتی تأکید کنند.

مفهوم اصلی دئونتولوژیک احترام، همراه با دو مفهومی است که از احترام بیرون می‌آیند: بی‌طرفی و امتناع از دیگران به عنوان ابزار صِرف یا چیز (منظور نگاه ابزاری به آدم‌ها است). در اینجا می توانیم به اختصار هر یک از این موارد را بررسی کنیم. انواع مختلفی از احترام وجود دارد، اما شکل مربوط به احترام اخلاقی توجه جدی‌ای، به نیازها و پروژه‌های دیگران است. چنین احترام اخلاقی‌ای می‌تواند و البته باید اغلب بر عمل تأثیر بگذارد: اگر ما به طور جدی نیازهای کسی را در نظر بگیریم، معمولاً به گونه‌ای عمل نمی‌کنیم که آن نیازها را تضعیف کنیم. با این حال، حتی زمانی که اقدامی نیز صورت نگیرد، ممکن است شکست‌هایی در احترام وجود داشته باشد، مانند خندیدن بی‌احترامانه به شکست‌های دیگران آن هم در صورتی که به آن آگاه نباشیم. رفتار اولیه‌ی ما با دیگران، به ندرت با احترام همراه است (اولین رفتار ما همیشه محترمانه نیست). در عوض، ما بی‌احترامی را ترجیح می‌دهیم و سعی می‌کنیم که بر اهداف و نیازهای خودمان ترمرکز کنیم تا اینکه بخواهیم نیازهای دیگران را در اولویت قرار دهیم؛ به این رفتار «جانبداری» می‌گویند. یعنی اهداف خودمان را بر دیگران ترجیح دهیم و برایشان ارزش بیشتری قائل شویم. مثلا اگر یک پلتفرم شبکه‌ی اجتماعی، تنها با هدف به حداکثر رساندن سود، کاربران خود را به شکل‌های تعامل مضر ترغیب کند، آن‌ها با کاربران خود به عنوان وسیله برای دستیابی به سود رفتار می‌کنند (برای اطلاعات بیشتر به تفسیر مورد ۶ - «بدافزار ذهنی» مراجعه کنید). به طور مشابه، اگر یک مزرعه یا کارخانه با حیوانات به‌عنوان منابع صِرف گوشت رفتار کند، آن‌ها را صرفاً وسیله می‌داند (به تفسیر مورد ۷ - «حیوانات و هوش‌مصنوعی» مراجعه کنید). چنین نگرشی به منزله‌ی شکست کامل احترام است. صرف‌نظر از اینکه دیدگاه دئونتولوژیک خوب است یا نه، مردم به طور پیش‌فرض به مسائلی مانند: احترام، حقوق و بی‌طرفی

اهمیت می‌دهند.

اخلاق فضیلت

نوشته‌ی جان هکر رایت

اخلاق فضیلت رویکردی به اخلاق یا به عبارت دقیق‌تر، خانواده‌ای از رویکردها است یک انسان برای خوب زیستن به آن نیاز دارد. این به ما می‌گوید که حالات خوب شخصیت به نام فضیلت را ایجاد و نشان دهیم، و از ایجاد و نشان دادن حالات بد شخصیت به نام رذایل اجتناب کنیم. برجسته‌ترین رشته‌ی اخلاق فضیلت در آکادمی غرب امروز توسط فیلسوف یونان باستان ارسطو (384-322 قبل از میلاد) ارائه شده است، اما نسخه‌های زیادی از اخلاق فضیلت وجود داشته و دارد. از این رو، برای مثال، می توان نسخه‌های کنفوسیوس و بودایی از اخلاق فضیلت را یافت. دیدگاهی که در مورد آنچه در ادامه می‌آید توضیح خواهم‌داد اخلاق فضیلت ارسطویی است. وقتی به یک فرد خوب فکر می‌کنید، ممکن است به فردی با ویژگی‌هایی مانند شجاعت، شفقت، صداقت و مانند آن فکر کنید. این‌ها فضایل فرضی است. هر ویژگی‌ای که فکر می‌کنیم کسی برای خوب زندگی کردن در حوزه خاصی از زندگی انسانی نیاز دارد، تصور ما از فضایل را شامل می‌شود. در حالی که فهرست قطعی از فضایل وجود ندارد، همگرایی قابل توجهی بر سر ویژگی‌هایی مانند شجاعت، صداقت، عدالت و خرد وجود دارد. اخلاق‌گرایان فضیلت می‌کوشند تا معیار درستی و نادرستی در عمل را از فضایل یا فرد نیکوکار استخراج کنند. یکی از فرمول‌بندی‌های برجسته می‌گوید: یک عمل درست است اگر و تنها اگر کاری باشد که یک فرد با فضیلت یا شخصیت، انجام می‌دهد. توجه داشته‌باشید که حتی اگر خودمان فاضل نباشیم، می‌توانیم از این امر پیروی کنیم، به شرط آنکه سطحی از بینش نسبت به

کاری که فاعل با فضیلت انجام می‌دهد و خویشتن‌داری کافی برای انجام آن‌گونه که فرد با فضیلت عمل می‌کند، داشته‌باشیم. اگر خواسته‌های ما بیش از حد بی‌نظم باشد [ضد و تقیض باشد، مثلا طرفداری از فمینیست به دلیل اینکه ما فرد روشن‌فکری هستیم یا به روشن‌فکران احترام می‌گذاریم]، ممکن است نتوانیم با نیات نیکو عمل کنیم و حتی ممکن است در نتیجه تلاش برای عمل به عنوان یک عامل نیکوکار، بدتر عمل کنیم! در این صورت، اختیار اخلاقی ما به دلیل ضعف اراده به خطر بیافتد. هدف ما همچنان این است که بتوانیم همانطور که عامل فاضل عمل می‌کند، عمل کنیم.

ممکن است قوانینی وجود داشته‌باشد که کلیات، الگوهای عمل، و ویژگی‌های استدلالی افراد با فضیلت را به تصویر بکشد، اما نمی‌توان آن‌ها را بدون تفکر به کار برد. به عبارت دیگر، سطحی از درک اخلاقی برای اعمال آن‌ها ضروری است. این ممکن است نقطه ضعف نظریه به نظر برسد، اما از سوی دیگر، نظریه‌های رقیب خود را متعهد به دیدگاه‌های عمیقاً ضد شهودی و گاه از نظر اخلاقی آزاردهنده درباره کنش درست بر اساس قوانین استثنایی می‌دانند: برای مثال، دیدگاه دین‌شناختی «امانوئل کانت» به طرز بدنامی به موضعی استثنایی متعهد است. هرگز دروغ نگفتن، حتی اگر این کار باعث نجات جان انسان‌ها شود. در مقابل، اخلاق‌دانان فضیلت ممکن است معتقد باشند که نیاز انسان به روابط اعتماد، صداقت را به یک فضیلت تبدیل می‌کند و در عین حال ادعا می‌کنند که ما می‌توانیم تعهد خود را به صداقت حفظ کنیم و در عین حال شرایطی را که دروغ را می‌طلبد مجاز بدانیم. به عنوان مثال، اگر از ما اطلاعات شخص خاصی را به منظور قتل خواستند، دروغ گفتن مناسب است. فقدان قوانین استثنایی نیز ممکن است یک مزیت برای اخلاق فضیلتی در برخورد با فناوری‌های نوظهور باشد.

از آنجایی که فضایل در مرکز اخلاق فضیلت قرار دارند، بسیار مهم است که بدانیم آن‌ها چیستند. برخی از فضایل برتری امیال و احساسات ما هستند، در حالی که برخی دیگر مانند حکمت عملی، در درجه اول برتری‌های فکری هستند. به عنوان مثال، شجاعت، به خواست ما به امنیت مربوط می‌شود و زمانی نشان داده می‌شود که احساس ترس و اعتماد به نفس ما به گونه‌ای باشد که فقط در مواجهه با چیزی که واقعاً خطرناک است، احساس ترس کنیم. ارسطو ایده‌ی فضیلت را با توسل به «آموزه

پست» معروف خود توضیح داد. در یک انسان شجاع، احساس ترس و اطمینان در حالتی میانی بین افراط و کمبود قرار دارد. کسی که احساس ترس بیش از حد می‌کند، از خطر فرار می‌کند و نمی‌تواند به چیزی ارزشمند دست یابد. ما به این افراد برچسب ترسو می‌زنیم زیرا آن‌ها رذیلت بزدلی را نشان می‌دهند.

کسی که احساس ترس بسیار کمی دارد ممکن است بی پروا عمل کند و در تلاش‌های بیهوده‌ای که باید از آن اجتناب می‌شد با جراحت یا مرگ مواجه شود. ویژگی رویکرد ارسطویی این است که ترس، در کنار سایر احساسات، چیزی است که برای خوب زیستن ضروری است. از این گذشته، وقتی احساس ترس می‌کنم، ارزش زندگی و تمامیت جسمی‌ام را به گونه‌ای ثبت می‌کنم که انگیزه‌ای برای عمل ایجاد کند. با این حال، من ممکن است برای زندگی و تمامیت جسمی خود بیش از حد ارزش قائل شوم. از نظر ارسطو، چیزهای مهمتری از زندگی و تمامیت جسمانی من وجود دارد، مانند آزادی شهرم و امنیت دوستان و خانواده‌ام. از این رو، از نظر او، در صورت وجود شانس غیرمعمول برای دستیابی به چنین هدفی، خطر مرگ چیز خوبی است. جنبه دیگری از دیدگاه ارسطو این است که شخص نمی‌تواند شجاعت نشان‌دهد مگر اینکه برای رسیدن به هدفی ارزشمند با ترس روبرو شود. دزدی که به خاطر دزدی با خطر روبرو می‌شود، شجاع نیست. اگرچه شخصیت آن‌ها به گونه‌ای است که مستعد احساس ترس نیستند، اما این حالت شخصیتی در آن‌ها برتری ندارد.

شرارت آن‌ها (دزدان) در حوزه دیگری، توانایی آن‌ها را برای رفتار شجاعانه تضعیف می‌کند. این جنبه دیگری از رفتار شناسان ارسطو است: او از ایده‌ای به نام «وحدت فضایل» دفاع می‌کند که در قوی‌ترین شکل خود بیان می‌کند که برای داشتن یک فضیلت باید همه آن‌ها را داشته‌باشیم. به بیان دیگر، ایده این است که هر رذیله‌ای، توانایی نشان دادن هر فضیلتی را تضعیف می‌کند. با فرض اینکه دولت‌هایی واسطه بین فضیلت و رذیلت وجود دارد، این امر فضایی را برای کمتر از فضیلت کامل بودن در برخی زمینه‌ها باز می‌کند بدون اینکه لزوماً فضیلت ما را در سایر زمینه‌ها تضعیف کند. با ماندن در فضیلت شجاعت به عنوان مثال، می توانیم تعجب کنیم که آیا شجاع بودن خوب است؟ به هر حال، اگر مستلزم این باشد که به خاطر دولت شهرم جانم را به خطر بیندازم، شاید بهتر باشد که

ترسو باشم. اما توجه داشته‌باشید که این دیدگاه بزدلانه جهان را می‌پذیرد: اینکه به هر قیمتی زنده ماندن بهتر است. انسان شجاع دنیا را متفاوت می بیند: بقا وقتی به قیمت آزادی شهر خود یا مرگ یا بردگی دوستان و خانواده‌اش تمام شود، خوب نیست.

پس آیا، ما در، کنارِ هم قرار گرفتن این دو دیدگاه گیج شده‌ایم یا اینکه دیدگاه شخص شجاع تطابق دارد؟ من معتقدم که دیدگاه افراد شجاع برتر است زیرا شجاعت یک ویژگی است که انسان برای زندگی خوب در دنیای خطر به آن نیاز دارد. ما انسان‌ها باید بتوانیم اهداف را، حتی در مواجهه با خطرات به پیش ببریم. این دیدگاه نسخه‌ای از اخلاقی است که بسیاری از ارسطویی‌ها آن را پذیرفته‌اند: اینکه خوبی در انسان، تابعی از نوع حیوانی است که آن‌ها هستند (که این حرف را فقط ارسطویی‌ها می‌گویند). فضائل قوای عقلانی و اشتهایی انسان را کامل می‌کند و این امری عینی است که صفات آن چنین است.

ارسطو در عصری با ساختار اجتماعی بسیار متفاوت و همچنین با فناوری‌های متفاوت زندگی می‌کرد. یقیناً امروزه هیچ یک از اخلاق‌شناسان فضیلت ارسطویی، نظرات او را بدون تعدیل نمی‌پذیرد. تأکید بیش از حد ارسطو بر فضیلت رزمی شجاعت در دیدگاه‌های سیاسی او، باعث چسباندن انگ زن‌ستیزی و نژادپرستی در زمان خود شد. اما چارچوب فلسفی او همچنان بینش را به همراه دارد. اخلاق فضیلت ارسطویی در پرداختن به سؤالات فناوری و علم داده، بر بررسی تأثیر فضیلت بر شخصیت ما تأکید می‌کند: چگونه استفاده از یک فناوری جدید بر تمایلات و تفکر ما تأثیر می‌گذارد؟ اگر یک فناوری ما را وادار می‌کند چیزی به عنوان ویژگی یک عامل شرور فکر یا احساس کنیم، پس این زمینه ای برای انتقاد اخلاقی از فناوری است. از این رو، تمرکز بر این است که چگونه با فناوری زندگی می‌کنیم. ما مجبور نیستیم برای ایجاد شک و تردیدهای اخلاقی در مورد یک فناوری، تأثیر چشمگیری بر جامعه یا نقض وظایف داشته‌باشیم. ما می‌توانیم با بررسی تحریف‌ها و تأثیرات آن بر افکار و احساسات خود به نقد اخلاقی فناوری نزدیک شویم (منظور اینکه فناوری چه تأثیرات بدی بر روی اخلاقیات ما داشته‌است). فناوری‌های جدید ممکن است خواسته‌های اخلاقی جدیدی از ما ایجاد کنند. در چنین مواردی، این پرسش مطرح می‌شود که آیا فضیلت جدیدی لازم است یا صرفاً

تفکر در مورد یک فضیلت سنتی در بستری جدید است. نظر من این است که تمایل بر این است که جنبه‌های فضایل سنتی را دوباره پیکربندی کنند، و انجام این کار ضرری ندارد و ممکن است فایده‌ای داشته باشد، زیرا ممکن است به ما کمک کند تا با دقت بیشتری در مورد موقعیت‌هایی که با آن روبرو هستیم فکر کنیم. به طور خلاصه، اخلاق فضیلت ارسطویی چارچوبی انعطاف‌پذیر برای اندیشیدن در مورد اینکه چقدر با فناوری‌های جدید زندگی می‌کنیم فراهم می‌کند، و نیازی نیست که آن را محکم با دیدگاه‌های باستانی ارسطو در مورد فضایل گره بزنیم.

اگر فرض شود که ما به‌عنوان افراد به تنهایی می‌توانیم ویژگی‌هایی را که برای خوب زندگی کردن در هر شرایطی به آن‌ها نیاز داریم، توسعه دهیم و از خود نشان دهیم، اخلاق فضیلت نادرست درک می‌شود. درعوض، اخلاق فضیلت، مربوط به سنجش شرایط اجتماعی است که برای خوب زیستن انسان‌ها ضروری است. این امر به ویژه در درنظر گرفتن تأثیر فناوری‌های جدید بسیار مهم است. آن‌ها ممکن است توانایی ما را برای تطبیق خواسته‌هایمان با اهداف آگاهانه‌مان تضعیف کنند (یا به‌عنوان خوش‌بین‌تر، تقویت کنند)، و در نتیجه تلاش‌های ما برای توسعه فضایل را تضعیف کنند. از دیدگاه ارسطویی، رشد فضایل مستلزم فرآیند عادت کردن است، یعنی فرآیندی از عمل به گونه‌ای که فاعل نیکوکار عمل می‌کند، شاید بر خلاف تمایلات ما، تا زمانی که از عمل به آن طریق لذت ببریم و بتوانیم آن را با اطمینان انجام دهیم (پس ارسطو میگوید که باید به رفتارهای خوب و نیکو، عادت کنیم).

اخلاق آفریقایی

نوشته‌ی جان مورانگی

در ادامه، باید انتظار دیدِ موقتی درباره‌ی اخلاق آفریقایی داشت. موقتی بودن اهمیت دارد زیرا جایی برای دیدگاه‌های دیگر باقی می‌گذارد. علاوه بر این، خواننده را متوجه این واقعیت می‌کند که آنچه در مورد اخلاق آفریقایی گفته می‌شود، همه‌ی آن نیست. چیزهای بیشتری برای گفتن وجود دارد؛ که از آن صرف نظر می‌کنم. اگر بخواهیم در مورد درک اخلاق آفریقایی عدالت را رعایت کنیم، کنار گذاشتن نژادپرستی بینش مهمی است. اخلاق آفریقایی مانند هر شاخه‌ی دیگری از اخلاق یک اخلاق منحصر به فرد است. نباید آن را با هیچ شاخه دیگری از اخلاق اشتباه گرفت. اخلاق، چه افریقایی یا غیرافریقایی، چه خاص و چه جهانی، در مورد رفاه است. در جوامع بومی آفریقا، رفاه اجتماعی، رفاه اجتماعی است. این بهزیستی است که جایگاهی برای رفاه فردی و همچنین رفاه گروهی دارد (منظور از رفاه که امروزه استفاده می‌شود، پول و جایگاه مادی است). یک جمله‌ی معروف در اخلاق آفریقایی و اوبونتو وجود دارد که برایتان آورده‌ام: ما هستیم، پس من هستم، این نشان‌دهنده‌ی این است که اخلاق آفریقایی برای ما (جمع انسان‌ها) ارزش بالایی قائل است. در اخلاق اوبونتو که زیرشاخه‌ی اخلاق آفریقایی است، هیچ‌وقت ارزش یک فرد، بالاتر از ارزش یک جمع نیست.

این مهم است که به خود یادآوری کنیم که اخلاق آفریقایی تابع قوم نگاری یا قوم شناسی نیست، این اخلاق قومی و قبیله ای نیست. همچنین این قضیه را باید به صورت محکم بیان نمود که کاشفان اروپایی در تاریخ مدرن میگفتند که آفریقایی‌ها وحشی هستند! این باور کاملا غلط و برخواسته از نژادپرستی است!

از آنجا که اخلاق در سعادت جامعه دخیل است، به نظر می‌رسد که جامعه شناسی در مطالعه اخلاق ضروری است. همانطور که جامعه شناسی مطالعه جامعه است، مطالعه اخلاق نیز در جامعه شناسی گنجانده شده‌است. علاوه بر این، از آنجایی که جامعه از نظر سیاسی امنیت دارد و منافع آن

توسط دولت (سیاسی) تبلیغ و پیگیری می‌شود، اخلاق اساساً سیاسی است. به گونه ای دیگر، اخلاق تابع جامعه شناسی سیاسی است. در اخلاق متعارف اروپایی-غربی، معماری چندلایه اخلاق به ندرت به رسمیت شناخته می‌شود. در بافت بومی آفریقا، این معماری به رسمیت شناخته شده‌است.

اخلاق بودایی

نوشته‌ی پیتر هرشوک

اخلاق می‌تواند شامل همه چیز باشد، از تبیین چیزی که به طور ایده‌آل در یک فرد «خوب» دخیل است، تا معنای عملی نمایندگی «قابل قبول» در یک حرفه یا شهروندان یک ملت یا جهان.

من به اخلاق به صورت عملیاتی برخورد می‌کنم و آن را حداقل به‌عنوان هنر ارزشیابی اصلاح مسیر انسانی تعریف می‌کنم: هنر اعمال هوشمندانه نتایج حاصل از تبعیض مشترک و کیفی بین ارزش‌ها، اهداف و علایق و ابزارهای ما برای تحقق آن‌ها. برای من، این هنری است که به طور اساسی با شرح و بسط معاصر مفاهیم و اعمال بودایی آشنا شده است.

بودیسم حدود 2600 سال پیش در دامنه‌های هیمالیا در جنوب آسیا ظهور کرد، تقریباً همزمان با سنت‌های فلسفی و سیاسی جهان مدیترانه و سینیتی. آن سنت‌ها با پرسش‌های بنیادینی دست و پنجه نرم می‌کردند: چه چیزی واقعی است؟ چی خوبه؟ جایگاه انسانیت در کیهان چیست؟ و جامعه چگونه باید اداره شود؟ بودیسم در پاسخ درمانی (به جای نظری) به دو سؤال متفاوت، اما به همان اندازه اساسی، پدید آمد. علل و شرایط ابتلا به دعا یا رنج و درگیری و گرفتاری چیست؟ و با چه وسیله‌ای می توانیم این علل و شرایط را از بین ببریم؟ پاسخ بوداییان به این سؤالات بر دو بینش کلیدی استوار است. اولاً، همه چیز به طور متقابل به وجود می‌آید و ادامه می یابد. به طور قوی بیان

می‌شود که رابطه گرایی اساسی تر از چیزهای مرتبط است. همه چیز تابعی از تمایز رابطه‌ای است، و هر چیز در نهایت همان چیزی است که برای دیگران معنا می‌کند. ثانیاً، کیهان ما خودسازمانده و دارای ساختار کرمی است. این کیهانی است که در آن الگوهای ثابت ارزش‌ها، نیات و اعمال منجر به الگوهای همخوانی از نتایج و فرصت‌های تجربی می‌شود.

هدف هنر بودایی اصلاح سیر انسانی، تحقق آزادی از درهم تنیدگی‌های رابطه‌ای است که دخخا ایجاد می‌کند، عمدتاً از طریق حل تعارضات بین ارزش‌ها، نیات و اعمال ما. این بستگی به ارزیابی انتقادی عادات فکر، گفتار، و رفتار، و تحقق آزادی توجه و آزادی نیت مورد نیاز برای تجدید نظر، مقاومت، یا انحلال آن عادات در صورت لزوم دارد تا دیگر توسط درهم تنیدگی‌های کارمایی و حضور اجباری محدود نشوند. به طور قابل توجهی، هدف تمرین بودایی (هدف نیروانا) تجویز یا تعریف نشده است. در عوض، به طور سنتی به صورت استعاری به عنوان خنک‌کننده یا خاموش‌کننده آتش ولع، بیزاری، و جهل تلقی می‌شد. این پیامدهای مهمی برای اخلاق بودایی دارد. به طور خلاصه، اخلاق بودایی هدف یا مقصد نیست. یک هنر بی پایان و بداهه است. اخلاق بودایی را می‌توان با برخی توجیهات، شامل عناصری از رویکردهای مبتنی بر فضیلت، وظایف (دئونتولوژیک) و مبتنی بر پیامد (فایده‌گرا) به اخلاقی دانست که در فلسفه غرب غالب شده‌اند، و همچنین رویکردهای مراقبت محوری مانند فمینیستی. با این حال، هستی شناسی رابطه ای بودایی به طور مشخص توجه ارزیابی را از عوامل اخلاقی، بیماران و اعمال مستقل و به سمت کیفیت رابطه ای سوق می‌دهد. علاوه بر این، در حالی که تأکید بودیسم بر فضیلت‌گرایی رابطه‌ای، اخلاق بودایی را متعهد به شرایط خاص می‌کند، با اخلاق موقعیتی غربی که اعمال را بر اساس نتایج نزدیک یا کوتاه‌مدت ارزیابی می‌کند، متفاوت است. آنچه از نظر اخلاقی اهمیت دارد صرفاً پیامدهای فوری یک عمل نیست، بلکه پیامدهای رابطه‌ای میان‌مدت و بلندمدت اجرای عمدی مجموعه‌های ارزش‌های خاص و شکل‌دهی آن‌ها به فرصت‌های ارادی و نیز نتایج تجربی است.

اخلاق بومی و فطری: کنش‌ها به مثابه تعامل

نوشته‌ی جوزف لن میلر و آندریا سالیوان کلارک

پاسخ به این سؤال که یک نظریه اخلاقی بومی چگونه است دشوار است. اول، مشکل «پان ایندیانیسم» وجود دارد. با توجه به تفاوت‌هایی که بین قبایل وجود دارد، اندیشیدن به مردم «بومی» به‌عنوان یک گروه همگن مشکل‌ساز است. دوم، از نظر تاریخی، اندیشه فلسفی مردم بومی به طور جدی دست کم گرفته شده‌است. اکثر متفکران غربی فرض کرده‌اند که مردم بومی آنقدر بدوی یا حتی «وحشی» بودند که نمی‌توانستند در مورد موضوعات یا پرسش‌های انتزاعی تأمل کنند. این تاریخ تأثیرات ماندگاری بر فلسفه بومی دارد. نه تنها ایده‌های بومی، حتی بنیادی‌ترین آن‌ها، باید بر اساس استانداردهای غربی «توجیه» شوند، بلکه این ایده‌ها باید در زمینه‌ای غیر از آنچه در آن شکل گرفته‌اند، توضیح داده‌شوند. همانطور که گفته شد، یکی از تمرکز مشترک مهم اخلاق بومی، به هم پیوستگی همه چیز است (مانند مردم، زمین، حیوانات غیر انسانی، نسل‌های گذشته و آینده و غیره). کیهان موجودی زنده است و درک می شود که در «گذار دائمی» است. این موضوع زمینه ای را برای مردم بومی فراهم می‌کند که «بر اساس اصول تعادل هماهنگی عمل می‌کند». مردم در اجتماع و روابط متولد می شوند. این‌ها شامل روابط غیرانسانی مانند ارواح، صخره‌ها، رودخانه‌ها، اعضای گونه‌های حیوانی غیرانسانی و غیره می‌شوند. هر موجودی که ما با آن رابطه داریم متفاوت است، و بنابراین اقدامات ما نسبت به روابطمان نیز متفاوت خواهدبود. به جای ارائه اصول جهانی برای هدایت رفتار، مفاهیم کلیدی وجود دارد که پایه و راهنمایی را برای تصمیم گیری اخلاقی فراهم می‌کند. این مفاهیم عبارتند از هماهنگی، متقابل، سپاسگزاری و فروتنی. درک چگونگی ارتباط این مفاهیم با یکدیگر می‌تواند به درک بهتر نحوه اجرای این مفاهیم در زمینه‌های مختلف کمک کند. روش صحیح زندگی، و عمل، سپس با آنچه ما از روابط خود و ارتباط ما با این مفاهیم می‌دانیم، آگاه می‌شود.

هماهنگی زمانی وجود دارد که بین مبادلات و تعاملات با محیط اطراف فرد تعادل وجود داشته‌باشد. تعادل و هماهنگی، ویژگی‌های دنیایی است که ما در آن متولد شده‌ایم، راهنمایی برای اطمینان از رفاه روابط ما و خودمان است. با توجه به وابستگی متقابل و روابط بین همه‌چیز، هر تعاملی بر رفاه یک فرد و محیط اطراف او تأثیر می‌گذارد. به عبارت دیگر، برای هر کنش، واکنشی است. برای ایجاد تعادل در این تعاملات، یک فرد باید بداند که چگونه متقابلا عمل کند. تعامل متقابل می‌تواند اشکال مختلفی داشته‌باشد (یعنی یک روش "درست" منحصر به فرد برای انجام متقابل وجود ندارد)، اما باید متناسب با موجودی باشد که فرد با او در تعامل است. هدف متقابل ایجاد تعادل در روابط است تا همه موجودات درگیر بتوانند به طور مسالمت آمیز با هم زندگی کنند. برای زندگی مسالمت آمیز با محیط اطراف، و رفتار متقابل مناسب، باید با عشق، سپاسگزاری و فروتنی رفتار کرد.

با در نظر گرفتن این مفاهیم، برای هر کنش (فعل) خاص باید سؤالات زیر را در نظر گرفت: چه عملی هماهنگی ایجاد می‌کند؟ چگونه باید آنچه را که به من داده‌اند جبران کنم؟ آیا با عشق، سپاسگزاری و فروتنی رفتار می‌کنم؟ توجه داشته باشید، پاسخ به این سؤالات به شدت به محیط و زمینه فرد بستگی دارد. پاسخگویی مناسب به این سؤالات مستلزم داشتن شناخت دقیق از محیط و روابط خود است. به عنوان مثال، دانستن چگونگی ایجاد هماهنگی (یعنی دانستن نحوه انجام رفتار متقابل) در رابطه با زمین مستلزم دانستن جزئیات دقیق در مورد خاک، زندگی گیاهی، بدنه‌های آبی، الگوهای آب و هوا، وابستگی متقابل بین گیاهان و حیوانات در منطقه است و غیره. برخی از مفاهیم سیاسی تا حدی به عنوان وسیله ای برای حفظ شیوه‌های زندگی که در حضور استعمار شهرک نشینان حول این مفاهیم شکل می گیرد، نقش برجسته تری در اخلاق بومی ایفا کردند. این شامل مفاهیم حاکمیت و احیاء است. از آنجایی که تمرکز این مجموعه اخلاق است، مفاهیم اساسی اخلاقی را که حاکی از تصمیم گیری اخلاقی در فلسفه بومی است، در اولویت قرار داده‌ایم. با این حال، با توجه به اهمیت و الهام‌بخش تبلیغات اخیر در مورد حاکمیت داده‌های بومی، ما از به اشتراک گذاشتن منابعی که نشان می‌دهند چگونه این مفاهیم (حاکمیت و احیا) در جمع‌آوری و استفاده از داده‌های مربوط به مردم بومی استفاده می‌شوند، خودداری می‌کنیم.

«کوکوتای» و «تیلور» اخیراً جلدی را ویرایش کرده‌اند که مقالاتی را در حمایت از «حقوق و منافع ذاتی و غیرقابل انکار مردمان بومی در ارتباط با جمع‌آوری، مالکیت، و کاربرد داده‌های مربوط به مردم، شیوه‌های زندگی و سرزمین‌هایشان» جمع‌آوری کرده‌اند. «رودریگز-لون بیر» و «مارتینز» استدلالی را در حمایت از «تغییر موقعیت اقتدار بر داده‌های بومی به مردم بومی» ارائه می‌کنند. «کارول» و همکاران نمونه‌هایی از اصول *(CARE)* برای حاکمیت داده‌های بومی (منافع جمعی، اختیار کنترل، مسئولیت و اخلاق) را بیان، توصیف و ارائه کنید.

به طور کلی، مردم بومی با فروتنی به این سؤال می پردازند که چگونه خوب زندگی کنند، زیرا می‌دانند که ما تنها بخش کوچکی از جهان هستیم. ما برای زنده ماندن به رفاه و سخاوت خویشاوندان خود (یعنی همه روابطمان) وابسته هستیم. اختلال در کار، هرج و مرج، بی نظمی و زوال رفاه بستگان ما ناهماهنگی ایجاد می‌کند و نشان دهنده این است که اعمال ما نادرست است و باید راه خود را تغییر دهیم.

فصل ۳

اخلاق تحقیق و روش علمی

اخلاق تحقیق و روش علمی برایان وانسینک اجازه نمی‌دهد شکست یک گزینه باشد. اگر داده‌های جالبی داشته باشد، به آن ادامه می‌دهد تا زمانی که چیزی پیدا کند، سپس منتشر می‌کند، منتشر می‌کند، منتشر می‌کند.

Andrew Gelman, Statistician

"یک ترفند ساده": آزمایشگاه غذا و برند کورنل

آیا می‌دانستید اگر در رستوران مورد علاقه خود کنار پنجره بنشینید، 80 درصد بیشتر احتمال دارد که سالاد انتخاب کنید؟ یا اینکه اگر نزدیک میله بنشینید (در نور کم و با پخش موسیقی بلند در پس زمینه) کالری بیشتری مصرف می کنید؟ آیا می‌دانستید افرادی که جعبه‌های غلات خود را بیرون پیشخوان نگه می‌دارند به طور متوسط 21 پوند وزن بیشتری نسبت به کسانی دارند که آن‌ها را در کمد پنهان می‌کنند؟ یا اینکه برند کردن سیب با شخصیت‌های کارتونی محبوب، مانند المو، باعث می‌شود که بچه‌ها با ناهار یکی از آن‌ها را به جای شیرینی انتخاب کنند؟ یا اینکه مردها وقتی خانم‌ها آن‌ها را تماشا می‌کنند بیشتر غذا می خورند (اما وقتی مردها آن‌ها را تماشا می‌کنند خانم‌ها کمتر می‌خورند؟ یا اینکه ایجاد یک «پوز قدرت» تأثیر مثبتی بر مصاحبه‌های شغلی، مذاکرات و سایر عملکردها دارد) به‌ویژه برای کسانی که موقعیت اجتماعی پایین‌تری دارند و منابع کمتری دارند؟

اگر به همه‌ی یا هر یک از این سؤالات «نه» پاسخ داده‌اید، می‌توانید به خودتان تبریک بگویید،

زیرا حق با شماست. ادعاهای مطرح شده توسط محققان در مطالعات فوق (که همگی زمانی به طور برجسته در رسانه‌ها تبلیغ می‌شدند) قابل تکرار نبودند و از آن زمان پس‌گرفته شدند. کار ایمی کادی روی ژست‌های قدرتی موضوع دومین سخنرانی پربیننده TED تا به حال بود، و حتی قبل از رد شدن، بخشی از حکمت عامیانه فرهنگی دریافتی ما شد. ادعاهای دیگر نیز راه خود را به عقل عامیانه علاقه مندان به آخرین اخبار رژیم غذایی و سلامتی (از جمله کسانی که مسئول تصمیم‌گیری در مورد برنامه‌های ناهار مدارس دولتی هستند) باز کرد. آن‌ها نیز پس از یافته‌های مربوط به تخلفات تحقیقاتی، همه‌ی آن‌ها پس‌گرفته شده‌اند.

