



مرکز ملی فضای مجازی
پژوهشگاه فضای مجازی

عصر
فضای
مجازی
نوزدهم



Google AI

اخلاق هوش مصنوعی
در گوگل

Ethics of artificial intelligence
in Google



مرکز ملی فضای مجازی
پژوهشگاه فضای مجازی

اخلاق هوش مصنوعی در گوگل

گزارش شماره ۱۹

آذر ماه ۱۳۹۸

تهیه شده در: پژوهشگاه فضای مجازی - گروه مطالعات اخلاقی فضای مجازی

تهیه کننده: مهدی خلیلی

دانشجوی دکترای فلسفه علم و فناوری دانشگاه صنعتی شریف

پژوهشگر وابسته مرکز CLUE+ دانشگاه VU Amsterdam

ناظر علمی: محمدرضا زارع میرک آباد، محمدمهدی نصرهرندی

نشانی: تهران، میدان آرژانتین، خیابان بیهقی، نبش خیابان ۱۶ غربی، پلاک ۲۰، کدپستی ۱۵۱۵۶۷۴۳۱۱

<http://www.majazi.ir>

شماره تماس: ۸۶۱۲۱۰۶۱

حقوق مادی و معنوی این اثر متعلق به مرکز ملی فضای مجازی است و استفاده از مطالب آن صرفاً با ذکر مأخذ بلامانع است.

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

سخن تحت

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

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

سید ابوالحسن فیروزآبادی

دبیر شورای عالی و رئیس مرکز ملی فضای مجازی

چکیده

در نوشتار حاضر، گزارشی از دیدگاه‌ها و فعالیت‌های گوگل در زمینه‌ی اخلاق هوش مصنوعی ارائه می‌شود. بایدها، نبایدها و ارزش‌های اخلاقی که واحد هوش مصنوعی گوگل برای توسعه فناوری‌هایش در نظر گرفته و اقداماتی که برای تحقق اهداف اخلاقی‌اش صورت داده است بیان می‌گردد. همچنین پیشنهادهای سیاست‌گذارانه‌ی گوگل در جهت توسعه‌ی هوش مصنوعی تشریح می‌شود؛ بخشی از این پیشنهادها ناظر به سیاست‌گذاری عمومی و برخی از آنها توصیه‌هایی به حکومت‌ها است. در پایان، ضمن ارائه‌ی خلاصه‌ای تحلیلی، ابعاد مختلف گزارش‌های گوگل در زمینه‌ی اخلاق هوش مصنوعی و جهت‌گیری آن مورد ارزیابی قرار می‌گیرد و برخی ملاحظات که لازم است حکومت‌ها در نظر داشته باشند، ارائه می‌شود. شایان ذکر است که این گزارش از روی منابع خود این شرکت تهیه شده و ارائه آن به معنای تأیید محتوای آن نیست.

واژگان کلیدی

اخلاق هوش مصنوعی (AI)، یادگیری ماشین (ML)، داده، گوگل، مسئولیت‌پذیری، انصاف، تفسیرپذیری، حریم خصوصی، امنیت

فهرست مطالب

مقدمه	۱
فصل ۱- اصول اخلاقی هوش مصنوعی گوگل	۲
۱-۱- مقدمه	۲
۱-۲- طراحی هوش مصنوعی مسئولیت پذیر برای همه	۲
۱-۳- کاربردهایی از هوش مصنوعی که دنبال نخواهند شد	۵
۱-۴- نتیجه	۵
فصل ۲- اجرای اصول هوش مصنوعی	۶
۱-۲- مقدمه: کنش‌های هوش مصنوعی مسئولیت پذیر	۶
۲-۲- انصاف	۷
۳-۲- تفسیرپذیری	۹
۴-۲- حریم خصوصی	۱۱
۵-۲- امنیت	۱۲
فصل ۳- به‌روزرسانی اصول هوش مصنوعی گوگل	۱۳
۱-۳- به‌روزرسانی اصول هوش مصنوعی گوگل - شش ماه بعد	۱۳
۲-۳- هوش مصنوعی مسئولیت پذیر: عملیاتی کردن اصول اخلاقی - یک سال بعد	۱۵
۳-۲-۱- آموزش داخلی	۱۵
۳-۲-۲- ابزارها و پژوهش	۱۵
۳-۲-۳- داوری فرآیند	۱۶
۳-۲-۴- تعامل با ذی‌نفعان خارجی	۱۷

فصل ۴- دیدگاه‌های سیاست‌گذارانه	۱۸
۱-۴- مقدمه	۱۸
۲-۴- توسعه‌ی مسئولانه‌ی هوش مصنوعی	۱۸
۱-۲-۴- سیاست‌گذاران چگونه می‌توانند کمک کنند؟	۱۸
۳-۴- دیدگاه‌هایی درباره‌ی مسائل حکمرانی هوش مصنوعی	۲۵
۱-۳-۴- حوزه‌های کلیدی برای وضوح‌بخشی و اقدامات پیشنهادی	۲۶
۴-۴- نقش آفرینی گوگل در به‌اشتراک‌گذاری مسئولانه‌ی داده‌های باز	۲۷
۱-۴-۴- بازکردن مسئولانه‌ی پایگاه‌های داده	۲۷
۲-۴-۴- یافتنی و سودمندکردن داده‌ها	۲۸
۳-۴-۴- ایجاد تعادل در سبک‌سنگین کردن‌ها	۲۹
فصل ۵- تحلیل و ارزیابی	۳۰
۱-۵- مقدمه	۳۰
۲-۵- خلاصه گزارش؛ الگوبرداری	۳۱
۳-۵- اراده‌گرایی یا جبرگرایی	۳۲
۴-۵- ذات‌گرایی یا ابزارگرایی	۳۳
۵-۵- نگاه آرمان‌شهرانه یا ویران‌شهرانه	۳۳
۶-۵- اخلاق هنجاری: نتیجه‌گرا	۳۴
۷-۵- مسئله‌های باز در اخلاق AI	۳۵
منابع	۳۷

مقدمه

با توسعه فناوری‌های هوش مصنوعی (AI) و ورود هر چه بیشتر آن‌ها به زندگی روزمره مردم، توجه به ابعاد اخلاقی AI اهمیت بیشتری یافته است. در این راستا، شرکت‌های بزرگی مانند گوگل، مایکروسافت و آمازون نسبت به دلالت‌های اخلاقی AI حساس شده و در این راستا دست به انجام پژوهش، برگزاری همایش و تدوین اسناد راهبردی زده‌اند. از آنجایی که این شرکت‌ها در بطن توسعه فناوری‌های AI هستند، توجه آن‌ها به ابعاد اخلاقی این فناوری‌ها مهم است. از سویی، این توجه به‌نحو انضمامی و با مشارکت مهندسان و محققان AI صورت گرفته است؛ یعنی در فضایی انتزاعی و به‌دوراز کاربرد عملیاتی نیست. از سوی دیگر، توجه به امور اخلاقی، به توسعه‌ی فناوری‌های آنان جهت می‌دهد، توسعه‌ی یک فناوری خاص را متوقف می‌کند یا سرمایه‌گذاری در حوزه‌ای را توصیه می‌کند. در نتیجه می‌توان تعاملی دوسویه را بین ابعاد فنی و اخلاقی توسعه فناوری مشاهده کرد. این تعامل می‌تواند هم به تعمق مسائل اخلاقی بینجامد هم توسعه‌ی فناوری را به‌نحو مطلوب هدایت کند.

در گزارش حاضر مباحث مربوط به اخلاق AI در یکی از این شرکت‌های بزرگ فناوری، یعنی گوگل (Google)، بررسی می‌شود. هدف این است که دیدی کلی از اقداماتی که گوگل در این جهت انجام داده است به دست آید. روش کار به این نحو است که در فصول مختلف، مطالب اخلاقی ارائه‌شده از سوی گروه AI گوگل گزارش می‌شود. این گزارش در اغلب موارد به‌طور خلاصه بیان می‌شود و در برخی به‌طور دقیق بازتاب‌دهنده‌ی متن منتشرشده‌ی آن‌ها است. در **فصل اول**، بخش مسئولیت‌پذیری^۱ AI گوگل معرفی و بایدها و نبایدهای اخلاقی آن بیان می‌شود. در **فصل دوم**، توصیه‌های گوگل برای حمایت از چهار ارزش «انصاف»^۲، «تفسیرپذیری»^۳، «حریم خصوصی»^۴ و «امنیت»^۵ بیان می‌شود. **فصل سوم** ناظر به دو گزارش است که اولی را گوگل در تاریخ ۱۸ دسامبر ۲۰۱۸، یعنی شش ماه بعد از تدوین اصول و توصیه‌های اخلاقی خود، منتشر کرده است که در آن روند پیشروی کار را در حوزه AI توصیف می‌کند. دومین گزارش نیز گوگل در تاریخ ۲۸ جولای، یعنی حدود یک سال پس از انتشار اصول و توصیه‌های اخلاقی‌اش، ارائه کرده است. **فصل چهارم** به دیدگاه‌های سیاست‌گذارانه‌ی گوگل در حوزه‌ی AI مربوط می‌شود. این فصل، هم دربردارنده‌ی سیاست‌گذاری‌های گوگل برای خود است و هم توصیه‌های گوگل را به حکومت‌ها بیان می‌کند. در **فصل پنجم**

^۱ Responsibility

^۲ Fairness

^۳ Interpretability

^۴ Privacy

^۵ Security

که فصل انتهایی است، تحلیلی درباره‌ی مطالب فصول پیشین ارائه می‌شود و جهت‌گیری سیاست‌ها و توصیه‌های گوگل مورد ارزیابی قرار می‌گیرد. مجدداً تأکید می‌کنیم که آن چه در این گزارش می‌آید، ادعاهای خود شرکت گوگل است و ما سخنی در صدق یا کذب این ادعاها بیان نمی‌کنیم.

فصل ۱- اصول اخلاقی هوش مصنوعی گوگل

۱-۱- مقدمه

در بخش «AI گوگل»^۱، پژوهشگران تحقیقاتشان را در سه حوزه‌ی توسعه AI، تسری آن به عرصه‌ها و محصولات جدید، و تولید ابزارهایی برای دسترسی همگانی به آن، دنبال می‌کنند. این بخش در زمینه‌های مختلفی چون پژوهش، آموزش، ابزار و مسئولیت‌پذیری فعال است. در ادامه، بخش مسئولیت‌پذیری که مرتبط با حوزه اخلاق است، تشریح می‌شود.

