رفتار واقعی هوش‌مصنوعی برای دانشمندان داده

انتشارات محمد رحیمی

۸ 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 واشنگتن

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 هوش مصنوعی 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 هوش مصنوعی است که از رویکردهای آماری برای یافتن الگوها در دنیا (که بهم ریخته است) استفاده می‌کند. در خیلی جهات، *(ML)* پاسخی برای شکست‌های زودهنگام هوش مصنوعی 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 داده ها برای به دست آوردن 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 گیری ها 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 دیپ‌فیک *(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 هوش مصنوعی تولید می‌شود! پدیده‌ای که اخیراً حتی در مجلات علمی با داوری مشابه نیز دیده می‌شود. یعنی هوش مصنوعی 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 هوش مصنوعی 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. هوش مصنوعی 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 هوش مصنوعی *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 هوش مصنوعی 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 هوش مصنوعی 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 – هوش مصنوعی 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 هوش مصنوعی 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 آسیب ها 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 هوش مصنوعی در مزارع و کارخانه‌ها بحث می کنند. مدل‌های ML در صنعت مزرعه‌ی کارخانه‌ای، برای جمع‌آوری اطلاعات در مورد حیوانات پرورشی، به منظور به دست آوردن 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 هوش مصنوعی 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 می تواند 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 هوش مصنوعی بازی می‌کند، بررسی می‌کند و اینکه آیا می‌توانیم

اخلاق هوش مصنوعی 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 هوش مصنوعی یا مادران و پدران هوش مصنوعی 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 هوش مصنوعی رهایی‌بخش به جای ظالمانه باز می‌کند (هوش مصنوعی‌ای 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 هوش مصنوعی 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 هوش مصنوعی می‌کند. اهمیت داده ها 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 هوش مصنوعی 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 آن ها 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 هوش مصنوعی 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 هوش مصنوعی» 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 انسان ها 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 الموِ، باعث می‌شود که بچه‌ها با ناهار یکی از آن‌ها را به جای شیرینی انتخاب کنند؟ یا اینکه مردها وقتی خانم ها آن‌ها را تماشا می‌کنند بیشتر غذا می خورند (اما وقتی مردها آن‌ها را تماشا می‌کنند خانم ها 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 انتخاب های 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 انتخاب هایشان 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، داده ها 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 تداعی های 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 داده ها اشاره می‌کند (به عنوان مثال با بازی با اندازه های نمونه) برای ایجاد یک نتیجه به ظاهر آماری مهم در جایی که هیچ کدام واقعا وجود ندارد. *p-hacking* نیز اعتبار مدل ها 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* می تواند 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 شیوه هایی می پیوندد که مدل های 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 پاسخ های 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 گیری های 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 نظریه ها، مدل ها 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 مدل ها و الگوریتم های 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 داده های 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 وابستگی ها.
* 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 پیکربندی هایپرپارامتر، و مشخصات تمام پارامترهای هایپر مورد استفاده برای تولید نتایج.
* تعداد دقیق دوره های 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 توصیه های ناآگاهانه در مورد عادات غذایی را داشت و در نتیجه به کسانی که از توصیه

های 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 می تواند 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.