این مطالعات محصول «برایان وانسینک» از دانشگاه «کرنل» *(Cornell University)* بود، جایی که او روانشناسی غذا خوردن را در آزمایشگاه غذا و برند «کورنل» خود مطالعه کرد. «وانسینک»، *(Food & Brand Lab)* را در سال 1997 در دانشگاه «ایلینویز» *(Illinois)* تأسیس کرد و در سال 2005 آن را به *«Ivy League»* منتقل کرد. آزمایشگاه *(Food & Brand)* بیشتر بودجه خود را از شرکت‌های مواد غذایی دریافت کرد. آزمایش‌های «وانسینک» نه تنها از بودجه خوبی برخوردار بودند، بلکه از محبوبیت بالایی برخوردار بودند. کتاب او با نام «غذا خوردن بی فکر: چرا بیشتر از آن چیزی که فکر می کنیم می خوریم» در سال 2006 در فهرست پرفروش ترین‌های نیویورک تایمز قرارگرفت. فلسفه او کاملاً با حکمت رایج در آن زمان متفاوت بود: وانسینک معتقد بود که به جای اینکه به مردم درباره فواید خوب آموزش دهد، با انتخاب‌های غذایی و خطرات افراد فقیر، او می توانست مردم را وادار کند تا ترفندها و عاداتی را به کار گیرند که آن‌ها را به سمت بهتر غذا خوردن سوق می‌دهد، بدون اینکه زیاد فکر کنند، یا مجبور باشند به هیچ وجه در مورد انتخاب‌هایشان منطقی باشند. او در سال 2015 به «کیرا باتلر» از «مادر جونز» گفت: «میلیون‌ها متخصص تغذیه وجود دارند که به شما می‌گویند به جای شکلات اسنیکرز، یک سیب بخورید، اگر واقعاً می‌خواهیم بهتر غذا بخوریم، باید مغزمان را فریب دهیم».

با این حال، دانشمندان دیگر شروع به ابراز نگرانی در مورد روش‌های تحقیق وانسینک کردند، از جمله «تناقض داده‌ها، غیرممکن‌های ریاضی، اشتباهات، تکراری‌ها، اغراق‌ها، تفسیرهای تعجب‌آور،

و مواردی از سرقت ادبی خود در 50 مطالعه او» که بسیاری از آن‌ها نشان داده‌شده و از آن زمان پس‌گرفته شده است. این‌ها شامل چندین مقاله است که توضیح می‌دهد چگونه ارائه جذاب غذاهای سالم در کافه تریاهای مدرسه باعث تشویق دانش آموزان به انتخاب میوه و سبزیجات بیشتر می‌شود. برنامه‌های مبتنی بر نشریات لغو شده وانسینک در 30000 مدرسه ایالات متحده اتخاذ شده است که میلیون‌ها دلار بودجه دولتی برای جنبش ناهارخوری‌های هوشمندتر جذب کرده‌است. این برنامه‌ها عمدتاً شامل دادن نام‌های تند و جذاب و برندهای رنگارنگ به غذای سالم بود، مانند «آب‌میوه‌گیر پرتقال»، «تلفن میمون یا موز»، «سیب لذیذ»، «برش‌های خنک خیار» و «پای شیرین سیب‌زمینی‌ها».

شکاف‌های تحقیق در اوایل قابل مشاهده بودند، اما به دلیل پست وبلاگی توسط خود وانسینک (که باید یکی از پیامدترین اقدامات غرورآفرین در تاریخ علم باشد)، به اوج خود رسیدند. در وبلاگ، وانسینک یک مجموعه داده‌ی اصلی را که طی چند هفته مشاهده در یک رستوران پیتزا در شمال نیویورک جمع‌آوری شده‌است، مورد بحث قرار می‌دهد. او خاطرنشان می‌کند که طرح تحقیق اولیه به نتیجه نرسید، بنابراین او به دنبال استخراج داده‌ها برای برخی از نتایج تحقیقات جدید "خوب" بود. او سپس به شدت از پسادکتری (با پول) خود به دلیل امتناع از کار با داده‌ها انتقاد کرد، در حالی که یک دکتر (بدون حقوق) از ترکیه، داده‌ها را استخراج کرد و در نهایت پنج مقاله مختلف منتشر کرد (که البته اکنون "مقاله‌های پیتزا" (اسم مقاله)، بدنام هستند. وانسینک به خلاقیت و ابتکار محقق ترک در تهیه‌ی این همه داده تبریک گفت و اظهار داشت: «با اینکه من با پسادکتری دانشگاه را ترک کردم، ولی به اندازه‌ی یک‌چهارم شما مقاله چاپ کردم». وانسینک به محقق ترکی، غبطه می‌خورد. تیم «ون درزی» از دانشگاه «لیدن در هلند»، یکی از اولین دانشمندانی بود که پست وبلاگ وانسینک را خواند و از سوء رفتار احتمالی در «مقاله‌های پیتزا» سخن گفت. مطالعات روی مقاله‌های پیتزا پس‌گرفته‌شده در یک رستوران بوفه‌ای به نام رستوران ایتالیایی *Aiello* در حدود 30 مایلی کرنل انجام شد. نمونه شامل حدود 130 بزرگسال بود که در یک دوره دو هفته‌ای در رستوران غذا خورده بودند.

نویسندگان خاطرنشان کردند که عدم بیان اینکه داده‌ها، همگی از یک مطالعه میدانی که قبلاً منتشر شده‌است، جمع‌آوری شده‌اند، باعث می‌شود که اعتبار آزمایشات از بین برود و درضمن این نکته، در هیچ‌یک از مقاله‌ها چاپ نشده بود! هنگامی که آن‌ها از وانسینک درخواست کردند، از دسترسی به داده‌های اصلی نیز محروم شدند. آن‌ها خاطرنشان کردند که حجم نمونه بین مقالات ناسازگار است، و نشان می‌دهد که برخی از شرکت کنندگان در برخی از مقالات گنجانده شده‌اند، و در برخی دیگر حذف شده‌اند! «ون درزی» همچنین به چندین اشتباه دیگر در این مقاله اشاره کرد:

> انواع خطاها عبارتند از: اندازه‌های نمونه غیرممکن در داخل و بین مقالات، آمارهای آزمایشی محاسبه‌شده و/یا گزارش‌شده نادرست و درجات آزادی، و تعداد زیادی از میانگین‌های غیرممکن و انحرافات استاندارد. در مجموع، ما تقریباً 150 تناقض و عدم امکان را در این چهار مقاله شناسایی کردیم. در مجموع، این مشکلات اعتماد به نتیجه گیری نویسندگان را دشوار می‌کند.

در ابتدا، «وانسینک» اشتباهات را جزئی و انتقادات را به عنوان "زورگویی سایبری" رد کرد، اما درخواست‌ها برای تحقیق کامل در مورد تحقیقات او افزایش یافت. «اندرو گلمن»، آماردان برجسته در دانشگاه کلمبیا، سپس در یک پست وبلاگی تند، وانسینک را صدا زد. گلمن اظهار داشت: «آنچه برایان را توصیف می‌کنید شبیه به *(p-hacking)* و *(HARKing)* است. مشکل این است که اگر فرضیه اصلی شما احتمالی کمتر از ۵۰ درصد داشت، احتمالاً تمام این تحلیل‌های زیر گروهی و داده‌های عمیق را انجام نمی‌دادید». در اینجا، «گلمن» به فرآیند «فرضیه‌سازی پس از مشخص شدن نتایج» *(HARKing)* اشاره می‌کند (در این مورد، به نظر می‌رسد که فرضیه اصلی وانسینک هیچ پشتیبانی پیدا نکرده است، بنابراین داده‌ها به سادگی توسط پست دکتر ترکیه استخراج شد تا ببیند آیا برخی تداعی‌های قابل قبولی پیدا شد). بل توصیه می‌کند که محققان می‌توانند با اعلام «فرضیه‌های با انگیزه‌ی واضح، در کنار پیش‌بینی‌های ابطال‌پذیر، قبل از آزمایش، از موارد *HARKing* اجتناب کنند». این امر در بسیاری از زمینه‌ها، از جمله یادگیری ماشینی، از طریق ثبت پیش‌ثبت آزمایش‌ها،

از جمله فرضیه‌ها، داده‌ها، تجزیه و تحلیل و طراحی آزمایشی انجام می‌شود. مخزن *OpenML* نمونه خوبی از حرکت به سمت علم باز است.

با استفاده از روش *p-hacking*، گلمن به روش بی اعتبار ماساژ داده‌ها اشاره می‌کند (به عنوان مثال با بازی با اندازه‌های نمونه) برای ایجاد یک نتیجه به ظاهر آماری مهم در جایی که هیچ کدام واقعا وجود ندارد. *p-hacking* نیز اعتبار مدل‌ها را به خطر می اندازد زیرا "فرض اصلی یک آزمون فرضیه آماری را باطل می‌کند: احتمال اینکه یک نتیجه منفرد به دلیل شانس باشد". *p-hacking* می‌تواند ما را به پذیرش نتایج معتبری که صرفاً تصادفی هستند سوق دهد. *p-hacking* به *HARKing*، لایروبی داده‌ها و گزارش نتایج بسیار مهم به عنوان شیوه‌هایی می پیوندد که مدل‌های نامعتبر را در یادگیری ماشین نیز تولید می‌کنند. مجموعه داده‌های بزرگی که در یادگیری ماشین استفاده می‌شوند، به‌ویژه در ایجاد نتایج مثبت نادرست هستند (برای تعاریف جنبه‌های اصلی روش علمی به کادر 3.1 مراجعه کنید). گلمن پست وبلاگ خود را با بیان این جمله به پایان رساند: «از جمله آخری که رزومه «همیشه پنج مقاله خواهد داشت» آزارم می‌دهد. وضعیت نهایی تحقیق رزومه نیست.

جعبه 3.1

روش علمی

تکرارپذیری:

نتایج به دست آمده در یک کارآزمایی یا آزمایش زمانی مشابه خواهد بود که در شرایط مشابه تکرار شود، که نیاز به مستندسازی توسط محققین به گونه ای کامل و همچنین شفاف دارد. همچنین به عنوان تکرارپذیری و تکرارپذیری شناخته می‌شود.

قابلیت اطمینان:

معیاری برای قابلیت اطمینان. همچنین به عنوان قابلیت اطمینان تست/آزمون مجدد شناخته می شود. فردی که چندین بار در آزمون شرکت می‌کند، تا حد زیادی پاسخ‌های مشابهی می‌دهد. سیستمی که چندین بار در شرایط یکسان اجرا می شود، نتایج تا حد زیادی در طول زمان ایجاد می‌کند.

دقت:

اندازه‌گیری‌ها یا آزمایش‌هایی که نتایجی شبیه به یکدیگر ایجاد می‌کنند.

صحت:

اندازه گیری خطا بین اندازه گیری‌های متوسط و مقدار واقعی.

اعتبار:

میزانی که یک مدل یا اندازه‌گیری ادعا شده دقیقاً آنچه را که ادعا می‌کند منعکس می‌کند.

این رسوایی پایان کار برای «وانسینک» بود. دانشمندان دیگر شروع به درخواست داده‌های اصلی در مطالعات ناهار مدرسه کردند، اما هیچ کدام یافت نشد. سپس نام تجاری و کاغذ ناهار مدرسه نیز پس گرفته‌شد. در سپتامبر 2018، وانسینک پس از تحقیقاتی که در کورنل انجام شد، بازنشسته‌شد و نشان‌داد که او واقعاً مرتکب تخلفات تحقیقاتی، از جمله گزارش نادرست داده‌ها، داده‌های از دست‌رفته، خطاهای آماری و اسناد نامناسب نویسندگی شده‌است. سال قبل، تحقیقات آن‌ها "خطا" پیدا کرده بود، اما "سوء رفتار" وجود نداشت.

انتقادات از تحقیقات وانسینک در زمان حساسی برای بحران تکرار مطرح شد و سینگال اظهار داشت که او یکی از تراژدی‌های بزرگ آن بحران بود. وانسینک و آزمایشگاه او، ناشران پرکارِ مطالعات جلب توجه بودند (روشی که اغلب منجر به خطاهای کنترل کیفیت می شود، مانند آنچه در اینجا دیدیم). بحران تکرارپذیری، البته، بسیار بیشتر از تکرارپذیری است. این در مورد ماهیت خود روش علمی کادر 3.1 و معنای تولید نظریه‌ها، مدل‌ها و دانشی است که تصویری عینی درست از واقعیت ارائه می‌دهد. بسیاری از نتایج در روانشناسی، پزشکی و علوم اجتماعی قابل تکرار نیستند (و بنابراین احتمالاً نیز نامعتبر هستند) (به کادر 3.2 مراجعه کنید).

جعبه 3.2

چک لیست تکرارپذیری

برای همه مدل‌ها و الگوریتم‌های ارائه شده، بررسی کنید که آیا شامل موارد زیر است:

* شرح واضحی از تنظیمات ریاضی، الگوریتم و/یا مدل.
* توضیح واضح در مورد هر فرضی.
* تجزیه و تحلیل پیچیدگی (زمان، مکان، اندازه نمونه) هر الگوریتم.

برای هر ادعای نظری، بررسی کنید که آیا شامل موارد زیر است:

* بیان واضح ادعا.
* اثبات کامل ادعا.

برای همه مجموعه داده‌های مورد استفاده، بررسی کنید که آیا شامل موارد زیر است:

* آمار مربوطه، مانند تعداد نمونه.
* جزئیات تقسیم قطار/ اعتبارسنجی/آزمایش.
* توضیحی در مورد هر داده ای که حذف شده است، و تمام مراحل قبل از پردازش.
* پیوندی به نسخه قابل دانلود مجموعه داده یا محیط شبیه سازی.
* برای داده‌های جدید جمع‌آوری شده، شرح کاملی از فرآیند جمع‌آوری داده‌ها، مانند دستورالعمل‌ها به حاشیه نویس‌ها و روش‌های کنترل کیفیت.

برای همه کدهای مشترک مرتبط با این کار، بررسی کنید که آیا شامل موارد زیر است:

* تعیین وابستگی‌ها.
* کد آموزشی.

* کد ارزیابی.
* مدل(های) (از قبل) آموزش دیده.
* فایل *README* شامل جدولی از نتایج است که با دستور دقیق اجرا برای تولید آن نتایج همراه است.

برای همه‌ی نتایج آزمایشی گزارش شده، بررسی کنید که آیا شامل موارد زیر است:

* محدوده فراپارامترهای در نظر گرفته شده، روش انتخاب بهترین پیکربندی‌هایپرپارامتر، و مشخصات تمام پارامترهای‌هایپر مورد استفاده برای تولید نتایج.
* تعداد دقیق دوره‌های آموزشی و ارزیابی.
* تعریف روشنی از معیار یا آمار خاص مورد استفاده برای گزارش نتایج.
* شرح نتایج با گرایش مرکزی (مثلاً میانگین) و تنوع (مثلاً نوارهای خطا).
* میانگین زمان اجرا برای هر نتیجه، یا هزینه انرژی تخمینی.
* شرح زیرساخت محاسباتی مورد استفاده.

منبع: *Pineau, Joelle*. چک لیست تکرارپذیری یادگیری ماشین (نسخه 2.0، 7 آوریل 2020).
www.cs.mcgill.ca/ jpineau/ReproducibilityChecklist-v2.0.pdf

بنابراین، بحران تکرارپذیری به روش‌شناسی ضعیف و همچنین فقدان اعتبار اشاره دارد: نتایج حاصل از روش‌شناسی غیراخلاقی منجر به مدل‌هایی می‌شود که معتبر نیستند (و بنابراین اطلاعات قابل اعتمادی در مورد دنیای واقعی به ما نمی‌دهند). این کتاب چندین مطالعه موردی را مورد بحث قرار می‌دهد که این می‌تواند منجر به آسیب‌های قابل توجهی شود (از جمله محکومیت‌های نادرست، بازداشت‌های غیرضروری، آزادی افراد خطرناک در شرایط نامناسب، و حتی نسل‌کشی،

پاک‌سازی قومی، و خشونت سیاسی). البته اخلاق تحقیق تنها زمانی به نتایج معتبر و مستحکمی منجر خواهدشد که خود، حوزه‌ی فرهنگی ایجادکند که به اخلاق علمی و دقت روش‌شناختی اهمیت می‌دهد. روانشناسان دریافته‌اند که پرورش فرهنگ تحقیق اخلاقی نه تنها به اطمینان از تکرارپذیری بلکه اعتبار واقعی نتایج منتشرشده کمک می‌کند. روش‌شناسی خوب همچنین باعث ایجاد اعتماد در بین محققان می‌شود. همانطور که «هایل» خاطرنشان می‌کند، "هیچ دانشمندی نمی‌تواند از هر مقاله‌ای که می‌خواند، نتایج را بازتولید کند" و تعداد بسیار کمی از مقالات منتشرشده حتی یک تلاش برای بازتولید مشاهده خواهندکرد. بقیه را ما اعتماد می‌کنیم.

جنچوغلو *(Gencoglu)* به این نکته اشاره می‌کند که ما به بسیاری از مطالعات موردی که در ادامه می‌آیند باز خواهیم گشت: یک فرهنگ تحقیقاتی دقیق در یادگیری‌ماشین باید «به نیازهای انسان و روانشناسی به شیوه‌ای واقع‌بینانه رسیدگی کند». برای انجام این کار، «متخصصان سطح بالا باید از ابتدا در تیم‌های مطالعاتی گنجانده شوند»، به‌ویژه به‌عنوان تلاش‌های یادگیری‌ماشین در زمینه‌هایی که مدت‌هاست تخصص خود را توسعه داده‌اند (شواهد پزشکی قانونی، ارزیابی خطر در جرم‌شناسی، بیومتریک، اثرات رسانه‌ها، و قوانین آزادی بیان، از جمله).

در پایان، هیچ ترفند ساده‌ای وجود ندارد که اطمینان حاصل کند که تحقیقات به ما دانش معتبر می‌دهد و تصویری دقیق و مفید از واقعیت ارائه می‌دهد که ما در تلاش برای درک و مدل سازی آن هستیم (همانطور که هیچ ترفند ساده ای برای یادگیری نحوه انجام آن وجود ندارد). سالم غذا بخورید، تصمیم بگیرید که چه محتوایی باید در رسانه‌های اجتماعی ممنوع شود، یا اینکه در یک محاکمه جنایی گناه و بی‌گناهی را تعیین کنید. در زمینه‌ای جوان و به سرعت در حال رشد مانند یادگیری ماشین، فرهنگی لازم است که روش‌های قوی و اعتبار مدل‌ها را ارج می‌نهد (فرهنگی که در مورد تولید دانشی که نیازهای مردم را برآورده می‌کند و در آزمون زمان مقاومت می‌کند تأمل می‌کند).

در ادامه، یک تفسیر از رفتار سودگرا را خدمت شما ارائه میدهیم.

تفسیر

اخلاق سودگرا

توسط «پیتر سینگر» و «ییپ فای تسه»

از دیدگاه فایده‌گرا، رفتار «وانسینک» غیراخلاقی است، زیرا خطر عواقب منفی بیشتری را نسبت به منافع بالقوه ایجاد می‌کند. حوزه علمی را تصور کنید که اکثریت یا حتی بخش قابل توجهی از پزشکان از نظر فکری صادق نیستند. تحقیق در آن زمینه قابل استناد نیست.

به نظر می‌رسد «وانسینک» یک دستور کار در پشت تحقیقات خود دارد: او می‌خواست مردم به روش خاصی (سالم، همانطور که او معتقد بود) غذا بخورند. این ممکن است دلیلی باشد که او فقط از نتایجی حمایت می‌کند که نظرات او را تأیید می‌کند. اینکه آرزو کنیم مردم به روش خاصی غذا بخورند، البته لزوماً بد نیست. و ممکن است، شاید به احتمال زیاد، نیت او خیر بوده باشد. اما نیت خوب، ناصادق بودن را توجیه نمی‌کند (به عبارت دیگر، حسن فاعلی داشته، ولی حسن فعلی نداشته).

داشتن نیت خیر به خودی خود برای اخلاقی عمل‌کردن کافی نیست. همچنین باید به روشی مبتنی بر شواهد، تجربی و نظری درست عمل کرد. یک فرد با نیت خوب، با یافتن شواهد یا دلایلی علیه دستور کار خود، نیاز به ارزیابی مجدد دارد، و شاید، اگر دلایل به اندازه کافی قوی باشد، دستور کار خود را رد کند.

نادیده گرفتن شواهد و استدلال‌ها علیه دستور کار خود، ممکن است نیت خوب را به خیال‌پردازی‌های خودفریبی تبدیل کند. همچنین می‌تواند آسیب جدی، احتمالاً در مقیاس وسیع، ایجاد کند. در مورد وانسینک، او بسیار بیشتر از شغل خود و شهرت رشته و موسسه خود به خطر انداخت. او همچنین خطر ارائه توصیه‌های ناآگاهانه در مورد عادات غذایی را داشت و در نتیجه به کسانی که از توصیه‌های

او پیروی می‌کردند آسیب می‌رساند.

صداقت فکری تنها شرط اخلاقی نیست. محققان، به‌ویژه آن‌هایی که روی پروژه‌هایی کار می‌کنند که به طور بالقوه می‌توانند به زندگی موجودات ذی‌شعور آسیب بزنند، از نظر اخلاقی مسئول تأثیرات قابل پیش‌بینی تحقیقات خود هستند. به عنوان مثال، تأثیر تحقیق در زیست شناسی می‌تواند قابل‌توجه باشد، زیرا اغلب پیامدهای عمده‌ای بر بسیاری از انسان‌ها و حیوانات دارد.

نگرانی اخیر در مورد استفاده از فناوری *(CRISPR)* برای قادر ساختن تروریست‌ها به اصلاح ویروس‌ها برای اهداف حمله، تنها نمونه‌ای از این است که چگونه بیوتکنولوژی می‌تواند تأثیرات عظیمی ایجاد کند.

علم داده، حداقل به اندازه‌ی زیست‌شناسی تأثیر مورد انتظار دارد. مهم است که محققان قبل از انتشار، یا حتی بهتر از آن، حتی قبل از انجام تحقیقات خود در زمینه‌های خاص، در مورد پیامدهای اخلاقی کار خود به دقت فکر کنند.

فصل ۴

مدل‌های ماشین در دادگاه

نتیجه‌گیری‌های علمی در معرض تجدیدنظر دائمی هستند. از سوی دیگر، قانون باید اختلافات را به سرعت و در نهایت حل کند. پروژه علمی با بررسی گسترده و گسترده انبوهی از فرضیه‌ها پیش می‌رود، زیرا فرضیه‌هایی که نادرست هستند، در نهایت نشان‌داده‌می‌شوند که چنین هستند، و این خود یک پیشرفت است. با این حال، حدس‌هایی که احتمالاً اشتباه هستند، در پروژه دستیابی به یک قضاوت حقوقی سریع، نهایی و الزام آور (اغلب دارای پیامدهای بزرگ) در مورد مجموعه خاصی از رویدادهای گذشته کاربرد چندانی ندارند.

(قاضی «بلکمون»، «دابرت» علیه داروسازی «مرل داو»، دادگاه عالی ایالات متحده، 1993)

محاکمه‌های شفاهی نیکلاس هیلاری

در «24 اکتبر 2011»، یک قتل عجیب و وحشتناک در پوتسدام، نیویورک (شهری کوچک در کنار رودخانه سنت لارنس و بسیار نزدیک به مرز با استان انتاریو کانادا) اتفاق افتاد. پسر 12 ساله‌ای به نام «گرت فیلیپس» حوالی ساعت 5 بعدازظهر، مدت کوتاهی پس از بازگشت از مدرسه به خانه، در اتاق خوابش خفه‌شد. همسایه‌ها غوغایی را شنیدند و با 911 (پلیس آمریکا) تماس گرفتند. پرده‌های پنجره اتاق خواب طبقه دوم «گرت» به سمت بیرون خم شده‌بود و باعث‌شد تا بازرسان شک کنند که قاتل از آن طرف بیرون‌پریده و فرارکرده است.

پلیس به سرعت در مورد «اورال نیکلاس (یا به اختصار نیک) هیلاری» به عنوان مظنون اصلی خود در این پرونده و به دلایلی که کاملاً واهی به نظر می رسد، مظنون شد. هیلاری مربی فوتبال تیم مردان در «دانشگاه کلارکسون» بود. هیلاری، بسیار موفق و محبوب بود به طوری که تقریباً همه در پوتسدام می‌دانستند او کیست. او، اخیراً با «تاندی سایرس» (مادر گرت) در رابطه‌ی عاشقانه بوده. او همچنین یکی از معدود آمریکایی‌های آفریقایی‌تبار ساکن در پوتسدام بود و رابطه‌ی او با «تندی موجی»، شوک در جامعه ایجاد کرده‌بود. آن‌ها اخیراً از هم جدا شده‌بودند زیرا دو پسر «تاندی» با نیک کنارنمی‌آمدند که باعث ایجاد مشکلاتی در خانواده شده‌بود. علاوه بر این، نیک و تاندی هر دو در زمان شروع رابطه با افراد دیگر در ارتباط بودند. در آن زمان، تاندی در حال دیدن جان جونز (یک کلانتر در پوتسدام) بود. «جان جونز» از اینکه نیک عاملی در جدایی او از تاندی بود بسیار ناراحت بود و به خانه نیک رفت تا با او مقابله کند و احتمالاً او را تهدید کند. آشکارا تنش‌هایی در جامعه وجود داشت و این احساس عمومی بود که جان نه تنها به خاطر از دست دادن دوست‌دخترش ناامید شده‌بود، بلکه به رقیبی که یک آمریکاییِ آفریقایی‌تبار و بسیار موفق بود شکست خورد.

پلیس خیلی سریع هیلاری را به عنوان یک مظنون دستگیر کرد، چیزی که به نظر می‌رسد یک مورد ناشی از خصومت‌های نژادی و شخصی است. او چندین ساعت بازداشت و مورد بازجویی و حتی بازرسی قرارگرفت تا ببینند آیا جراحاتی مطابق با پرش از پنجره طبقه دوم دارد یا خیر! او به

هیچ‌وجه مقصر نبود. بدون هیچ مدرکی دالِ بر ارتباط او با قتل گرت، او آزاد شد و بعداً یک شکایت حقوق‌مدنی علیه پلیس تنظیم کرد.

این قضیه، تازه آغاز مشکلات حقوقی هیلاری بود؛ نه پایان آن! وکیل مدافع در برابر دعوی حقوق‌مدنی، راهبردی را برای اثبات اینکه نیک در واقع مرتکب قتل شده است، ایجاد کرد! (و او از شهادت هیلاری در هنگام تسلیم شدن علیه خود استفاده کرد). با وجود شواهد بسیار متزلزل و کم احتمال، دادستان منطقه‌ی «مری راین» در «12 می 2014» کیفرخواستی را برای قتل درجه‌ی دوم علیه هیلاری به دست‌آورد! این کیفرخواست، در «اکتبر 2014» به دلیل سوءرفتار دادستانی از جانب راین *(Rain)* ردشد. «رین» در «2 فوریه 2015»، دومین هیئت منصفه را تشکیل‌داد و کیفرخواست دیگری را برای قتل علیه هیلاری دریافت‌کرد.

این قضیه، تازه آغاز مشکلات حقوقی هیلاری بود؛ نه پایان آن! وکیل مدافع در برابر دعوی حقوق‌مدنی، راهبردی را برای اثبات اینکه نیک در واقع مرتکب قتل شده است، ایجاد کرد! (و او از شهادت هیلاری در هنگام تسلیم شدن علیه خود استفاده کرد). با وجود شواهد بسیار متزلزل و کم احتمال، دادستان منطقه‌ی «مری راین» در «12 می 2014» کیفرخواستی را برای قتل درجه‌ی دوم علیه هیلاری به دست‌آورد! این کیفرخواست، در «اکتبر 2014» به دلیل سوءرفتار دادستانی از جانب راین *(Rain)* ردشد. «رین» در «2 فوریه 2015»، دومین هیئت منصفه را تشکیل‌داد و کیفرخواست دیگری را برای قتل علیه هیلاری دریافت‌کرد.

جامعه در آشفتگی بود زیرا پرونده بدون هیچ راه‌حل روشنی به طول انجامید. شایعات گسترده‌ای مبنی بر وجود شواهد قوی *DNA* علیه هیلاری و سرکوب آن به دلیل "فنی" وجودداشت.

خطرناک ترین دادستان نیویورک

آن شواهد *DNA* در فضای‌نادری از خصومت‌نژادی علیه هیلاری و شواهد روشنی از سوءرفتار دادستانی از سوی *«DA Mary Rain»* جمع‌آوری و تفسیر شد. «راین» در ابتدا در سکویی برای حل قتل گرت شرکت‌کرد و به انتقاداتی مبنی بر استفاده او از قتل برای منافع سیاسی منجرشد (او اغلب در کنار تاندی سایرس در رویدادهای مبارزات انتخاباتی ظاهر می شد). «راین» به سرعت از موقعیت جدید خود برای آزار و اذیت مقاماتی استفاده کرد که در زمانی که او یک مدافع عمومی بود او را به دلیل بی کفایتی اخراج کرده‌بودند. او در سال 2017 دفتر را زیر ابری از سوء ظن ترک‌کرد بدون اینکه به دنبال انتخاب مجدد باشد. او بعداً به مدت دو سال از وکالت تعلیق‌شد (یک اتفاق بسیار نادر و گواهی بر شدت و تداوم سوء رفتار او به عنوان دادستان).

در واقع، رفتار نادرست «راین» در دوران تصدی او به عنوان یک *DA* بدنام بود. در پرونده علیه هیلاری، چندین اقدام غیرقانونی برای افشای اطلاعات دربرمی‌گرفت. او این واقعیت را که شاهدی گزارش داده بود جان جونز را در حال ورود به آپارتمان گرت در نزدیکی زمانی که او کشته‌شده‌بود، ردکرد. «تاندی سایرس» در ژانویه 2011 از «جونز» شکایت کرده بود و اظهار داشت که «جونز» به گونه‌ای عمل می‌کند که باعث ترس او برای امنیت خود و فرزندانش می‌شود، از جمله اینکه جونز بدون اعلام قبلی و بدون دعوت به آپارتمانش می‌رود، علی‌رغم اینکه مکرراً به او گفته‌شده‌بود که نکند.

اگرچه او یک مظنون معقول در این پرونده بود، «جونز» به «راین» حقایقی داد که او به راحتی آن را پذیرفت. «راین» با این ادعا که اظهارات شاهد «با نظریه [دادستان] پرونده مطابقت ندارد» و بنابراین دلیلی برای افشای آن به دفاع وجود ندارد، سرکوب شواهد را توجیه کرد. این احتمالاً مقصرترین و غیراخلاقی‌ترین دلیل ممکن برای ناتوانی دادستان در افشای شواهد تبرئه‌کننده است (به‌ویژه زمانی که مظنون مورد بحث یک کلانتر محلی است که فعالانه در تحقیقات قتل شرکت داشته است).

اولین پرونده هیئت منصفه علیه «هیلاری» به دلیل رفتار غیراخلاقی «راین»، رد شد. «قاضی ریچاردز» حکم داد که راین این روند را لکه‌دار کرده‌است، از جمله با نشان‌دادن دختر 17 ساله «هیلاری» برای افشای ارتباطاتی که توسط امتیاز وکیل-موکل محافظت می‌شود. در همان زمان، *FBI* در حال تحقیق از راین برای تماس با زندانیان بدون رضایت وکیلشان بود تا آن‌ها را متقاعدکند که علیه سایر زندانیان شهادت‌دهند. خبرچینان زندان که برای شهادت تحت فشار قرار گرفته‌اند در حالی که از حقشان برای داشتن وکیل محروم شده‌اند، شواهد بسیار غیرقابل اعتمادی ارائه‌می‌کنند، و نشان‌داده‌شده‌است که این عامل مهمی در محکومیت‌های نادرست است.

شواهد DNA

محاکمه‌ی هیلاری برای قتل «گرت فیلیپس» تنها در مقابل یک قاضی در شهرستان سنت لارنس برگزارشد. هیلاری ممکن است یک محاکمه روی نیمکت را انتخاب کرده باشد، زیرا فکرمی‌کرد (احتمالاً به درستی) ممکن است هیئت منصفه محلی، با او منصف نباشد. در محاکمه شواهد فیزیکی کمی در دسترس‌بود. چهار اثر انگشت نهفته روی پنجره طبقه دوم و اطراف آن پیداشد که گمان‌می‌رود عامل جنایت از آنجا فرار کرده است. اثر انگشت‌ها متعلق به هیلاری نبودند و هرگز با کسی که به این پرونده مرتبط است یا کسی در پایگاه داده ایالت نیویورک *SAFIS* مطابقت نداشت.