۱-۲- طراحی هوش مصنوعی مسئولیت‌پذیر برای همه

در بخش مسئولیت‌پذیری، اعضای AI گوگل در پی توسعه‌ی آن به نحو مسئولیت‌پذیرانه هستند تا برای تمام مردم و جامعه منفعت داشته باشد. AI مسئولیت‌پذیرانه می‌تواند در جهت پیشرفت پزشکی، فهم زبان و تسریع اکتشافات علمی به کار آید. در راستای طراحی مسئولیت‌پذیرانه، اصولی مبنای قرار داده شده‌اند که راهنمای برنامه‌های گوگل و درک مشترک با دیگر شرکت‌ها باشد. این اصول که جنبه‌ی اخلاقی دارند، با هدف توسعه و استفاده از AI در جهت تأثیرگذاری مثبت بر جامعه و زندگی مردم تدوین شده‌اند. این اصول، AI را به عنوان عرصه‌ای که دارای منفعت عمومی است معرفی می‌کند و بیانگر تعهدات گوگل در توسعه مسئولانه فناوری‌های مفید و عدم توسعه برخی حوزه‌ها است. طبق این اصول، یک سامانه AI مطلوب، باید ویژگی‌های زیر را دارا باشد:

۱- از نظر اجتماعی مفید باشد.

AI بر سلامت، امنیت، انرژی، حمل‌ونقل، ساخت‌وساز و تفریحات اثرگذار است. لذا باید در توسعه‌ی بالقوه و استفاده از فناوری‌های آن، عامل‌های اجتماعی و اقتصادی را لحاظ کرد و بالتبع حوزه‌هایی را پی گرفت که منافع آن به طور ملموسی از مخاطرات آن بیشتر باشد. به علاوه، ضمن تحلیل اطلاعاتی که با رضایت و بدون مشکل فراهم آمده، باید به تمایزهای فرهنگی، اجتماعی و هنجارهای حقوقی در کشورهای مختلف احترام

¹ Google AI

گذاشت. گوگل همچنین در نظر دارد چه زمانی چه فناوری‌هایی را بدون نگاه اقتصادی در دسترس همگان قرار دهد.

۲- از ایجاد یا تقویت سوگیری^۱ غیرمنصفانه پرهیز کند.

الگوریتم‌ها و پایگاه‌های داده^۲ می‌توانند انواع سوگیری‌های غیرمنصفانه را بازنمایی یا تقویت کنند. اما تمایز سوگیری منصفانه از غیرمنصفانه کار راحتی نیست و در فرهنگ‌ها و جوامع مختلف متفاوت است. باید تلاش کرد از تأثیرات ناعادلانه روی مردم اجتناب شود، خصوصاً وقتی نسبت به ویژگی‌هایی چون نژاد، قومیت، جنسیت، ملیت، درآمد، تمایل جنسی، توانایی و عقیده‌ی مذهبی یا سیاسی حساسیت وجود دارد.

۳- در ساخت و آزمون^۳ سامانه، ایمنی لحاظ شود.

باید از نتایج ناخواسته از نظر ایمنی و امنیت که ممکن است مخاطره‌ای ایجاد کنند، جلوگیری کرد. از این جهت باید محتاطانه دست به طراحی سامانه‌های AI زد و آن‌ها را طوری توسعه داد که با ایمنی پژوهش در تناسب باشد. مثلاً در بحث آزمون فناوری‌های AI باید از محیط‌های محصور و محافظت‌شده استفاده کرد و بر عملکرد آن نظارت داشت.

۴- دارای عملکرد توضیح‌پذیر و قابل تجدیدنظر باشد.

سامانه‌های هوشمند باید طوری طراحی شوند که بازخورد، توضیح‌های مرتبط و تجدیدنظر^۴ را ممکن سازند. فناوری‌های AI باید تحت نظارت و راهبری انسانی باشند.

۵- اصول حریم خصوصی را رعایت کند.

در توسعه و استفاده از فناوری‌های AI اصول حریم خصوصی رعایت شوند. برای جلب توجه و رضایت کاربر، فرصت کافی داده شود، معماری‌هایی که الزامات امنیتی دارند حمایت شوند، شفافیت مناسبی فراهم شود و در استفاده از داده، نظارت صورت گیرد تا حریم خصوصی نقض نشود.

۶- استانداردهای سطح بالا در مورد تعالی علمی^۵ رعایت شود.

نوآوری در تولید و توسعه‌ی فناوری، ریشه در استفاده از علم و روش علمی دارد. علم در محیطی تولید می‌شود که پژوهشگران بتوانند ایده‌های خود را آزادانه بیان و از آن دفاع کنند. محیط علمی باید فراهم‌کننده‌ی همبستگی و همکاری بین دانشمندان باشد. بنابراین ایجاد یک محیط علمی برای نوآوری و توسعه فناوری، مهم است. همچنین ابزارهای AI این ظرفیت را دارند که عرصه‌های جدیدی از تحقیق و دانش علمی را در

¹ Bias

² Datasets

³ Test

⁴ Appeal

⁵ Scientific Excellence

حوزه‌هایی حساس چون زیست‌شناسی، شیمی، پزشکی و علوم بوم‌شناختی ایجاد کنند، و بدین ترتیب به رشد علم بینجامند تا پیشرفت علم نیز در گام‌های بعد به توسعه‌ی فناوری‌ها کمک کند. در این راستا AI باید طوری توسعه یابد که استانداردهای بالای تعالی علمی را رعایت کند.

گوگل قصد دارد با ذی‌نفعان مختلف کار کند تا با استفاده از رویکردهای چندرشته‌ای و جامع، رهبری اندیشمندانه در این حوزه ارتقا یابد. همچنین به‌طور مسئولانه دانش AI را به اشتراک بگذارد و به‌وسیله‌ی انتشار مواد آموزشی، بهترین عملکردها و تحقیقاتی که افراد بیشتری را برای توسعه کاربردی AI توانمند می‌سازد، در دسترس همگان قرار دهد.

۷- راه توسعه برای کاربردهایی که با این اصول همخوانند فراهم شود.

فناوری‌ها می‌توانند کاربردهای گوناگونی داشته باشند؛ یعنی هم می‌توانند در جهت مثبت و هم منفی به کار گرفته شوند. با توجه به این نکته، طراح یک فناوری می‌تواند محدودیت‌هایی را در فرآیند طراحی لحاظ کند که کاربرد منفی محدود شود و کاربرد مثبت به‌عنوان استفاده‌ی اصلی مطرح شود. لذا باید در جهت محدود کردن کاربردهای بالقوه مخرب و قابل سوءاستفاده پیش رفت. برای چنین کاری باید مؤلفه‌هایی را در نظر داشت و بر اساس آن‌ها استفاده‌ی محتمل از یک فناوری را ارزیابی کرد. برای نمونه می‌توان مؤلفه‌های زیر را مورد بررسی قرار داد:

هدف و استفاده اصلی^۱: هدف عمده و استفاده محتمل از یک فناوری و برنامه کاربردی بررسی شود و این ارزیابی صورت گیرد که چقدر این هدف عمده به استفاده مخرب می‌انجامد؟

سرشت و اختصاصی بودن^۲: آیا فناوری اختصاصی است یا در دسترس عموم است؟ به عبارت دیگر آیا فناوری برای نهاد، سازمان یا کاربرانی خاص تولید می‌شود یا ابزاری است که به‌طور کلی در خدمت عموم خواهد بود؟

گستره^۳: آیا استفاده از این فناوری تأثیر قابل توجهی خواهد داشت؟ به‌طور مثال آیا توسط اقشار مختلف مردم استفاده می‌شود یا گستره‌ی استفاده از آن محدود است؟ روشن است که هر چه گستره‌ی استفاده از یک فناوری بزرگ‌تر باشد آن فناوری می‌تواند تأثیرات اخلاقی بیشتری داشته باشد.

¹ Primary Purpose and Use

² Nature and Uniqueness

³ Scale

سرشت درگیر شدن گوگل^۱: ارزیابی شود که آیا ابزارهایی که تولید می‌شوند چندمنظوره‌اند، ابزارهای ترکیبی برای مشتری^۲ هستند، یا راه‌حلی خاص برای آن^۳ هستند؟
با توجه به این مؤلفه‌ها می‌توان دیدی نسبت به استفاده از فناوری AI به دست آورد.^۵

۱-۳- کاربردهایی از هوش مصنوعی که دنبال نخواهند شد

گوگل قصد دارد در حوزه‌های کاربردی زیر، سامانه‌های AI را طراحی و استفاده نکند:
فناوری‌هایی که حتماً یا احتمالاً به ضرر عمومی می‌انجامند. اگر مخاطره‌ای شناسایی شود، باید تنها فناوری‌ای را توسعه داد که درباره آن این باور وجود دارد که فایده‌های آن به‌طور ملموس بیش از مخاطرات آن است و می‌توان مخاطرات را با محدودیت‌های ایمن و مصونیت‌ساز همراه کرد.
سلاح‌ها یا دیگر فناوری‌هایی که اهداف اصلی یا کاربرد اصلی‌شان ایجاد یا کمک مستقیم به صدمه‌زدن و آسیب‌رساندن به مردم است.
فناوری‌هایی که به جمع‌آوری و استفاده اطلاعات برای تجسس و پاییدن^۶ می‌پردازد و از این جهت هنجارهای پذیرفته‌شده بین‌المللی را زیر پا می‌گذارد.
فناوری‌هایی که هدف اصلی‌شان در تضاد با اصول پذیرفته‌شده‌ی قوانین بین‌المللی و حقوق بشر است.
هر چه تجربه ما در این فضا گسترده‌تر شود، این فهرست کامل و احتمالاً متحول می‌شود.

۱-۴- نتیجه

گوگل بر این عقیده است که اصول پیش‌گفته مبنایی درست برای این شرکت و توسعه‌ی آینده AI است. به‌علاوه، این قلمرو پویا و در حال دگرگونی است و باید با تواضع، تعهد به تعامل‌های داخلی و خارجی، و تمایل نسبت به به‌روزرسانی رویکردها، با توجه به آنچه در طی زمان می‌آموزیم، فعالیت کنیم و پیش برویم.

¹ Nature of Google's Involvement

² General-Purpose Tools

آچار فرانسه یا انبردستی ابزارهای چندمنظوره‌اند یعنی چندین کاربرد می‌توان به آن‌ها نسبت داد. در فضای نرم‌افزارها و اپلیکیشن‌های کامپیوتری نیز خیلی از ابزارها چنین ویژگی‌ای دارند. به‌طور مثال Alerts, Picasa, Chrome جزء ابزارهای چندمنظوره‌ی گوگل هستند.