همچنین مقادیر کمی از شواهد *DNA* وجود داشت که در نهایت به اثبات این پرونده منجرشد. مشخصات *DNA* از تراشیدن ناخن‌های دست جمع‌آوری‌شده در کالبد شکافی گرت ایجادشد. محققان نظریه ای را ارائه‌کردند که ممکن است گرت قبل از مرگ با مهاجم خود مبارزه‌کرده باشد. با این حال، *DNA* را فقط می‌توان در مقادیر کمی بازیابی کرد، که نشان‌می‌دهد ممکن است دستکاری قابل‌توجهی در *DNA* وجود داشته‌باشد، یا ممکن است بخشی از پس‌زمینه یا آلودگی به

واسطه محقق بوده‌باشد. از آنجایی که *DNA* تعداد کپی پایینی داشت، کمتر از میزان توصیه‌شده برای تجزیه و تحلیل قرارگرفت. این به نوبه‌ی خود، تفسیر مشخصات *DNA* را بسیار دشوارتر می‌کند و ارزش اثباتی شواهد را زیر سؤال می‌برد. اثبات شده‌است که تفسیر پروفایل‌های *DNA* تخریب‌شده، کم کپی و مختلط برای دانشمندان پزشکی قانونی و دادگاه‌ها، یک چالش است. با توجه به پیچیدگی محاسبه احتمال اینکه مشخصات *DNA* یک فرد معین در نمونه گرفته‌شده از صحنه جرم گنجانده‌شود، الگوریتم‌های متعددی برای تخمین احتمالات گنجاندن و نسبت‌های احتمال ایجاد شده‌است. روش‌های سنتی تخمین احتمالات برای نمونه‌های کم‌کپی و مخلوط، احتمال یکسانی را به همه ژنوتیپ‌ها اختصاص می‌دهند که ارزش اثباتی این شواهد را محدود می‌کند. الگوریتم‌های تفسیر *DNA* وزن‌های آماری را به ژنوتیپ‌های مختلف اختصاص می‌دهند (از جمله احتمال اینکه آلل‌های خاصی ممکن است «از بین بروند» و در نمونه ظاهر نشوند)، یا اینکه یک آرتیفکت ممکن است به‌عنوان یک آلل ظاهر شود و در نتیجه زمانی که در واقع گنجانده نشده‌است، «افت کند». نمایه‌ی *DNA* که از زیر ناخن‌های گرت ایجاد شد، یک نیمرخ جزئی بود، به این معنی که چندین آلل از بین‌رفته‌بودند و در الکتروفروگرام قابل تشخیص نبودند. بنابراین، این الگوریتم‌ها کار بسیار بهتری را برای تخمین اینکه آیا یک فرد معین در یک نمونه پیچیده گرفته‌شده از صحنه‌ی جرم گنجانده‌شده‌است یا خیر، انجام می‌دهند. دو تا از محبوب‌ترین مدل‌های نرم‌افزار تفسیر مخلوط *DNA* تجاری موجود، *TrueAllele* و *STRmix* هستند. هر دو در مورد هیلاری مورد استفاده قرارگرفتند و هر دو به نتایج متفاوتی رسیدند! که آیا *DNA* او در نمایه‌ی ایجاد شده از خراش‌دادن ناخن گرت گنجانده شده است یا خیر.

«جان باکلتون»، متخصص ژنتیک پزشکی قانونی پیشرو که نقشی کلیدی در توسعه *STRmix* داشت، بیان می‌کند که الگوریتم‌های مخلوط *DNA* می‌توانند نمونه‌های پیچیده *DNA* را با سرعت و دقت بیشتری تجزیه و تحلیل کنند. نرم‌افزار تفسیر *DNA* به طور کلی با استفاده از روش‌های زنجیره مارکوف مونت کارلو *(MCMC)* برای حل مخلوط‌ها و توسعه‌ی احتمالات مشروط گنجاندن کار می‌کند. روش‌های *(MCMC)* مدت‌هاست که در مدل‌های یادگیری ماشین مورد استفاده قرار

می‌گیرند (زیرا مدت‌هاست که در بسیاری از زمینه‌ها از جمله فیزیک، اقتصاد سنجی و علوم کامپیوتر برای حل مسائل با ابعاد بالا مورد استفاده قرار گرفته‌اند). این مهم است که دانش دامنه سطح بالا را در مدل *(MCMC)* برای ایجاد فهرست مناسبی از فرضیه‌های نامزد وارد کنید.

فرضیه‌های نامزد در این مورد شامل اینکه آیا *DNA* هیلاری در نمونه گنجانده شده‌است یا خیر. آیا او به عنوان مشارکت‌کننده در نمونه حذف شده‌است یا خیر. آیا نمونه *DNA* از طریق آلودگی پس‌زمینه زیر ناخن‌های گرت قرار گرفته‌است و دلیل اثباتی در قتل او نیست (معمولی نیست که به دلیل آلودگی ناشی از فعالیت‌های روزمره، *DNA* افراد دیگر را در زیر ناخن‌هایمان به مقدار کمی پیدا کنیم). و اینکه آیا *DNA* پس از جنایت از طریق نوعی آلودگی با واسطه بازپرس معرفی شده‌است یا خیر.

برای تفسیر *DNA*، این دانش حوزه همچنین شامل قوانینی برای یافتن شواهد قابل‌قبول در دادگاه (کادر 4.1) و همچنین دستورالعمل‌های *SWGDAM* برای تأیید سیستم‌های ژنوتیپ احتمالی است. قبل از استفاده، یک تحلیلگر *DNA* پزشکی قانونی باید پیک‌ها را تفسیر کند، افت و ریزش را تخمین بزند، و در واقع آلل‌ها را فراخوانی کند. نرم‌افزار ابتدا باید به‌صورت داخلی توسط آزمایشگاه تحت شرایطی مشابه شرایط نمونه صحنه‌ی جرم تأییدشود. این کنترل‌های کیفی اولیه در مورد هیلاری استفاده نشدند! «قاضی فلیکس کاتنا» یک جلسه استماع (*Frye*، اسم آزمونی در دادگاه‌های ایالات متحده) فرای برگزارکرد تا مشخص کند که آیا شواهد *STRmix* با توجه به اینکه از *DNA* با تعداد کم کپی گرفته‌شده بود و فقط از یک نمایه *DNA* جزئی تشکیل‌شده‌بود، قابل‌قبول است یا خیر. آزمون‌فرای یکی از دو استاندارد اصلی است که دادگاه‌ها در ایالات متحده برای تعیین اینکه آیا شواهد کارشناسی در دادگاه قابل‌قبول هستند یا خیر (کادر 4.1) استفاده می‌کنند. ماهیت آزمون فرای این است که اگر علمی که نظر مبتنی بر آن است به طور کلی در آن جامعه علمی قابل اعتماد تلقی‌شود، شواهد کارشناسی پذیرفته می‌شود. آزمون کلیدی دیگر پذیرش، که در دابرت ارائه شده‌است، نه تنها به این می‌پردازد که آیا تکنیک یا نظریه مورد قبول است یا خیر، بلکه به این موضوع می‌پردازد که آیا می‌توان آن را آزمایش کرد و آزمایش شده‌است، آیا میزان خطای شناخته‌شده‌ای

برای رویه وجود دارد، آیا کنترل کیفیت وجود دارد یا خیر. و سایر استانداردهای حاکم بر رویه، و اینکه آیا مورد بررسی همتایان قرارگرفته‌است (کادر 4.2).

کادر 4.1

قواعد اولیه مدرک

قواعد شواهد حجیم هستند و هر حوزه قضایی تغییرات خاص خود را خواهد داشت. ادله به طور کلی در دادگاه قابل پذیرش است اگر موارد زیر باشد:

مربوط:

شواهد در صورتی مرتبط هستند که به دادگاه کمک‌کنند تا به سؤالی که مورد اختلاف است یا تمایل به اثبات یا رد واقعیتی مهم دارد، پاسخ دهد. ارزش اثباتی شواهد به میزان تمایل یک مدرک به اثبات یا رد واقعیت مورد اختلاف اشاره دارد.

قابل اعتماد:

شواهدی که غیرقابل اعتماد هستند، یا می‌خواهند یک داور حقیقت را گمراه‌کنند (هیئت منصفه یا قاضی که به تنهایی نشسته است). شواهدی که از دانش دست اول به دست می‌آیند، یا مطابق با روش‌های کنترل کیفیت یا توسط یک مرکز آزمایشگاهی معتبر جمع‌آوری شده‌اند، اغلب قابل اعتمادتر در نظر گرفته می‌شوند. وزنی که باید توسط محاکم کننده واقعیت به شواهد داده‌شود، اغلب به میزان قابل اعتماد بودن شواهد بستگی دارد.

لازمه:

دلیل برای اثبات یا نفی یک امر یا موضوع مورد اختلاف ضروری است. اگر شواهد و

مدارک دیگری را کپی کنند، غیرضروری خواهندبود. از سوی دیگر، اگر هیچ راه دیگری برای طرفین وجود نداشته‌باشد که آن شواهد را در دادگاه ارائه کند، ممکن است شواهد لازم باشد، که دادگاه هنگام ارزیابی قابل اعتماد بودن و منصفانه بودن آن مدارک مورد توجه قرار خواهد گرفت. شواهد جمع‌آوری شده توسط سیستم یادگیری ماشین در غیاب هر اپراتور انسانی ممکن است ضروری باشد.

مستثنی نشده:

بسیاری از قواعد استثنایی وجود دارد که ممکن است دادگاه را ملزم به حذف شواهدی کند که در غیر این صورت قابل پذیرش هستند. به عنوان مثال، قواعد شنیده‌ها، یا ارتباطات ممتاز، می‌تواند منجر به حذف شواهد قابل اعتماد و اثباتی شود.

منصفانه:

پذیرش شواهد اغلب منصفانه تلقی می‌شود اگر ارزش اثباتی آن بر پیش داوری ناعادلانه ای که ممکن است برای یک طرف ایجاد کند، بیشتر باشد. بسیاری از حوزه‌های قضایی نیز قوانین اساسی خود را خواهند داشت که بر تحقیقات پلیس نظارت می‌کند و به دادگاه‌ها کمک می‌کند تا تعیین کنند که چه زمانی مدارک به شیوه‌ای غیرمنطقی جمع‌آوری شده‌اند، مانند متمم چهارم قانون اساسی ایالات متحده.

این مورد، شواهد DNA پزشکی قانونی با استانداردهای «فرای» یا «دابرت» مطابقت نداشت. نتایج اولیه از شواهد مخلوط DNA توسط TrueAllele مورد تجزیه و تحلیل قرار گرفت، اما به دلیل کیفیت ضعیف مشخصات ،DNA هیچ نتیجه‌ای بدست نیامد: هیلاری نه می‌تواند شامل شود و نه از نمونه حذف می‌شود. TrueAllele در آن زمان به این نتیجه رسید که آن‌ها «هیچ پشتوانه آماری» پیدا نکردند که هیلاری در ترکیب DNA گرفته‌شده از زیر ناخن‌های «گرت» مشارکت داشته است. TrueAllele بیان می‌کند که بیش از 100 مورد علاقه را در این پرونده بررسی کرده

است، "و نشان داد که هیلاری به شواهد DNA در این پرونده مرتبط نیست."

کادر 4.2

پذیرش مدارک علمی و کارشناسی در ایالات متحده

آزمون فرای:

آزمون پذیرش عمومی نیز نامیده می‌شود. دادگاه در صورتی که مدارک علمی یا کارشناسی را در جامعه علمی مربوطه پذیرفته‌باشد، می‌پذیرد. این آزمون در *Frye v.United States, 293 F.1013 (D.C. Cir.1923)* در موردی تنظیم شده‌است که شواهد چاپ گراف را کنار گذاشته است زیرا به طور کلی به عنوان قابل اعتماد پذیرفته نشده است. برخی از حوزه‌های قضایی ایالات متحده از این آزمون استفاده می کنند، مانند واشینگتن، کالیفرنیا، ایلینوی، مینه سوتا، نیویورک، نیوجرسی و پنسیلوانیا. اکثر ایالات دیگر آزمون دابرت را پذیرفته‌اند.

آزمون دابرت:

این استاندارد از شواهد کارشناسی در دابرت علیه «مرل داو» داروسازی، ایالات متحده، شماره 579 (1993) تنظیم شده است و قانون 702 قوانین فدرال شواهد را به این معنا تفسیر می‌کند که قضات باید یک وظیفه نگهبانی برای اطمینان از اینکه علمی و شواهد کارشناسی مرتبط و قابل اعتماد است. دانش علمی، دانشی است که بر اساس روش علمی گردآوری شده باشد و این بستگی به خیلی بیشتر از مقبولیت عمومی دارد. دادگاه همچنین می‌تواند بررسی کند که آیا روش‌ها آزمایش و تأیید شده‌اند، آیا میزان

خطای شناخته‌شده‌ای وجود دارد، آیا این روش توسط همتا بررسی شده است، آیا به‌طور خاص برای پرونده حاضر تولید شده است یا اینکه مورد قبول و استفاده قرار گرفته است.

قانون 702. **شهادت شهود خبره:** شاهدی که از نظر دانش، مهارت، تجربه، آموزش یا تحصیل صلاحیت کارشناس را داشته باشد، می‌تواند به صورت نظر یا در موارد دیگر شهادت دهد:

الف)

دانش علمی، فنی، یا سایر دانش‌های تخصصی کارشناس، به آزموده حقیقت کمک می‌کند تا شواهد را درک کند یا واقعیت مورد بحث را تعیین کند.

ب)

شهادت بر اساس حقایق یا داده‌های کافی است.

ج)

شهادت محصول اصول و روشهای قابل اعتماد است.

د)

کارشناس اصول و روش‌ها را به طور قابل اتکایی در مورد واقعیات پرونده اعمال کرده است.

منبع: قواعد شواهد فدرال ایالات متحده، اصلاح شده در 17 آوریل 2000، لازم الاجرا در 1 دسامبر 2000. و همانطور که در 26 آوریل 2011 اصلاح شد، از 1 دسامبر 2011 لازم الاجرا شد.

«ویلیام فیتزپاتریک»، یک DA از «شهرستان اونونداگا»، سپس با «جان باکلتون» تماس گرفت تا

ببیند آیا الگوریتم *DNA* او، *STRmix*، می تواند نتیجه متفاوتی به دست‌آورد. آن‌ها به *DNA* یافت شده در زیر ناخن‌های «گرت» نگاه کردند و مشخصات جزئی آن شامل نیک هیلاری بود. آزمایشگاه جنایی ایالت نیویورک ابتدا *DNA* موجود در نمونه بافت را از طریق *PCR* (واکنش زنجیره ای پلیمراز) تکثیر کرد، اما به نظر می‌رسد که واکنش را بیشتر از آنچه توصیه می‌شود در تلاش برای جمع‌آوری مقدار بیشتری از *DNA* ردیابی انجام داده است (این باعث افزایش اثرات تصادفی می شود. و می‌تواند منجر به "افتادن" شود که در آن نویز به عنوان یک آلل در پروفایل *DNA* ظاهر می‌شود). همچنین در روشی که تحلیلگر در ابتدا نمایه جزئی را ارزیابی می‌کرد، سوگیری وجودداشت، زیرا او نه تنها می‌دانست که هیلاری مظنون است، بلکه با ارجاع به نمایه خود هیلاری، آلل‌های موجود در نمایه را نام می‌برد. «لیون» می‌گوید که «یادداشت‌های کاری او نشان می‌دهد که او مشخصات *DNA* هیلاری را در حالی که سعی می‌کرد آن را با شواهد مطابقت دهد، بررسی کرده است.» این در تضاد با بهترین شیوه‌ها در علم پزشکی قانونی است که به موجب آن یک تحلیلگر باید نسبت به اینکه مظنونین بالقوه چه کسانی هستند و مشخصات *DNA* مظنون چیست، کور باشد تا سوگیری‌های شناختی و زمینه‌ای را در تجزیه و تحلیل خود به حداقل برسانند. آزمایشگاه یک قطع دلخواه *50 rfu* را برای فراخوانی آلل‌ها انتخاب کرد. به نظر می‌رسید که این پایه جز کمک به حذف برخی از آلل‌های موجود در نمایه هیلاری ندارد، و این امر مستلزم این بود که تحلیلگر به این نتیجه برسد که او در نمونه مشارکت ندارد. مارک پرلین از *TrueAllele* شهادت داد که قله‌هایی درست زیر این آستانه وجود دارد که هیلاری را حذف می کند، و بنابراین نمونه *DNA* تبرئه کننده بود.

همانطور که توسط *STRmix* توصیه شده است، مطالعات اعتبارسنجی مناسب توسط آزمایشگاه انجام نشد. زمانی که باکلتون الگوریتم *STRmix* را اجرا کرد، فقط الکتروفروگرام به او داده‌شد (که توسط تحلیلگر به شکلی مغرضانه تهیه شده بود) و بنابراین مجبورشد «داده‌ها را از «منابع معتبر» مختلف انتخاب و انتخاب کند و پارامترهای ورودی را در برنامه به گونه‌ای وارد کند که او معتقد بود که نظام تحمل خواهد کرد». آزمایشگاه جنایی ایالت نیویورک مجاز به استفاده از *STRmix* بدون مطالعات اعتبار سنجی نبود و همانطور که خود باکلتون توصیه کرده‌بود. همچنین

آزمایشگاه از دستورالعمل‌های خود *SWGDAM* پیروی نمی‌کرد، که نیاز به اعتبارسنجی داخلی کامل توسط آزمایشگاه از نمونه‌های پیچیده، کم کپی و نمونه‌های ترکیبی داشت. همچنین آزمایشگاه از دستورالعمل‌های خود *SWGDAM* پیروی نمی‌کرد، که نیاز به اعتبارسنجی داخلی کامل توسط آزمایشگاه از نمونه‌های پیچیده، کم کپی و نمونه‌های ترکیبی داشت. به این دلایل، «قاضی کاتنا» شواهد *DNA* را رد کرد. این منجر به تبرئه هیلاری شد، زیرا شواهد کمی علیه او وجود داشت. «مری رین» پس از تبرئه اظهار داشت که علیرغم کمبود شواهدی که این موضوع را تأیید می‌کند (و تعداد مظنونان دیگری که توسط شایعات شهر و خبرنگارانی که این پرونده را پوشش می دادند، *100%* از گناهکار بودن هیلاری مطمئن بودند). «راین» اظهار داشت که برای هیچ کس دیگری جستجو نمی‌شود، زیرا هیچ کس دیگری نمی‌توانست مرتکب جنایت شود. *DA* جدید، گری پاسکوا، به دنبال سرنخ‌های جدید است، اما قتل گرت فیلیپ حل نشده باقی مانده است.

تفسیر

اخلاق یهود

توسط «ساموئل جی لوین»

پرسش‌های اخلاقی حول استفاده و سوءاستفاده‌ی احتمالی از اشکال شواهد DNA محرمانه در محاکمات جنایی، اگرچه برخاسته از پیشرفت‌های علمی کنونی است، اما نمایانگر اخیر پرسش‌های فلسفی همیشگی است که به دل ماهیت حقوقی و اخلاقی می‌پردازد. از زمان‌های بسیار قدیم، نظام‌های حقوقی با مفاهیم هنجاری برداشت‌های علمی و فلسفی جدید دست و پنجه نرم کرده‌اند. با این حال، سرعت پیشرفت تکنولوژی نیاز به در نظر گرفتن کاربردهای عملی موضوعاتی را که تا

همین اواخر به نظر می‌رسید در محدوده بحث نظری یا شاید علمی تخیلی باقی می‌ماند، برجسته کرده است.

به عنوان یک سیستم فکری که هم قانون و هم الهیات را دربر می‌گیرد، اخلاق یهودی مفاهیم به هم پیوسته حقیقت متعالی و واقعیت عملی را بررسی می‌کند. به عنوان مثال، فیلسوفان حقوقی یهودی به هزاران سال قبل، تنش، اگر نگوییم تناقض، ضمنی در مفاهیم اراده آزاد و جبر، تصدیق کرده‌اند. قراردادن دانای کل خداوند، از جمله آگاهی از آینده، این پرسش را تشدید می‌کند که آیا مردم باید بر اساس اعمالی که هنوز انجام نداده اند مورد قضاوت قرار گیرند یا خیر؟ این معماها که در منابع متعدد تفکر یهودی به آن‌ها پرداخته شده‌است، گاه با این پذیرش بدیهی حل می‌شود که داوری خداوند ذاتاً عادلانه است، و بنابراین، پاداش و مجازات الهی باید در قلمرو اعمال اراده آزاد انسان صورت گیرد. شاید تعجب‌آور نباشد که فیلسوفان یهودی تحلیلی از این موضوعات را بر اساس این اصل که قوانین خدا ذاتاً عادلانه هستند، فرض کنند. با این حال، شاید شگفت‌انگیزتر این باشد که بسیاری از قضات و متفکران حقوقی آمریکایی نیز با کمال میل دکترین اراده آزاد را به عنوان یک نوع ایمان، به جای اینکه نظریه‌های اراده آزاد را فقط در معرض انواع بحث‌های سخت‌گیرانه اعمال شده در سایر حوزه‌های پیچیده حقوق آمریکا قرار دهند، می‌پذیرند. همانطور که پیداست، قضات آمریکایی که به مسائل اراده آزاد و جبرگرایی می پردازند تقریباً همیشه به این اذعان می‌پردازند که نتایج آن‌ها بر اساس اصول و مفروضاتی استوار است که ممکن است با پیشرفت‌های علمی و برداشت‌های فلسفی از حقیقت مرتبط نباشد و البته شاید نیازی هم ندارد.

اگرچه شاید از برخی جهات رضایت بخش نباشد، اما این رویکرد به سؤالات اراده آزاد ممکن است به طور متناوب نشان دهنده‌ی عنصر تازه‌ای از صراحت و فروتنی از سوی سیستم عدالت کیفری و قضاتی باشد که مجازات را تعیین می‌کنند. قضاوت دیگران یک تعقیب مخاطره آمیز است، اگر اجتناب ناپذیر باشد، به ویژه در زمینه قوانین کیفری، که مجرمیت اخلاقی را به کسانی که گناهشان ثابت شده است نسبت می‌دهد. اگرچه ممکن است مجرمان اغلب مستحق محکومیت اخلاقی باشند، ارزیابی کامل و دقیق ارزش اخلاقی یک فرد خارج از قلمرو اجرای عدالت انسانی و فراتر از درک توانایی‌های

انسانی محدود باقی می‌ماند. در اینجا نیز، اندیشه‌ی یهودی مدت‌هاست تصدیق کرده‌است که علیرغم نیاز جامعه به حفظ نظم از طریق اجرای قواعد و پیامدهای قانونی، قضاوت اخلاقی نهایی برای ولایت خداوند محفوظ است.

امتناع قضات آمریکایی از اتخاذ رویکردهای فلسفی یا علمی در قبال جبرگرایی و اراده آزاد ممکن است موجب پایبندی اساسی به استقلال قانون به عنوان نماینده ارزش‌ها و باورهای جامعه شود. برای اطمینان، قانون باید در نظر داشته‌باشد و در صورت اقتضا، باید از پیشرفت‌های درک بشری برای اطلاع‌رسانی و بهبود عملکرد یک سیستم حقوقی استفاده کند. در حالت ایده آل، قانون در کنار ظهور پیشرفت علمی پیشرفت خواهد کرد. با این حال، این قانون به طور جدایی‌ناپذیری با جامعه مرتبط است و منعکس‌کننده ماهیت انسان است، که اغلب ناتوانی در مهار اکتشافات علمی را به گونه‌ای نشان داده است که ارزش‌های اساسی پیشرفت انسانی را ترویج می‌کند. در میان درس‌های دیگر، سوءاستفاده قضایی از فناوری DNA به‌عنوان یادآوری وسوسه‌ها و تمایلات برای بهره‌برداری از فناوری در تعقیب و اعمال قدرت است، به نحوی که ممکن است از مرزهای اخلاق و عدالت فراتر رود. برای پیشرفت در کنار پیشرفت‌های علمی، قانون باید تعهدی هم‌زمان با پیشرفت‌های اخلاقی جاری داشته‌باشد.

اخلاق دئونتولوژیک

نوشته «کالین مارشال»

پرونده «مدل‌های ماشین در دادگاه» سؤالات اخلاقی مختلفی را از دیدگاه ریشه‌شناسی مطرح می‌کند. دو اقدام مربوط به فناوری، به‌ویژه، مستلزم یک تحلیل ریشه‌شناختی است: (1) دادستان منطقه «فیتزپاتریک» به دنبال نتیجه‌ای متفاوت از آنچه توسط *TrueAllele* ارائه شده‌بود (به شرطی که

قصد او این بوده باشد) و (2) تحلیل‌گری که از *STRmix* در حین ارجاع به نمایه هیلاری استفاده می‌کند (به عنوان مظنون شناخته شده پرونده).

یکی از تمرکزهای سنتی در اخلاق دئونتولوژیک بر رد اشکال مشکل ساز جانِبداری بوده است. اقدامات جزئی مشکل ساز باعث مزیت نامناسب برخی افراد نسبت به دیگران می‌شود. کسی را تصور کنید که در حال فکر کردن است که آیا یک سوار آزاد باشد یا خیر، یعنی به این فکر می‌کند که آیا از همکاری دیگران در یک سیستم سود می‌برد در حالی که خودشان همکاری نمی‌کنند. نمونه‌هایی از سواری رایگان شامل استفاده از حمل و نقل عمومی بدون پرداخت کرایه و استفاده از خدمات دولتی و اجتناب از پرداخت مالیات است. چنین اقداماتی به علاقه آزاد سوار بر دیگران امتیاز می‌دهد، و بنابراین (مگر اینکه عوامل کاهش دهنده وجود داشته باشد) جانبداری نامناسب را نشان می‌دهد.

در حالی که هیچ سواریِ رایگانی در مدل‌های ماشین در پرونده دادگاه رخ نمی‌دهد، ما همچنان می‌توانیم بپرسیم که آیا اقدامات (1) و (2) همانطور که توضیح داده شد، جزئی بودن مشکل‌ساز را نشان می‌دهند یا خیر؟ متأسفانه، هیچ راه کاملاً دقیق و بدون مناقشه‌ای برای شناسایی جایی که جانبداری وجود دارد یا زمانی که مشکل‌ساز است وجود ندارد. با این حال، بسیاری از متخصصان اخلاق *deontology* استفاده از روشی به نام «آزمون جهانی‌سازی» را مفید دانسته‌اند.

ایده اصلی پشت «آزمون جهانی‌سازی» یک ایده‌ی آشنا است و در این سؤال منعکس شده است: "چه می شود اگر همه این کار را انجام دهند؟". کمی دقیق‌تر، «آزمون جهانی‌سازی» به‌صورت زیر اجرا می‌شود: یک عامل از نظر اخلاقی یک عمل احتمالی را با این سؤال از خود ارزیابی می‌کند که آیا سیستمی را تأیید می‌کند که در آن همه عوامل در موقعیت‌های مشابه به طور مشابه عمل کنند. به عنوان مثال، فروشنده‌ای که تصمیم به دروغ گفتن به منظور تضمین یک قرارداد پرسود دارد، ممکن است در نظر داشته‌باشد که آیا مایل است سیستمی را تأیید کند که در آن همه فروشندگان به منظور تضمین قراردادهای پرسود دروغ بگویند. در چنین سیستمی، فروشندگان عموماً غیرقابل اعتماد شناخته می‌شوند. بنابراین چنین دروغ‌هایی در آزمون جهانی‌سازی مردود می‌شوند. اگرچه «آزمون جهانی‌سازی» برای یک سناریوی خیالی جذاب است، اما به آشکار شدن جانبداری واقعی

فروشنده در پشت دروغ کمک می‌کند و به طور نامناسبی منافع خود را بر دیگران برتری می‌دهد.

چگونه «آزمون جهانی‌سازی» برای اقدام (1) اعمال می‌شود؟ سؤالی که «دی فیتزپاتریک *(DA Fitzpatrick)*» باید از خود می‌پرسید چیزی شبیه به این بود: آیا او سیستمی را تایید می‌کرد که در آن «وکلای دادگستری» همیشه به دنبال منبع تکنولوژیکی دیگری برای حمایت از دیدگاه‌قبلی خود بودند، در صورتی که اولین منبع چنین نبود؟ پاسخ این سؤال نسبت به فروشنده دروغگو کمتر واضح است. با این حال، اگر همیشه (یا تقریباً همیشه) امکان یافتن منبعی تکنولوژیکی وجود داشته‌باشد که از هر حکم مورد نظر پشتیبانی می‌کند، مشکل مشابهی پیش می‌آید: در چنین سیستمی، هر گونه توسل به یک منبع ارزش متقاعدکننده خود را از دست می‌دهد. هر کس که تلاش می‌کند به یک منبع تکنولوژیکی خاص متوسل شود، نمی تواند آن سیستم کلی را تایید کند (البته منظور بیشتر این است که هیچ کس نمی‌تواند). غیرقابل‌قبول بودن این سناریوی خیالی نشان می‌دهد که اقدام «دی فیتزپاتریک» نشان دهنده‌ی جانبداری مشکل‌ساز است. از سوی دیگر، اگر همیشه (یا تقریباً همیشه) امکان یافتن یک منبع فن‌آوری که هر حکم مورد نظر را پشتیبانی می‌کند ممکن نباشد، چنین سیستمی ممکن است مشکل‌ساز نباشد، که نشان می‌دهد «دی فیتزپاتریک» جانبداری مشکل‌ساز نشان نداده است.

چگونه آزمون جهانی‌سازی برای اقدام (2) اعمال می‌شود؟ در اینجا، سؤالی که تحلیلگر باید مطرح می‌کرد، در این راستا بود: آیا آن‌ها سیستمی را تأیید می‌کنند که در آن، کاربرد الگوریتم‌ها در ارزیابی احساس گناه همیشه (یا تقریباً همیشه) از باورها و سوء ظن‌های پیشین تحلیلگر مطلع باشد؟ با توجه به نقش بزرگی که تحلیلگران در کاربرد الگوریتم‌ها دارند، این تهدیدی است که جذابیت فناوری مانند *STRmix* را کم ارزش می‌کند و بنابراین پشتیبانی از اتهامات نادرست را آسان می‌کند. احتمالاً هیچ کس نمی‌تواند سیستمی را تأیید کند که در آن هرگونه اتهام نادرست با استفاده از فناوری به راحتی قابل پشتیبانی باشد. این نشان می‌دهد که تحلیلگر در اقدام (2) جانبداری غیرقابل‌قبول از خود نشان می‌دهد.

در حالی که آزمون جهانی‌سازی در ارزیابی جزئی بودن، مفید است، نمی‌توان آن را به صورت

الگوریتمی اعمال کرد. هنگامی که آزمون برای یک عمل معین اعمال می‌شود، سؤال اصلی همیشه این است که کدام جنبه از عمل باید تعمیم داده‌شود. به عنوان مثال، با اقدام (2)، این سؤال نباید این باشد که آیا تحلیلگر سیستمی را تأیید می‌کند که در آن هر کسی که به «نیک هیلاری» مشکوک بود، مجاز است از باورهای پیشینه خود در به کارگیری الگوریتم‌ها استفاده کند (این سؤال به شناسایی موارد مرتبط کمک نمی کند. اشکال جانبداری در این مورد). از این رو، آزمون همیشه باید با قضاوت‌های غیر پیش پا افتاده در مورد اینکه کدام جنبه از اعمال از نظر اخلاقی مرتبط هستند هدایت شود، و هیچ فرمول ساده ای برای تعیین اینکه آن جنبه‌ها چیست، وجود ندارد. با این وجود، مواردی مانند فروشنده دروغگو نشان می دهد که این قضاوت‌های غیر پیش پا افتاده گاهی نسبتاً آسان و غیرقابل بحث هستند. در حالی که اکثر مردم مستعد انواع خاصی از جانبداری هستند، با کمی فاصله، بسیاری از ما می توانیم جانبداری مشکل ساز را تشخیص دهیم.

فصل ۵

رسانه‌های‌مصنوعی و خشونت سیاسی

بنابراین رسانه‌های ترکیبی ظرفیت تخریب سرمایه‌ی اجتماعی و اعتبار را در پایه خود دارند، و این به همان اندازه صادق است که آیا محتوای آن‌ها را درست می‌پذیریم یا نه. مشکل اساسی‌تر این است که ما هیچ معیار عینی‌ای برای تعیین اینکه چه چیزی شایسته باور است نداریم. علاوه بر این، جلب توجه به پدیده رسانه‌های‌مصنوعی تنها تأثیرات آن را تقویت می‌کند و برای ما چاره‌ای جز بازگشت به ترجیحات، تعصبات و ایدئولوژی‌های سیاسی خود باقی نمی‌گذارد.

تریسی داودزول و شان گولتز *(Tracey Dowdeswell and Sean Goltz)*

کودتا در گابن

«گابن»، کشوری نسبتاً باثبات در سواحل غربی آفریقا است، در روزهای ابتدایی سال 2019 توسط یک کودتا لرزید. جرقه‌ی این کودتا تا حد زیادی توسط یک سخنرانی معمولی در شب سال‌نو توسط «رئیس جمهور علی بونگو اودیمبا» آغاز شد که در شبکه‌های اجتماعی منتشر شده‌بود. رئیس جمهور مدتی بود که از انظار عمومی دور بود (در واقع او بیش از دو ماه بود که در کشور نبود) و شایعاتی مبنی بر اینکه او به شدت بیمار است یا حتی مرده است، در شبکه‌های اجتماعی پیچیده بود.

پیام سال‌نو به طور گسترده‌ای به عنوان یک «*deepfake*» محکوم شد. در حالی که پیام سال

نو به خودی خود غیرقابل توجه بود، نحوه نمایش علی بونگو در این ویدئو کاملاً عجیب است: چهره رئیس جمهور بالای دهانش به طرز عجیبی بی حرکت است. او به مدت 1 دقیقه و 39 ثانیه پس از فیلم، چشمانش اصلاً پلک نمی‌خورند و گفتار و حرکات او غیرطبیعی و مصنوعی به نظر می‌رسد. نظرات به این ویدیو نشان می‌دهد که بینندگان احساس می‌کردند که این ویدیو «وحشتناک» و «*deepfake*» است. در 3 ژانویه 2019، یک رسانه‌ی خبری برجسته در گابن، مقاله‌ای را منتشر کرد که آشکارا این ویدئو را یک *deepfake* محکوم کرد.

"رسانه مصنوعی" اصطلاحی فنی است که برای توصیف آنچه اغلب "*deepfake*" نامیده می‌شود استفاده می‌شود. رسانه‌های مصنوعی از یادگیری ماشین و شبکه‌های عصبی برای ایجاد رسانه‌های صوتی، عکس، ویدیو یا حتی متن مصنوعی استفاده می‌کنند که معتبر به نظر می‌رسند. فیلم ویدئویی را می‌توان تغییر داد تا گفتار و حرکات یک نفر را بر روی صحبت‌های شخص دیگر قرار دهد. رسانه‌های مصنوعی فراتر از فیلم‌های عکس و ویدیو هستند و شامل گفتار و نوشتار تولید شده توسط هوش‌مصنوعی می‌شوند. این روش‌ها برای تولید حجم فزاینده‌ای از محتوایی که در اینترنت می‌بینیم استفاده می‌شوند و توسط «مزدوران» برای تولید ایمیل‌ها، متن‌ها یا پیام‌های نادرست که به نظر می‌رسد از سوی افرادی هستند که شما را به خوبی می‌شناسند، استفاده می‌شوند. متن مصنوعیِ ساخته‌شده توسط هوش‌مصنوعی، حتی برای تولید مقالات آکادمیک نیز استفاده شده است؛ و جالب است که بدانید آن مقاله به چاپ نیز رسیده‌است! چندین ادعا در مجله *Arabian Journal of Geosciences* به دلیل بسیار غیرمعمولی، به 10 ادعای برتر ایوان اورانسکی در سال 2021 تبدیل شد؛ آن‌ها چرندیات محض بودند! [توضیح ادعای اورانسکی: Watch، Retraction وبلاگی است که در مورد ادعای مقالات علمی و موضوعات مرتبط گزارش می‌دهد. این وبلاگ در آگوست 2010 راه‌اندازی شد و توسط نویسندگان علم، ایوان اورانسکی (معاون سابق، *Editorial Medscape*) و آدام مارکوس (ویرایشگر اخبار گوارش و اندوسکوپی) تولید می‌شود]. یک مقاله باید از مجله *Arabian Journal of Geosciences* به این دلیل غیرعادی که «محتوای این مقاله مزخرف است» پس گرفته شود. به نظر می‌رسد این مشکل بسیار

گسترده شده است: بیش از 400 مقاله در مجلات متعلق به *Springer Nature* و صدها مقاله‌ی دیگر در *Elsevier* نیز آلوده به متن مصنوعی هستند! فناوری‌های رسانه‌ی‌مصنوعی سریع‌تر از فناوری‌هایی که برای شناسایی آن‌ها استفاده می‌شوند توسعه می‌یابند و *deepfake* را تقریباً نامرئی می‌کنند.

نیروهای مسلح گابن (که مدت‌ها در مخالفت با حزب حاکم بونگو بودند) با ویدیوی سال نو رئیس جمهور نیز احساس کردند که به تمسخر گرفته شده‌اند. در ساعات اولیه صبح 7 ژانویه 2019، «ستوان کِلی اوندو اوبیانگ» از گارد جمهوری خواه اعلام کرد که سخنرانی سال نو نشان داد که «علی بونگو» برای اداره کشور مناسب نیست. بر این اساس، ارتش او را از سمت خود برکنار کرد و می‌خواست «شورای بازسازی» ملی را برای حکومت به جای او تشکیل دهد. تانک‌ها وارد لیبرویل (پایتخت) شدند. اینترنت و برق قطع شد. حدود 300 معترض به حمایت از کودتا آمدند و توسط نیروهای دولتی با گاز اشک آور مورد حمله قرار گرفتند. صدای تیراندازی در پایتخت شنیده شد. تا ساعت 10 صبح، نیروهای دولتی دوباره کنترل شدند و رهبران کودتا یا مرده بودند یا در بازداشت بودند.

کارشناسان هوش‌مصنوعی شروع به آزمایش کردند که آیا ویدیوی سال نو ساخته‌شده‌است یا خیر، و به سرعت یک اجماع واضح ظاهر شد. «استیو گروبمن (*Steve Grobman*)» (مدیر ارشد فناوری *McAfee*) ویدیو را از طریق الگوریتم‌های آن‌ها اجرا کرد و با احتمال بسیار بالا (92) تشخیص داد که ویدیو واقعی است. سیوی لیو (*Siwei Lyu*)، پروفسور علوم کامپیوتر در «*SUNY Albany*»، همچنین ویدیو را با استفاده از الگوریتم *deepfake* خود بررسی کرد و تائید کرد که به احتمال 99 ویدئو واقعی است. الگوریتم‌ها، هیچ مدرکی مبنی بر استفاده از روش‌های شناخته‌شده‌ی مصنوعی‌کردن رسانه، برای تولید ویدئو پیدا نکردند. اگر این یک *deepfake* بود، باید یک فرضیه بسیار پیچیده بوده باشد (فرضیه‌ای که توسط عجیب بودن خود ویدیو پشتیبانی نمی‌شود).

صحت این ویدیو با مشاهده خود علی بونگو در حضورهای عمومی بعدی تأییدشد که نشان

می‌دهد او زنده است (اما همچنین بسیار تغییر یافته است). «الکساندر درومریک» (*Alexander Dromeric*)، متخصص مغز و اعصاب واشینگتن پست، اظهار داشت که حرکات و بی‌تحرکی صورت علی بونگو مشخصه افرادی است که سکته کرده یا دچار نوعی آسیب مغزی شده‌اند. از آن زمان دفتر رسمی مطبوعاتی دولت از تایید یا تکذیب اینکه آیا علی بونگو دچار سکته مغزی شده است خودداری کرده است. دولت گابن در واقع به اطلاعات نادرست و بی اعتمادی در مورد سلامت رئیس جمهور دامن می زد (اما به دلیل عدم شفافیت آن‌ها و نه با تولید رسانه‌های مصنوعی). مشکل *deepfake* فقط این نیست که اطلاعات نادرست تولید می‌کنند، بلکه وجود آن‌ها مردم را به سمت بی اعتبار کردن گزارش‌هایی سوق می‌دهد که در واقع درست هستند.

«آویو اودیا» (*Aviv Ovadya*)، متخصص هوش‌مصنوعی و رسانه‌های مصنوعی، بیان می‌کند که *deepfake* می‌تواند بسیار خطرناک باشد، دقیقاً به این دلیل که بی اعتمادی گسترده نسبت به همه‌ی رسانه‌ها ایجاد می‌کند. این باعث ایجاد نوعی «بی‌تفاوتی واقعیت» می‌شود که اودیا آن را «*Infocalypse*» نامیده است (از دست دادن اساسی اعتماد به نهادهای اجتماعی). یک *Infocalypse* زمانی به وجود می‌آید که متوجه می‌شویم استانداردهایی برای حقیقت و عینیت نداریم و کنترلی بر فناوری‌های در حال تکامل سریع نداریم، همانطور که در گابن اتفاق افتاد. اودیا می‌گوید که «مخاطره‌ها زیاد است و پیامدهای احتمالی فاجعه‌بارتر از مداخله خارجی در انتخابات است («*Infocalypse*»: تضعیف یا فروپاشی نهادهای اصلی).

بی‌تفاوتی واقعی که توسط رسانه‌های‌مصنوعی ایجاد می‌شود پرهزینه است (همانطور که اودیا می‌گوید، «هم برای سازمان‌های رسانه‌ای که مجبور به صرف زمان و منابع برای بررسی و شناسایی جعلی بودن این شکل ویدیوها هستند هم برای جوامعی که در بحث‌هایی درباره اصالت آن‌ها صحبت می‌کنند»). برای گابن، این هزینه‌ها در خشونت سیاسی، افزایش ترس، گسست اجتماعی و مرگ دو عضو نیروهای مسلح شورشی دیده‌شد.

بولی بای *(Bulli Bai)*: فروش زنان به صورت مصنوعی در هند

در جای دیگر، ما *deepfake*ها را به‌عنوان شکل مخصوصاً موذیانه‌ای از تبلیغات محاسباتی توصیف کرده‌ایم، که عمدتاً به دلیل پتانسیل آن‌ها در دامن زدن به تنش بین کشورها، به خطر انداختن امنیت ملی و تضعیف سیاست خارجی و دیپلماسی بین‌المللی است. رویدادهای اخیر در هند نشان می دهد که چگونه می توان از رسانه‌های مصنوعی برای تعمیق خصومت‌های قومی و مذهبی (در این مورد با هدایت خشونت جنسی با انگیزه‌های سیاسی علیه زنان مسلمان) استفاده کرد.

«رعنا ایوب» یک روزنامه نگار مشهور تحقیقی در هند است. او نه تنها یک زن هندی است که در حال مذاکره در مورد زندگی عمومی در یک کشور محافظه کار اجتماعی است، بلکه یکی از اعضای اقلیت مسلمان نیز است (و یکی که به دلیل انتقاد از اعضای حزب حاکم بهاراتیا جاناتا *(BJP)* شهرت پیدا کرده است). او در آوریل 2018 گزارشی جنجالی درباره تجاوز جنسی به دختری 8 ساله در کشمیر نوشت. رعنا ایوب حتی به *BBC* رفت و اعضای حزب ملی‌گرا *BJP* را به راهپیمایی در حمایت از متهم دعوت کرد. ضربه برگشتی به ایوب سریع اما غیرقابل پیش‌بینی بود. افراد ناشناس شروع به پخش پیام‌های جعلی در توییتر کردند که ادعا می‌شد از طرف ایوب آمده است و نظراتی مانند «من از هند متنفرم»، «من از هندی‌ها متنفرم»، «من عاشق پاکستان هستم»، و «من عاشق متجاوزین به کودکان هستم و اگر آن‌ها این کار را به نام اسلام انجام می‌دهند، من از آن‌ها حمایت می‌کنم».

اما بدتر از آن هنوز در راه بود! شخصی از داخل *BJP* به ایوب هشدار داد که ویدیویی در *WhatsApp* به اشتراک گذاشته شده‌است که دیدن آن برای او بسیار دشوار است. آن‌ها به او گفتند:"من آن را برای تو می‌فرستم، اما به من قول بده که ناراحت نخواهی شد". چیزی که ایوب دریافت کرد یک ویدیوی مستهجن بود که در آن صورت او بر روی بدن برهنه یک زن (بسیار جوان) نقش بسته بود! او می‌گوید که این ویدیو در «تقریباً همه تلفن‌های هند» پخش شد. شما می‌توانید خود را روزنامه‌نگار خطاب کنید، می‌توانید خود را یک فمینیست بنامید، اما در آن لحظه،

من نمی‌توانستم این تحقیر را ببینم. ایوب در پایان گفت: حتی با وجود اینکه هیچ کس فکر نمی‌کرد (یا قرار بود فکر کند) این پورنوگرافی جعلی واقعی است، اما تأثیر موردنظر خود را داشت: "من از روی ناچاری کمی خودسانسور شده ام".

این تنها یکی از بسیاری از حوادث مشابه در هند است. در پایان سال 2021، چندین زن (همگی مسلمان مانند ایوب) در یک سایت حراجی جعلی به نام «بولی بای» ظاهر شدند که یک موقعیت فوق‌العاده تحقیرآمیز برای زنان مسلمان است. این زنان نیز در زندگی عمومی هند برجسته بودند: روزنامه نگاران، فعالان، و وکلا. این سایت آن‌ها را در شرایط توهین آمیز، اغلب جنسی صریح یا تحقیرآمیز به تصویر کشیده است. این تصاویر از حساب‌های رسانه‌های اجتماعی گرفته‌شده و سپس دستکاری شده‌اند تا زنان را در موقعیت‌های زننده به تصویر بکشند.

حدود 6 ماه قبل، سایت مشابهی به نام *"Sulli Deals"* در فضای مجازی منتشر شده‌بود. همانند *Bulli Bai*، این تصاویر با تعریف *deepfake* مطابقت ندارند. آن‌ها *«shallow fakes»* بودند (رسانه‌های مصنوعی که قرار نیست باورشان شود، اما با این وجود، تأثیرات آن‌ها بر اهدافشان به شدت محسوس است). یکی از زنانی که در سایت‌های حراج جعلی هدف قرار گرفته، دانشجوی ۲۶ ساله دانشگاه کلمبیا به نام «هیبا بگ» است. بگ نیز مانند ایوب از حزب حاکم و سیاست ملی گرایانه آن انتقاد کرده است. او اظهار داشت که این "ارعاب با هدف مجبور کردن زنان مسلمانی است که صدای خود را علیه بی‌عدالتی بلند می‌کنند تا از زندگی عمومی کناره گیری کنند. اما شما عقب نشینی نمی‌کنید، حتی اگر همه چیز طاقت‌فرسا شود".

«عصمت آرا»، یکی دیگر از قربانیان «بولی بای»، زمانی که در فهرست «بولی بای روز» قرار گرفت، تصویری از «فروخته شدن» خود در حراج منتشر کرد و خاطرنشان کرد که سال جدید 2022 او با «حس ترس» و انزجار آغاز شده‌است". در مورد او نیز، این سایت از «تصویر اصلاح‌شده من در زمینه‌ای نامناسب، غیرقابل قبول و آشکارا زننده» استفاده کرده‌است.

یکی دیگر از قربانیان سایت‌های حراج وحشی، «قراتولین رهبر» (روزنامه‌نگار اهل کشمیر تحت مدیریت هند و همسر یک قاضی دادگاه عالی در دهلی است). رهبر اظهار داشت: وقتی عکسم را

دیدم گلویم سنگین شد، بازوهایم شل شده بود و بی حس شده بودم. تکان دهنده و تحقیرکننده است. رهبر اظهار داشت که سایت حراج جعلی "برای تحقیر زنان مسلمان" بسیارخوب بوده است.

سایت‌های حراج‌جعلی «حسیبه امین» را که به عنوان هماهنگ کننده‌ی رسانه‌های اجتماعی برای حزب مخالف کنگره کار می‌کند نیز هدف قرار داده‌اند. او نگران است که استفاده از این سایت‌ها برای ترویج خشونت و تهدید علیه زنان اقلیت عواقبی داشته‌باشد که فراتر از توانایی آن‌ها برای تحقیر و سانسور زنان برجسته هندی است. او می‌ترسد که تهدید به مرگ و ارعاب آنلاین به خشونت جنسی در دنیای واقعی دامن بزند. او می‌پرسد: «ما چه تضمینی از سوی دولت داریم که فردا تهدید و ارعاب آنلاین به خشونت جنسی واقعی در خیابان‌ها تبدیل نشود؟».

تفسیر

اخلاق فضیلت

توسط پیتر سینگر و ییپ فای تسه

ملاحظات کلیدی - ارزش حقیقت

ما برای حقیقت ارزش زیادی قائل هستیم و نادیده‌گرفتن این ارزش را بسیار دشوار می‌دانیم. بر این اساس، ما همچنین به شیوه‌هایی که حقیقت و صداقت فکری را در بر می‌گیرند، مانند توسعه روش‌های علمی که به شواهد و به روز کردن باورها در پرتو بهترین شواهد موجود بستگی دارد، ارزش می‌گذاریم. در مقابل، وقتی حقیقت را نمی‌پذیریم، برای مثال، اگر جامعه استفاده نادرست زیادی از فناوری‌های *deepfake* مشاهده کند، احتمالاً بی اعتمادی عمومی نسبت به صحت صدا، فیلم و تصاویر افزایش می‌یابد. این در مورد «کودتای گابن» نشان داده شده‌است. و نه تنها بر سیاست‌های

محلی تأثیر می‌گذارد. این نگران‌کننده است که با فناوری‌های *deepfake* که برای فریب دادن یا گمراه کردن افراد و سهولت دسترسی به آن‌ها طراحی شده‌اند، ممکن است نقطه‌ای وجود داشته‌باشد که هرکسی بتواند هر کاری را در هر مکان (یا در مکان‌های خیالی) انجام دهد. عواقب آن ممکن است شامل استفاده از هر فیلم یا عکسی به عنوان مدرک در دادگاه باشد، اما محدود به آن نیست. دولت‌ها و سازمان‌های غیردولتی قادر به شناسایی موارد نقض حقوق بشر نیستند. و با تغییر تفکر، کسانی که مرتکب جنایت می‌شوند از محکومیت فرار می‌کنند، زیرا می‌توانند صحت مدارک علیه خود را به طور قابل‌قبولی انکار کنند.

آیا کل قضیه‌ی *deepfake* همین است؟

برای برخی، ممکن است به نظر برسد که *deepfake* یک مشکل کاملاً جدید است، اما اینطور نیست. «نانسی پلوسی»، رئیس مجلس نمایندگان ایالات متحده، هدف چند ویدیوی تغییریافته بود تا صدای او را مبهم یا مست کند، از جمله سخنرانی او روی صحنه در یک رویداد مرکز پیشرفت آمریکا. آن ویدئوها به سرعت پخش شد و به عنوان مدرکی علیه شایستگی و اخلاق کاری او استفاده شد. مشخص شد که سرعت فیلم‌ها فقط تا 0.75 کاهش یافته است، عملکردی که در طیف گسترده‌ای از نرم‌افزارهای اصلی پخش یا ویرایش ویدیو موجود است. از این رو، ما نباید وسوسه شویم که فکر کنیم فناوری‌های *deepfake* تمام مسئولیت را بر عهده می‌گیرند. همچنین نباید فکر کنیم که *deepfake* "چیزی بیش از فتوشاپ" نیست. استفاده از نرم‌افزارهایی مانند فتوشاپ به آموزش و تجربه، شاید استعداد نیز نیاز دارد، اما فناوری‌های *deepfake* به رابط‌های برنامه‌نویسی کاربردی (بیش از حد آسان) یا API تبدیل شده‌اند که این امکان را برای میلیون‌ها نفر فراهم می‌کند که با چند کلیک *deepfake* بسازند. همچنین، همانطور که نتایج کنونی نشان داده‌است، فناوری‌های *deepfake* مسلماً قوی‌تر از هر روش قبلی‌ای برای مصنوعی‌سازی رسانه‌ها هستند. و مهم‌تر از همه،

آن‌ها پیچیده‌تر خواهندشد و تشخیص آن‌ها سخت‌تر می‌شود، زیرا الگوریتم‌های پشت آن‌ها می‌توانند به طور مداوم با تحقیقات بیشتر و آموزش بیشتر با داده‌ها بهبود یابند.

شاید بتوان به فناوری‌های تشخیص عمقی به عنوان دلیلی اشاره کرد که چرا نباید زیاد نگران باشیم. اما آن‌ها نیز مشکلاتی دارند. اول، ممکن است *100%* قابل اعتماد نباشند. دوم، حتی اگر آن‌ها بتوانند *deepfake* ها را از رسانه‌های واقعی با اطمینان بالا شناسایی کنند، تشخیص را از دست افرادی که نمی‌توانند به نرم‌افزار تشخیص دسترسی داشته‌باشند، رها می‌کند. سوم، ممکن است ما را به رد شواهد معتبر به دلیل استفاده نادرست گسترده از *deepfake* ها و بی‌اعتمادی که این امر ایجاد می‌کند، سوق دهد. چهارم، گاهی اوقات تشخیص نمی‌تواند آسیبی را که قبلاً وارد شده است، از بین ببرد، مثلاً در مورد پورنوگرافی بدون رضایت. پنجم، روش‌های تشخیص ممکن است خود به بهبود فناوری‌های *deepfake* کمک کنند، یا با وادار کردن طراحان یا الگوریتم‌های *deepfake* به بهتر شدن، یا حتی به‌طور مستقیم برای تبدیل شدن به ابزاری برای آموزش این الگوریتم‌ها (مثلاً به‌عنوان متمایزکننده شبکه‌های آموزشی متخاصم مولد).

برخی ممکن است این نکته را مطرح‌کنند که فایده‌گرایی استفاده از *deepfake* را توجیه می‌کند، حتی اگر به برخی افراد آسیب وارد کند، تا زمانی که افراد بیشتری از آن سود ببرند. مثالی که در آن مردم (به ظاهر سودمندگرا نیستند) چنین استدلالی را مطرح کرده‌اند، پورنوگرافی مصنوعی بدون رضایت است. ما اذعان می‌کنیم که تعداد افرادی که از چنین موادی لذت می‌برند بسیار بیشتر از تعداد قربانیان است. اما این استدلال، سخنی مضحک از آن چیزی است که سودگرایان ادعا می‌کنند، هر چند که شاید با شعار گمراه‌کننده «بزرگترین خوشبختی از بیشترین تعداد» ایجاد شده‌باشد. درست است که «جرمی بنتام»، بنیانگذار فایده‌گرایی، از این شعار استفاده کرد، اما بعداً وقتی متوجه شد که این شعار به این معناست که هر چیزی که به نفع 51 درصد جمعیت باشد، درست است، آن را رد کرد، حتی اگر 51 درصد از آن سود ببرند. فقط اندکی و 49 درصد آسیب‌های بزرگی را متحمل می‌شوند. در موردی که در اینجا مورد بحث قرارگرفت، اگر پورنوگرافی مصنوعی ساخته‌شود، زنانی که توسط آن به تصویر کشیده می‌شوند آسیب‌هایی را متحمل می‌شوند که از نوع کاملاً متفاوت، و

بسیار جدی‌تر از دست دادن «منافع» ناتوانی در مشاهده چنین مواردی است.

علاوه بر ارزش بلندمدت حقیقت که قبلاً به آن اشاره کردیم، پیامدهای مهم و بلندمدت دیگری نیز از پورنوگرافی مصنوعی غیرقانونی وجود دارد: این ایده را تقویت و تداوم می‌بخشد که زنان، اشیایی هستند که ممکن است، بدون رضایت آن‌ها، مورد استفاده قرارگیرند. لذت بردن از دیگران، و این باعث ترویج و ایجاد نگرش‌های قابل‌قبول تر می‌شود که برای زنان در بسیاری از جنبه های مختلف زندگی مضر است.

اخلاق آفریقایی

نوشته جان مورانگی

فکر کردن به جایگاه اخلاق در جهان که توسط علم داده یا، به طور خاص، توسط هوش مصنوعی ایجاد شده‌است، دشوار است. من گمان می کنم که اکثر دانشمندان داده و افراد هوش مصنوعی باور ندارند که در کاری که انجام می‌دهند یک جهانی خلق می‌کنند یا ساکنان چنین دنیایی هستند. بسیاری از آن‌ها ممکن است از این دنیا بی‌خبر باشند و معمار آن باشند. همچنین ممکن است برخی از آن‌ها بدانند که معماران آن هستند و در آن زندگی می‌کنند. اما حتی اگر آن‌ها چنین دانشی داشته‌باشند، ممکن است که درک کاملی از پیامدهای آن نداشته‌باشند.

به نظر من آنچه برای تعیین معنادار جایگاه اخلاق در جهان ایجاد شده توسط علم داده و هوش‌مصنوعی لازم است، تعمیق و گسترش آگاهی از پیامدهای آن است. در طی انجام این کار، آگاهی از وجود دیدگاهی برخاسته از دنیایی متفاوت، جهانی که عمدتاً اخلاقی است، می‌تواند به عنوان پادزهری برای جنبه‌های غیرانسانی جهان که توسط علم داده و هوش‌مصنوعی ایجاد می‌شود،

عمل کند که در چارچوب چنین دیدگاهی است که من به پیوند بین رسانه‌های مصنوعی و خشونت سیاسی فکر می‌کنم.

در مورد گابن، رئیس جمهور علی بونگو که به طور مصنوعی تولید شده است، به سختی می توان از رئیس جمهور بونگو که ساخته‌نشده تشخیص داد. به طور مشابه، تشخیص زن روزنامه نگار هندی ساخته‌شده از روزنامه نگار زن هندی ساخته‌نشده دشوار است. حتی اگر وسیله‌ای برای تشخیص یکی از دیگری وجود داشته باشد، خود وسیله نمی‌تواند آسیبی را که قبلاً وارد شده است جبران کند و هیچ تضمینی وجود ندارد که در آینده آسیب بیشتری ایجاد نشود. دشواری تشخیص واقعیت از جعلی همچنان ما را آزار می‌دهد.

علاوه بر این، تأیید ممکن است در معرض دستکاری هوش‌مصنوعی باشد. در مواقعی، مشروعیت تأیید بستگی به چیزی دارد که شخص می‌خواهد تأیید کند و آنچه که می‌خواهد تأیید کند می‌تواند ساخته‌اش باشد! نباید قدرت متقاعدسازی رسانه‌های مصنوعی را دست‌کم گرفت. هوش‌مصنوعی آینده‌ای دست‌نخورده دارد. پیشرفت آن آینده قابل پیش‌بینی مشخصی ندارد. قدرت بی وقفه‌ای برای خلق آینده‌ی خود دارد. علاوه بر این، ممکن است آنچه در نظر بیننده واقعی است و همچنین مصنوعی باشد.

دنیای رسانه‌های‌مصنوعی می‌تواند در ظاهر واقعی‌تر از دنیای واقعی باشد. برای برخی، به طور فزاینده‌ای در حال تبدیل شدن به دنیای‌واقعی است. همچنین باید توجه داشت که اخلاق خود از دستکاری رسانه‌های‌مصنوعی مصون نیست. خوبی که هدف آن است می‌تواند خیری باشد که توسط هوش‌مصنوعی تعیین می‌شود. آسیبی که تصور می‌شود محصول هوش‌مصنوعی است، ممکن است به عنوان مخالف آن تلقی شود: به عنوان خوب. می‌توان تصور کرد که کسانی که رئیس‌جمهور جعلی را در گابن تولید کردند، همانطور که در مورد کسانی که یک روزنامه‌نگار زن هندی جعلی تولید کردند، چیزی مضر تولید نمی‌کردند. تا این حد، به نظر نمی‌رسد که لزوماً ارتباط علی بین هوش‌مصنوعی و خشونت سیاسی وجود داشته‌باشد.

پیوند بین رسانه‌های‌مصنوعی و خشونت سیاسی توجه را به پیوند بین علمِ داده و هوش‌مصنوعی

در سیاست جلب می‌کند. نه پیوند و نه آنچه که مستلزم آن است بدیهی است. آموزش دانشمندان داده یا آموزش بازیگران هوش‌مصنوعی شامل مطالعه سیاست نمی‌شود. ظاهراً غیرسیاسی یا از نظر سیاسی بی‌طرف به نظر می‌رسد. حتی اگر مطالعه سیاست نیز گنجانده شود، به احتمال زیاد علم سیاست به معنای پوزیتیویستی آن خواهد بود. به احتمال زیاد، مطابق با سایر علوم اجتماعی، به ظاهر یک علم سیاسی غیرسیاسی خواهدبود.

جوامع عمدتاً برای رفاه آن‌ها از نظر سیاسی تأسیس شده‌اند. این رفاه در درجه اول موضوع اخلاق است. اگر این را پذیرفت، نمی‌توان امر سیاسی را از امر اخلاقی و امر اخلاقی را از امر سیاسی جدا کرد. امر اخلاقی امر سیاسی است و امر سیاسی امری اخلاقی است. وقتی پوزیتیویسم مطالعه سیاست را تعیین می‌کند، علم سیاسی که چنین تعیین شده‌است جایی برای اخلاق ندارد. در اینجا، امر سیاسی بدون احساس امر اخلاقی در علم داده غیرسیاسی و در غیرسیاسی در هوش مصنوعی جایگاهی دارد. معرفی اخلاق در علم داده و در آموزش هوش‌مصنوعی به عنوان یک حواس پرتی در این آموزش ظاهر می‌شود. این آموزش اغلب به عنوان آموزش بدون ارزش پیش بینی می‌شود (آموزش بدون اخلاق، به عنوان آموزش غیرسیاسی). در جوامع بومی آفریقا، رفاه اجتماعی یک امر اشتراکی است. این رفاه هم سیاسی و هم اخلاقی است. این بهزیستی است که جایگاهی برای رفاه فردی و همچنین رفاه گروهی دارد. در هر دو صورت، چنین رفاهی به بهای رفاه جامعه نیست. این حس گسترده تر از رفاه اجتماعی بر مفهوم آفریقایی اوبونتو استوار است. در حالت اوبونتو انسان، ادعا می‌شود که "ما هستیم، پس من هستم." به رسمیت شناخته شده‌است که در جستجوی رفاه، جایی برای یک فرد یا گروهی برای بهزیستی وجود دارد، اما نه به قیمت سعادت جامعه. همچنین مشخص شده است که یک فرد یا یک گروه می‌تواند بر خلاف رفاه اجتماعی عمل کند، اما کنترل های اجتماعی برای به حداقل رساندن تهدید برای رفاه جامعه وجود دارد. این کنترل‌ها به دلیل تغییرات مدرن در جامعه ضعیف شده‌است.

تغییرات جامعه از درون و بیرون، ساخت رفاه اجتماعی آفریقا را پیچیده کرده‌است. «ما» در اوبونتو پیچیده شده است. دیگر نمی‌توان «ما» واقعی را از «ما» جعلی تشخیص‌داد. هر چیزی که به

عنوان «ما» واقعی درک می‌شد، توسط «ما» مصنوعی واژگون شده است. تشخیص آسیب به جامعه به دلیل نقش پیچیده‌ای که هوش‌مصنوعی در دستکاری ادراک بازی می‌کند، دشوار شده‌است. یکی از پیامدهای هوش‌مصنوعی، تجاوز به مرزهای اجتماعی است. امروزه، هیچ جامعه‌ای در آفریقا یا هر جای دیگری از تجاوزات روزافزون هوش‌مصنوعی مصونیت ندارد. تشخیص آنچه واقعی از جعلی است یا تشخیص بی‌ضرر از آنچه مضر است به طور فزاینده‌ای دشوار است. پلیس در فضای مجازی درست است، اما چه کسی بر پلیس نظارت دارد؟ آیا بازیگران هوش‌مصنوعی می‌توانند به عنوان افسرپلیس خدمت کنند؟ چه کسی بر آن‌ها نظارت خواهدکرد؟ آن‌ها باید چه آموزش سیاسی/اخلاقی داشته‌باشند تا بتوانند این کار را در راستای تحقق عدالت اجتماعی انجام دهند؟ آیا خود عدالت اجتماعی تبدیل به یک محصول هوش‌مصنوعی نشده‌است؟ آیا ما در عصر استبداد هوش‌مصنوعی زندگی نمی‌کنیم؟

امروزه به نظر می‌رسد که بازیگران هوش‌مصنوعی کشیشان عصرما هستند. آن‌ها به عنوان الهیات سیاسی عمل می‌کنند. آن‌ها در جامعه به عنوان ناجی نوع بشر شناخته می‌شوند. مراکز هوش‌مصنوعی هاله‌ی تقدس را به خود گرفته‌اند. آیا این مستلزم الحاد جدید نیست؟ کافران جدید؟

اخلاق بومی

توسط جوی میلر و آندریا سالیوان کلارک

سؤال اخلاقی اصلی که مایلیم به آن بپردازیم در پایان این مطالعه موردی مطرح شده‌است. آیا "هدف قرار دادن زنان فعال سیاسی در هند از طریق پورنوگرافی عمیق، "سخنرانی خطرناک" را تشکیل می‌دهد؟ در یک درک بومی از اخلاق، کاملا بله.

برای درک چرایی آن، باید حداقل دو دسته از مسائل اخلاقی را که در پاسخ به این سؤال مطرح می‌شوند، در نظر بگیریم. اول، این موضوع وجود دارد که چه چیزی مورد استفاده قرار می‌گیرد. در

این مورد، استفاده از پورنوگرافی‌عمیق‌جعلی نگرانی‌های اخلاقی را در مورد چگونگی و چرایی ایجاد و به دست آوردن چنین تصاویری ایجاد می‌کند. دوم، این موضوع وجود دارد که چه چیزی از استفاده از پورنوگرافی‌عمیق‌جعلی حاصل می‌شود. در این صورت استفاده از چنین تصاویری منجر به اجبار، فریب و انقیاد زنان فعال سیاسی در هند می‌شود. برای توضیح اینکه چرا *deepfake* ها در این زمینه گفتار خطرناکی را تشکیل می‌دهند، به نوبه‌ی خود هر دوی این نگرانی‌ها را بیشتر توضیح خواهیم‌داد. با توجه به موضوع اخلاقی چگونگی و چرایی خلق و به دست آمدن چنین تصاویری، ایجاد چنین تصاویری نقض حاکمیت و خودمختاری است. همانطور که در مورد رسانه‌های مصنوعی و خشونت سیاسی مشخص است، داده‌ها را می‌توان به سلاح تبدیل کرد. حتی اگر مصنوعی باشد، می توان از آن برای آسیب رساندن، کنترل، سرکوب و سلب حق رای دیگران استفاده کرد. برای مردم بومی، تاریخچه‌ای از سلاح سازی داده‌ها وجود دارد. به این ترتیب، جنبشی از سوی محققان بومی برای جمع‌آوری و کنترل داده‌های مربوط به مردمانشان وجود دارد. جنبش حاکمیت داده‌ها شاهدی بر اهمیت حاکمیت و خودمختاری برای مردم بومی است. این فقط به این نیست که چگونه داده‌ها را می‌توان تسلیحاتی کرد، بلکه این است که چه کسی بهتر می‌داند با داده‌ها چه کند. مردم بومی در موقعیت بهتری برای استفاده (به عنوان مثال، جمع‌آوری و پیاده‌سازی) داده‌های مربوط به خود برای درک بهتر و زندگی خود نسبت به افراد خارجی که اغلب از چنین داده‌هایی برای برنامه‌های خود استفاده می‌کنند، هستند.