³ Integrating Tools for Customers

افزایش تعداد اپلیکیشن‌ها، سرویس‌ها و انواع مختلف داده‌ها در یک اکوسیستم این امر را ضروری کرده است که ابزارهایی برای جمع‌آوری و ترکیب داده‌ها به‌وجود بیاید. این ابزارها را ابزارهای ترکیب‌کننده می‌گویند که متناسب با هدف و نیاز مشتریان به ترکیب داده‌ها می‌پردازد. برخی ابزارهای ترکیب‌کننده‌ی داده‌ها عبارت‌اند از: Centerprise Data Integrator, Microsoft SQL, IBM InfoSphere.

⁴ Custom Solution

این ابزارها ناظر به مسئله‌ای خاص که سازمان یا کاربران خاص با آن دست‌به‌گریبان هستند طراحی می‌شوند.
^۵ باید توجه داشت که این مؤلفه‌ها کامل نیستند. اولاً آن‌ها ممکن است با هم هم‌پوشانی داشته باشند. ثانیاً جامع نیستند یعنی مؤلفه‌های دیگری وجود دارد که دید ما را نسبت به استفاده‌ی احتمالی از یک ابزار عمیق‌تر می‌کند. در ضمن، طراح ابزار پس از بررسی و ارزیابی استفاده‌ی احتمالی از یک ابزار باید شرایطی را فراهم آورد که استفاده‌ی اخلاقی بد آن، محدود شود. در مباحث «اخلاق طراحی» روش‌ها و سازوکارهای چنین کاری بررسی می‌شوند.

^۶ Surveillance

فصل ۲- اجرای اصول هوش مصنوعی

۲-۱- مقدمه: کنش‌های هوش مصنوعی مسئولیت‌پذیر

گوگل قصد دارد برنامه‌های پیشگام^۱، فرآیندها و ساختارهای حکمرانی را به‌طور مداوم طوری توسعه دهد که «اصول AI» را عملیاتی کند و این اصول را دائماً با دیگران به اشتراک بگذارد. گوگل خود را متعهد می‌داند تا در استفاده و توسعه مسئولانه AI و به‌اشتراک‌گذاری منابع با جامعه گسترده‌تر، پیشرفت داشته باشد. توسعه‌ی AI فرصت‌های جدیدی در زمینه‌های تجارت، سلامتی و آموزش، برای زندگی انسان‌ها در اقصا نقاط جهان ایجاد کرده‌است. درعین‌حال، سؤالاتی درباره بهترین راه رسیدن به انصاف، تفسیرپذیری، حریم خصوصی و امنیت در سامانه‌ها مطرح شده‌است. تابه‌حال به این سؤالات پاسخی کامل داده نشده و درواقع زمینه‌های فعالی از تحقیق و توسعه به آن‌ها اختصاص یافته است. گوگل در توسعه مسئولانه AI و به اشتراک‌گذاری دانش، پژوهش، ابزارها، پایگاه‌های داده و منابع دیگر با جامعه بزرگ‌تر بیرونی، خود را متعهد می‌داند. در گزارش AI گوگل، چنین بیان می‌شود: «ما مثل بقیه پژوهش‌هایمان، آخرین یافته‌هایمان را مبنا قرار می‌دهیم و کار می‌کنیم تا به تلقی یکپارچه از آن‌ها برسیم. همچنین نتایج خود را متناسب با آنچه در طی زمان یاد می‌گیریم به‌روز می‌کنیم.» در ادامه چند کنش در این راستا بیان می‌شود.

توصیه‌ها

سامانه‌های قابل‌اعتماد، کارا و انسان‌گرای^۲ AI باید متناسب با بهترین کنش‌های عمومی برای سامانه‌های نرم‌افزاری و کنش‌های ناظر به ملاحظات مختص یادگیری ماشین^۳ (ML) طراحی شوند. توصیه‌های مهم در این زمینه از این قرارند:

- از رویکرد طراحی انسان‌گرا استفاده کنید.
- شاخص‌هایی چندگانه برای ارزیابی آموزش (ترینینگ)^۴ و نظارت^۵ شناسایی کنید.
- اگر ممکن بود، به‌طور مستقیم داده‌های خام را بررسی کنید.
- متوجه محدودیت‌های پایگاه‌های داده و مدل باشید.
- سامانه‌ها را بیازمایید، بیازمایید و بیازمایید.
- بعد از راه‌اندازی و استفاده، همچنان به سامانه نظارت کنید و آن را به‌روزرسانی کنید.

¹ Initiatives

² Human-Centered

³ Machine Learning

⁴ Metric

⁵ Training

⁶ Monitoring

۲-۲- انصاف

سامانه‌های AI تجربه‌ها و قابلیت‌های جدیدی برای مردم در همه جای دنیا فراهم می‌کنند. آن‌ها می‌توانند برای وظایفی حساس مانند پیش‌بینی، حضور و سختی یک بیماری در شخص، رساندن انسان‌ها به شریک و همدم مناسب یا کار متناسب، یا مثلاً در مورد خودروهای بدون سرنشین، شناسایی عبور یک شخص از خیابان، به کار گرفته شوند. سامانه‌های کامپیوتری یا سامانه‌های تصمیم‌گیری خودکار، نه‌تنها این ظرفیت را دارند که منصفانه‌تر و دربرگیرنده‌تر از قواعد دلخواه و بی‌مبنا عمل کنند، بلکه می‌توانند با رعایت بی‌طرفی در مواردی که انسان‌ها عموماً پیش‌قضاوت و سوگیری دارند، عملکردی عادلانه داشته باشند. البته خطر وجود سوگیری و عدم انصاف در چنین سامانه‌هایی می‌تواند ابعاد بزرگی داشته باشد. بنابراین هر چه تأثیر AI در بخش‌ها و جوامع مختلف افزایش می‌یابد، باید در جهت ایجاد سامانه‌هایی که انصاف و دربرگیری^۲ بیشتری دارند، تلاش کرد.

طراحی سامانه با درجه انصاف و دربرگیری بالا، بسیار سخت و پیچیده است. اولاً در فرآیند یادگیری ماشین (ML)، داده‌های مبنای آموزش، از سامانه‌های واقعی جهان جمع‌آوری شده‌اند. بنابراین یک مدل دقیق، ممکن است سوگیری‌های موجود در داده‌های مبتنی بر نژاد، جنسیت، دین یا ویژگی‌های دیگر را مبنای عملکرد خود قرار دهد یا حتی سوگیری‌ها را تشدید کند. برای مثال، یک سامانه کاریابی^۳ ممکن است از روی داده‌های فعلی یاد بگیرد که رؤسای شرکت‌ها اغلب مرد هستند؛ پس در انتخابش مردها را در اولویت قرار دهد. یا مثلاً ممکن است در ترجمه‌ی واژه‌ها به زبان اسپانیایی ضمایر مؤنث را برای کارهایی چون پرستاری یا نگه‌داری بچه پیش فرض بگیرد صرفاً به این خاطر که بیشتر داده‌های موجود، این الگو را تأیید می‌کنند.

ثانیاً حتی با وجود آموزش یک سامانه‌ی هوشمند توسط داده‌های متنوع و سپس آزمودن چندباره‌ی محصول، همچنان نمی‌توان مطمئن شد که یک سامانه در تمام شرایط بدون سوگیری عمل خواهد کرد. برای مثال، سامانه‌ی تشخیص صدایی که روی بزرگسالان آمریکایی آموزش دیده است، در آمریکا احیاناً منصفانه و دربرگیرنده، عمل خواهد کرد. حال اگر این سامانه توسط نوجوانان استفاده شود، احتمالاً در شناخت برخی اصطلاحات کوچه‌خیابانی یا برخی عبارات باز می‌ماند. اگر در انگلستان استفاده شود، ممکن است در شناختن لهجه محلی انگلیسی‌ها دچار مشکل شود. حتی اگر همین سامانه توسط بزرگسالان آمریکا استفاده شود، همچنان این امکان وجود دارد که برخی بخش‌های پیش‌بینی‌نشده‌ی آمریکا (مثل جمعیتی که لکنت زبان دارند) به‌طور ضعیفی پردازش شوند. بنابراین بعد از شروع به استفاده از سامانه، معلوم می‌شود که چه نقاط کور غیرمنصفانه و پیش‌بینی‌نشده‌ای وجود دارد.

^۱ Inclusive

^۲ Inclusion

^۳ Job-Matching

ثالثاً هیچ تعریف استانداردی از انصاف و بی‌طرفی در اختیار نداریم؛ چه در موردی که عامل، انسان است و چه زمانی که عامل، ماشین است. معرفی معیارهای درست بی‌طرفی برای یک سامانه، مستلزم این است که تجربه‌ی کاربران و ملاحظات فرهنگی، اجتماعی، تاریخی، سیاسی، حقوقی و اخلاقی - که در عمل خیلی از آن‌ها سبک‌سنگین^۱ و جایگزین هم می‌شوند - فهم و تبیین شوند. مثلاً آیا منصفانه است که وام‌های یکسانی به گروه‌های مختلف که نحوه بازپرداخت متفاوتی داشته‌اند داده شود، یا باید براساس بازپرداخت به‌موقع، این وام تعلق گیرد؟ یا شاید هیچ‌کدام از این دو منصفانه نباشد؟ پس باید در چه سطحی از دقت، گروه‌ها را تعریف کرد؟ آیا کافی است که افراد را به دو گروه «وام‌گیرنده‌ی خوش‌قول» و «وام‌گیرنده‌ی بدقول» تقسیم کنیم؛ یا بهتر است تقسیم‌بندی دقیق‌تری انجام شود؟ مثلاً براساس ماه‌های تأخیر، به چهار گروه تقسیم شوند: «یک ماه تأخیر و کمتر»، «از دو تا شش ماه تأخیر»، «از هفت تا بیست و چهار ماه تأخیر» و «بیش از دو سال تأخیر». اینکه تا چه حد لازم است دقت در تقسیم‌بندی داده‌ها انجام شود که سامانه در عین کارآمدی، منصفانه نیز باشد، مسئله‌ای است که در اینجا محل بحث است. سؤال دیگر این است که چطور مرز بین گروه‌ها تعریف شود؟ اصلاً چه زمانی منصفانه است که گروه تعریف کرد به‌جای اینکه براساس تمایزهای فردی شناخت حاصل کرد؟ حتی برای موقعیت‌هایی که به نظر ساده می‌رسد، آدم‌ها برسر اینکه چه چیزی منصفانه است ممکن است با هم اختلاف نظر داشته باشند و روشن نباشد که چه دیدگاهی باید در تعیین خط‌مشی‌ها مبنا قرار گیرد، خصوصاً در مورد تنظیمات جامع^۲ سامانه‌ها (تنظیماتی که قرار است بر همه سرورهای محلی اعمال شود).