همین امر را می‌توان در مورد افرادی در هند نیز گفت که از شباهت هایشان برای ایجاد پورنوگرافی‌عمیق‌جعلی استفاده می‌شود. در این مورد، هیچ رضایتی برای استفاده و ایجاد این موارد پورنوگرافی‌عمیق‌جعلی داده‌نشد. داده‌ها (یعنی *deepfake* ها) ناشی از استفاده از شباهت یک فرد بدون رضایت اوست. این از نظر اخلاقی اشتباه است زیرا مقدار کافی حاکمیت (توانایی کنترل زندگی و تصمیم گیری شخصی) برای خوب زیستن ضروری است. *deepfake* ناشی از بی توجهی کامل به حاکمیت است.

برای درک این موضوع که چه چیزی از استفاده از پورنوگرافی‌عمیق‌جعلی حاصل می شود، یک

ایده کلیدی از فلسفه‌ی بومی باید درک شود: کلمات قدرت دارند. با توجه به ارتباط همه چیز، عمل صحبت کردن و کلمات گفته شده، بر محیط اطراف فرد تأثیر می‌گذارد. این بدان معنی است که آن‌ها نه تنها بر تعاملات فرد تأثیر می‌گذارند، بلکه خود آن‌ها کنش متقابل هستند (یعنی (فعل) کنش. برای انسان، کلمات روشی برای تعامل با محیط اطراف خود هستند. بنابراین در وجود هارمونی تأثیر دارند).

واضح است که در مورد زنان فعال سیاسی در هند، توانایی آن‌ها برای خوب زندگی کردن تحت تأثیر استفاده از *deepfake* قرار گرفته‌است. عمل ایجاد *deepfake*، و همچنین نحوه‌ی استفاده از آن‌ها، یک عمل دستکاری و اجباری برای وادار کردن زنان به رفتاری است که "در خط" یا برای صاحبان قدرت (مانند مردان و *BJP*) مفید باشد. این کار برای زندگی در هماهنگی با یکدیگر انجام نمی‌شود (این تلاشی است برای تحمیل اراده یک گروه به گروه دیگر و در عین حال ترویج ناهماهنگی).

فصل ۶

بیومتریک و تشخیص چهره

افراد صرفاً با مشارکت در جهان به گونه‌ای که ممکن است چهره خود را برای دیگران آشکار کند یا امکان ثبت تصویر آن‌ها در دوربین را فراهم کند، حق حریم خصوصی خود را نادیده نمی‌گیرند. حریم خصوصی برای کرامت، استقلال، رشد شخصی و مشارکت آزاد و آزاد افراد در زندگی دموکراتیک حیاتی است. هنگامی که نظارت افزایش می‌یابد، افراد می‌توانند از اعمال این حقوق و آزادی‌ها منصرف شوند.

دانیل ترین *(Daniel Therrien)*، **کمیسر حریم خصوصی کانادا**

شرکت *Clearview AI*

استثمار جنسی آنلاین از کودکان (چیزی مانند پورنوگرافی برای کودکان) یک مشکل جدی و رو به رشد در سراسر جهان است. بین سال‌های 2014 تا 2019، گزارش‌ها به پلیس سواره سلطنتی کانادا *(RCMP)* از عکس‌ها یا ویدیوهایی که آزار جنسی کودکان را به تصویر می‌کشند، حدود *1106%* افزایش یافته است. در سال 2019، *RCMP* بیش از 102967 گزارش از سوءاستفاده جنسی آنلاین از کودکان دریافت‌کرد. با توجه به مقیاس مشکل، مکان یابی، شناسایی و حذف مطالب ممکن است دشوار باشد.

این می‌تواند به ویژه برای قربانیان سخت باشد. مطالب مستهجن که سوءاستفاده از آن‌ها را به تصویر می‌کشد می‌تواند سال‌ها در اینترنت باقی‌بماند. در فصل 5 ما بحث کردیم که چگونه فیلم‌های مستهجن جعلی برای روزنامه نگاران زن در هند آسیب زا بوده است و اینکه ارسال چنین مطالبی به عنوان ابزاری برای خشونت سیاسی علیه زنان در آن کشور استفاده شده‌است. برای قربانیان سوءاستفاده جنسی از کودکان، ویدیوهایی که آن‌ها را در شرایط تحقیرآمیز و توهین‌آمیز جنسی نشان می‌دهد واقعی است و به نظر نمی‌رسد که هرگز ازبین نرود.

چندین سازمان غیرانتفاعی شروع به استفاده از فناوری‌های یادگیری‌ماشین برای مشکل شناسایی و حذف مطالبی که آزار جنسی کودکان را به تصویر می‌کشد، کرده‌اند. در ایالات متحده، سازمان غیرانتفاعی *Thorn* ابزاری به نام *Spotlight* ایجاد کرده است که از فناوری تشخیص چهره *(FRT)* برای شناسایی قربانیان سوءاستفاده جنسی و قاچاق کودکان استفاده می‌کند. افسران مجری قانون می‌توانند عکس یک کودک گم شده یا مورد استثمار را آپلود کنند و سپس به جستجوی ویدیوهایی بپردازند که کودک را به تصویر می‌کشد، یا تبلیغات آنلاینی را که به کودک پیشنهاد رابطه جنسی می‌دهد. در کانادا، *Project Arachnid* از هوش‌مصنوعی *(AI)* برای جستجوی تصاویر سوءاستفاده استفاده جنسی از کودکان در اینترنت عادی و حتی دارک‌وب *(dark web)* استفاده می‌کند و سپس درخواست‌های حذف را برای حذف مطالب صادر می‌کند. آن‌ها حدود 6 میلیون عکس و ویدیو را از وب حذف کرده‌اند و هر روز تعداد بیشتری پست می‌شود.

در اکتبر سال 2019، مرکز ملی جرایم بهره‌کشی از کودکان *(NCECC)* (بخشی از *RCMP*) شروع به استفاده از فناوری تشخیص‌چهره برای شناسایی کودکان قربانی استثمار جنسی آنلاین کرد. آن‌ها دو مجوز از یک شرکت آمریکایی به نام *Clearview AI* خریداری کردند که به آن‌ها امکان دسترسی به الگوریتم‌های تشخیص چهره *Clearview* و دیتابیس‌های عظیم عکس‌ها را می‌داد. آن‌ها همچنین از چندین حساب آزمایشی رایگان استفاده کردند که توسط *Clearview* به سازمان‌های مجری قانون ارائه شده‌بود. *RCMP* از این فناوری در مقر ملی و همچنین در بریتیش کلمبیا، آلبرتا، مانیتوبا و نیوبرانزویک استفاده کرد.

مرکز *NCECC* بیان می‌کند که از فناوری تشخیص چهره *Clearview* در 15 مورد استفاده کرده و 2 کودک را نجات داده‌است. علاوه بر این، *Clearview* حدود 14 بار برای شناسایی عاملی که از اجرای قانون فرار می‌کرد استفاده شد. *RCMP* بیان می‌کند که در غیر این صورت از این فناوری به صورت آزمایشی استفاده می‌کند تا ببیند فناوری تشخیص‌چهره چه کاربردی می‌تواند در پیشبرد تحقیقات جنایی به طور کلی داشته‌باشد. کمیسیونر حریم خصوصی بیان می‌کند که *RCMP* هدف اکثر صدها جستجوی انجام شده را فاش نکرده است. پلیس تورنتو همچنین از *Clearview* در چندین تحقیق در همان زمان استفاده کرد. مانند *RCMP*، استفاده مجاز نبود یا مشمول هیچ گونه کنترل‌داخلی نبود. در پایان، در 84 تحقیق بین اکتبر 2019 و فوریه 2020، که بیشتر آن‌ها مربوط به قتل و جنایات جنسی بود، استفاده شد:آن‌ها 12 قربانی (10 نفر ازآن‌ها کودک) و همچنین 2 شاهد و 4 مظنون را شناسایی کردند. هیئت خدمات پلیس تورنتو بیان می‌کند که برنامه‌ای برای استفاده مجدد از *Clearview* ندارد و اخیراً سیاستی را برای استفاده از فناوری تشخیص چهره وضع کرده‌است. مانند سرویس پلیس تورنتو، *RCMP* در ابتدا استفاده از *Clearview* را برای کمیسر حریم خصوصی کانادا رد کرد. سپس دفتر کمیسر حریم خصوصی تحقیقاتی را در مورد اینکه آیا استفاده از *Clearview* قوانین حریم خصوصی کانادا را نقض می‌کند، آغاز کرد. چندین نگرانی توسط کمیسر حریم خصوصی مطرح شد، از جمله این واقعیت که فناوری تشخیص چهره ممکن است علیه فعالان و معترضان به کار گرفته‌شود، اینکه این فناوری پتانسیل "فناوری نظارتی بسیار تهاجمی" را دارد، و اینکه استفاده از آن ممکن است سایر انسان‌های اساسی را نقض کند. حقوق، از جمله با ترویج تبعیض نژادی در سیستم عدالت کیفری.

بیومتریک و فناوری تشخیص چهره

یکی از انواع فناوری بیومتریک است (که امروزه بسیاری از آن‌ها با یادگیری ماشین تسهیل شده‌اند). شناسایی‌های بیومتریک از سایر اشکال شناسه، مانند رمزهای عبور، کارت‌ها و سایر نشانه‌ها، ایمن‌تر هستند همه‌ی آن‌ها ممکن است گم شوند، دزدیده شوند، جعل شوند یا در معرض حملات آزمایشی و خطا قرار گیرند (که خود توسط سیستم‌های هوش مصنوعی که توسط بازیگران بد به کار می‌روند آسان‌تر می‌شوند). یک بیومتریک قابل اعتماد هم بسیار فردی خواهد بود (باید به طور دقیق بین یک فرد و فرد دیگر تمایز قائل شود) و هم در طول زمان پایدار است، به طوری که یک فرد بتواند در صورت نیاز به استفاده از هویت خود تکیه کند. دو مورد از قابل اعتمادترین و منحصربه‌فردترین بیومتریک‌ها، «اثرانگشت» و «اسکن عنبیه چشم» هستند و احتمالاً به همین دلیل برای برنامه *Aadhaar* هند انتخاب شده‌اند. با این حال، قابل اعتماد بودن به معنای بی خطا بودن نیست. اسکن عنبیه برای افراد مسن و افراد مبتلا به آب مروارید دشوار است. اثر انگشت برای حدود 1 تا 3 درصد از جمعیت یک شکل غیرقابل اعتماد از شناسه است و همانطور که بسیاری از هندی‌ها به ضرر خود کشف کرده‌اند، باید با تغییر اثر انگشت ما در طول عمرمان به‌روزرسانی شوند، در حالی که برای دیگران ممکن است به طور کامل تغییر یابند یا از بین‌بروند. در عین حال، چیزی که بیومتریک را به یکی از مطلوب‌ترین و مطمئن‌ترین شکل‌های شناسایی تبدیل می‌کند، این واقعیت است که آن‌ها بخشی جدایی‌ناپذیر از شخصیت ما به‌عنوان افراد منحصربه‌فرد هستند؛ دقیقاً همان چیزی است که اطلاعات آن‌ها را بسیار حساس می‌کند.

فناوری تشخیص‌چهره در مقایسه با اسکن اثر انگشت و عنبیه، روشی کمتر قابل اعتماد برای شناسایی است. فناوری‌های تشخیص‌چهره به طور مستقیم چهره ما را اندازه گیری یا تجزیه و تحلیل نمی‌کنند. آن‌ها "هیچ تصوری از یک شخص خاص ندارند" و "برای شناسایی افراد خاص ساخته نشده‌اند." در عوض، آن‌ها فواصل مشخصی را بین اجزای صورت اندازه می‌گیرند: فاصله بین چشم‌های ما، عرض بینی، عمق حدقه‌های چشم، طول خط فک و غیره. این اندازه‌گیری‌ها یک مقدار

برداری تولید می‌کنند که به عنوان یک پروکسی (نماینده) برای یک فرد معین عمل می‌کند. سپس باید بین مقادیر برداری تصاویر مختلف مقایسه‌شود. مقایسه را می‌توان برای اهداف ثبت‌نام انجام داد، به عنوان مثال، هنگام گرفتن عکس برای ثبت‌نام در یک برنامه تشخیص چهره که به فرد امکان دسترسی به یک مکان امن، یک دستگاه الکترونیکی یا دریافت خدماتی مانند یک داروی تجویزی را می‌دهد. مقایسه همچنین می‌تواند بر اساس «یک به یک» باشد، مانند زمانی که یک فرد از تصویری از چهره‌ی خود استفاده می‌کند تا بعداً به برنامه دسترسی پیدا کند. جستجوها همچنین می‌توانند «یک به چند» باشند که شامل مقایسه عکس فرد با پایگاه‌داده عکس‌های مشابه است. سپس مقدار برداری ایجاد شده از تصویر فرد با تصویر موجود در پایگاه داده مقایسه می‌شود تا احتمال اینکه تصاویر مربوط به همان شخص باشد، مشخص شود. شناسایی موقت زمانی انجام می‌شود که احتمال از مقدار آستانه معینی فراتر رود. این توسط یک اپراتور انسانی که تصاویر را مقایسه می‌کند بررسی می‌شود، اما این می‌تواند نوعی سوگیری تایید را در فرآیند شناسایی معرفی کند. اپراتور انسانی به احتمال زیاد یکی از نامزدها را انتخاب می‌کند و کاندیدایی را که بالاترین رتبه را دارد انتخاب می‌کند زیرا معتقد است الگوریتم بسیار مؤثر است. استفاده پلیس از تصاویر برای جستجوی مظنونان، و استفاده *RCMP* از *Clearview AI* برای جستجوی قربانیان جرایم جنسی آنلاین شامل جستجوهای یک به چند از این نوع است.

تشخیص چهره و حفظ حریم خصوصی

کمیسیونر حریم خصوصی دریافت که *Clearview AI* با جمع‌آوری اطلاعات خصوصی کانادایی‌ها بدون رضایتآن‌ها با جمع‌آوری یک پایگاه داده تصویری از حدود 10 میلیارد عکس برای مقایسه، قوانین حریم خصوصی کانادا را نقض کرده است. *Clearview* پاسخ داد که این عکس‌ها که از

اینترنت و سایت‌های رسانه‌های اجتماعی حذف شده‌اند، هنگام انتشار در دسترس عموم قرار گرفته‌اند. *RCMP*، به نوبه خود، بر اظهارات *Clearview* مبنی بر اینکه عکس‌ها "عمومی" شده اند، تکیه کرد. کمیسیونر حریم خصوصی دریافت که *RCMP* باید رضایت افراد را برای استفاده از تصاویر آن‌ها در تحقیقات خود جلب می‌کرد.

در پایان، *RCMP* موافقت کرد که سیاست‌های خود را تغییردهد و از توصیه‌های کمیسر حریم خصوصی تبعیت کند. در آن زمان، نرم‌افزار در ونکوور، ادمونتون، کلگری و اتاوا آزمایش شده‌بود. همچنین در فرانسه، ایالات متحده، استرالیا، و بریتانیا مورد استفاده قرار گرفته‌است، جایی که موضوع بحث و جدل‌های فراوان و چندین اقدام قانونی در مورد پتانسیل آن برای نظارت جمعی و جستجوها و توقیف‌های غیرمنطقی بوده است. این واقعیت که داده‌های مورد استفاده برای آموزش و اجرای الگوریتم بیومتریک توسط کاربران عمومی شده است، به دانشمندان داده این حق را نمی‌دهد که فرض کنند صاحبان آن داده‌ها با استفاده از آن موافقت کرده‌اند.

تفسیر

اخلاق بومی

توسط جوی میلر و آندریا سالیوان کلارک

بیومتریک و فناوری تشخیص‌چهره این پتانسیل را دارند که ابزاری برای سرکوب شوند. این امر به ویژه زمانی اتفاق می‌افتد که هیچ نظارت یا سیاستی در مورد استفاده از آن‌ها وجود نداشته‌باشد. بدون چنین راهنمایی، این فناوری ممکن است به طور خودسرانه علیه جوامع به حاشیه رانده شده، به ویژه جوامعی که وضعیت موجود را به چالش می‌کشند، استفاده‌شود. در غیاب نظارت خارجی و

ایجاد سیاست‌های مربوط به استفاده از بیومتریک، نگرانی اصلی افراد و جوامع بومی، امکان هدف قرار گرفتن برای نظارت است. این یک نگرانی مشروع است (این اولین بار نیست که از واکنش‌های نظامی علیه مردم بومی استفاده می‌شود). تاریخ‌های اخیر ایالات متحده و کانادا مملو از نمونه‌هایی از پاسخ‌های نظامی شده به اعتراضات بومی است، مانند بحران اوکا، سنگ ایستاده، و محاصره‌ی *«Wet'suwet'en»*. استعمار را می‌توان به عنوان جنگ توصیف کرد و روش‌هایی که دولت‌های فدرال برای حفظ وضعیت موجود به کار می‌برند، نشان‌دهنده این موضوع است. با این حال، نگرانی در مورد نظارت و از دست دادن حریم خصوصی فرد در مطالعه موردی، چارچوب غربی این موضوع است.

برای درک اخلاقیات استفاده از بیومتریک و فناوری تشخیص‌چهره به گونه‌ای که با جهان‌بینی بومی سازگار باشد، مستلزم چارچوب بندی مجدد بحث بر حسب روابطی است که به جای بحث حقوق فردی درگیر است. به عنوان مثال، استفاده از بیومتریک و فناوری تشخیص‌چهره بر روابط بین مردم بومی و دولت فدرال تأثیر می‌گذارد. وقتی دولت‌ها بر حقیقت و آشتی تأکید می‌کنند در حالی که آژانس‌های آن‌ها از فناوری برای نظارت بر افراد و جوامع بومی استفاده می‌کنند، بر روابط پرتنش تاریخی تأثیر منفی می‌گذارد. زمانی که اقدامات و سخنان هر یک از طرفین سازگار نباشد، هر دو طرف در روابط متضرر می‌شوند: کار برای استعمارزدایی برای مردم بومی سخت‌تر می‌شود و دولت‌ها (مانند ایالات متحده و کانادا) نه تنها چهره‌ی خود را در صحنه جهانی از دست می‌دهند. و برای مردم بومی، اما آن‌ها همچنین نمی‌توانند از رویکردهای متنوع حل مسئله که ممکن است یک جهان‌بینی بومی ارائه دهد، بهره‌مند شوند.

بیومتریک و فناوری تشخیص‌چهره، از دیدگاه بومی، به خودی خود غیراخلاقی نیستند. در عوض، نحوه استفاده از فناوری (به عنوان مثال، آیا استفاده از آن به تعادل و هماهنگی کمک می‌کند؟ یا استفاده از آن باعث ارتقای روابط خوب می‌شود؟) نحوه ارتباط فرد با آن را تعیین می‌کند. فقدان نظارت و سیاست‌های مربوط به نحوه استفاده از فناوری، امکان استفاده خودسرانه علیه جوامع حاشیه‌نشین را فراهم می‌کند. نه تنها باید یک سیاست استفاده ایجاد شود، بلکه باید با مشورت مردم

و جوامع بومی انجام شود. از دیدگاه غربی، *Clearview AI* یک ابزار است و کاربرد آن به استفاده از آژانس‌های پلیس دولتی محدود می‌شود. مشورت با جوامع بومی ممکن است کاربردهای دیگری نیز داشته‌باشد، مانند بازگرداندن تعادل با رسیدگی به موارد متعدد زنان و دختران بومی گم‌شده و کشته‌شده *(MMIWG)*. کار در همکاری با مردم بومی به رابطه اعتماد کمک می‌کند و در عین حال حاکمیت و خودمختاری ملت‌های بومی را حفظ می‌کند.

موضوع دیگری که نیاز به بررسی دارد این است که برخی از اقدامات بیومتریک، مانند اثر انگشت و فناوری تشخیص‌چهره، ممکن است باعث ایجاد اعتماد کاذب در شناسایی افرادی شود که گمان می‌رود مرتکب جرم شده‌اند. اعتماد بیش از حد با توجه به استفاده از این فناوری نشان‌دهنده عدم تواضع (یک ارزش بومی) است. داشتن چنین اعتمادی بر مردم بومی تأثیر منفی خواهد گذاشت، زیرا آن‌ها در حال حاضر بیش از حد در میزان زندانی شدن در ایالات متحده و کانادا حضور دارند. جستجوهایی که بر مقایسه یک عکس با عکس‌های موجود در پایگاه‌داده تکیه می‌کنند، عینیت فردی را که مقایسه می‌کند، زیر سؤال می‌برد. افرادی که دارای ویژگی‌های فنوتیپی مرتبط با گروه‌های نژادی هستند در برابر تعصبات و تعصبات ضمنی کسانی که تصمیم می‌گیرند آسیب پذیرتر می‌شوند. فروتنی ممکن است برای کاهش مسئله اعتماد به نفس بیش از حد مورد استفاده قرار گیرد.

بشردوستی و قوانین درگیری مسلحانه

نوشته تریسی داودزول

علیرغم چالش‌های قانونی، هوش مصنوعی *Clearview* به طور فعال به دنبال اولین قراردادهای بزرگ دولت ایالات متحده است (به ویژه با سازمان‌های مجری قانون فدرال مانند *FBI*، اداره مهاجرت و گمرک، و خدمات ماهی و حیات وحش). آن‌ها همچنین در حال تحقیق در مورد استفاده از

تشخیص‌چهره و واقعیت افزوده برای ایمن‌سازی پست‌های بازرسی پایگاه نیروی هوایی هستند.

علیرغم آزمایش نشده‌بودن در درگیری‌های مسلحانه، هوش‌مصنوعی *Clearview* به طور رسمی فناوری خود را برای استفاده در زمان جنگ در 10 مارس 2022 در اوکراین عرضه کرد. وزارت دفاع اوکراین استفاده از هوش‌مصنوعی *Clearview* را ظاهراً برای شناسایی کشته‌شدگان، مبارزه با اطلاعات نادرست و «دامپزشکی افراد موردعلاقه در پست‌های بازرسی» آغاز کرد.

علیرغم آزمایش نشده بودن در درگیری‌های مسلحانه، هوش‌مصنوعی *Clearview* به طور رسمی فناوری خود را برای استفاده در زمان جنگ در 10 مارس 2022 در اوکراین عرضه کرد. وزارت دفاع اوکراین استفاده از هوش‌مصنوعی *Clearview* را ظاهراً برای شناسایی کشته‌شدگان، مبارزه با اطلاعات نادرست و «دامپزشکی افراد مورد علاقه در پست‌های بازرسی» آغاز کرد. با توجه به مستعد بودن سیستم‌های تشخیص‌چهره به خطا و سوگیری، استفاده از آن علیه غیرنظامیان به‌ویژه نگران‌کننده است، زیرا مثبت کاذب (منظور از مثبت کاذب، این است که یک فرایند یا موضوع که کاملا صحیی و بی‌آزار است، به خاطر شباهت ظاهری به موارد غلط و نادرست، به عنوان مورد مشکوک شناسایی بشود) ممکن است منجر به بازداشت‌های نادرست یا حتی قتل در نقض قوانین بین‌المللی جنگ شود. اگر این سیستم غیرنظامیان را در ایست‌های بازرسی، یا در داخل و اطراف سایت‌های نبرد شناسایی کند، ممکن است منجر به ارتکاب جنایات جنگی شود.

تمایز غیرنظامیان از جنگجویان در زمان جنگ (از جمله غیرنظامیانی که ممکن است به عنوان مبارزان مقاومت، جاسوسان، شورشیان یا سایر نیروهای چریکی عمل کنند) و اینکه چگونه باید با این موضوع در قوانین بین المللی جنگ برخورد شود، مشکلی دیرینه است. وضعیت فعلی قانون این است که نیروهای نظامی تنها تا زمانی می‌توانند چنین غیرنظامیانی را هدف قراردهند که آن‌ها به طور فعال تهدید مسلحانه باشند. هر گونه فعالیت دیگری باید با یک محاکمه عادلانه بر اساس قوانین داخلی برای افرادی که حق ندارند به عنوان اسیران جنگی رفتار شوند، رسیدگی‌شود. در 150 سال گذشته، کشورها از اسناد بین‌المللی بشردوستانه مانند کنوانسیون‌های لاهه و ژنو و پروتکل‌های الحاقی آن‌ها عقب‌نشینی کرده‌اند تا به آن‌ها آزادی عمل بیشتری برای شناسایی و کشتن غیرنظامیانی بدهند که

فکر می‌کنند ممکن است تهدیدی برای تلاش‌های جنگی آن‌ها باشد.

استفاده از فناوری تشخیص‌چهره در یک منطقه‌ی جنگی قطعاً این پتانسیل را دارد که این مشکل طولانی مدت را تشدید کند و جنایات جنگی را تسهیل کند. یک غیرنظامی که به یک ایست بازرسی نزدیک می‌شود، یا به دنبال کمک‌های بشردوستانه است، ممکن است به دلیل تطابق مثبت کاذب از پایگاه‌داده تشخیص‌چهره، مورد هدف نیروهای مسلح قرار گیرد. هنگامی که این فناوری به ارتش اوکراین ارائه شد، هیچ تلاشی برای اطمینان از عدم استفاده از آن برای ارتکاب جنایات جنگی از این نوع انجام نشد. در واقع، خود *Clearview AI* خاطرنشان کرد که "هدف دقیقی که وزارت دفاع اوکراین از این فناوری برای آن استفاده می‌کند نامشخص است". مدیر عامل *Clearview*، *Hoan Ton-That*، اظهار داشت که او هرگز نمی‌خواهد ببیند که این فناوری در تضاد با کنوانسیون ژنو است و هرگز نباید از آن به عنوان تنها منبع شناسایی استفاده شود، اما او هیچ قاعده یا حفاظتی را وضع نکرده است، که از این اتفاق جلوگیری کند.

قوانین بین‌المللی جنگ به طور کلی «اکتساب یا اتخاذ» روش‌ها و ابزارهای جنگی جدید را ممنوع می‌کند، مگر اینکه دولت بتواند ابتدا تعیین کند که «به‌کارگیری آن در برخی شرایط یا در همه‌ی شرایط، توسط این پروتکل یا هر قاعده‌ی دیگری منع می‌شود یا خیر". استفاده از فناوری تشخیص‌چهره در درگیری مسلحانه بین‌المللی جدید است و با توجه به خطرات بسیار بالای استفاده از فناوری‌های تشخیص‌چهره برای شناسایی افراد ناشناس در زمان واقعی، همراه با پتانسیل اثبات شده‌ی آن‌ها برای خطاها و سوگیری جمعیتی، ما استدلال می‌کنیم که به طور کلی باید به عنوان یک ابزار ممنوعه جنگ در نظر گرفته‌شود. ما توصیه می‌کنیم که قوانین بین‌المللی به گونه‌ای تفسیر شود که شناسه‌های بیومتریک مانند فن‌آوری‌های تشخیص‌چهره را فقط می‌توان در درگیری‌های مسلحانه بین‌المللی برای اهداف بشردوستانه، مانند شناسایی متوفیان و افراد آواره و پیوستن مجدد آن‌ها به خانواده‌هایشان، به کار برد.

فصل ۷

تعدیل محتوا: سخنان نفرت انگیز و نسل کشی در میانمار

این یک شوخی نژادپرستانه است. اینجا مردی است که با یک حیوان مزرعه رابطه جنسی دارد! در اینجا یک ویدیوی گرافیکی از قتل ضبط شده توسط یک کارتل موادمخدر است.

کیسی نیوتن، در مورد تعدیل‌محتوا در *Facebook*

شرکت *Facebook* و پاکسازی قومی در برمه

در سال 2017، نیروهای نظامی در میانمار سرکوب وحشیانه علیه روهینگیا، یک اقلیت قومی مسلمان ساکن در منطقه غربی این کشور را تشدید کردند. حدود 9000 روهینگیایی توسط نیروهای نظامی به قتل رسیدند و نزدیک به یک میلیون نفر از مرز به بنگلادش گریختند. حدود سه چهارم روهینگیایی‌هایی که در آن زمان در منطقه زندگی می‌کردند شخصاً شاهد یک قتل بودند، یک پنجم شاهد کشتار دسته جمعی بیش از 100 نفر بودند و اکثریت شاهد استفاده نیروهای نظامی از خشونت جنسی علیه زنان روهینگیا به عنوان بخشی از یک کارزار گسترده و سیستماتیک از پاکسازی قومی بودند. نیروهای نظامی که مرتکب این خشونت شدند، دولت منتخب دموکراتیک را در فوریه

2021 سرنگون کردند و به عملیات نظامی خود و همچنین سرکوب مخالفان و آزادی بیان در میانمار ادامه می‌دهند. ایالات متحده رسما خشونت علیه روهینگیا را نسل‌کشی نامیده‌است و دولت نظامی میانمار از همکاری با تحقیقات دادگاه کیفری بین المللی خودداری کرده است.

از حدود سال 2016، سخنان نفرت‌پراکنی علیه روهینگیاها در *Facebook* افزایش یافت که بیشتر آن به حساب‌های کاربری نیروهای نظامی در میانمار مرتبط بود. بیشتر این سخنرانی شبیه تحریکات خشونت‌آمیز بود که در نسل‌کشی‌های قبلی، از جمله در رواندا در سال 1994 دیده شده‌بود. هزاران پست وجودداشت که به ترویج غیرانسانی کردن مسلمانان روهینگیا و تحریک خشونت علیه آن‌ها، از جمله "شبیه کردن روهینگیاها به حیوانات، و تماس با آن‌ها پرداخته بود." برای کشته شدن روهینگیا، توصیف روهینگیا به عنوان مهاجمان خارجی و به دروغ متهم کردن روهینگیا به جنایات فجیع». پست‌های دیگر در *Facebook* مستقیماً به قتل، تجاوز و آوارگی اجباری روهینگیاها دامن زد.

شواهد واضح بود: حتی «ناتانیل گلیچر، رئیس سیاست امنیت سایبری خود *Facebook*»، اعتراف کرد که *Facebook* پست‌هایی را تبلیغ می‌کرد که «تلاش‌های واضح و عمدی برای پخش مخفیانه تبلیغاتی بود که مستقیماً به ارتش میانمار مرتبط بود». ارتش حساب‌های کاربری منتشر کرد و تبلیغاتی را پخش کرد که مخصوصاً برای برانگیختن نفرت طولانی‌مدت قومی علیه روهینگیا طراحی شده بود، از جمله «عکس‌های ساختگی از اجساد که به گفته آن‌ها شواهدی از قتل عام روهینگیا بود». هدف همه این‌ها توجیه سرکوب نظامی علیه اقلیت قومی مسلمان و بیرون راندن آن‌ها از میانمار بود (که البته با موفقیت بسیار زیادی انجام شد).

نفرت قدیمی و فناوری جدید

نقش *Facebook* در نسل کشی میانمار از آن زمان توسط سازمان ملل، تائید شده‌است. این موضوع همچنین موضوع یک شکایت دسته جمعی توسط گروهی از مسلمانان روهینگیا علیه شرکت *Meta*، شرکت مادر *Facebook*، در کالیفرنیا بوده است. در این شکایت 150 میلیارد دلار به عنوان غرامت و خسارات تنبیهی درخواست شده‌است. این شکایت ادعا می‌کند که *Facebook* در حذف سخنان تنفرآمیز که خشونت قومی را تحریک می‌کند، سهل‌انگاری کرده‌است، اما *Meta* همچنین ادعا می‌کند که یک ادعای جدید در مورد مسئولیت محصول، به دلیل طراحی معیوب الگوریتم‌های تعدیل‌محتوای *Facebook* بوده. تعدیل‌محتوا یکی از سخت‌ترین کارهایی است که شرکت‌های رسانه‌های اجتماعی باید انجام دهند. *Facebook* روزانه میلیاردها پست را تقریباً به هر زبان و فرهنگی در سراسر جهان تعدیل می‌کند. آن‌ها باید تفاوت‌های ظریف زبانی و فرهنگی را در یک حوزه فرهنگی در حال تغییر در نظر بگیرند (حوزه ای که خود توسط رسانه های اجتماعی و شیوه های تعدیل‌محتوای آن شکل گرفته‌است). *Facebook* سخنان نفرت پراکنی را حذف می کند (حدود 7 میلیون پست در سه‌ماهه سوم سال 2019). بیش از 80 درصد از سخنان نفرت‌انگیز حذف شده توسط الگوریتم‌های هوش‌مصنوعی (AI) شناسایی شد. بقیه توسط خود کاربران انجام می‌شوند. *Facebook* بیان می‌کند که تصمیم نهایی برای حذف یک پست به دلیل سخنان مشوق نفرت همیشه توسط یک ناظر انسانی گرفته می‌شود که شناسایی الگوریتم را بررسی می‌کند.