انصاف و دربرگیری در AI حوزه‌ای فعال از پژوهش را به خود اختصاص داده است. این پژوهش‌ها شامل چنین مواردی می‌شوند: «توسعه‌ی نیروی کار دربرگیرنده‌ای که دانشی عمیق و متنوع را محقق کند»، «ارزیابی عملکرد سامانه در قبال آموزش از پایگاه‌های داده، برای منابعی که سوگیری بالقوه دارند»، «آموزش هرچه بیشتر مدل‌ها برای حذف یا اصلاح سوگیری‌های مسئله‌دار»، «ارزیابی مدل‌های ML در راستای تشخیص نابرابری در اجرا» و بالاخره «آزمودن مداوم سامانه‌های نهایی برای تشخیص خروجی‌های غیرمنصفانه‌ی احتمالی». درواقع خود مدل‌های ML می‌توانند در جهت شناسایی برخی از سوگیری‌ها و موانع دربرگیری، به کار آیند. مواردی که ممکن است در انسان‌ها، خودآگاه یا ناخودآگاه وجود داشته باشد، در طول تاریخ رشد کند و ریشه بدواند. انصاف در AI مسئله‌ای است که تاکنون حل نشده است. این می‌تواند هم فرصت به شمار آید و هم چالش. گوگل خود را متعهد می‌داند که در همه این حوزه‌ها پیشرفت کند و در جهت خلق ابزارها، پایگاه‌های داده و دیگر منابع موردنیاز برای جامعه بزرگ‌تر، بیشتر و بیشتر اقدام کند. در ادامه تفکر فعلی حاکم در گوگل به‌طور خلاصه بیان می‌شود.

توصیه‌ها

^۱ Trade Off

^۲ Global Settings

برای هر وظیفه‌ی خاص، باید بررسی کرد که آیا ML می‌تواند راه‌حلی منصفانه برای انجام آن باشد؟ اگر می‌تواند، باید توجه داشت که یک مدل واحد و صحیح برای همه‌ی وظایف ML وجود ندارد. به همین صورت، یک تکنیک واحد برای تضمین انصاف در همه راه‌حل‌ها وجود ندارد. در عمل، محققان و توسعه‌دهندگان باید از روش‌های مختلف و خلاقانه برای بهبود انصاف در سامانه استفاده کنند. در این راستا، موارد زیر توصیه می‌شوند:

- مدلی طراحی کنید که اهدافی مشخص برای انصاف و دربرگیری داشته باشد.
- از پایگاه‌های داده‌ای که نمونه‌ی نمایانگر^۱ جامعه هستند استفاده کنید. نمونه‌ی نمایانگر منصفانه‌ترین بازنمایی از جامعه است.
- سامانه را دائماً برای تشخیص سوگیری‌های غیرمنصفانه بیازمایید. در این باره فهرستی از سوگیری‌ها تهیه کنید و آن را بررسی کنید.
- سامانه را در مرحله‌ی اجرا، از نظر انصاف تحلیل کنید.

۲-۳- تفسیرپذیری^۲

پیش‌بینی و تصمیم‌گیری خودکار می‌تواند موجب ارتقای زندگی از جهات مختلف شود. از موارد ساده‌ای مانند پیشنهاد موسیقی‌هایی که شخص ممکن است دوست داشته باشد تا نظارت بر علائم حیاتی بیمار، توسط سامانه‌های خودکار انجام می‌شود. تفسیرپذیری امری کلیدی است در این جهت که بتوان از چرایی عملکرد سامانه‌های AI پرسید، آن را فهمید و به آن اعتماد کرد. تفسیرپذیری همچنین دانش محلی و ارزش‌های اجتماعی را بازنمایی می‌کند، برای دانشمندان و مهندسان شرایطی را فراهم می‌کند تا بهتر طراحی کنند، توسعه دهند و مدل‌ها را عیب‌زدایی^۳ کنند. علاوه‌بر آن کمک می‌کند تا سامانه‌های AI همان‌طور که موردنظر است کار کنند.

مسائل حوزه تفسیرپذیری، هم در مورد انسان‌ها و هم درباره‌ی سامانه‌های هوشمند مطرح است. حتی برای یک انسان این که تصمیم‌هایش را تبیین کند کار راحتی نیست. برای مثال، برای متخصص غدد^۴ کار اصلاً راحتی نیست که همه دلایلش را برای اینکه چرا فکر می‌کند سرطان یک بیمار عود کرده به‌طور کمی و دقیق توضیح دهد. ممکن است بگوید شهودی دارد که سبب‌شده آزمایش‌های تکمیلی برای نتایج مطمئن‌تر را در دستور کار قرار دهد. برخلاف انسان، سامانه AI می‌تواند فهرستی از اطلاعات مختلف را که به این پیش‌بینی

¹ Representative

^۲ منظور از تفسیرپذیری، توضیح‌پذیری، تبیین‌پذیری یا فهم‌پذیری (Interpretability) این است که مدلی که از یادگیری ماشین حاصل می‌شود این قابلیت را داشته باشد که توسط اصطلاحاتی که برای انسان قابل فهم‌اند توضیح داده شود. از آنجاکه اعتماد به سامانه‌ای که قابل فهم نیست مسئله‌برانگیز است، تفسیرپذیری به مباحث اخلاق ارتباط پیدا می‌کند.

³ Debug

⁴ Oncologist

انجامیده ارائه کند. مثلاً نشانگرهای زیستی در سطوح مختلف و همچنین اسکن‌هایی را از ۱۰۰ بیمار مختلف طی ۱۰ سال به‌عنوان اطلاعات مبنا بیان کند. البته این‌که چطور این داده‌ها با هم ترکیب شده‌اند تا مثلاً به این تخمین برسند که در مورد یک بیمار ۸۰ درصد شانس سرطان هست و به همین خاطر او باید اسکن PET شود، کار مشکلی است. فهم مدل‌های پیچیده AI، مانند شبکه‌های عصبی عمیق (DNN)^۱ می‌تواند حتی برای خود خبرگان حوزه ML هم کار سختی باشد.

فهم و آزمون سامانه‌های AI همچنین چالش‌های جدیدی (در مقایسه با چالش‌های نرم‌افزارهای سنتی) ایجاد کرده است. نرم‌افزار سنتی^۲ سلسله‌ای از قواعد اگر-آنگاه‌ها هستند که در آن‌ها تفسیر و عیب‌زدایی اجرا عموماً از طریق دنبال کردن یک مسئله در چندراهه‌ها پی گرفته می‌شود. اگرچه چندراهه‌ها می‌توانند پیچ‌درپیچ باشند لافل انسان می‌تواند مسیرها را بر اساس کدشان پیگیری کند و نتیجه را بفهمد.

در سامانه‌های AI «مسیر کد»^۴ ممکن است شامل میلیون‌ها پارامتر و عملگر ریاضی باشد و به‌مراتب سخت‌تر از این است که یک عیب^۵ مشخص را که به تصمیمی اشتباه منجر شده پیدا کرد. البته در طراحی سامانه‌ی خوب، ممکن است آن میلیون‌ها پارامتر به داده‌هایی که برای آموزش استفاده شده‌اند، نحوه توجه مدل به داده، یا ویژگی‌های خاص برگردد که با شناسایی چنین چیزی می‌توان عیب را پیدا کرد. این با یکی از مسائل کلیدی در نرم‌افزار تصمیم‌گیری سنتی یعنی «اعداد جادویی»^۶ تفاوت دارد. آن‌ها قواعد یا آستانه‌های تصمیمی هستند که بدون هیچ تبیینی توسط یک برنامه‌نویس تعیین شده و اغلب مبتنی بر شهود او یا مجموعه‌ای محدود از نمونه‌های آزمایشی هستند.

به‌طور کلی، سامانه AI به‌وسیله بررسی داده‌های آموزش^۸ و فرآیند آموزش، تفسیر و فهمیده می‌شود. اگرچه تفسیرپذیری سامانه، چالشی جدید است؛ اما تلاش جمعی جامعه فناوری این است که با فرمول‌بندی رهنمودها، بهترین ابزارها را برای فهم، کنترل و عیب‌زدایی سامانه‌ها فراهم کنند. این حوزه‌ای فشرده از دانش و توسعه در گوگل است. تصمیم بر این است که برخی از کارها و دیدگاه‌های این حوزه به اشتراک عمومی گذاشته شود.

توصیه‌ها

با علم به اینکه رسیدن به راه‌حلی کامل برای تفسیرپذیری و پاسخ‌گویی^۹ هنوز میسر نشده و این عرصه، حوزه‌ای پویا در بحث‌های تحقیقاتی گوگل و دیگر جوامع ML است؛ در ادامه برخی موارد توصیه می‌شود:

- برای تفسیرپذیری بخش‌های مختلف سامانه، برنامه‌ریزی داشته باشید.

¹ Deep Neural Networks

² Traditional Software

³ If-Then Rules

⁴ Code Path

⁵ Bug

⁶ Magic Numbers

⁷ Threshold

⁸ Training Data

⁹ Accountability

- تفسیرپذیری را به عنوان بخشی مهم و محوری از تجربه کاربر، بدانید.
- مدل را طوری طراحی کنید که تفسیرپذیر باشد.
- شاخص‌هایی انتخاب کنید که هدف نهایی و وظیفه نهایی را بازنمایی کنند.
- مدل آموزش دیده را بفهمید.
- تفسیرها را با کاربران مدل در میان بگذارید.
- آزمایش کنید، آزمایش کنید و آزمایش کنید.

۲-۴- حریم خصوصی

مدل‌های ML از داده‌های آموزش یاد می‌گیرند و بر اساس داده‌های ورودی، پیش‌بینی انجام می‌دهند. گاهی داده‌های آموزش، داده‌های ورودی یا هر دو می‌توانند بسیار حساس^۱ باشند. گرچه ممکن است منافع زیادی در ساختن مدلی باشد که بر اساس داده‌های حساس عمل می‌کند (مثلاً ردیاب سرطان که بر اساس پایگاه‌های داده‌ی عکس از بافت‌ها، آموزش دیده است و برای بررسی اسکن‌های یک بیمار، به کار گرفته می‌شود)، حتماً باید عواقب استفاده از داده‌های حساس را از نظر حریم خصوصی بررسی کرد. این موضوع علاوه بر در نظر گرفتن الزامات حقوقی و قانونی، شامل لحاظ کردن هنجارهای اجتماعی و تجربه‌های فردی نیز می‌شود. برای مثال، ایمن‌بان‌ها یا الزامات امنیتی‌ای که لازم است به کار گرفته شود تا حریم خصوصی افراد تضمین شود چیست؟ با توجه به اینکه ممکن است مدل‌های ML جنبه‌هایی از داده‌ها را ضبط، یادآوری یا آشکار کنند. چه اقداماتی نیاز است تا بتوان مطمئن شد که شفافیت کافی وجود دارد و کاربران نظارت کافی بر داده‌هایشان دارند؟

خوشبختانه این امکان نامطلوب یعنی قابلیت آشکارسازی داده‌های آموزش مدل‌های ML، می‌تواند با به کارگیری تکنیک‌های مختلف، به روشی اصولی و دقیق، محدود شود و از بین برود. گوگل به‌طور مداوم چنین تکنیک‌هایی را توسعه می‌دهد تا بتواند از آشکارسازی داده‌ها در فرآیند آموزش جلوگیری و حریم خصوصی را در سامانه‌های AI حفاظت کند. این یک حوزه از پژوهش ML است و فضای زیادی برای رشد دارد. در ادامه برخی نکات که گوگل روی آن‌ها تأکید دارد بیان می‌شود.

توصیه‌ها

همان‌طور که پیش‌تر گفته شد، یک مدل صحیح برای وظایف ML وجود ندارد و به همین صورت یک رویکرد صحیح به حفاظت حریم خصوصی، در تمام سناریوهای ممکن، کارایی ندارد. در عمل، محققان و توسعه دهندگان باید با تکرار و آزمودن مدل‌های مختلف، رویکردی را که بین حریم خصوصی و مطلوبیت^۳، تعادل

¹ Sensitive

² Safeguard

³ Utility

ایجاد می‌کند، برای انجام وظیفه‌ی موردنظر انتخاب کنند. برای اینکه این فرآیند، موفقیت‌آمیز پیش برود، به تعریفی روشن از حریم خصوصی نیاز است. چنین تعریفی می‌تواند هم شهودی و هم دارای صورت منطقی دقیق باشد.

- داده را به‌نحو مسئولانه در حفظ حریم خصوصی، جمع‌آوری و مدیریت کنید.
- از پردازش بر-وسیله (on-device) هر جا که مناسب بود حداکثر استفاده را ببرید.
- به نحو مناسب، حریم خصوصی مدل‌های ML را حفاظت کنید.

۲-۵- امنیت

ایمنی و امنیت، سامانه‌های AI را از مداخله‌ی حمله‌گرها مصون می‌کند. در نتیجه می‌توان اطمینان داشت که سامانه آن‌طور که انتظار می‌رود عمل می‌کند. مهم است که امنیت سامانه، پیش از آنکه به‌طور گسترده به‌کار گرفته شود، تأمین گردد. امنیت سامانه‌های AI چالش‌های زیادی را در پی دارد. بنابراین بسیار بعید است بتوان همه سناریوها را پیش‌بینی کرد. خصوصاً در زمانی که ML در جهت مسئله‌هایی که حلشان برای انسان ها سخت است، استفاده می‌شود. همچنین سخت است سامانه‌هایی طراحی کرد که هم الزامات ضروری را برای امنیت داشته باشند، هم دارای انعطاف بالا در برخورد با راه‌حل‌های خلاقانه باشند و هم خودشان را با ورودی‌های غیرمعمول به‌روزرسانی کنند. علاوه‌براین، همان‌طور که فناوری AI توسعه می‌یابد، حمله‌گرها راه‌های جدیدی پیدا می‌کنند و در مقابل آن، باید راه‌حل‌های جدید توسعه یابند. در ادامه توصیه‌هایی مبتنی بر آنچه تاکنون در گوگل تجربه شده است بیان می‌شود.

توصیه‌ها

پژوهش درباره امنیت در ML تهدیدهای مختلفی را دربرمی‌گیرد. این تهدیدها شامل سمی‌سازی داده‌های آموزش، بازیابی داده‌های آموزش حساس، دزدیدن مدل و مثال‌های خصمانه دیگر می‌شود. گوگل محققان را برای پژوهش در این حوزه‌ها دعوت کرده است. یک تمرکز پژوهش در گوگل، درباره امنیت روی یادگیری خصمانه^۲ بوده است؛ یعنی استفاده از شبکه عصبی برای تولید مثال‌های خصمانه که می‌تواند یک سامانه و شبکه دوم را که تلاش دارد تقلب را ردیابی کند سر کار بگذارد یا احمق کند.

هم‌اکنون راه‌حل‌های دفاعی علیه مثال‌های خصمانه هنوز به‌اندازه کافی برای استفاده در محیط‌های تولید، قابل‌اعتماد نیستند. در این زمینه پژوهش‌های زیادی در حال انجام است. اما چون هنوز دفاع مؤثری وجود ندارد، توسعه‌دهندگان باید نتایج محتمل حمله‌ی موفق به سامانه‌ی خود را در نظر بگیرند. همچنین نباید سامانه را جایی بسازند که چنین حمله‌هایی محتمل است چون آثار منفی زیادی به بار می‌آورد. در این راستا توصیه‌های زیر ارائه می‌شوند:

¹ Attacker

² Adversarial Learning

- تهدیدهای بالقوه به سامانه را شناسایی کنید.
- رویکردی نسبت به تهدیدهای مهلك اتخاذ کنید.
- به یادگیری ادامه دهید تا پیشرفته و به روز بمانید.

فصل ۳- به روزرسانی اصول هوش مصنوعی گوگل

۳-۱- به روزرسانی اصول هوش مصنوعی گوگل - شش ماه بعد^۱

اصول AI گوگل که توسعه و استفاده اخلاقی از AI در پژوهش و محصولات را بیان می کند، شش ماه پیش رونمایی شد. به عنوان تکمله‌ی آن، کنش‌های AI مسئولانه منتشر شد. آن‌ها مجموعه‌ی تکنیک‌های توصیه‌شده‌ای هستند که فصل به فصل به روز می‌شوند و قرار است با اکوسیستم وسیع‌تر AI به اشتراک گذاشته شوند. از آن موقع، برنامه‌ها و فرآیندهای بیشتری در جهت اطمینان از تحقق اصول اخلاقی، در دستور کار گوگل قرار گرفته است.

اول اینکه از تمام گروه‌های گوگل خواسته شده است بررسی کنند که آیا و چطور اصول اخلاقی AI بر پروژه‌شان تأثیر می‌گذارد؟ برای این مقصود، چندین اقدام به انجام رسیده است:

- **آموزش:** براساس پروژه‌ی «اخلاق در کنش فناوری» که در مرکز مارکولا برای اخلاق کاربردی^۲ در دانشگاه سانتا کلارا^۳ توسعه یافت، پروژه آموزش به انجام رسید. محتوای تولیدشده در جهت کمک به کارمندان فنی و غیرفنی گوگل تدوین شد تا آن‌ها بتوانند با موضوعات چندجانبه‌ی اخلاقی که در کارشان برخورد می‌کنند به خوبی مواجه شوند. تاکنون، بیش از ۱۰۰ کارمند گوگل از کشورهای مختلف این دوره را گذرانده‌اند و در آینده قرار است این آموزش‌ها در دسترس هرکسی که در شرکت هست قرار گیرد.
- **سلسله سخنرانی‌های اخلاق AI:** با استفاده از خبرگانی از کشورها، مناطق و رشته‌های مختلف اجرا شد. تا به حال ۸ برنامه با حضور ۱۱ سخنران در موضوعاتی چون «سوگیری در پردازش زبان طبیعی»^۴ و «استفاده از AI در سامانه حقوقی کیفری»^۵ برگزار شده است.
- **پودمان^۶ (ماژول) فنی انصاف** به دوره فشرده ML که به ۱۱ زبان آماده شده و برای یادگیری بیش از ۲۱۰۰۰ کارمند گوگل به کار می‌رود، اضافه شده است. **پودمان انصاف** که هم‌اکنون به زبان انگلیسی

^۱ این گزارش در تاریخ ۱۸ دسامبر ۲۰۱۸ نوشته شده است. نویسنده آن کنت واکر (Kent Walker) مشاور ارشد گوگل در مسائل حقوقی است.

^۲ Markkula Center for Applied Ethics

^۳ Santa Clara University

^۴ Bias in Natural Language Processing

^۵ The Use of AI in Criminal Justice

^۶ Module

آماده شده و بقیه زبان‌ها نیز به‌زودی به آن اضافه می‌شود، به دانش‌پژوه آموزش می‌دهد که چطور متوجه شود داده‌ی آموزش، قابلیت سوگیری دارد، چطور آن را بشناسد و با آن مقابله کند.

• به همراه این تلاش‌ها که ناظر به کارمندان گوگل بود، یک **ساختار داوری رسمی** (فرمال) ایجاد شد تا پروژه‌ها، محصولات و قراردادهای براساس آن ارزیابی شوند. این ساختار در تلاش است با تصمیم‌های مدیرانه و ملاحظات دقیق و جزئی، این مسائل را حل کند: چگونه اصول AI عملیاتی شود؟ چگونه وقتی با هم تعارض دارند سبک‌سنگین و جایگزین هم شوند؟ و چگونه مخاطرات در شرایطی خاص کاهش یابد؟ ساختار داوری رسمی شامل سه گروه مرکزی است:

۱- گروه نوآوری مسئولانه که عملیات روزانه و ارزیابی‌های ابتدایی را راه‌اندازی می‌کند. این گروه متشکل از پژوهشگر رفتار کاربران^۱، دانشمندان علوم اجتماعی، حقوق‌دانان، دانشمندان اخلاق، متخصصان حقوق بشر، مشاوران سیاستی و حریم خصوصی است که برخی به‌طور تمام‌وقت و برخی پاره‌وقت کار می‌کنند. چنین گروهی در جهت تحقق تنوع، شمول و دربرگیری رشته‌ها و عرصه‌های مختلف کمک می‌کند.

۲- گروهی از خبرگان ارشد رشته‌های مختلف که تخصص فناوریانه، کارکردی و کاربردی دارند.

۳- گروهی از مدیران اجرایی ارشد که برای راه‌اندازی مشکل‌ترین و پیچیده‌ترین مسائل فکر می‌کنند و تصمیم می‌گیرند. تصمیمات آن‌ها محصولات و فناوری‌های گوناگون را تحت تأثیر قرار می‌دهد. ساختار داوری رسمی گوگل، تا به حال بیش از ۱۰۰ داوری به انجام رسانده و در این راستا شاخص، سختی و احتمال بهترین و بدترین سناریو برای هر محصول و قرارداد را ارزیابی کرده‌است. بیشتر این موارد، مانند یکپارچه‌سازی راهنمای آموزش دربرگیرنده‌ی ماشین در محصولات Cloud AutoML، با اصول اخلاقی همخوان شده‌اند. برخی اقدامات، نظیر پژوهش در شناسایی مکالمه تصویری، طوری اصلاح شده است تا منافع به‌روشنی محقق شوند و همچنین محدودیت‌های مدل، طوری شفاف شوند که ظرفیت سوءاستفاده به حداقل برسد. در تعدادی از موارد استفاده از محصولات - مانند شناسایی اهداف عمومی مربوط به ساختار API - نیز چنین تصمیمی گرفته شده است که پیش از آنکه گوگل روی سؤالات مهم مربوط به فناوری و سیاست کار کند، پیشنهاد عملیاتی در این زمینه ندهد.