شرکت *Facebook*، الگوریتم‌هایی را برای شناسایی سخنان مشوق عداوت و تنفر به بیش از ۴۰ زبان توسعه داده‌است. در زمان نسل‌کشی روهینگیا، آن‌ها از هوش‌مصنوعی یا ناظر محتوای انسانی برای شناسایی و حذف سخنان نفرت‌انگیز به هیچ یک از زبان‌های رایج در برمه استفاده‌نکردند. *Facebook* همچنین فاقد مدیران محتوای انسانی است که به زبان‌ها و شیوه‌های فرهنگی بسیاری از کشورهای در حال توسعه مسلط هستند (از جمله بسیاری از کشورهای ضعیف که خشونت‌های قومی و سیاسی در آن‌ها شایع است). این بدان معناست که سخنان نفرت‌انگیز و تحریک به خشونت

بیشتر در این کشورها نسبت به کشورهایی مانند اروپا و آمریکای شمالی سریع‌تر گسترش می‌یابد و باعث ایجاد نابرابری‌های سیستماتیک در افرادی می‌شود که در معرض سخنان تنفرآمیز قرار می‌گیرند. *Facebook* از آن زمان برای رفع این مشکل تلاش کرده است: اکنون الگوریتم‌های تعدیل‌محتوا به زبان برمه‌ای را توسعه داده و حدود 100 ناظر محتوای برمه‌زبان (برای کشوری با بیش از 50 میلیون نفر از لحاظ زبانی و قومیتی) استخدام کرده است که به توسعه داده‌های آموزشی بهتر و طبقه بندی گفتارهای نفرت‌انگیز برای *Facebook* کمک می‌کنند.

وقتی کسی به تعدیل‌محتوا فکر می‌کند، معمولاً به الگوریتم‌هایی فکر می‌کند که برای بررسی حجم عظیمی از مطالب آنلاین و حذف زیرمجموعه‌ای از محتوای توهین‌آمیز طراحی شده‌اند. اما این امر از الگوریتم‌های فراگیر که تقریباً همه‌ی مطالب ارسال شده در تمام سایت‌های رسانه‌های اجتماعی را تبلیغ، توصیه و کاهش می‌دهند، در تلاشی بی‌پایان برای جلب مشارکت (و دلارهای تبلیغاتی) نادیده می‌گیرد. حتی زمانی که مطالب اعتراض آمیز این هدف را ترویج می کند (اما نه زمانی که کاربران را از خود دور می کند، استقبال می‌شود). یکی از مدیران محتوا برای *Cognizant*، یکی از پیمانکاران فرعی *Facebook*، به «کیسی نیوتن» گفت که کار مدیران محتوا برای برند *Facebook* اساسی است، و اظهار داشت: «اگر ما در آنجا نبودیم و این کار را انجام نمی‌دادیم، *Facebook* بسیار زشت بود. ما داریم می‌بینیم. ما همه‌ی چیزهایی که از طرف آن‌ها می‌آید را می‌بینیم».

حقیقت این است که همیشه همه‌ی محتوا تعدیل می‌شود. شکایت علیه *Meta* با این ادعا که الگوریتم‌های *Facebook* از بازاریابی، روان‌شناسی و علوم اجتماعی برای سوءاستفاده از آسیب‌پذیری ما در برابر محتوای عاطفی، هیجان‌انگیز، و تفرقه‌انگیز سیاسی استفاده می‌کنند، تلاش می‌کند تا این موضوع را با این ادعا که الگوریتم‌های *Facebook* را بیشتر و بیشتر از آن تغذیه می‌کند، جلب کند. همانطور که «رز استوکول» می‌گوید، «این همان کاری است که رسانه‌های اجتماعی مرتباً با ما انجام می‌دهند: ما را تشویق می‌کند تا درگیری‌ها را مشاهده کنیم و در موضوعاتی که در غیر این صورت نظرات کمی درباره آن‌ها داشتیم، طرف‌هایی را انتخاب کنیم. در هسته خود، این

یک دستگاه خدمات دهی به افکار است. و در رسانه‌های اجتماعی، همه نظرات به یک اندازه ارائه نمی‌شوند». نفرت، خشونت (حتی «اخبار جعلی» و انواع اطلاعات نادرست) در مورد آنچه که در رسانه‌های اجتماعی با آن درگیر می‌شویم، مزیت قابل توجهی دارند. از سوی دیگر، هرچه بیشتر به فیلتر کردن محتوا تشویق می‌کنیم، جریان آزاد گفتار و ایده‌ها را بیشتر زیر پا می‌گذاریم و همان «حباب‌های فیلتر» را که در وهله اول باعث ایجاد مشکل می‌شوند، بیشتر تبلیغ می‌کنیم.

به دلیل تازگی نسبی رسانه‌های اجتماعی در میان مردم، همراه با سانسور شدید میانمار و کمبود منابع اطلاعات، ممکن است نفرت قدیمی و فناوری جدید به شکلی سمی و خطرناک در میانمار با هم برخورد کرده‌باشند. در سال 2014، کمتر از 1 درصد از مردم به اینترنت دسترسی داشتند و این تعداد تا سال 2018 به حدود 15 میلیون نفر رسید (بیش از یک چهارم جمعیت). این اتفاق به این دلیل رخ داد که گوشی‌های هوشمند ارزان قیمت با سیم‌کارت‌های 1 دلاری پس از سال 2014 به بازار برمه سرازیر شدند (و تقریباً هر یک از این تلفن‌ها با *Facebook* از پیش نصب شده عرضه شدند). همانطور که ارزیابی حقوق‌بشر از نقش *Facebook* در نسل‌کشی روهینگیا بیان کرد، سواد دیجیتال و حاکمیت قانون در برمه بسیار ضعیف بودند.

میانمار درگیر «محدودیت‌های شدید آزادی بیان»، از جمله بازداشت خودسرانه روزنامه‌نگاران، و قوانین سرکوبگر افترا جنایی که برای سرکوب مخالفان طراحی شده‌اند، شناخته شده‌است. هوگان و سافی گزارش می‌دهند که یک تحلیلگر امنیت سایبری در یانگون اظهار داشت که این منجر به وضعیتی شده‌است که در آن *Facebook* مسلماً تنها منبع اطلاعات آنلاین برای اکثریت در میانمار است. *Facebook* می‌دانست که مردم از نظر دیجیتالی ساده لوح هستند، دولت درگیر سرکوب شدید اطلاعات است، و فضای سیاسی مملو از اختلافات قومی است و به شدت مستعد سخنان نفرت‌انگیز و خشونت‌آمیز است. آن‌ها از این موقعیت برای تقویت تعامل و افزایش درآمدهای تبلیغاتی در برمه استفاده کردند. همانطور که در شکایت علیه *Facebook* آمده است، "*Facebook* تصمیم شرکتی گرفت تا به سمت نفرت متمایل شود".

چه شکایت دسته معی علیه *Meta* موفقیت‌آمیز باشد یا نه، این بخشی از نقطه‌ی عطف در

نحوه نگرش ما به سیستم‌های هوش‌مصنوعی برای تعدیل محتوا است (در نحوه تشخیص و پاسخ به نقص طراحی در قلب این الگوریتم ه)ا. تعامل، مدل کسب‌وکار رسانه‌های اجتماعی را هدایت می‌کند، زیرا تعامل به معنای لایک، اشتراک‌گذاری و در نتیجه درآمدهای تبلیغاتی است. پست‌هایی با تعامل بالاتر در فیدهای خبری رسانه‌های اجتماعی بالاتر قرار می‌گیرند. محتوای نفرت‌انگیز و خشونت‌آمیز توسط تعداد زیادی حساب‌های جعلی تولید و تبلیغ می‌شود، که تعامل بالایی ایجاد می‌کند، و بنابراین الگوریتم‌های *Facebook* «آن را در فیدهای خبری کاربران واقعی اولویت‌بندی می‌کنند». در برمه، این شکایت ادعا می‌کند که الگوریتم‌های فیس‌بوک نه تنها در شناسایی و حذف سخنان نفرت‌انگیز علیه روهینگیا شکست خورده‌اند، بلکه از آن بهره‌برداری کرده و در فیدهای خبری کاربران تبلیغ کرده است. این تأثیر رادیکالیزه شدن کاربران و «تحمل، حمایت و حتی مشارکت در آزار و اذیت و خشونت قومی» علیه مسلمانان روهینگیا داشت.

تفسیر

اخلاق بودایی

نوشته پیتر هرشوک

اتصال دیجیتالی با واسطه محاسباتی، تبدیل تدریجی داده‌های منتقل‌شده با توجه را به جریان‌های درآمد و قدرت برای پیش‌بینی و تولید افکار و رفتار انسانی ممکن می‌سازد. این پتانسیل‌ها از تسریع مصرف مد سریع گرفته تا تقویت اتحادهای سیاسی پوپولیستی، تأثیرگذاری بر رای‌دهندگان نوسان، و دامن زدن به خشونت قومی را شامل می‌شود. گستره این پتانسیل ها از بیهوده تا قاتل به عنوان شاهدی است که اخلاق تعدیل محتوا ساده نیست و نمی‌تواند باشد.

به طور مثال، این مطالعه موردی روشن می کند که *Facebook* به طور همزمان در میانمار به عنوان یک سرویس خواستگاری غیرسیاسی و درآمدزا برای تولیدکنندگان و مصرف کنندگان، به عنوان بستری برای تحریک خشونت قومی، به عنوان مجرای برای به اشتراک گذاشتن شواهدی از همدستی دولت و ارتش در آن، خشونت، و به عنوان وسیله‌ای برای سازماندهی اعتراض‌ها و مبارزه با سلاح‌سازی احساسات ناامنی شخصی و جمعی، خدمت کرده‌است.

تلاش برای مسئول دانستن *Facebook* در قبال خشونتی که روهینگیا متحمل شده است بر اساس منطقی است که به راحتی قابل درک است. *Facebook* در جهت منافع شخصی تجاری خود و با ناآگاهی یا بی توجهی فعال به پتانسیل غم انگیز غفلت از نظارت بر محتوای منتشر شده از طریق پلت فرم خود عمل کرد. از مسئولیت های اخلاقی خود شانه خالی کرد.

اما آیا شرکت‌ها وظایف اخلاقی دارند؟ اگرچه شرکت‌ها «اشخاص حقوقی» در نظر گرفته می‌شوند، اما کارگزاران اخلاقی سنتی نیستند. *Facebook* شرکتی برای تجاری سازی یک رسانه‌ی دیجیتال یا حوزه ارتباطی است. ممکن است منطقی باشد که چنین شرکتی مسئولیت فنی جهانی را برای تعدیل محتوا بپذیرد. اما انتساب مسئولیت‌های اخلاقی محلی موضوع دیگری است زیرا هنجارهای اعتدال مطلوب و مجاز در بین 62.1 میلیارد کاربر روزانه آن بسیار متفاوت است. می‌توان این بحث را مطرح کرد که همانطور که کشاورزان (و نه مزارع‌شان) هستند که تعیین می‌کنند کدام محصولات را بکارند و بفروشند، این *Facebook* نیست که مسئول بذرهای خشونت کاشته‌شده در پلتفرم آن است. این کسانی هستند که سخنان نفرت‌انگیز را منتشر می‌کنند.

باز هم کشاورزان به تقاضاهای بازار پاسخ می‌دهند، و این قیاس نشان می‌دهد که مسئولیت انتشار سخنان نفرت‌انگیز رفع انسداد در *Facebook* را نمی‌توان تنها به کسانی نسبت داد که پست‌های نفرت‌انگیز نوشته‌اند. بدون اقدامات کسانی که آن‌ها را "لایک" و "به اشتراک گذاشتند"، آن پست ها نمی‌توانست چنین پیامدهای گسترده و خشونت‌آمیزی داشته‌باشد. آژانس با میانجیگری پلتفرم‌های رسانه‌های اجتماعی، حداقل مسئولیت توزیع شده را پیش‌فرض می‌گیرد.

مسائل زمانی پیچیده‌تر می‌شود که در نظر گرفته شود، در حالی که همه کسانی که به‌عنوان

عوامل مستقیم یا غیرمستقیم خشونت و اختلال در زندگی روزمره ناشی از سخنان نفرت‌انگیز منتشر شده در *Facebook* درگیر هستند، بیماران آن اقدامات ارتباطی و همه موارد دیگر نیز شده‌اند. پس از آن‌ها، از جمله تحریم‌های بین المللی، آسیب‌های آبروی و محکومیت اخلاقی، در مجموع، زیرساخت جهانی اتصال دیجیتال نه تنها مرزهای ملی را متخلخل می‌کند، بلکه مرزهای مفهومی را بین عوامل اخلاقی، اعمال و بیماران را نیز محو می‌کند.

هوش‌مصنوعی اغلب به عنوان یک فناوری همه‌منظوره شناخته می‌شود و به طور کلی فرض می‌شود که از نظر اخلاقی، خنثی است. هر گونه آسیب ناشی از فناوری هوشمند بر عهده‌ی کسانی است که ابزارهای الگوریتمی و یادگیری عمیق را طراحی، استقرار و استفاده می‌کنند. یعنی این آسیب‌ها به عنوان کارکرد تصادفی طراحی یا استفاده نادرست توسط طراحی در نظر گرفته می‌شوند. هستی شناسی رابطه‌ای بودایی چیز دیگری را نشان می‌دهد.

برای شروع درک چرایی آن، ابتدا تمایز بین ابزار و فناوری مفید است. ابزارها مصنوعات قابل بومی‌سازی هستند که ظرفیت‌های ما را برای عمل گسترش می‌دهند یا افزایش می‌دهند و ما به صورت جداگانه آزاد هستیم که از آن‌ها استفاده کنیم یا نه. فن‌آوری‌ها رسانه‌های رابطه‌ای غیرقابل بومی‌سازی هستند که مقاصد و ارزش‌های انسانی را افزایش می‌دهند و به‌طور انتخابی محیط‌هایی را که در آن تصمیم‌گیری می‌کنیم و عمل می‌کنیم تغییر می‌دهند و بر چگونگی و چرایی این کار تأثیر می‌گذارند. به این معنا که فناوری‌ها از رفتار انسان سرچشمه می‌گیرند و به رفتار انسان‌ها اطلاع می‌دهند/ ساختار می‌دهند، همان‌طور که اکوسیستم‌ها از روابط گونه‌ها بیرون می‌آیند و به صورت بازگشتی اطلاعات/ساختار روابط گونه‌ها را می‌دهند. ما «حق خروج» از فناوری را نداریم.

حوادث طراحی و استفاده‌ی نادرست از طریق طراحی، خطرات ابزار هستند. آسیب‌هایی که توسط الگوریتم‌های تحلیل محتوای محدود یا معیوب *Facebook* ممکن شده است، نمونه‌ای از موارد اول است. استفاده از *Facebook* برای ترویج سخنان نفرت‌انگیز و تحریک خشونت‌های قومی نمونه‌هایی از موارد اخیر است. هر دو اساساً ریسک‌های عامل هستند. ریسک‌های تکنولوژیکی ساختاری و رابطه‌ای هستند. برخلاف خطرات ابزار، آن‌ها از شرطی‌سازی پیچیده و بازگشتی روابط

انسان-فناوری-جهان سرچشمه می‌گیرند. خطرات تکنولوژیکی بسیار بیشتر از آسیب‌های محلی است که در مراحل علّی نهایی استفاده از ابزار رخ می‌دهد (نقطه‌ای که در آن عوامل، با یا بدون توجه به بیماران آن اقدام، نیات خود را عملی می‌کنند). بنابراین، وقتی لابی‌گران حقوق اسلحه استدلال می‌کنند که «اسلحه نمی‌کشد؛ مردم این کار را انجام می‌دهند.» آن‌ها در حال انجام تدبیر مفهومی هستند که به طرز ماهرانه‌ای توجه انتقادی را به سمت ابزارها (تفنگ‌ها) و طراحان و کاربران آن‌ها نادرست هدایت می‌کنند، و از فناوری سلاح دور می‌شوند (یک رسانه‌ی رابطه‌ای که محیط‌های تصمیم‌گیری را به روش‌های مساعد بازسازی می‌کند). منطقی کردن ایجاد آسیب از راه دور در پاسخ به تهدیدها، توهین‌ها یا تضاد منافع.

اخلاق بودایی، مبتنی بر شناخت منشأ همه چیز، که به یکدیگر وابسته هستند، نشان می‌دهد که پرسیدن اینکه کدام عوامل مسئول نتایج معین هستند، بسیار مهم‌تر از این است که بپرسیم چه ارزش‌ها و نیاتی در شکل‌دادن به پویایی گردش رسانه‌های اجتماعی نقش دارند. چه الگوهایی از نتایج تجربی و فرصت‌های ارادی توسط رسانه‌های ارتباطی فن‌آوری هوشمند افزایش یافته و به صورت بازگشتی تقویت می‌شوند؟

شرکت *Facebook* با انعکاس خاستگاه آمریکایی و لیبرالیسم دره سیلیکون *(Silicon Valley)*، آزادی انتخاب و بیان را ارزشمند می‌داند و حکمت اعتدال محتوای حداقلی و صرفاً واکنشی را فرض می‌کند. پیامدهای استفاده از رسانه‌های اجتماعی در میانمار این موضوع را زیر سؤال می‌برد. در مقابل، در حالی که چین از تکنیک‌های هوش‌مصنوعی و غربالگری انسانی مشابه *Facebook* استفاده می‌کند، سیاست‌ها و شیوه‌های مدیریت محتوای دیجیتال آن بر ارزش‌های ثبات سیاسی و هماهنگی اجتماعی متمرکز است. علاوه بر این، در حالی که تعدیل محتوای آن هنوز پیشگیری از آسیب را هدف قرار می‌دهد، همچنین هدف آن ایجاد عادات خوب شهروندی از طریق مشوق‌های رفتاری است. مدیریت محتوا به طور فعال جهت ارتقای رفاه اجتماعی است، همانطور که توسط دولت حزب تعریف شده‌است.

از منظر بودایی، چه رویکرد پیشگیرانه چین به مدیریت محتوا یا رویکرد مینیمالیستی و

واکنش‌گرایانه *Facebook* برای تعدیل محتوا قابل تحسین یا تأسف باشد، نباید به تأیید یا رد استفاده آن‌ها از ابزارهای یادگیری‌ماشین، صرفه‌جویی در مسئولیت آن‌ها یا حتی آن‌ها بسنده کرد. تأثیرات کوتاه مدت بر تک تک کاربران رسانه‌های اجتماعی. این باید به پیامدها و خطرات رابطه میان مدت و بلندمدت آن‌ها بستگی داشته‌باشد.

برخی از راهنمایی‌ها برای این ارزیابی توسط «راه هشت‌گانه بودایی» ارائه می‌شود، که مسیری را به سمت حضور روشنگرانه و روشنگرانه از طریق پرورش دیدگاه‌ها، نیات، گفتار، رفتار، معیشت، تلاش، توجه و تمرکز درست یا اصلاح‌کننده ترسیم می‌کند. به طور سنتی، گفتار درست، اصلاحی شامل پرهیز از دروغ، غیبت، تهمت، زبان تند، یا توهین آمیز، و همچنین پچ پچ و شایعات بیهوده است. بسیاری از آنچه توسط رسانه‌های اجتماعی منتشر می‌شود، به وضوح واجد شرایط نیستند.

یک ویژگی مثبت تر از خوب بودن در تمرین بودایی این است که منجر به آغشته کردن تمام موقعیت فرد به ویژگی های رابطه‌ای شفقت، مهربانی، متانت و شادی در خوشبختی دیگران می‌شود. در حال حاضر، این‌ها توابع هدفی نیستند که سیستم‌های هوش‌مصنوعی در حال حاضر برای بهینه‌سازی هدایت می‌شوند. اما آیا آن‌ها می‌توانند باشند؟

شرکت *Facebook* با موفقیت در دستکاری فیدهای رسانه‌های اجتماعی برای تأثیرگذاری بر احساسات کاربران آزمایش کرده‌است. خلبانان سیستم اعتبار اجتماعی چین ابزارهای قابل‌قبولی برای تقویت مدنیت عمومی به اثبات رسانده‌اند و محققان ژاپنی سیستم‌های هوش‌مصنوعی را توسعه داده‌اند که به‌طور دقیق احساسات را می‌خواند و انسان‌ها را به روش‌هایی درگیر می‌کند که به عنوان مراقب تجربه می‌شوند. همه‌ی این‌ها نشان می‌دهد که هیچ مانع فنی برای ایجاد شفقت، مهربانی، متانت، و شادی همدردی به عنوان کارکردهای عینی مدیریت محتوای رسانه‌های اجتماعی مبتنی بر هوش‌مصنوعی وجود ندارد. اتصال دیجیتال، شاید با طراحی آزاد کننده باشد.

اخلاق فضیلت

نوشته جان هکر رایت

این مورد عواقب واقعاً وحشتناکی را که می‌تواند ناشی از احتمالات جدید برای دستکاری افکار عمومی از طریق رسانه‌های اجتماعی باشد، نشان می‌دهد. نارسایی‌های آشکاری در رهبری شرکت Meta وجود دارد که نوعی بی‌عدالتی را در کسب سود بالاتر از امنیت روهینگیایی که هدف سخنان نفرت‌انگیز قرار گرفته‌اند، به نمایش گذاشت. در این نظر من از این موضع نسبتاً بدبینانه شروع می‌کنم که با وجود تلاش‌های فزاینده از سوی شرکت‌های رسانه‌های اجتماعی، ما به احتمال زیاد شاهد پایان سخنان نفرت‌انگیز و سایر اشکال محتوای دستکاری در پلتفرم‌های رسانه‌های اجتماعی نیستیم. به هر حال، همانطور که در این فصل اشاره شد، انگیزه‌ی قوی‌ای برای شرکت‌های شبکه‌های اجتماعی برای حفظ و ترویج محتوای تحریک‌آمیز با توجه به اینکه باعث تعامل بیشتر می‌شود، وجود دارد. اما حتی جدای از آن، این غیرواقعی است که تصور کنیم همه‌ی چنین محتوایی را می‌توان شناسایی و حذف کرد، حتی با بهترین نیت و بودجه‌ی قوی. با توجه به آن، چه فضیلت‌هایی را می‌توانیم به عنوان کاربران ایجاد کنیم که در برابر چنین دستکاری‌ها محافظت کرده و به طور بالقوه جان انسان‌ها را نجات دهد؟

توجه به این نکته مهم است که حساسیت به چنین دستکاری‌هایی با برخی از ویژگی‌های خوب معامله می‌شود که ما نباید در این فکر کنیم که چگونه خود را در برابر چنین پیامدهای غم انگیزی که در میانمار رخ می‌دهد، قربانی کنیم. تنها افرادی هستند که نگران شرایط جوامع و ملت خود هستند که می‌توانند بر اساس این نگرانی دستکاری شوند. دوستی مدنی، که متشکل از احساس مشترک هویت و نگرانی متقابل برای اعضای جامعه است، و میهن پرستی از آنجایی که در کارگزاران با فضیلت دیگر، اعضای یک جامعه را برمی‌انگیزد تا برای یک خیر عمومی اقدام کنند، مسلماً فضیلت هستند.

میهن‌پرستی اغلب با اشتیاق کور به کشور خود همراه است که انگیزه وفاداری بی‌فکر، نگرانی تنگ‌نظرانه‌ی زنجیروارانه و بدرفتاری با بیگانگان است. اگر میهن‌پرستی به عنوان یک اصطلاح فضیلت به کار می‌رود، این یک اشتیاق نیست، بلکه یک ویژگی مشخص از شخصیت است که عشق ما به کشورمان را مطابق با ویژگی‌های آن تنظیم می‌کند، چیزی مانند غرور مناسب در اعمال خود که با شایستگی‌های واقعی آن‌ها مطابقت دارد. در کمک به کشورهای همسایه، استقبال از تازه واردان، و مقاومت در برابر عوام فریبی می‌توان غرور میهن‌پرستانه‌ی مناسبی داشت. میهن‌پرستی همچنین باعث ایجاد احساس شرم در زمانی که کشور فرد ناعادلانه عمل می‌کند. از این رو، قابل‌قبول است که میهن پرستی را به عنوان یک اصطلاح فضیلت تلقی کنیم، در حالی که اذعان می‌کنیم که این اصطلاح اغلب به این شکل استفاده نمی‌شود.

اما در غیر این صورت می‌توان از ویژگی‌های خوب مانند دوستی مدنی و میهن‌پرستی برای ایجاد انگیزه در اعمال بد از طریق اطلاعات نادرست استفاده کرد. کسی که عمیقاً نگران جامعه خود است، هنگامی که اطلاعات نادرست در مورد تهدیدی برای جامعه خود دریافت می‌کند، ممکن است در برابر تهدید درک شده ناعادلانه عمل‌کند. استفاده از نیروی کشنده برای دفاع از جامعه خود در برابر تهدید عموماً از نظر اخلاقی مجاز است. «توماس آکویناس، فیلسوف و متکلم قرون وسطایی»، که نظریه اخلاقی ارسطو را به تفصیل و نظام مند ساخت، نظریه جنگ عادلانه را توسعه داد که بر اساس آن، تهدید برای جامعه به وضوح شرط کافی برای وارد شدن به درگیری نظامی است (البته نه برای ارتکاب جنایات یک بار علیه دشمن خود، نبرد آغاز شده است). با این حال آکویناس آنچه را که آشکارا به عنوان پیش‌زمینه در نظریه‌اش فرض می‌شود به صراحت بیان نمی‌کند: قضاوت در مورد تهدید باید از نظر معرفتی صحیح باشد.

در این راستا، فضیلت حکمت عملی بسیار مهم است. خرد عملی یک فضیلت فکری است که تفکر ما را در مورد آنچه انجام دهیم تنظیم می‌کند. شخصی با خرد عملی در مورد عمل به خوبی فکر می‌کند. بخش قابل‌توجهی از استدلال خوب در مورد آنچه که شامل استدلال از مقدمات واقعی است، و دستیابی به مقدمات واقعی، در زندگی عملی آسان تر از زمینه‌های علمی نیست، اما شامل

چالش‌های متمایز است. طبق اخلاق‌فضیلت ارسطویی، یکی از جنبه‌های استدلال از پیش‌فرض‌های واقعی، شامل داشتن فضایل اخلاقی مانند شجاعت است که احساسات ترس و اطمینان ما را تنظیم می‌کند، و اعتدال که اشتهای ما را برای غذا و رابطه جنسی تنظیم می‌کند.

این را در نظر بگیرید: برای کسی که شجاعت ندارد، به نظر می‌رسد یک تهدید بسیار بزرگتر از آن چیزی است که واقعاً هست (ترس او با هدف نامتناسب است). برای کسی که معتدل است، غذایی که ناسالم است یا متعلق به شخص دیگری است برای خوردن خوب به نظر می‌رسد. در اینجا ما یک تعهد اساسی ارسطویی را می‌بینیم: در عمل قضاوت‌های ما در مورد جهانی که در آن قرار داریم بر اساس احساسات ما است. بنابراین، فقدان فضیلت، ادراک ما را تحریف می‌کند و در نتیجه، مقدمات نادرستی را به وجود می‌آورد، به عنوان مثال، «آن مرد، آنجا، بسیار خطرناک است» یا «آن کیک کوچک برای خوردن خوب است». از آن فرض‌های نادرست، احتمالاً به نتایجی می‌رسیم که منجر به اقدامات بدی می‌شود، مانند «فرار می‌کنم» یا «آن را می‌خورم». در مورد خشم، کسی که زود عصبانی می‌شود ممکن است به دروغ درباره کسی که شایسته قصاص است قضاوت کند. البته، استدلال ما همیشه به این صراحت بیان نمی‌شود، و اغلب به روشی سریع و ضمنی، سیستم 1 رخ می‌دهد، اما اگر املا شود، ممکن است چیزی شبیه به بازسازی به‌نظر برسد. نتیجه این است که در غیاب فضایل اخلاقی، استدلال ما در مورد چگونگی عمل مخدوش خواهد شد. از این رو، داشتن فضیلت برای پاسخگویی مناسب به موقعیت‌هایی که در آن قرار داریم، از جمله موقعیت‌هایی که از طریق رسانه‌های اجتماعی به ما ارائه می‌شود، مهم است.

تا اینجا، تصویری از استدلال عملی که ارائه کردم تا حد زیادی «ادراکی» است، به این معنا که ما را در حالی نشان می‌دهد که از نقطه نظر خواسته‌هایمان به دنیا نگاه می‌کنیم، و اگر خواسته‌هایمان مرتب باشد، از طریق یک تربیت خوب (ما تمایل داریم که خوب عمل کنیم). اما داشتن حکمت عملی بیش از داشتن فضایل اخلاقی است. به گفته ارسطو، افراد دارای خرد عملی «درباره آنچه برای خود و برای انسان خوب است، درک نظری دارند». این تا حدودی به داشتن دانش‌واقعی اولیه در مورد آنچه در حوزه‌هایی مانند تغذیه خوب است، مربوط می‌شود، اما مهم‌تر از آن، داشتن بینشی در

مورد آنچه که از نظر یک زندگی خوب خوب است، است. به عبارت دیگر، بینش در مورد اینکه چه اعمالی برای یک انسان بهتر است. ارسطو فکر می کرد که «پریکلس، دولتمرد مشهوری که آتن» را در بخشی از «جنگ پلوپونز» رهبری کرد، چنین مردی است. کاملاً مشخص نیست که ارسطو چه چیزی را در حمایت از پریکلس می‌ستود، اما در یک سخنرانی معروف در مراسم تشییع جنازه، همانطور که توسیدید گزارش می‌دهد، پریکلس زندگی با مشارکت اجتماعی فعال، دنبال شرافت و به خطر انداختن مرگ به خاطر آزادی را می‌ستاید. حتی اگر با این تصور از خیر انسانی مخالف باشیم، در برابر برخی دیدگاه‌های کلی در مورد اینکه بهترین نوع فعالیت برای یک انسان چیست، پیشنهادهای خاص مورد ارزیابی قرارمی‌گیرد و این ادعاهای کلی نیز از جمله مقدمات استدلال ما این است که «برای انسان شایسته است که. . .» یا «برای انسان بهتر است که. . .».

یکی دیگر از مؤلفه‌های حکمت عملی، که در این مورد اهمیت ویژه‌ای دارد، یک ظرفیت‌فکری است که ارسطو آن را «درک» می‌نامد، که ظرفیتی است که به وسیله آن موقعیتی را که در آن قرار داریم درک می‌کنیم. در توضیح این ویژگی، اخلاق‌دان فضیلت ارسطویی «روزالیند هرست هاوس» به موارد زیر اشاره می‌کند:

> «موقعیتی» که به انجام کاری نیاز دارد، ممکن است اصلاً با من روبرو نباشد و منتظر باشد تا آن را بخوانم، بلکه چیزی است که باید جزئیات آن را از آنچه دیگران در مورد آن می‌گویند بررسی کنم. و تا زمانی که نتوانم قضاوت درستی در مورد گزارش‌های آن‌ها در مورد مسائل مربوطه داشته‌باشم، هر نتیجه عملی‌ای که در مورد اینکه در «این وضعیت» چه کنم، در تاریکی انجام می‌شود.

گزارش هرست هاوس از این جزء از خرد عملی، این را به عنوان یک واقعیت در مورد زندگی انسان تصدیق می کند که ما اغلب در موقعیتی قرار داریم که تصویری از آنچه در جامعه ما اتفاق می‌افتد بر اساس گزارش‌های دیگران از آن بسازیم. و اغلب دیگران اطلاعات نادرست دارند یا قصد

دارند عمداً اطلاعات نادرست را ارائه دهند. در تشخیص اینکه چه کسی قابل اعتماد است باید درجاتی از زرنگی به دست آوریم. چنین حساب‌هایی ظاهر یک معضل یا مسیرهای عمل ضروری را در جایی که گزینه‌های بیشتری وجود دارد ایجاد می کند. ارزیابی حساب‌ها و توانایی رد کامل آن‌ها و انجام تحقیقات بیشتر به تنهایی ظرفیتی حیاتی برای خرد عملی است که توسعه آن به تجربه نیاز دارد.

آنچه در این مورد می‌بینیم، نیاز به بسط «فنی اخلاقی» مفهوم ارسطو از درک است. بدیهی است که رسانه‌های جمعی و رسانه‌های اجتماعی چالش‌های جدیدی را برای درک مطلب ایجاد می‌کنند که ما را ملزم به توسعه مهارت‌های مناسب برای شناخت تحریف‌ها می‌کند. ما می‌توانیم سوگیری‌های سیاسی و حس‌گرایی را در رسانه‌های جمعی تشخیص دهیم و تمایلات خود را برای عمل بر اساس آن‌ها قطع کنیم. به همین ترتیب، می‌توانیم گرایش رسانه‌های اجتماعی به ما را به چیزهایی که قبلاً به آن اعتقاد داریم، ببندیم، دیدگاه‌های متضادی را که ممکن است قضاوت‌های ما را به چالش بکشند، و پناه دادن به حساب‌های ربات جعلی که ما را به نتایجی راهنمایی می‌کنند که در غیر این صورت به آن‌ها نمی‌رسیدیم، برسانند.

در زمینه فناوری‌های نوظهور، مؤلفه حکمت عملی، که ارسطو تنها چند سطر از رساله خود را صرف آن کرده‌است، اهمیت فوق‌العاده‌ای پیدا می‌کند. درک فنی اخلاقی به عنوان ظرفیتی برای تشخیص اطلاعات قابل‌اعتماد در رسانه های اجتماعی، ظرفیتی است که، به نظر من، حتی در مکان‌هایی که از همان ابتدا رسانه‌های اجتماعی داشته‌اند، به طور گسترده‌ای به نمایش گذاشته نمی‌شود. توسعه‌ی آن مطمئناً یک فرآیند ناهموار خواهدبود، اما به همان اندازه مطمئن است که فضیلت حیاتی که ما در مواجهه با دستکاری گسترده در سیستم عامل‌های رسانه‌های اجتماعی به آن نیاز داریم. و این چیزی نیست که بتوان آن را در معماری رسانه‌های اجتماعی طراحی کرد. در هر صورت، اعتماد بیش از حد به الگوریتم‌ها مانع توسعه آن خواهد شد. پرچم‌هایی که روی داستان‌هایی که مشکوک تلقی می‌شوند قرار می‌گیرند ممکن است ما را به اعتماد نابجا نسبت به داستان‌هایی که چندان پرچم‌دار نیستند، جلب کند. اگر طراحان پلتفرم‌های رسانه‌های اجتماعی می‌خواهند کمک کنند، شاید بهتر است با تشویق و تأمین مالی تلاش‌های مستقل و تحت رهبری مربیان برای ایجاد

مهارت‌های رسانه‌ای تفکر انتقادی که درک فنی اخلاقی را تقویت می‌کنند، انجام‌شود.

اخلاق بومی

توسط جوی میلر و آندریا سالیوان کلارک

ارتباط بین اعتدال محتوا و پاکسازی قومی در میانمار (یعنی برمه) اهمیت اخلاقی این ایده را در فلسفه‌ی بومی نشان می‌دهد که همه‌چیز به هم مرتبط است. الگوریتم‌ها به خودی خود باعث آسیب نمی‌شوند. با این حال، نحوه‌ی طراحی و نحوه‌ی استفاده از آن‌ها مطمئناً می‌تواند باعث آسیب شود. این به دلیل تعداد بی‌شماری از حقایق در مورد نحوه ارتباط و تعامل انسان با محیط اطراف است (به عنوان مثال، ویژگی‌های مختلف روان‌شناختی، بیولوژیکی، اجتماعی، فیزیولوژیکی و غیره انسان و نحوه استفاده از این ویژگی‌ها برای حرکت در محیط اطراف خود). این بدان معناست که مفاهیم اخلاقی برای نحوه طراحی و استفاده از الگوریتم‌ها وجود دارد. در حالی که *Facebook* خود مستقیماً یا عمداً در قتل و فرار روهینگیایی ها مشارکت نداشته‌است، *Facebook* بدون شک مسئولیت این نسل‌کشی را بر عهده دارد.

بر اساس درک بومی از اخلاق، مسئولیت *Facebook* ناشی از عدم به کارگیری افراد کافی که به زبان برمه صحبت می کنند نیست. به این ترتیب، استخدام تعداد بیشتری از مدیران محتوای برمه‌زبان این مشکل را برطرف نمی‌کند. در عوض، استفاده از الگوریتم‌های *Facebook* که به‌گونه‌ای طراحی شده‌اند که با بهره‌برداری از ویژگی‌های روان‌شناختی انسان (یعنی بیولوژیکی، اجتماعی و غیره) تعامل را ارتقا دهند، مشکل‌ساز است. به طور خاص، در مورد عواقب الگوریتم‌هایی که *Facebook* استفاده می‌کند، پیش‌بینی یا توجهی وجود ندارد، به این معنی که ارتباط بین این الگوریتم‌ها و تأثیرات آن‌ها (یعنی به هم پیوستگی همه چیز) یا در نظر گرفته نمی‌شود، نادیده گرفته می‌شود یا نادیده گرفته می‌شود. کنش‌ها به‌عنوان کنش‌های (مقابله) شناخته نمی‌شوند.

حداقل روشن است که توجه کافی به آسیب‌هایی که ممکن است از این نوع الگوریتم‌ها به‌وجود بیاید، صورت نگرفته‌است. با توجه به افزایش آسیب ناشی از انتشار اطلاعات نادرست، و همچنین ترویج سخنان نفرت‌انگیز و خشونت، واضح است که *Facebook* درک کافی از مشتریان خود یا پیامدهای الگوریتم‌های خود نداشته‌است. از این نظر، *Facebook* به برهم‌زدن (یا حداقل برهم زدن بیشتر) تعادل و رفاه در میان مردم و گروه‌های درگیر در نسل‌کشی در میانمار کمک کرد. به عبارت دیگر، *Facebook* به ترویج ناهماهنگی کمک کرد.

با توجه به اینکه، در فلسفه‌ی بومی، (تعامل) کنش‌ها تا حدی درست یا نادرست هستند که هماهنگی را ترویج یا مختل کنند، واضح است که استفاده *Facebook* از این الگوریتم‌ها اشتباه است. با این حال، شیوه‌های «غیرمستقیم» یا «غیر عمدی» *Facebook* همچنان به برهم‌زدن هماهنگی کمک می‌کنند. حتی اگر ناهماهنگی قبلاً در میانمار وجود داشته‌باشد، برهم‌زدن بیشتر هماهنگی همچنان اشتباه است. این لزوماً استفاده از همه الگوریتم‌ها را به منظور تعدیل محتوا رد نمی‌کند. این فقط به این معنی است که اگر قرار است از الگوریتم‌ها استفاده‌شود، افرادی که از آن‌ها استفاده می‌کنند باید درک کافی از محدودیت‌ها، دریافت و پیامدهای آن‌ها داشته‌باشند. بدون چنین درک، تمایل بیشتری برای برهم‌زدن هارمونی وجود دارد.

هیچ قانون قابل اجرا جهانی در مورد چگونگی ارتقای هماهنگی در همه‌ی شرایط وجود ندارد. در حالی که یک ایده‌ی کلی وجود دارد که هماهنگی باید ترویج شود، چگونگی ارتقای هماهنگی به محیط اطراف فرد (یعنی موقعیت و شرایط آن‌ها) بستگی دارد. بخشی از انگیزه‌ی پشت این ایده این است که متغیرهای زیادی وجود دارد که در هنگام طراحی نظریه‌های اخلاقی نمی توان آن‌ها را درنظر گرفت. این موضوع در مورد طراحی الگوریتم‌ها نیز صادق است. الگوریتم‌ها، مانند نظریه‌های اخلاقی، نمی‌توانند برای توضیح همه‌ی موارد احتمالی اشتباه طراحی شوند. در واقع، همچنین مانند نظریه‌های اخلاقی، می توان از آن‌ها برای ترویج اشتباه استفاده کرد. این دو ویژگی یا الگوریتم، (1) محدودیت‌های آن‌ها و (2) استفاده از آن‌ها برای ترویج اشتباه، همچنین می‌تواند به تأکید بر اینکه چرا در فلسفه‌ی بومی، هیچ اصل اخلاقی وجود ندارد که در همه موقعیت‌ها یکسان اعمال شود، کمک

کند.

شرایط، موقعیت‌ها و زمینه‌ها اهمیت دارد. بیان فروتنی فکری با آگاهی مناسب از محدودیت‌های یک الگوریتم و اینکه چگونه می‌توان از آن‌ها برای اشتباه استفاده کرد، باید به راهنمایی در مورد نحوه طراحی الگوریتم‌ها یا تصمیم‌گیری از کدام الگوریتم‌ها کمک کند. در فلسفه‌ی بومی کلمات قدرت دارند. با تنزل دادن تعدیل (مثلاً ترویج و تنظیم) کلمات به الگوریتم‌ها، نمی‌توان محدودیت‌های این الگوریتم‌ها و/یا قدرت کلمات را تشخیص داد. به عبارت دیگر، آن‌ها فروتنی نشان نمی‌دهند و این منجر به برهم خوردن هماهنگی یا ترویج ناهماهنگی می‌شود.

فصل ۸

بدافزارهای ذهنی: الگوریتم‌ها و معماری انتخاب

ما نمی‌خواهیم از مردم بپرسیم که قرار است چه کاری انجام دهند. زیرا می‌دانیم که این امر چندان پیش‌بینی‌کننده نحوه اجرای یک تبلیغ نیست، زیرا افراد به مغز چپ خود می‌روند و بیش از حد شروع به فکر کردن می‌کنند.

کری کالینگ، مدیر بازاریابی در شرکت بازاریابی *«System 1 Group»*

رسوایی داده های کمبریج آنالیتیکا

رسوایی داده‌های کمبریج آنالیتیکا ریشه در سال 2010 داشت که *Facebook* برنامه *Open Graph* خود را راه‌اندازی کرد. *Open Graph* به توسعه دهندگان برنامه‌های شخص ثالث اجازه می‌دهد تا به اطلاعات شخصی کاربران *Facebook* و همچنین همه‌ی داده‌های «دوستان» خود دسترسی داشته‌باشند. در سال 2013، «محقق دانشگاهی الکساندر کوگان»، در همکاری با شرکت بازاریابی و تجزیه و تحلیل داده‌ها، کمبریج آنالیتیکا، اپلیکیشنی به نام «این زندگی دیجیتال شماست» راه‌اندازی کرد. این اپلیکیشن از کاربران دعوت کرد تا در یک مسابقه شخصیت‌شناسی رایگان شرکت کنند و حدود 300000 کاربر *Facebook* این کار را انجام‌دادند. این برنامه داده‌های مربوط به پروفایل‌های روان‌سنجی آن‌ها را از آزمون جمع‌آوری کرد (که پنج ویژگی شخصیتی بزرگ کاربران را اندازه‌گیری می‌کرد) اما همچنین به‌طور آزادانه داده‌های *Facebook* را از همه‌ی دوستان

آن‌ها جمع‌آوری کرد. کمبریج آنالیتیکا در تلاش بود تا مجموعه‌ای از رای‌دهندگان آمریکایی را تا حد امکان جمع‌آوری کند.

در سال 2015، اولین گزارش‌ها منتشر شد مبنی بر اینکه کمپین سیاسی تد کروز میلیون‌ها نفر از این پروفایل‌های روان‌سنجی را در تلاش برای کسب مزیت در انتخابش به مجلس سنای ایالات متحده تجزیه و تحلیل کرده است (افشاگری که کاملاً نامحبوب بود و منجر به تضمین‌هایی از سوی *Facebook* و کمبریج آنالیتیکا که داده‌های مورد نظر حذف شده است).

با این حال، در سال 2018 اخبار منتشر شد مبنی بر اینکه داده‌های ده‌ها میلیون کاربر *Facebook* (شاید به 87 میلیون نفر) توسط کمبریج آنالیتیکا جمع‌آوری‌شده و در انتخابات ریاست‌جمهوری آمریکا در سال 2016 توسط ستاد انتخاباتی دونالد ترامپ استفاده شده‌است. بسیاری از این افراد تست شخصیت را انجام نداده بودند، اما از زمانی که یکی از دوستانشان شرکت کرده بود، محققان می‌توانستند آزادانه به داده‌های آن‌ها دسترسی داشته‌باشند. سپس کمبریج آنالیتیکا داده‌های *Facebook* را با سایر داده‌هایی که خریداری کرده بود و همچنین فهرست‌های انتخاباتی محلی ارجاع داد. بنابراین، کمبریج آنالیتیکا توانست پرونده‌های گسترده‌ای را در مورد ده‌ها میلیون رای‌دهنده، از جمله ویژگی‌های جمعیتی، ویژگی‌های شخصیتی، شبکه‌های اجتماعی، تاریخچه خرید، لایک‌ها، عضویت در احزاب سیاسی و غیره جمع‌آوری کند. «مک‌نامی» تخمین می‌زند که این پرونده‌ها در نهایت شامل حدود 13 درصد همه‌ی رای دهندگان واجد شرایط ایالات متحده بوده.

کمبریج آنالیتیکا از مشخصات رأی دهندگان خود برای کسب برتری در انتخاب دونالد ترامپ به عنوان رئیس جمهور ایالات متحده در سال 2016 و همچنین کمپین «خروج» رفراندوم برگزیت در بریتانیا استفاده کرد. همانطور که *Cadwalladr* توضیح می‌دهد، کمبریج آنالیتیکا از نتایج آزمایش و داده‌های *Facebook* برای ساخت الگوریتمی استفاده کرد که می‌تواند پروفایل‌های فردی *Facebook* را تجزیه و تحلیل کند و ویژگی های شخصیتی مرتبط با رفتار رأی‌گیری را تعیین کند. این الگوریتم به‌ویژه مؤثر بود زیرا به دانشمندان داده اجازه می‌داد تا رأی‌دهندگان نوسان را شناسایی

کنند و سپس آن‌ها را با تبلیغات و پیام‌های خاصی که به احتمال زیاد رأی آن‌ها را «تحریک» می‌کرد، هدف قرار دهند.

معماری انتخابی و فناوری متقاعد کننده: "جعبه ای برای انسان مدرن"

کمبریج آنالیتیکا از داده‌ها برای آموزش الگوریتم‌های توصیه و ترویج محتوای رسانه‌های اجتماعی «قادر به حرکت دادن افکار عمومی در مقیاس» استفاده‌کرد. الگوریتم‌های کمبریج آنالیتیکا این کار را با دسته‌بندی خرد افراد در گروه‌هایی که با ویژگی‌های جمعیت‌شناختی، سیاسی و روان‌سنجی تعریف می‌شوند انجام می‌دهند: برای مثال، رای دهندگان زن محافظه کار اجتماعی در حومه‌های مرفهِ دی سی که فرزندانشان تحت تاثیر تعطیلی مدارس مرتبط با کووید قرارگرفته‌اند، رای دهندگان طبقه کارگر در کمربند زنگ زده که درازمدت کم کار هستند، بازنشستگان طبقه پایین از فلوریدا مرکزی که نگران افزایش هزینه‌های مراقبت‌های بهداشتی هستند. هدف از این الگوریتم‌ها هدف قرار دادن «رای‌دهندگان نوسان» بسیار مورد علاقه است تا بتوانند رأی خود را در مناطق مهم میدان نبرد تحت‌تأثیر قراردهند. بیش از این، آن‌ها می‌توانند الگوریتم‌های خود را در زمان واقعی در گروه‌های متمرکز آموزش‌دهند و بهبود بخشند.

اگرچه *Facebook* به دلیل نقض داده‌ها توسط کمیسیون تجارت فدرال 5 میلیارد دلار جریمه شد، خطرات آنچه به عنوان "فناوری متقاعد کننده" شناخته می‌شود بسیار فراتر از رسوایی داده‌های کمبریج آنالیتیکا است. مربیان و روانشناسان برای سال‌ها نگرانی‌هایی را در مورد فناوری متقاعدکننده مطرح کرده‌اند، اما این تأثیر کمی بر صنایع بازاریابی مصرف‌کننده و سیاسی داشت. این رسوایی بسیار بیشتر از نقض حریم خصوصی کاربران است. همانطور که مک نامی بیان می‌کند، این در مورد این است که چگونه "داده‌های ما هوش‌مصنوعی را تغذیه می‌کند که هدف آن‌ها دستکاری توجه و رفتار

کاربران بدون اطلاع یا تایید آن‌ها است".

در حالی که بازاریابی مبتنی بر گروه‌های جمعیتی سابقه طولانی دارد، این عمل با حجم عظیمی از داده‌های ایجاد شده توسط رسانه‌های اجتماعی افزایش یافته‌است. به عنوان مثال، *Facebook* گروه‌هایی به نام *"lookalikes"* ایجاد کرده‌است که کاربران را به گروه‌هایی با پروفایل‌های مشابه طبقه‌بندی می‌کند تا به شرکت در هدف‌گیری خرد آن‌ها کمک کند. «کریستوفر ویلی، دانشمند سابق داده که در کمبریج آنالیتیکا افشاگر شد»، اظهار می‌دارد که به کاربران محتوایی بر اساس گروه مشابه خود ارائه می‌شود که سایر کاربران نمی‌بینند. این امر باعث ایجاد حباب‌های فیلتر شده و شکاف‌های اجتماعی را عمیق‌تر می‌کند. او بیان می‌کند که «خط ظریفی بین الگوریتمی وجود دارد که شما را تعریف می‌کند تا نشان دهد واقعاً چه کسی هستید و الگوریتمی که شما را برای ایجاد یک پیش‌گویی خودشکوفایی از اینکه فکر می‌کند باید تبدیل شوید، تعریف می‌کند».

شبیه‌سازی‌ها و ریزهدف‌گذاری از قدرت علم‌داده برای درگیرشدن (هرچند بسیار مؤثرتر) در رویه‌ی قدیمی تبلیغات استفاده می‌کنند. بازاریابان، برای شرکت‌های مصرف‌کننده و کمپین‌های سیاسی، به طور معمول محتوای اخلاقی و بسیار احساسی را ارائه می‌دهند که به سرعت در رسانه‌های اجتماعی پخش می‌شود. در واقع، رسانه‌های اجتماعی و دیگر پلت‌فرم‌های آنلاین اکنون «منابع اولیه محرک‌های اخلاقی مرتبطی هستند که افراد در زندگی روزمره خود تجربه می‌کنند». کسانی که از رسانه‌های اجتماعی برای تأثیرگذاری بر افکار عمومی استفاده می‌کنند، از یادگیری پاداش اجتماعی نیز استفاده می‌کنند (معمولاً به شکل «اشتراک‌گذاری»، «کلیک»، «لایک»، «فالوور» و دیگر اشکال تقویت‌کننده تعامل). این رفتارها نه تنها بسیار پاداش دهنده هستند، بلکه می‌توانند توسط سیستم‌های یادگیری‌ماشین استخراج شوند تا رفتار آینده‌ی ما، دوستان و پیروان ما و گروه‌های مشابه ما را پیش‌بینی کنند. فردی را که با محتوای آنلاین درگیر می‌شود مانند موش در جعبه اسکینر که اهرم پاداش را فشار می‌دهد تصور کردند و به این نتیجه رسیدند که رسانه‌های اجتماعی مانند "جعبه اسکینر برای انسان مدرن است".

مقایسه بین تعامل در رسانه‌های اجتماعی و یک موش آزمایشی در یک جعبه عمیق است. در

سال 2014، مطالعه ای با همکاری *Facebook* و دانشگاه کرنل منتشر شد)این آزمایش شامل آزمایشگاه غذا و برند نیست، بلکه دپارتمان علوم ارتباطات و اطلاعات بود(. محققان محتوای احساسی پست‌هایی را که کاربران *Facebook* در فیدهای خبری خود دریافت می‌کردند، دستکاری کردند، به‌ویژه از افرادی که به آن‌ها اعتماد داشتند، مانند دوستان و کسانی که دنبال می‌کردند. محققان می‌خواستند ببینند آیا محتوای عاطفی مثبت در مقابل منفی بر خلق و خوی پست‌های بعدی کاربران تأثیر می‌گذارد یا خیر (به عنوان مثال، آیا شواهدی از «سرایت عاطفی» در رسانه‌های اجتماعی وجود دارد یا خیر). وجود داشت، اما این مهم‌ترین جنبه مطالعه نبود. به شرکت کنندگان اطلاع داده‌نشد که از آن‌ها به عنوان موضوع تحقیق استفاده می‌شود. در واقع، رضایت آن‌ها برای شرکت در تحقیقات تجربی هرگز دریافت‌نشد. از آنجایی که داده‌ها توسط *Facebook* جمع‌آوری شده‌بود، محققان حتی به دنبال تأیید هیئت بررسی اخلاق پژوهشی کورنل نبودند. معلوم نیست اگر می‌گرفتند تایید می‌شدند.

هنگامی که اخبار عدم‌رضایت منتشر شد، واکنش‌های منفی علیه این مطالعه وجود داشت. اصل اصلی اخلاق تحقیق مستلزم کسب رضایت آگاهانه از شرکت کنندگان در تحقیق است. این اصل در زمینه اخلاق پزشکی ایجاد شد (به کادر 8.1 مراجعه کنید) اما از آن زمان به تمام زمینه‌های تحقیقات آکادمیک مربوط به موضوعات انسانی گسترش یافته است. بسیاری از متخصصان اخلاق زیستی استدلال می‌کردند که این تحقیق «به‌طور فاحش» هیچ‌یک از اصول قانون یا اخلاقی را نقض نمی‌کند، و اگر این کار را انجام می‌داد، به این معناست که شیوه‌های استاندارد *Facebook* نیز از نظر اخلاقی مشکوک هستند. «کاترین فلیک» با این استدلال پاسخ داد که اصول اخلاقی شرکت‌هایی که به طور معمول کاربران را بدون اطلاع یا رضایت آن‌ها در معرض دستکاری آزمایشی و آزمایشی قرار می‌دهند، دقیقاً موضوع مورد بحث است.

در حالی که دانشمند اصلی در مطالعه سرایت عاطفی عذرخواهی کرد، این نقض اخلاقی رسوایی کمبریج آنالیتیکا را که به زودی دنبال می‌شود، پیش‌بینی کرد. اساتید برجسته در دانشگاه‌های معتبری مانند کمبریج و هاروارد از توسعه الگوریتم کمبریج آنالیتیکا اطلاع داشتند و آن را هیجان‌انگیز

و نوآورانه می‌دانستند. به نظر می‌رسد که بحثی در مورد اخلاق تحقیق وجود نداشته‌است. همانطور که «ویلی» می‌گوید، «با توجه به اینکه دانشمندان دانشگاه‌های برجسته جهان به من می‌گفتند در آستانه «انقلاب‌سازی» علوم اجتماعی هستیم، من حریص شده‌بودم و جنبه‌های تاریک کاری را که انجام می‌دادیم نادیده می‌گرفتم».

کادر 8.1

اصول قانون نورنبرگ

اصول نورنبرگ در مورد آزمایش انسان عبارتند از:

۱ رضایت داوطلبانه سوژه انسانی کاملاً ضروری است.

۲ آزمایش باید به گونه ای باشد که نتایج مثمر ثمری برای صلاح جامعه داشته باشد، غیرقابل تهیه با روش ها یا وسایل مطالعه دیگر باشد و ماهیت تصادفی و غیر ضروری نداشته باشد.

۳ آزمایش باید به گونه ای طراحی و بر اساس نتایج آزمایش بر روی حیوانات و آگاهی از تاریخچه طبیعی بیماری یا سایر مشکلات مورد مطالعه باشد که نتایج پیش بینی شده انجام آزمایش را توجیه کند.

۴ آزمایش باید به گونه ای انجام شود که از همه رنج ها و آسیب های جسمی و روحی غیر ضروری جلوگیری شود.

۵ هیچ آزمایشی نباید در جایی انجام شود که دلیل پیشینی وجود داشته باشد که باور شود مرگ یا آسیب ناتوان کننده رخ خواهد داد. به جز، شاید، در آزمایش هایی که پزشکان تجربی نیز به عنوان سوژه خدمت می کنند.

۶ درجه خطری که باید متحمل شود هرگز نباید بیشتر از میزانی باشد که با اهمیت انسان دوستانه مشکلی که باید توسط آزمایش حل شود تعیین می شود.

۷ باید آماده‌سازی مناسب و امکانات کافی برای محافظت از آزمودنی آزمایشی در برابر احتمالات دوردست آسیب، ناتوانی یا مرگ فراهم شود.

۸ آزمایش فقط باید توسط افراد واجد شرایط علمی انجام شود. بالاترین درجه مهارت و مراقبت باید در تمام مراحل آزمایش کسانی که آزمایش را انجام می دهند یا درگیر آن هستند، لازم باشد.

۹ در طول آزمایش، آزمودنی انسانی باید آزاد باشد که آزمایش را به پایان برساند، اگر به وضعیت جسمی یا روانی رسیده باشد که ادامه آزمایش به نظر او غیرممکن است.

۱۰ در طول آزمایش، دانشمند مسئول باید آمادگی داشته باشد که آزمایش را در هر مرحله خاتمه دهد، اگر احتمالاً دلایلی برای باور داشته باشد، به اعمال حسن‌نیت، مهارت برتر و قضاوت دقیق که از او مستلزم ادامه است. این آزمایش احتمالاً منجر به آسیب، ناتوانی یا مرگ آزمودنی می شود.

فناوری متقاعدکننده بخشی از حوزه وسیع‌تر «معماری انتخاب» است. یک معمار انتخابی «زمینه‌ای را که مردم در آن تصمیم می‌گیرند» سازمان‌دهی می‌کند. چیزی به نام زمینه‌ی خنثی

وجود ندارد: همه‌ی شرایط حداقل مقداری فشار برای تصمیم‌گیری به یک روش یا روش دیگر اعمال می‌کنند. همه‌ی ما در یک حوزه‌ی اجتماعی پیچیده و پویا تازه‌کار هستیم که توسط متخصصان، متخصصان و الگوریتم‌های بسیار آموزش دیده پر شده‌است (که هدف ترکیبی آن عمدتاً فروش چیزی به ما یا تشویق ما برای پذیرش یک عقیده یا نامزد سیاسی بر دیگری است). الگوهای توصیه‌ای از سوگیری‌های شناختی و استعدادهای ناخودآگاه بهره می‌برند و از «افراد پرمشغله‌ای که سعی می‌کنند در دنیای پیچیده‌ای که در آن نمی‌توانند در مورد هر انتخابی که باید عمیقاً و طولانی فکر می‌کنند، کنار بیایند» سود می‌برند.

بازاریابان می‌توانند با تجزیه و تحلیل حالات چهره‌ی ما در زمان واقعی، به واکنش‌های احساسی (ناخودآگاه و چند ثانیه‌ای ما نگاه کنند). سپس آن‌ها می‌توانند از این داده‌ها برای هدف قرار دادن پیام‌ها و تأثیرگذاری بر رای دادن و رفتار مصرف‌کننده استفاده‌کنند. این از چیزی که «دانیل کانمن» روانشناس آن را تفکر «سیستم 1» می‌نامد بهره می‌برد (تفکر خودکار، احساسی و ناخودآگاه ما که تفکر منطقی و عمدی «سیستم 2» را دور می‌زند). به این ترتیب، شرکت‌های بازاریابی مانند کمبریج آنالیتیکا می‌توانند اطمینان حاصل کنند که انتخاب‌های ما خودکار هستند، به راحتی توسط محتوای احساسی و اخلاقی دستکاری می‌شوند، و طوری طراحی شده‌اند که مطمئن شوند هیچ زمانی را صرف «فکر کردن بیش از حد» نمی‌کنیم!!!

سیاستمداران «محبوب»، فعالیت «غیر اصیل» و اثر متیو

معماری انتخابی و فناوری متقاعدکننده توسط سیاستمداران و کمپین‌های سیاسی در سراسر جهان، از جمله توسط چندین رژیم استبدادی، بسیار مورد استفاده قرار گرفته‌است. «خوان اورلاندو هرناندز، رئیس جمهور هندوراس»، با ایجاد صفحات و پروفایل‌های کاربری جعلی، صدها هزار دنبال‌کننده

و لایک در *Facebook* جمع‌آوری کرد. همه‌ی این صفحات توسط همان شخصی اداره می‌شد که حساب‌های شبکه‌های‌اجتماعی خود «هرناندز» را مدیریت می‌کرد. هرناندز یک حاکم ملی‌گرا و خودکامه است که از کودتای 2009 در هندوراس حمایت کرد. او متهم شده‌است که با استفاده از تاکتیک‌هایی مشابه آنچه روسیه در انتخابات ۲۰۱۶ آمریکا به کار گرفته است، پیروزی خود در انتخابات ۲۰۱۷ را دستکاری کرده است.

«سوفی ژانگ»، دانشمند داده در *Facebook* که به افشاگر تبدیل شد، یادداشتی 6600 کلمه ای برای افشای این کلاهبرداری نوشت. کار او مبارزه با مشارکت جعلی از این نوع در *Facebook* بود. او توضیح می‌دهد که "مدیر می‌تواند با نشستن پشت صفحه رایانه، پستی در مورد اینکه هرناندز چقدر خوب کارش را انجام می‌دهد در صفحه *Facebook* رئیس‌جمهور منتشر کند، سپس از صدها صفحه‌ی ساختگی خود برای محبوب نشان‌دادن پست استفاده کند" این (معادل یک اتوبوس ساختگی از افراد برای سخنرانی به صورت دیجیتالی) است. (نوعی فعالیت غیراصیل که به نام «آستروتورفینگ» نیز شناخته می شود).

این نوع تعامل جعلی که در استانداردهای اجتماعی *Facebook* به عنوان «رفتار غیراصیل هماهنگ» شناخته می‌شود، از «اثر متیو» استفاده می‌کند. این به تمایل شخصی که دارای مزیت اولیه در یک سیستم است برای انباشت بیشتر در طول زمان اشاره دارد. این نام از انجیل به روایت متی آمده است، که می‌گوید: "زیرا به کسی که دارد، بیشتر داده می‌شود و فراوانی خواهد داشت، اما از کسی که ندارد، حتی آنچه دارد گرفته می‌شود".

اثرات «مَتیو» گاهی اوقات سودمند است، اما اغلب باعث بی‌عدالتی می‌شود. مشخص شده‌است که آن‌ها نقش کلیدی در حفظ و گسترش اقشار اقتصادی و اجتماعی از همه نوع دارند. از آنجایی که همه سیستم‌های پیچیده پویا هستند، گاهی اوقات اثر معکوس رخ می‌دهد (هر چند وقت یک‌بار فقرا ثروتمندتر می‌شوند) اما این نادرتر است و اثرات ضعیف‌تر هستند. این حلقه‌های بازخورد بخشی از بسیاری از سیستم‌های طبیعی و اکولوژیکی و گونه‌های زنده هستند. در سراسر جهان طبیعی، به نظر می‌رسد که ثروتمندان تمایل (منظور میل و خواست نیست) آشکاری به ثروتمند شدن و فقرا

به فقیرتر شدن دارند. سیاستمدارانی که لایک‌ها و فالوورهای زیادی دارند، در شبکه‌های اجتماعی مشارکت بیشتری دارند و آن‌هایی که تعداد کمتری دارند، کمتر.

تیم ارزیابی تهدیدات *Facebook*، یافته‌های ژانگ مبنی بر اینکه رئیس‌جمهور هندوراس درگیر یک فعالیت غیراصیل هماهنگ شده بود را تأیید کرد. یک گزارش داخلی از *Facebook* بیان کرد که مبارزات انتخاباتی او "به طور مداوم یک رئیس جمهور غیرقانونی احتمالی را در یک *ARC* [کشور در معرض خطر] تقویت کرده است" و این احتمالاً "تأثیر *IRL* [در زندگی واقعی] داشته است". نزدیک به 1500 صفحه و صدها حساب تا جولای (ژانویه) 2019 حذف شدند!

حذف‌ها کمی تأثیر بلندمدت داشتند. وقتی حساب‌ها و صفحات جعلی حذف می‌شوند، حساب‌های جدید روز بعد دوباره بالا می‌روند. هرناندز تنها رهبر خودکامه‌ای نبود که از اثر متیو در کمپین‌های رسانه‌های اجتماعی خود استفاده می‌کرد. ژانگ بیان می‌کند که فعالیت‌های غیراصولی مشابه توسط شبکه‌هایی در کشورهای سراسر جهان از جمله افغانستان، آلبانی، آذربایجان، بولیوی، جمهوری دومینیکن، اکوادور، السالوادور، هند، اندونزی، عراق، ایتالیا، مکزیک، مغولستان، پاراگوئه، فیلیپین، لهستان، کره جنوبی، تایوان، تونس، ترکیه و اوکراین به کار گرفته‌شده‌است.

فعالیت غیرواقعی آذربایجان به‌ویژه به دلیل سابقه ضعیف حقوق بشر و تمایل دولت به استفاده از تحریم‌های اقتدارگرایانه و خشونت برای سرکوب روزنامه‌نگاران و منتقدان دولت و محدود کردن آزادی‌های اینترنتی و دسترسی به اطلاعات، نگران‌کننده است. «وانگ» گزارش می‌دهد که «الهام علی‌اُف رئیس جمهور و حزب آذربایجانِ نوینِ او»، از حساب‌های جعلی در *Facebook* به عنوان بخشی از کمپین هدف قرار دادن روزنامه نگاران و صداهای مخالف استفاده کردند. او می‌گوید که در یک دوره سه‌ماهه در سال 2019، «تقریباً 1.2 میلیون کامنت منفی و آزاردهنده در صفحات *Facebook* رهبران مخالف و رسانه‌های مستقل منتشر کرد و آن‌ها را به خائن بودن متهم کرد».

ژانگ بیان می‌کند که *Facebook* تهدیدهایی را که مستقیماً بر منافع ژئوپلیتیکی آمریکای شمالی و اروپای غربی تأثیر نمی‌گذارد، اولویت‌بندی می‌کند و جبران ناچیزی برای شهروندانی که تحت حاکمان خودکامه در کشورهایی مانند هندوراس و آذربایجان رنج می‌برند، باقی می‌گذارد.

مانند سخنان نفرت‌انگیزی که *Facebook* در برمه مجاز کرد، سوءاستفاده‌هایی که در کشورهای غیرغربی و فقیرتر انجام می‌شد به طور کلی نادیده‌گرفته‌شد. کسانی که آزادی‌های کمتری دارند کمتر می‌گیرند. *Facebook* با این ادعا که سیاست آن‌ها اولویت‌بندی فوری‌ترین تهدیدها است، به مقابله پرداخته است، اما ژانگ می‌گوید که مشکل جدی است و *Facebook* منابع کافی را برای مشکلی که آن‌ها در ایجاد آن نقش داشته‌اند، اختصاص نمی‌دهد.