تنوع و گستره‌ی مواردی که تا به حال در نظر گرفته شده‌اند وجود چارچوبی برای یکسان‌سازی فرآیند محصولات و فناوری‌های گوگل را ضروری می‌کند. این چارچوب شامل ساختن یک گروه مشاوره‌ی خارجی متشکل از متخصصان رشته‌های مختلف است تا حاکمیت و فرآیندهای داخلی بیان شده در بالا را تکمیل کنند. گوگل خود را متعهد به ارتقای ملاحظات عمیق درباره‌ی مسائل مهم می‌داند و کارگروه‌هایی را که در جهت داوری فرآیند، مشارکت دارند ارج می‌نهد.

^۱ پژوهشگری که روی رفتار، نیاز و انگیزه‌های کاربران مطالعه می‌کند = User Researcher

۳-۲- هوش مصنوعی و مسئولیت‌پذیری: عملیاتی کردن اصول اخلاقی - یک سال بعد^۱

در تاریخ ۲۸ جولای ۲۰۱۹، در گزارشی اقدامات یک سال اخیر گوگل در ارتباط با اصول AI منتشر شد. با توجه به اینکه این اصول در حدود یک سال پیش تدوین شد و شش ماه بعد از آن، گزارش پیشروی آن‌ها منتشر شد، گزارش فعلی را می‌توان نحوه‌ی عملیاتی کردن اصول اخلاقی گوگل در حوزه‌ی AI در یک سال اخیر تلقی کرد. در ادامه، این اقدامات عملیاتی تشریح می‌شوند.

۳-۲-۱- آموزش داخلی

گوگل تلاش کرد کارمندانش را آموزش دهد و توانمند سازد تا مسائل مهم AI را درک کنند و بتوانند درباره این‌که چطور AI را در عمل به کار گیرند به‌طور منتقدانه فکر کنند. در سال گذشته، هزاران گوگل‌گر^۲ درباره بی‌طرفی در بحث ML، آموزش دیدند. همچنین در چهار دفتر گوگل، گروه‌های آموزش اخلاقی تشکیل شد و یک سلسله سخنرانی در سه قاره‌ی آمریکا، اروپا و آسیا برگزار شد.

۳-۲-۲- ابزارها و پژوهش

در سال گذشته، گوگل در راستای به‌اشتراک‌گذاری دانش، ساختن ابزارها، به‌روزرسانی محصولات و بنانهادن چارچوبی برای توسعه‌ی AI اخلاقی و مسئولانه که به همگان سود برساند حرکت کرد. بیش از ۷۵ مقاله‌ی تحقیقاتی در ارتباط با AI مسئولانه با موضوع انصاف در ML، تفسیرپذیری، حریم خصوصی و امنیت منتشر شد. علاوه‌براین، ۱۲ ابزار جدید توسعه داده شد و به نحو منبع‌باز^۳ در اختیار همه قرار گرفت. برخی از این ابزارها به عبارت‌اند از:

- ابزار «چه می‌شود-اگر» (What-If Tool): ابزاری است که اجازه می‌دهد کاربران، یک مدل ML را توسعه دهند بدون آنکه نیازی به نوشتن کد داشته باشند. این ابزار کاربران را قادر می‌سازد تا سوگیری‌ها و تأثیرات محدودیت‌های مربوط به بحث انصاف را تصویرسازی کنند و عملکرد مدل‌های مختلف را با هم مقایسه کنند.
- گوگل آموزش یکپارچه‌ی مدل‌ها را توسعه داد. این رویکرد جدید نسبت به ML، به توسعه‌دهنده‌ها^۴ اجازه می‌دهد مدل‌های AI را آموزش دهند بدون اینکه داده‌ها از دستگاه کاربر خارج شوند. این رویکرد هم‌اکنون به نحو منبع‌باز در اختیار همگان است: TensorFlow Federated.

^۱ این گزارش در تاریخ ۲۸ جولای ۲۰۱۹ منتشر شده است. نویسندگان گزارش، جف دین (Jeff Dean)، از مدیران ارشد هوش مصنوعی گوگل و کنت واکر (Kent Walker)، از مشاوران حقوقی گوگل هستند.

^۲ Googler = کارمند گوگل

^۳ Open Source

^۴ Developer

- «مردم + کتاب راهنمای AI» جعبه‌ابزاری از روش‌ها و چارچوب‌های تصمیم‌گیری است که برای ساخت محصولات انسان‌محور تهیه شده‌است. این ابزار در ماه مه امسال آغاز به کار کرد و از چهل گروه محصول گوگل در آن مشارکت دارند.
- گوگل مدعی است با در نظر گرفتن آخرین ایده‌ها و کارهای تکنیکی، به‌طور فصلی در حال به‌روزرسانی کنش‌های AI مسئولانه است.

۳-۲-۳- دآوری فرآیند

دآوری فرآیند تحقق اصول اخلاقی AI را تسهیل می‌کند. همه‌ی کارمندان گوگل تشویق شده‌اند به این موضوع توجه کنند که چطور اصول اخلاقی AI می‌تواند بر پروژه‌شان اثر بگذارد. همچنین فرآیندها طوری تکامل یافته‌اند تا این اطمینان حاصل شود که حتماً پروژه‌ها، محصولات و قراردادهای جدید، به‌طور کامل بررسی و ارزیابی شده‌اند. تصمیم بر این است که در هر پروژه، منافع در نظر گرفته شود و تلاش شود تا مخاطرات از بین برود. در این رابطه دو مثال بیان می‌شود.

۱- منبع AI ابری^۱

به کمک منبع AI ابری، شرکت‌ها و سازمان‌ها می‌توانند تعداد زیادی مدل ML را که تاکنون آموزش دیده‌اند به اشتراک بگذارند و به بقیه نیز دسترسی داشته باشند. بیشتر محتوای منبع AI توسط سازمان‌های بیرون از گوگل منتشر می‌شود. این باعث می‌شود نتوان آن‌ها را از جهت نسبتی که با اصول اخلاقی AI گوگل دارند ارزیابی کرد. به همین خاطر، گوگل به ارزیابی ملاحظات اخلاقی که پیرامون انتشار منبع AI است پرداخته و مثلاً آن را از این دید که بالقوه ممکن است استفاده‌های دوگانه‌ی مضر داشته باشد، یا در جهت سوءاستفاده یا ارائه اطلاعات گمراه‌کننده به کار برده شود ارزیابی کرده است.

در نتیجه‌ی دآوری انجام‌شده، گوگل، راهبردی دوسطحی^۲ برای محتوایی که بالقوه مخاطره‌آمیز و مضر است ارائه داد:

- تشویق اعضای جامعه برای وزن‌دهی به مسائلی مثل سوگیری غیرمنصفانه. برای حمایت از جامعه، AI ابری منابعی فراهم کرد (مانند [inclusive ML guide](#)) تا از کاربران برای تشخیص محتوای قابل اعتماد، کمک بگیرد.
 - تدوین ضوابط برای منبع AI ابری، مخصوصاً بخش‌هایی از آن که با محدودیت‌های محتوا و رفتار مرتبط است.
- این الزامات ایمنی، اکوسیستم محتوای منبع AI را به کارایی و حفاظت بیشتر، نزدیک کرده و باعث شده است که گوگل با خیالی آسوده به سوی راه‌اندازی منبع AI پیش برود.

¹ Cloud AI Hub

² Two-Tired Strategy

۲- مقاله پژوهشی متن به گفتار (TTS)^۱

گروهی در گوگل مقاله‌ای آکادمیک نوشتند که به چالشی اساسی در پژوهش AI می‌پرداخت: سامانه‌ها اغلب نیاز دارند از ابتدا و بدون هیچ دانشی آموزش داده شوند که این توسط داده‌های خیلی زیاد قابل انجام است و برای هر وظیفه، حتی اگر اندکی متفاوت با دیگر وظایف باشد، باید جداگانه صورت گیرد. این مقاله به‌طور جزئی یک شبکه پربازده متن به گفتار را بررسی می‌کند طوری که یک بار آموزش ببیند و پس‌از آن با سخنگوهای مختلف که زمان یا داده‌های کمتری در اختیار دارند سازگار شود.

گرچه چنین سامانه‌ی هوشمندی می‌تواند برای افراد با مشکلات گفتاری به کار بیاید، اما مدیران گوگل متوجه شدند که این سامانه می‌تواند در جهت کاربردهایی مضر، مانند تقلید صدای یک نفر برای اهداف گول‌زننده، به کار گرفته شود.

درنهایت گوگل به این نتیجه رسید که فناوری توصیف‌شده، ظرفیت محدودی برای سوءاستفاده دارد؛ بنا به دلایل مختلف، مثلاً کیفیت بالای صدا برای راه‌اندازی سامانه شرط است؛ در صورتی که ضبط کردن‌های ساختگی از اینترنت چنین شرطی را ندارد. به‌علاوه، به‌اندازه کافی بین صداهای تقلیدشده توسط سامانه و صدای خود سخن‌گو، تمایز هست که بتوان مورد واقعی را از غیرواقعی تشخیص داد. سرانجام گوگل به این نتیجه رسید که این مورد با اصول اخلاقی AI همخوان است. البته فایده‌ی این بررسی افزایش حساسیت گوگل برای شناسایی و جلوگیری از سوءاستفاده‌های بالقوه بود.

۳-۲-۴- تعامل با ذی‌نفعان خارجی

برای ایجاد AI مسئولانه به لحاظ اجتماعی، تعامل مستمر با جامعه بزرگ‌تر لازم است. اعضای بخش مسئولیت‌پذیری گوگل با سیاست‌گذاران و جامعه‌ی فناوری در تعامل بوده‌اند. آن‌ها در بیش از ۱۰۰ کارگاه، کنفرانس پژوهشی و جلسات مدیریتی شرکت کرده‌اند و به‌طور مستقیم با بیش از ۴۰۰۰ ذی‌نفع در اقصا نقاط جهان تعامل داشته‌اند.