او می‌گوید: «در سه سالی که در *Facebook* گذرانده‌ام، چندین تلاش آشکار از سوی دولت‌های خارجی برای سوءاستفاده از پلتفرم ما در مقیاس وسیع برای گمراه کردن شهروندان خود پیدا کردم و در موارد متعدد باعث ایجاد اخبار بین‌المللی شدم». او همچنین بیان می کند که *Facebook* در مورد حذف آن‌ها برای فعالیت‌های غیراصیل هماهنگ شده شفاف نیست.

شرکت *Facebook* به طور فزاینده‌ای نقش پیشرو در شکل‌دادن به سیاست، افکار عمومی، و بحث‌های مربوط به سیاست‌های عمومی (حتی تأثیرگذار بر نتایج انتخابات) در سراسر جهان، و نه فقط برای 8.2 میلیارد نفر از اعضای بشریت که مستقیماً از خدمات آن استفاده می‌کنند، ایفا می‌کند. این به رهبری شرکت‌های شبکه‌های مجازی مانند *Facebook* نقشی بزرگ و غیرقابل پاسخگویی در سیاست جهانی می‌دهد. وونگ بیان می‌کند که این به برخی از کارکنان *Facebook* اجازه می‌دهد تا «به‌عنوان نوعی شعبه قانون‌گذاری در تقریب *Facebook* با یک دولت جهانی عمل کنند»، در حالی که «بقیه بیشتر شبیه یک هیئت دیپلماتیک خصوصی‌شده هستند، دفاتر کارکنان در سراسر جهان برای ارتباط با مشاغل محلی، جامعه مدنی. گروه‌ها، تنظیم کننده‌های دولتی و سیاستمداران». بیشتر این قدرت از الگوریتم‌های غیرشفاف ناشی می‌شود که محتوایی را که در پلتفرم‌هایشان ظاهر می‌شود (و مجاز به نمایش آن نیستند) تبلیغ، توصیه، فیلتر و واسطه می‌کنند و می‌توانند توسط بازیگران بد، دولت‌ها و خود شرکت‌ها دستکاری شوند.

«اداره‌ی فضای‌سایبری چین» اخیراً اقداماتی را برای تنظیم الگوریتم‌ها برای بهبود شفافیت و جلوگیری از برخی از این مشکلات انجام داده است (و تلاش‌های آن‌ها توسط تنظیم‌کننده‌ها در سراسر جهان مشاهده می‌شود). این مقررات برای منع تبعیض توسط الگوریتم‌ها، جلوگیری از

اعتیاد و استفاده بیش از حد، محافظت از مصرف‌کنندگان در برابر افزایش قیمت، و بهبود شفافیت و کنترل کاربر بر روی الگوریتم‌های توصیه و فیلتر برای تقویت «عدالت و عدالت اجتماعی» طراحی شده‌اند. آن‌ها همچنین هدفشان ترویج یک "جهت گیری ارزشی اصلی"، "انتشار فعال انرژی مثبت" و ممنوعیت اطلاعات نادرست است که بر منافع ملی یا بازارهای اقتصادی چین تأثیر منفی می‌گذارد. این تهدید بسیار واقعی وجود دارد که دولت-ملت‌ها از این الگوریتم‌ها برای سانسور دیدگاه‌های مخالف استفاده کنند و در عین حال تبلیغات خود را تقویت کنند، همراه با این احتمال که این الگوریتم‌ها و کسانی که آنها را کنترل می‌کنند امور جهانی را در قرن بیست و یکم شکل دهند!

منشور اخلاق پزشکی نورنبرگ

در «دادگاه جنایات جنگی نورنبرگ» که پس از جنگ جهانی دوم انجام شد، تقریباً 22 پزشک و دانشمند به دلیل انجام آزمایشات علمی غیرقانونی بر روی زندانیان، که بسیاری از آن‌ها یهودی بودند، در اردوگاه‌های کار اجباری نازی‌ها محاکمه شدند. این آزمایش‌ها، از جمله آزمایش‌هایی که روی کودکان یهودی انجام می‌شد، شامل شکنجه، کشتار دسته جمعی و اتانازی بود. این دادگاه که به عنوان "محاکمه پزشکان" شناخته می شود، از 25 اکتبر 1946 تا 20 اوت 1947 برگزار شد.

در نتیجه آزمایش پزشکان، استاندارد جدیدی ایجاد شد که براساس آن آزمایش‌های انسانی بر اساس اصول رضایت آگاهانه داوطلبانه افرادی که روی آن‌ها آزمایش می‌شوند، کنترل می‌شد و آزمایش‌ها به نفع جامعه به عنوان یک کل هدایت می‌شد. این کتاب به یکی از مهمترین اسناد هدایت کننده اخلاق علمی در عصر مدرن تبدیل شده‌است. این به هدایت اخلاق تحقیق و آزمایش در مورد موضوعات انسانی در سراسر جهان ادامه می‌دهد. کدهای اخلاق تحقیق هنوز در مورد آزمایشات انجام شده توسط شرکت‌های خصوصی اعمال نمی‌شود.

تفسیر

اخلاق بودایی

نوشته پیتر هرشوک

سیستم‌های هوش‌مصنوعی که در پیش‌بینی علایق، دوست نداشتن‌ها، احساسات، انتخاب‌ها و اعمال انسان مهارت دارند نیز می‌توانند آن‌ها را تولید کنند. در مورد کمبریج آنالیتیکا، هدف شرکتی اعلام شده آن شکل دادن به افکار عمومی و انتخاب با بهره‌برداری از ظرفیت سیستم‌های یادگیری‌ماشین برای تبدیل منابع معرفتی تولید شده توسط رسانه‌های اجتماعی، تجارت الکترونیک و جستجوی دیجیتال به قدرت هستی‌شناختی است. به صورت دیجیتالی فردی برای دستکاری اطلاعاتی.

قدردانی از خشم اخلاقی در «طرح تجاری» کمبریج آنالیتیکا سخت نیست. همانطور که مطالعه موردی اشاره می‌کند، با این حال، هیچ زمینه‌ی انتخاب واقعاً خنثی وجود ندارد. کل محتوای اینترنت نمی‌تواند به طور همزمان برای هیچ کاربری ارائه شود، و بنابراین ارزش ها و مقاصد لزوماً در معماری انتخابی اتصال دیجیتال وارد می‌شوند. این نشان می‌دهد که سؤالات اخلاقی اساسی که باید پرسیده‌شود این است که کدام ارزش ها، چه کسانی و چرا انتخاب شده‌اند.

تابع هدف الگوریتم‌های کمبریج آنالیتیکا ساده و قابل‌فروش است: ایجاد رفتار رأی‌دهی مطابق با خواسته‌های مشتریانش. سیستم اعتبار اجتماعی چین و سیاست‌ها و شیوه‌های گسترده‌تر اداره فضای سایبری این کشور، عملکرد هدفی ظاهراً مطلوب‌تر و ارزش‌آفرین‌تری را ارائه می‌کند (ترویج انصاف و عدالت اجتماعی). ظاهراً، به نظر می‌رسد که ما با دو رویکرد کاملاً متفاوت برای شکل‌دهی الگوریتمی انتخاب‌ها و رفتار، و، اساساً، با دو سیستم به ظاهر متضاد حاکمیت اتصال روبرو هستیم.

یکی مبتنی بر منطق میانجی‌گری اجتماعی مبتنی بر انتخاب است که هم جلب توجه و هم استقلال تجربه‌شده را با تقویت محیطی الگوهای مبتنی بر فردیت اتصال دیجیتال به حداکثر می‌رساند

(سیستمی که با رویکرد «بازار» آمریکا در حاکمیت داده‌ها تجسم یافته‌است که پیگیری منصفانه و رقابتی را امکان‌پذیر می‌سازد). از یک جامعه ظاهراً خودسازمانده و پر جنب و جوش "چند صدایی". دیگری مبتنی بر منطق مهندسی اجتماعی مبتنی بر کنترل سوگیری است که توسط سیستم اعتبار اجتماعی "مدیریتی" چین تجسم یافته است که پتانسیل‌های تعاونی و یکپارچگی رابطه‌ای را به حداکثر می‌رساند (سیستمی که حول دستکاری متمرکز تمرکز جمعیت و پویایی عمدی در پیگیری ترکیبی یک پایدار طراحی شده است. جامعه "سمفونیک").

برخلاف ابزارگرایی بی‌سابقه‌ی کمبریج آنالیتیکا، تعهد اعلام‌شده چین به توسعه معماری انتخاب دیجیتال برای افزایش رفاه‌اجتماعی جذابیت قابل‌توجهی دارد. اگر افراد به طور تقلیل‌ناپذیری رابطه‌ای فرض می‌شوند، اگر رفاه‌شخصی تابعی از رفاه اجتماعی درنظر گرفته‌شود، و اگر این مسئولیت دولت است که شرایط رفاه اجتماعی را تضمین کند، در این صورت می‌توان استدلال کرد که استفاده از ابزارهای الگوریتمی برای شکل‌دهی رفتار شهروندان و روابط مدنی. مسئولیت اخلاقی دولت اگر بتوان معماری انتخابی را برای بهبود روابط انسان-انسان، انسان-جهان و انسان-تکنولوژی-جهان طراحی کرد، باید طراحی و اجرا شود. این، قطعاً موضع حزب کمونیست چین است.

البته، اگر واحد اساسی تحلیل اخلاقی، فرد و انسان ایده‌آل مستقل باشد، سیستم اعتبار اجتماعی چین و تلاش‌های اداره فضای‌سایبری آن برای القای رفتار شهروندان و مصرف‌کنندگانی که دولت چین آن را مطلوب می‌داند، مهندسی اجتماعی اجباری است (نقض آشکار حقوق اطلاعاتی و ارتباطی با توجه به اینکه همین امر در مورد دستکاری‌های انتخاباتی کمبریج آنالیتیکا نیز صادق است، ممکن است تصور شود که بهترین و بدیهی ترین جایگزین این است که اطمینان حاصل شود که ترجیحات کاربر (و فقط ترجیحات کاربر) معماری انتخابی را شکل می‌دهند که با ادغام ارزش‌ها و تصمیمات آن‌ها به صورت بازگشتی در محاسبات سفارشی که در تنظیم تجربیات پیوندی آن‌ها نقش دارند).

این طرح اولیه‌ی الگوریتم‌های جستجو و توصیه است که در حال حاضر برای مثال توسط گوگل و آمازون استفاده می شود. با توجه به جذب و ارضای خواسته‌ها و خواسته‌های فردی به عنوان ابعاد اولیه عملکرد هدف سیستم‌های یادگیری‌ماشین آن‌ها، یک معماری انتخابی در حال بهبود بازگشتی

پدیدار می‌شود که برای پیش‌بینی دقیق‌تر و ارائه‌ی آنچه که مردم می‌خواهند (از نظر اخلاقی قابل ستایش است). "سیستم‌هایی برای افزایش آزادی‌های شخصی در انتخاب".

آزادی انتخاب بدون شک بر نبود آن ارجح است. اما آزادی انتخاب به تنهایی عامل ناقصی برای آزادی در مفهوم بودایی تحقق پویایی‌های رابطه‌ای است. کارما شامل الگوهای چرخه‌ای درهم تنیدگی است که نتایج تجربی ارزش‌ها و مقاصد اعمال‌شده نیز به‌عنوان فرصت‌های ارادی عمل می‌کنند. خطر رابطه‌ای از الگوریتم‌های جستجو و توصیه‌های خودبهبود این است که کاربران گروگان رفتارهای گذشته‌شان خواهند بود، زیرا معماری انتخاب شخصی‌شده‌شان آنقدر مؤثر می‌شود که همیشه «دقیقاً» چیزی را که به دنبال آن هستند (اخبار، سرگرمی‌ها، محصولات، پیدا می‌کنند). خدمات و ارتباط اجتماعی طوری زندگی‌ها را خواهند ساخت که، در آن هرگز لازم نیست از اشتباهات درس بگیریم یا درگیر رفتار انطباقی باشیم. تصحیح دوره هرگز ضروری یا مطلوب به نظر نمی‌رسد.

این به اندازه‌ی‌کافی ناراحت‌کننده است. اما با توجه به ماهیت رابطه‌ای همه‌چیز، خطرات تکنولوژیکی بسیار عمیق‌تر می‌شوند. در میان علل کشمکش، مشکل و رنج، بودیسم این باور را محوری می‌داند که هر یک از ما به طور مستقل وجود داریم و دارای یک "خود" واحد و ماندگار هستیم که می‌تواند مستقل از بدن وجود داشته‌باشد. غیردوآلیسم رابطه‌ای (و نه تقلیلی) بودایی مستلزم این است که همه چیز را به‌طور قابل توجهی به هم وابسته بدانیم (ببینیم که هر چیز چگونه برای همه چیزها معنی دارد). ذهن و بدن، پدیده‌ای و جسمانی، پیامدهای سازنده یکدیگر هستند. آگاهی عبارت است از تمایز منسجم حضورهای محسوس و حسی یا تمایز منسجم ماده و آنچه مهم است. به این معنا که توضیح این که چگونه پدیده‌های پدیدار از جسم فیزیکی یا چگونه انگیزه‌های انسانی از حرکات مولکولی عصبی ناشی می‌شوند، «مشکل سخت» وجود ندارد.

مغز عامل آگاهی نیست. در عوض، روابط بین مغزها، بدن‌ها و محیط‌هایی که با آن‌ها و درون آن‌ها تکامل یافته‌اند، زیرساخت آگاهی را تشکیل می‌دهند. آن‌ها نتیجه کاری هستند که آگاهی انجام می‌دهد (به طور منسجمی تفاوت‌ها را توضیح می‌دهد) به همان شکلی که زیرساخت‌های حمل و نقل نتیجه شیوه‌های حمل و نقل گذشته است که سپس به صورت بازگشتی شیوه‌های حمل و نقل

بعدی را شکل می‌دهند. اهمیت اخلاقی این موضوع این است که زیرساخت آگاهی انسان هم درون جمجمه‌ای و هم برون جمجمه‌ای است، هم شخصی و هم بین فردی و فراتر از بدن ما به محیط‌های طبیعی ما گسترش می‌یابد، اما همچنین محیط‌های اجتماعی، فرهنگی و تکنولوژیکی ما (از جمله اتصال دیجیتالی).

آزمایش میدانی که از طریق زیرساخت محاسباتی اتصال دیجیتال برای تأثیرگذاری بر انتخاب‌های انسان، احساسات و رفتار انجام می‌شود، مشابه قرار دادن الکترودها در زیرساخت عصبی مغز به منظور تولید تجربیات خارق‌العاده یا اعمال بدنی است. چیزی که کمبریج آنالیتیکا، *Facebook* و اداره فضای‌سایبری چین در حال آزمایش آن هستند، قرار دادن "الکترودهای" الگوریتمی در بافت‌همبند زیرساخت اجتماعی تجسم یافته و اجرا شده آگاهی مشترک انسان است (فرآیندی که به همان اندازه تهاجمی و از نظر اخلاقی مملو از درج است). الکترودها به مغز آزمایش‌های انبوه در جذب دیجیتالی و بهره‌برداری از توجه انسان برای شکل‌دهی به احساسات، باورها، تصمیم‌گیری و اجتماعی‌سازی ارزشی خنثی نیستند و هرگز نمی‌توانند ارزشی داشته‌باشند.

اگر ارزش‌های انسانی رقیب به فناوری هوشمند تزریق شود، این تضادها را افزایش داده و عمیق‌تر می‌کند. بداهه اخلاقی و فضیلت رابطه‌ای که برای حل این تعارض‌ها مورد نیاز است، طبق آیین بودا، به مشارکت در اعمالی بستگی دارد که اساسی‌ترین حقوق انسانی ما ـحق آزادی توجه) را تضمین می‌کند. بدون آزادی توجه، آزادی قصد وجود نخواهد داشت، و بدون آزادی نیت، ما درگیر درهم تنیدگی‌های کارمایی گذشته خواهیم بود و بنابراین نمی‌توانیم با تغییر دادن آنچه که قصد داریم و برای آن چیزی که هستیم، تغییر دهیم برای یکی دیگر. به عبارت دیگر، ما نمی‌توانیم در مهم‌ترین هنرهای بشری (هنر اخلاقی تصحیح مسؤولانه‌ی درس) شرکت کنیم.

اخلاق فضیلت

نوشته جان هکر رایت

اخلاق فضیلت می‌تواند به ما کمک کند تا اهمیت شفافیت را در الگوریتم‌هایی که برای متمرکز نگه داشتن تمرکز بر محتوا و تغییر خواسته‌ها و باورهای ما طراحی شده‌اند، کمک کند. دیدگاه‌های ارسطو در مورد شخصیت، به‌ویژه، می‌تواند به ما کمک کند تا بفهمیم یک انتخاب معتبر چه می‌تواند باشد و از این طریق به ما کمک کند تا در مورد هنجارهایی که چنین انتخابی را ترویج می‌کنند، توافق کنیم، و همچنین به ما کمک کند که چه فضایل کاربران برای تعامل با فناوری‌های رسانه‌های اجتماعی در راهی که خود تضعیف‌کننده نیست.

در کتاب «هفتم اخلاق نیکوماخوس»، ارسطو یک گونه شناسی مفید از شخصیت را ترسیم می‌کند. بیشتر انسان‌ها در یکی از شرایط زیر قرار می گیرند: فضیلت، خودداری یا قدرت اراده، ضعف یا ضعف اراده و رذیلت. بدیهی است که فضیلت یک حالت خوب از شخصیت است که در آن اهداف درستی را هدف قرار می‌دهیم و ما فاقد امیال سرکشی هستیم که ما را به سمت چیزهایی که بد می‌دانیم می‌کشاند. این یک شرایط غبطه‌انگیز است، زیرا یک عامل کاملاً با فضیلت مجبور نیست برای انجام کاری که بهترین تصمیم را دارد با خواسته‌های خود مبارزه کند. این موضوع در مورد عامل قاره صادق نیست. او خواسته‌هایی خواهد داشت که بر خلاف آنچه او می‌داند بهتر است انجام دهد، اما ویژگی او این است که بتواند به طور قابل اعتمادی بر آن خواسته‌ها غلبه کند و به قضاوت خود عمل کند.

از نظر ارسطو، هم عامل با فضیلت و هم فاعل قاره، ظرفیت دست نخورده‌ای برای انتخاب دارند، زیرا امیال کلی آن‌ها در نهایت با استدلال آن‌ها در مورد اینکه چه باید بکنند، همخوانی دارد، حتی اگر با کشمکشی از عامل پاکدامنی در پرونده رخ دهد. از سوی دیگر، عامل بی‌اختیاری، در برابر خواسته‌هایی که بر خلاف آنچه که آن‌ها بهترین ارزیابی می‌کنند، شکست می‌خورد. عوامل بی اختیار می‌یابند که کارها را برخلاف قضاوت بهترشان انجام می‌دهند. آن‌ها احتمالاً در نتیجه احساس شرم می‌کنند. استدلال و قضاوتی که بر اساس تصور صحیح آن‌ها از آنچه باید انجام شود هدایت می‌شود،

بی اثر یا ناتوان است، زیرا در نهایت توسط تمایلاتی هدایت می‌شوند که آن‌ها را تأیید نمی‌کنند.

سرانجام، عامل شرور وجود دارد. ارسطو دو تصویر متمایز از عامل شرور ارائه می‌دهد. از یک دیدگاه، خواسته‌های عامل شرور در پشت ایده نادرست آن‌ها از آنچه باید انجام شود ردیف می‌شود، بنابراین آنچه که عمدتاً با عامل شرور اشتباه می‌کند، تصور نادرست آن‌ها در مورد خوب است. در دیدگاه دیگری که او ارائه می‌کند، تمایلات عامل شرور چنان بی‌نظم است که عوامل شرور، از دیدگاه روان‌شناختی، یک ویرانگر هستند. دیدگاه ارسطو ممکن است این باشد که در حالی که اصولاً هماهنگی خاصی در روح شرور وجود دارد، آن‌ها به دلیل ماهیت چیزی که می‌خواهند به نابودی ختم می‌شوند، که عامل شرور را تباه می‌کند.

اگر عامل شرور لزوماً یک خرابی روانی است، پس ما مبنای بسیار روشنی برای توصیه به فضیلت داریم. باید پیگیری شود، زیرا هرکسی که این کار را نکند، در نهایت در شرایط روحی ناخوشایند و شکنجه‌ای قرار می‌گیرد. اما ممکن است افراد بتوانند کاملاً به هدفی متعهد باشند که به نظر می‌رسد کاملاً بدون ارزش است. بنابراین، ممکن است تصور کنیم ممکن است کسی وجود داشته‌باشد که هر ساعت بیداری هر روز را در رسانه‌های اجتماعی یا بازی‌های ویدیویی سپری کند، و هیچ دلسردی یا تمایلی به انجام هیچ کار دیگری احساس نمی‌کند. در این صورت، ممکن است بخواهیم بگوییم که یک رذیله وجود دارد، زیرا فرد متصور، احمقانه عمر خود را صرف چیزی می‌کند که پوچ است.

این بدیهی است که یک قضاوت اخلاقی اساسی است که عامل ظاهراً شرور آن را رد می‌کند. ممکن است تصور کنیم که فردی در این شرایط با تلاش‌های غیرشفاف برای دستکاری او در مجموعه‌ای متفاوت از باورها و خواسته‌ها، آشفته می‌شود و احساس می‌کند که مورد ظلم قرار می‌گیرد. اگر تصدیق کنیم و عمداً تصمیم بگیریم از کمک برای تغییر رفتار استفاده کنیم، مثلاً با استفاده از برنامه‌های ردیابی زمان یا برنامه‌هایی که دسترسی به رسانه‌های اجتماعی را مسدود می‌کنند، یک چیز است، اما اگر در معرض تلاش‌هایی برای تغییر رفتار خود باشیم، این یک چیز دیگر است. دانش نتیجه‌ی این است که هم عاملان با فضیلت و هم عوامل شرور ممکن است در رد تلاش‌ها برای دستکاری باورها و خواسته‌هایشان بدون اطلاع آن‌ها توافق کنند.

با این حال، این همان چیزی است که بسیاری از رسانه‌های اجتماعی و فناوری‌های بازی عمداً به دنبال آن هستند، در نتیجه بسیاری از افرادی که قدرت اراده را در حوزه‌های دیگر نشان می‌دهند، در مواجهه با این طرح‌ها ضعف اراده نشان می‌دهند. بسیاری از ما، حداقل در ابتدا، از این تلاش عمدی برای تضعیف کنترل خود از جانب طراحان بی اطلاع بودیم. احتمالاً هیچ پیشرفتی در فضیلت انسانی ما را قادر نخواهد ساخت که در مواجهه با چنین تلاش‌های عمدی و پیچیده، خویشتن‌داری را حفظ کنیم. از این رو، والور معتقد است که بخش اساسی راه حل، درخواست ابزارهایی است که ما را ضعیف نمی‌کنند. به نظر من این یک تقاضای حداقلی است. در حالت ایده‌آل، ما ابزارهایی می‌خواهیم که به جای محو کردن عاملیت ما را تقویت کنند.

رسیدن به نقطه‌ی دستیابی به چنین ابزارهایی مطمئناً همه ما را ملزم می‌کند که درک خود از روانشناسی و فناوری را تا آنجا که لازم است برای درک اینکه چگونه توسط فناوری دستکاری می‌شویم عمیق‌تر کنیم. به نظر من این دانش بخشی از آنچه برای ما درک نظری کلی از خیر انسانی را تشکیل می‌دهد و بخشی از حکمت عملی است. ارسطو مدعی است که برای اینکه بتوانیم خوب عمل کنیم، باید درک کلی از خیر انسان داشته‌باشیم. شرط این نیست که هر یک از ما نیاز به داشتن سطح تخصص توسط پزشک، روانشناس، متخصص تغذیه و مربی بدن داشته‌باشیم. این دانش تخصصی از خیر انسانی است که در شرایط خاص مورد نیاز است. نوع دانشی که هر انسانی به آن نیاز دارد عمومی‌تر است: دانستن وضعیت خوب یک انسان تا به عنوان مثال اگر تشخیص دادم از آن شرایط کوتاهی می‌کنم، می‌توانم با پزشک یا روانشناس مشورت کنم.

من در اینجا، فراتر از ارسطو، ادعا می‌کنم که محتوای درک کلی از خیر انسانی که برای ما ضروری است، ممکن است در طول زمان تغییر کند، در پاسخ به شرایط جدید و به طور گسترده‌ای مشترک که با آن مواجه هستیم. قابل‌قبول است که فکر کنیم برای عملکرد خوب در جامعه معاصر به آگاهی اولیه از فناوری نیاز داریم، از جمله آگاهی از اینکه چگونه فناوری از آسیب پذیری‌های روانی ما سوءاستفاده می‌کند. این دانش به ما دفاعی در برابر تکنیک‌های بدافزار ذهنی می‌دهد که بدون شک حتی اگر در قانون‌گذاری مقررات بهتری که نیازمند شفافیت در طراحی الگوریتم‌ها هستند، موفق

شویم، همچنان ادامه خواهند داشت. در واقع، به اشتراک گذاشتن یک درک کلی از نیاز به چنین قانونی مستلزم نوعی درک از خیر انسانی است که من در اینجا به آن‌ها اشاره می‌کنم. ما باید درکی اساسی از اینکه یک تصمیم انسانی اصیل و دستکاری نشده چگونه به نظر می‌رسد داشته‌باشیم تا بخواهیم قانونی که از امکان چنین انتخابی در برابر فناوری‌هایی که برای تضعیف آن طراحی شده‌اند دفاع کند.

اخلاق دئونتولوژیک

نوشته کالین مارشال

همانطور که مورد بدافزار ذهنی توضیح می‌دهد، *Facebook* (و دیگر پلت‌فرم‌های شبکه‌های اجتماعی) هم بر دیگران تأثیر می‌گذارند و هم به آن‌ها کمک می‌کند تا بر تعداد زیادی از مردم تأثیر بگذارند. کدام راه‌های تأثیرگذاری بر دیگران از نظر اخلاقی مجاز است؟ یک رویکرد دئونتولوژیک دو شرط لازم برای مجاز بودن اخلاقی را پیشنهاد می‌کند. اولاً، هرگونه تلاش برای تأثیرگذاری بر دیگران باید با احترام محدود شود، یعنی با توجه جدی به اهداف و پروژه های شخصی آن‌ها. دوم، اگرچه همه‌ی ارتباطات باید شامل برخی از عناصر غیرعقلانی باشد، اما دین شناسی مستلزم آن است که تأثیرگذاری بر ارتباطات صادقانه‌ای که از ظرفیت‌های عقلانی ما می‌گذرد، اولویت بندی کند. به نظر می‌رسد هیچ یک از این شرایط توسط بازیگران کلیدی پرونده بدافزار ذهنی رعایت نشده باشد. ما می‌توانیم هر شرط را به نوبه‌ی خود در نظر بگیریم.

اولین شرط دئونتولوژیک در مورد تأثیر مجاز را نیز می‌توان به عنوان الزام به عدم برخورد با دیگران صرفاً به عنوان وسیله بیان کرد. به عبارت دیگر، وقتی هدفی را دنبال می‌کنیم، باید از راه‌هایی برای دستیابی به آن که نیازها و پروژه‌های دیگران را زیر پا می‌گذارد، اجتناب کنیم و با آن‌ها به

عنوان ابزاری صرف برای رسیدن به آنچه می‌خواهیم رفتار کنیم. این شرایط می‌تواند توضیح دهد که چرا از نظر اخلاقی کسب درآمد از محصولات اعتیادآوری که زندگی افراد را از مسیر خارج می‌کند، غیرمجاز است. برای مثال، رهبران کارتل‌های مواد مخدر، با محصولاتی که می‌دانند می‌توانند توانایی مردم را برای زندگی به گونه‌ای که می‌خواهند داشته‌باشند، به اهداف مالی خود می‌رسند.

در مورد بدافزار ذهنی، هم *Facebook* و هم سایر نهادهایی که از این پلتفرم استفاده می‌کردند (مانند کمبریج آنالیتیکا و محققان کورنل) به نظر می‌رسید که کاربران *Facebook* را صرفاً به عنوان وسیله‌ای در نظر می‌گرفتند. هیچ مشکلی ذاتی در مورد ارائه *Facebook* به کاربران برای حفظ و ایجاد دوستی، ارائه‌ی فرصت‌هایی برای سرگرمی (مانند آزمون‌های شخصیتی) یا جمع آوری‌داده‌ها در مورد رفتار کاربران وجود ندارد. اما همه چیز از نظر اخلاقی مشکل‌ساز می‌شود، طبق اولین شرط ریشه‌شناختی، زمانی که پلتفرم تلاش می‌کند کاربران را به روش‌هایی جذب کند که می‌تواند بر زندگی آن‌ها تأثیر منفی بگذارد، و زمانی که پلت فرم جمع‌آوری و استفاده از داده‌ها را صرفاً به منظور ایجاد سود و تولید سیاسی تسهیل می‌کند. نتایجی که (حداقل در برخی موارد) به وضوح به نفع کاربران نبود.

دومین شرط ریشه‌شناختی تأثیر مجاز این است که صداقت را در اولویت قرار دهد و به ظرفیت‌های عقلانی دیگران متوسل شود. نقض آشکار این شرایط شامل دروغ، شستشوی مغزی و استفاده از الکل برای "نرم کردن افراد" است. با این حال، دئونتولوژیست‌ها همیشه تشخیص‌داده‌اند که تمام تعاملات بین افراد شامل چیزی است که «تالر و سانستاین» آن را «معماری انتخاب» می‌نامند، که به چیزی که کانمن «سیستم 1» می‌نامد جذاب است. واضح است که هنگام ارائه‌ی گزینه‌ها، ابتدا باید یک گزینه داده‌شود و این سفارش می‌تواند بر تصمیمات افراد تأثیر بگذارد. با این حال، حتی اگر نفوذ باید با توسل به سیستم 1 شروع شود، شرایط دئونتولوژیک ایجاب می‌کند که همچنان از سیستم 2 (استدلال آگاهانه و آگاهانه) عبور کند، و به افراد حداقل گزینه ارزیابی منطقی و به طور بالقوه رد تأثیر تلاش‌شده را می‌دهد.

به نظر می‌رسد نقض‌های مختلفی از این شرط در پرونده بدافزار ذهنی صورت گرفته‌است. رای

دهندگانی که با تبلیغات هدفمند ارائه شده بودند معمولاً نمی‌دانستند که تبلیغات بر اساس داده‌های شخصی آن‌ها است. رای دهندگان در هندوراس و سایر کشورها نمی دانستند که تعداد قابل توجهی از "لایک ها" در مورد پست‌های خاص غیرواقعی است یا اینکه "لایک"های دیگر از طریق اثر متیو به دست آمده است. فعالیت دستکاری و غیراصولی که به رازداری و فریب بستگی دارد و به کاربران امکان ارزیابی منطقی تأثیر تلاش شده را نمی‌دهد، شرط دوم دئونتولوژیک مجاز بودن را برآورده نمی‌کند.

برای بهبود اوضاع در آینده چه کاری می‌توان انجام داد؟ افزایش شفافیت، حداقل در اصل، می‌تواند *Facebook* و نهادهایی را که از پلتفرم آن استفاده می‌کنند به انطباق با دو شرط ریشه‌شناسی نزدیک‌تر کند. به هر حال، شفافیت نوعی صداقت است و می‌تواند به افراد اجازه دهد تا به طور منطقی ارزیابی کنند که آیا مایل به پذیرش تأثیر برخی فعالیت‌ها بر نیازها و پروژه های شخصی خود هستند یا خیر. برای مثال، کاربران *Facebook* که می‌دانند *Facebook* فعالیت‌های آن‌ها را ردیابی می‌کند و تبلیغاتی را که می‌بینند شخصی‌سازی می‌کند، می‌توانند ارزیابی منطقی از ادامه استفاده از *Facebook* (در پرتو نیازها و پروژه‌های شخصی خود) انجام دهند و تصمیم بگیرند که آیا روی آن تبلیغات کلیک کنند یا خیر.

حداقل دو محدودیت قابل توجه برای میزان پیشرفت در شفافیت در این زمینه وجوددارد. اولاً، برای اینکه شفافیت از نظر اخلاقی مهم باشد، اطلاعات ارائه‌شده باید قابل‌فهم باشد (ارائه قراردادهای کاربری طولانی و پر از اصطلاحات تخصصی حقوقی به افراد کافی نیست، و همچنین برای ایجاد الگوریتم‌های کلید عمومی که اکثر کاربران فاقد ویژگی‌های فنی هستند کافی نیست). توانایی‌های پردازش دوم، برخی از تکنیک‌های دستکاری می‌توانند به تضعیف فرآیندهای عقلانی افراد ادامه دهند، حتی زمانی که ما از آن‌ها آگاه شویم. با توجه به ساختار اجتماعی عمیق روانشناسی انسان، فریب پاداش‌های اجتماعی خوش ساخت (مانند لایک‌ها و فالوورها) بسیار قوی است و می‌تواند ما را به سمت فعالیت‌هایی بکشاند که برخلاف نیازها و پروژه‌های ما هستند، حتی زمانی که از آنچه آگاه هستیم. اتفاق می‌افتد. بنابراین، برای اینکه شفافیت از نظر اخلاقی بر اساس استانداردهای

دئونتولوژیکی مهم باشد، باید به گونه‌ای برای کاربران تنظیم شود که انتخاب‌های آگاهانه و معنادار را بر اساس آنچه برای آن‌ها مهم است تسهیل کند.