همان‌طور که پیشرفت در AI ادامه دارد، گوگل دیدگاه‌های خود را به اشتراک می‌گذارد و با دانشگاه‌ها، صنایع و سیاست‌گذاران تعامل می‌کند تا پیشرفت مسئولانه AI را تبلیغ و ترویج کند. گوگل از قواعد هوشمندانه که متناسب با واحدها و کاربردهای خاص است حمایت می‌کند. او در ابتدای امسال (۲۰۱۹) گزارش جامعی منتشر کرده است تا رویکردهای روبه‌جلو و عمل‌گرایانه به حکمرانی AI را ترویج کند. این گزارش شامل پنج حوزه^۲ می‌شود. حوزه‌هایی که حاکمیت باید در تعامل با کاربران و کارورزان AI تلاش کند تا چارچوبی برای آن تعیین کند.

^۱ Text-to-Speech

^۲ این حوزه‌ها در بخش ۴-۳-۱ با عنوان «حوزه‌های کلیدی وضوح‌بخشی و اقدامات پیشنهادی» مطرح می‌شوند.

این بخش چنین به پایان می‌رسد: «ما یاد گرفته‌ایم که همواره کارهای بیشتری برای انجام داریم و باید با رهبران، سیاست‌گذاران، دانشگاهیان و دیگر ذی‌نفعان صنایع همکاری کنیم تا مسائل مهم را حل و فصل کنیم. گفتگوهای این‌چنینی، تلاش زیاد و دوندگی، و بررسی دیگر دیدگاه‌ها، لازم و ضروری است. تا اینکه مطمئن شویم AI به فهرست فناوری‌هایی می‌پیوندد که زندگی را برای بهتر شدن تغییر می‌دهد.»

فصل ۴ – دیدگاه‌های سیاست‌گذاران

۴-۱ – مقدمه

همان‌طور که ظهور فناوری‌های مختلف سؤال‌آلاتی را مطرح می‌کند، استفاده از AI نیز سؤال‌آلاتی درباره سیاست‌گذاری، چارچوب حاکمیتی، حوزه‌های کاربرد، تأثیر اقتصادی و موارد دیگر مطرح کرده است. در ادامه دیدگاه‌های گوگل درباره این مسائل و دیگر مسائل انضمامی این‌چنینی بیان می‌شود.

۴-۲ – توسعه‌ی مسئولانه‌ی هوش مصنوعی

در این مقاله، رویکرد گوگل به توسعه و استفاده‌ی مسئولانه از AI توضیح داده می‌شود و پیشنهادهایی برای چارچوب سیاست‌گذاری حکومت‌ها ارائه می‌شود.

۴-۲-۱ – سیاست‌گذاران چطور می‌توانند کمک کنند؟

مانند دیگر فناوری‌ها، هیچ‌چیز به‌طور پیشینی تأثیر AI را کاملاً روشن نمی‌کند. گرچه محققان AI می‌توانند زمینه‌هایی به لحاظ تکنیکی، ایجاد کنند؛ اما تأثیرات AI به تمایلاتی که صنعت و جامعه از خود نشان می‌دهند و همچنین به رهنمودها و حدودمرزهایی که حکومت‌ها برای کاربرد آن وضع می‌کنند بستگی دارد. به همین خاطر سیاست‌گذاران در این جهت از اهمیتی بسزا برخوردارند تا چشم‌انداز را مشخص کنند و چارچوبی بنا گذارند که پیشرفت AI را پشتیبانی کند.

در حالی که جزئیات این حوزه و مباحث عملیاتی، در هر کشور متفاوت است، موضوعات مشترکی وجود دارد که جا دارد سیاست‌گذارانی که طرفدار استفاده مسئولانه از AI هستند در آن‌ها کندوکاو کنند. اطمینان‌آفرینی، تشویق مردم و حمایت از تجارت‌هایی که مرتبط با AI هستند ضروری است. در کنار این، شفافیت چارچوب‌های حکمرانی اهمیت ویژه‌ای دارد. همچنین تبدیل حکومت به الگویی برای AI مسئولانه منافی به دنبال دارد؛ ضمن این‌که بهترین راهنما و الگو برای دیگران است. علاوه‌براین‌ها، چالش‌های به‌خصوصی وجود دارد که به دسترسی به داده و پشتیبانی مالی از پژوهش‌ها مربوط می‌شود که شایسته است سیاست‌گذاران در جهت رفع آن‌ها کمک کنند.

در ادامه جزئیاتی بیشتر درباره هر کدام از این موضوعات بیان می‌شود. در این گزارش، با کمک مثال‌هایی از جای‌جای جهان، پیشنهادهایی درباره تاکتیک‌های مواجهه با چالش‌ها بیان می‌شود. این مثال‌ها برای این است که بتواند جرقه‌ای در جهت ایده‌های مربوط به سیاست ایجاد کند تا شاهد تلاش‌های بیشتر در جهت توسعه AI باشیم. موارد زیر توصیه‌های کلی در زمینه‌ی حکمرانی است.

۱- کمک کنید تا اطمینان و فهم عمومی درباره AI ارتقا یابد.

AI با تقویت دسترسی به مراقبت پزشکی و افزایش بهره‌وری اقتصادی، فرصتی بزرگ در جهت منافع عمومی ایجاد کرده‌است. اما تحقق چنین وعده‌هایی به پذیرش جامعه وابسته است. هم‌اکنون، اطمینان عمومی درباره AI توسط تصاویر علمی-تخیلی مخدوش شده و تیتیر جراید به این‌ها اختصاص یافته‌است و باعث شده که خطرات مهم‌تر ولی غیرواضحی مانند سوگیری و استفاده بدخواهانه از AI کمتر مورد توجه واقع شود. همچنین ترس از AI با نگرانی‌های عمومی درباره‌ی آینده‌ی کار و افزایش ناعدالتی توأم شده است. حکومت‌ها در کنار صنایع، در این باره نقشی کلیدی دارند تا ضمن تصدیق این نگرانی‌ها به آن‌ها بپردازند و یک تصویر متوازن و مبتنی بر واقعیت از فرصت‌ها و چالش‌های AI ارائه کنند؛ تصویری که برآمده از دیدگاه‌های متخصصان رشته‌ها و مشاغل مختلف باشد. در حوزه‌هایی که اعتماد عمومی خیلی کم است، برگزاری گردهمایی‌های مشورتی نخبگانی که چارچوبی برای تعامل و نظارت مستمر و رسمی فراهم می‌کنند، می‌تواند اعتمادبخش باشد.

فعالیت‌هایی که می‌توان در این زمینه در نظر گرفت از این قرارند:

- برگزاری گردهمایی برای جمع و ترکیب کردن نظرات بخش‌های مختلف جامعه
- برگزاری برنامه‌هایی با حضور متخصصان که ورود به آن برای عموم آزاد باشد؛ تا بر روی موضوعات کلیدی مباحثه صورت بگیرد.
- تشکیل هیئت منصفه‌ای از شهروندان^۱ برای گفتگو درباره مسائل کلیدی AI و تدوین توصیه‌هایی در این زمینه.
- بررسی‌های رسمی برای جمع‌آوری توصیه‌های مردم به منظور نشان دادن این که نسبت به نظرات عموم حساسیت و پیگیری وجود دارد.
- تأسیس کمیته‌های مشورتی تخصصی برای فراهم آوردن محلی برای تعامل عمومی.
- جلب نظر سازمان‌های ملی فعال در حوزه علم و سازمان‌های غیردولتی (NGO) در تعامل عمومی خلاقانه؛ مثلاً توسط سخنرانی‌های دانشمندان محبوب و جذاب، هنرهای جذاب و رخدادهای علمی.
- راه‌اندازی پویای آموزش‌های آموزشی که روی کاربردهای AI در زندگی مردم -مثلاً در بخشی‌های خدمات عمومی یا در مراقبت و سلامتی- تأکید کنند و عموم مردم و نه فقط متخصصان، به آن بپیوندند.
- کمک‌هزینه‌های آموزشی پرداخت شود که مردم با پیشینه‌های مختلف، درباره AI آموزش ببینند تا اینکه دیدگاه‌های تازه‌ای به میدان بیاید و به نماینده‌های مختلفی از جامعه، اجازه ظهور و بروز داده شود.

¹ Citizen Juries

۲- بخش‌های اولویت‌دار صنعت را برای به‌کارگیری AI حساس کنید.

شکوفایی اقتصادی AI تنها زمانی محقق خواهد شد که با روشی هدفمند توسط صنعت عملیاتی شود. چنین چیزی نیازمند فهمی جامع از این موارد است: انواع مسائلی که AI در حل آن خوب عمل می‌کند، محدودیت‌های جاری و ذاتی، منابع (ابزارها، داده‌ها، تخصص‌ها و نیروی محاسباتی که راه‌حل‌های AI را اجرایی می‌کنند) و مواردی دیگر از این قبیل. حکومت‌ها می‌توانند به‌عنوان تابلواعلاناتِ کارا در تلنگرزدن^۱ به کسب و کارها عمل کنند تا در فرصت‌های AI کندوکاو و سرمایه‌گذاری صورت گیرد. اهرم مهم دیگر حکومت، حمایت است تا آموزش‌ها در زمینه‌ی AI عملی، تسریع شود و بهترین کنش‌ها و استانداردها ترویج شود.

فعالیت‌هایی که می‌توان در نظر گرفت این موارد هستند:

- مطالعه و پایش در زمینه‌ی آگاهی صنعت از AI، علاقه به استفاده از آن و موانع درک آن، صورت پذیرد.
- گروهی متخصص برای مشاوره به کسب و کارها درباره اولویت‌ها و منافع AI تشکیل شود. حتی می‌توان با فراهم کردن هزینه‌ی اولیه‌ی سرمایه‌گذاری در منابع تخصصی پرورش AI^۲ از کسب و کارها حمایت کرد.
- امتیازات و اقدامات تشویقی برای کسب و کارهایی که AI را پذیرفته‌اند تا در حرکت به سمت علایق شان تسریع شود.
- اعطای یارانه برای حمایت از سرمایه‌گذاری در زیرساخت‌های فیزیکی AI در مناطقی که چنین امکاناتی نیست؛ مثلاً تخفیف در هزینه برق و تسریع در کاهش بهای هزینه‌های سرمایه (capex)^۳
- حمایت مالی از به‌کارگیری AI در کاربردهایی که نیاز اجتماعی زیادی به آن هست؛ مثلاً در ستاد مدیریت بحران.
- تشویق پژوهشگران و تجار برای خلق و به اشتراک‌گذاری پایگاه‌های داده‌ی مربوط به مناطق کلیدی، با در نظر داشتن هنجارهای حریم خصوصی.
- یافتن و اجرای راه‌حل‌های ملموس ساده‌سازی دسترسی کسب و کارها به ابزارهای AI از جمله استفاده از خدمات ابری پایه^۴. (مثلاً با وضع قوانین منعطف‌تر درباره محلی‌سازی داده)
- ارائه‌ی مشوق‌هایی برای تسهیل همکاری و هم‌افزایی فکری در رشته‌های علوم کامپیوتر و دیگر رشته‌های دانشگاهی مرتبط با این صنعت. (مثل پزشکی با زمینه‌های مطالعاتی علوم کامپیوتر)

¹ Nudge

تلنگر مفهومی در سیاست‌گذاری، اقتصاد و روان‌شناسی رفتار است که طبق آن می‌توان کاری کرد که بدون اجبار یا حذف گزینه‌ها، شخص به‌طور غیرمستقیم به تصمیم‌گیری یا رفتاری خاص سوق پیدا کند.

² AI Incubator

³ Capital Expenditure

⁴ Cloud-Based Services

- تشویق دانشگاه‌ها برای ارائه‌ی مواد درسی درباره‌ی به‌کارگیری AI (نه فقط در رشته‌های مهندسی) تا فارغ‌التحصیلان همراه با ورود به بازار کار، آمادگی لازم را داشته باشند.
- اعطای کمک‌هزینه برای حمایت از ارائه و توسعه‌ی آموزش‌های ضمن خدمت برای کسانی که در حوزه AI استخدام شده‌اند یا دنبال کار در بخش‌های اولویت‌دار آن هستند.

۳- از پژوهش‌های معطوف به رفع موانع اجرایی AI حمایت کنید.

به‌رغم پیشرفت‌های اخیر، بسیاری از چالش‌های پژوهشی AI نیازمند توجه‌اند و این باید تا قبل از تحقق تمام ظرفیت‌های AI ادامه پیدا کند. برای مثال، سامانه‌های AI نیاز دارند که بیشتر تفسیرپذیر شوند و از نظر ابعاد داده‌ها و محاسباتی که برای آموزش مدل‌ها نیاز است، پربازده‌تر و راحت‌تر شوند تا افراد بیشتری بتوانند از آن‌ها استفاده کنند یا آن‌ها را بسازند. علاوه‌براین‌ها، برای داشتن اکوسیستم پررونق در AI، مهم است که پژوهش‌های بنیادین صرفاً توسط بخش‌های خصوصی انجام نشود، چراکه موفقیت در امور با ریسک بالا و پاداش بالا^۱ تنها توسط پشتیبانی مالی عمومی امکان‌پذیر است. همچنین در مواردی که کاربردها و منافع بزرگ برای جامعه مطرح است، دانشگاه‌ها و سازمان‌های پژوهشی دولتی، وظایفی در این زمینه دارند.

در این راستا فعالیت‌هایی که می‌توان در نظر گرفت از این قرارند:

- تخصیص بودجه برای پژوهش‌های AI و مطالعه پیشرفته در نهادهای ملی
- پایه‌گذاری مراکز محلی تعالی پژوهش و کاربردهای AI
- دسترسی آزاد به پژوهش‌هایی که از محل بودجه عمومی حمایت شده‌اند و همچنین دسترسی به پایگاه‌های داده‌ی آن.
- اجرای پژوهش AI در حوزه‌های کلیدی و ملی در آزمایشگاه‌ها و مراکز دولتی.
- طراحی چارچوب‌هایی برای تسریع و تسهیل همکاری بخش دولتی و خصوصی در پژوهش AI

۴- به‌منظور ارتقای دسترسی به داده‌های آموزش، به اشتراک‌گذاری مسئولانه‌ی داده را تشویق کنید.

مدل‌های ML برای آموزش، به پایگاه داده نیاز دارند. پایگاه‌های داده باید به‌دقت سازگار و متناسب شده باشند تا به بهترین نحو، مسئله‌ای که قرار است حل شود، روشن گردد و مواجهه با آن صورت گیرد. حکومت‌ها می‌توانند مراکز تحقیق و توسعه در AI را با ایجاد چارچوبی که خلق، به‌اشتراک‌گذاری و استفاده مجدد را از پایگاه‌های داده‌ی مرتبط با حوزه‌های مهم کاربردی، فراهم می‌کند، تسهیل کنند. البته این کار باید با رعایت احترام به حریم خصوصی کاربران صورت پذیرد. به‌موازات آن، همواره به شفافیت بیشتر در تفسیر مقررات و تنظیماتی که بین نفع جامعه و حقوق فردی موازنه ایجاد می‌کند، نیاز است.

فعالیت‌هایی که می‌توان در نظر گرفت ذیل ارائه می‌شود:

¹ High-Risk, High-Reward

- استانداردها و سازوکارهایی برای به اشتراک گذاری داده توأم با حفظ حریم خصوصی، تدوین کنید. این کمک می‌کند که بار حقوقی و اداری مذاکره درباره‌ی انتقال داده، کاهش یابد.
- پایگاه‌های داده‌ی عمومی بیشتری فراهم کنید؛ به‌خصوص برای حوزه‌های نوآورانه و اولویت‌دار.
- محققانی را که حمایت مالی عمومی دریافت می‌کنند تشویق کنید تا پایگاه‌های داده‌ی خود را در فرمت قابل خواندن برای ماشین، منتشر کنند. در عین حال، اصول حریم خصوصی را رعایت کنید.
- راه‌هایی در نظر بگیرید تا بخشی از مردم که می‌خواهند داده‌های شخصی‌شان برای پژوهش یا کاربردهای دیگر استفاده شود، به آسانی آن‌ها را به اشتراک گذارند.
- مجموعه‌ای تخصصی راه‌اندازی کنید که قادر به فراهم آوردن ورودی‌های به‌موقع باشد.

۵- چارچوب‌های حاکمیت سازنده را ارتقا بخشید و مهارت و تخصص را در مجموعه‌های ناظر/حاکمیتی نهادینه کنید.

AI می‌تواند به شیوه‌های مختلف، جامعه را تحت تأثیر قرار دهد، به همین خاطر حکومت نقشی مهم در کنار صنعت برای اطمینان از خروجی‌های AI ایفا می‌کند. هم‌اکنون بخشی از مقررات موجود که مربوط به بخش‌هایی مانند سلامتی، حمل‌ونقل و ارتباطات هستند، ابعاد عملی AI را راهبری می‌کنند. با اینکه متخصصان هر بخش، دانش و مهارت کافی برای ارزیابی کاربردهای خاص و تأثیرات فناوری‌های جدید دارند، اما همچنان به حمایت از پرورش متخصصان AI نیازمندیم. همین‌طور که AI پیشرفت می‌کند، حکومت‌ها باید تخصصشان را در این حوزه گسترش دهند و چارچوب‌های همکاری مختلف را بشناسند تا مباحث مسئله‌برانگیز به حداقل، و ظرفیت AI به حداکثر برسد. کنش‌های مبتنی بر اجماع و مجموعه‌های خودتنظیم‌گر^۱ می‌توانند در جهت ایجاد رویکردهای منعطف و دقیق گام بردارند.

فعالیت‌هایی که می‌توان در نظر گرفت موارد زیر هستند:

- بخش‌های مختلف صنعت را تحلیل کنید تا معلوم شود چگونه طرح و رویه‌های تنظیم‌گری موجود به سامانه‌های AI در آن بخش اعمال می‌شود.
- موانع استفاده‌ی مسئولانه از AI و رسیدن به راه‌حل‌های آن را شناسایی کنید.
 - مثلاً استنتاج نژاد^۲ از این جهت اهمیت دارد که بررسی شود آیا سامانه‌ها از نظر نژادپرستی سوگیری دارند یا نه؟ البته قوانین موجود درباره تبعیض و حریم خصوصی می‌تواند در این موضوع اخلاقی ایجاد کند.
 - مثلاً AI بر-وسیله (on-device) مخاطرات و ویژگی‌های متفاوتی با AI ابری^۳ دارد و ظرایف این

¹ Self-Regulatory Bodies

² Inferring Race

³ Cloud AI

موضوع ممکن است توسط قاعده‌ی کلی «یک سایز برای همه مناسب است»^۱ بازنمایی نشود.

○ مثلاً، قوانین کپی‌رایت می‌تواند داده‌ی موردنیاز برای آموزش سامانه‌ها را محدود کند. بنابراین در جریان تلاش‌های در جهت کاهش سوگیری، اختلال ایجاد شود. (اگر داده‌های مربوطه در اختیار قرار نگیرند.)

- تخصص فنی ناظران بخش‌هایی را که با ظرفیتی عظیم از اختلال و خراب‌کاری سروکار دارند، ارتقا دهید و از آن حمایت مالی کنید.
- کمیته‌ی مشورتی استخدام کنید یا مسئول اداره تولید^۲ را به سمت همکاری درباره مسائل مربوط به حکمرانی AI سوق دهید. در این کمیته نمایندگان از پژوهشگران، صنعتگران و شهروندان هم حضور داشته باشند.
- با مجموعه‌های حکومتی در سراسر جهان تعامل کنید، از تجارب آن‌ها بهره‌مند شوید و تجارب خود را به اشتراک بگذارید.
- با صنعت تعامل کنید تا بهترین راهکارها را با آن‌ها در میان بگذارید و کدهای اخلاقی را ارتقا بخشید.
- از سوی حاکمیت به پژوهشگران آموزش اخلاقی داده شود. (شبيه به کارآموزی الزامی در حوزه‌ی اخلاق پژوهش. این کارآموزی برای پژوهشگرانی که در علوم زیستی کار می‌کنند توسط مؤسسه‌ی ملی سلامتی آمریکا^۳ الزامی شده است.)

۶- حکومت را به الگویی برای پذیرش مسئولانه‌ی AI تبدیل کنید.

ظرفیت AI برای افزایش بهره‌وری و کیفیت خدمات، همان‌قدر به بخش عمومی قابل‌اعمال است که به تجارت است. حکومت‌ها می‌توانند طوری راهبری کنند که بهترین کنش در جهت اولویت‌سازی فرصت‌های پذیرش AI اجرا شود. همچنین می‌توانند نشان دهند چگونه AI می‌تواند به‌طور عملی و جزئی به کار گرفته شود. به‌طور کلی، حکومت‌ها نقشی ارزشمند در جهت سرعت‌بخشی به توسعه ایفا می‌کنند. این می‌تواند از طریق فراهم‌سازی پایگاه‌های داده‌ی عمومی صورت پذیرد تا دیگران بتوانند برای توسعه‌ی خدماتشان از آن‌ها استفاده کنند. (شبيه به راهی که [ImageNet](#) در سرعت‌بخشی به پیشرفت مشاهده توسط کامپیوتر^۴ طی کرد.)

فعالیت‌هایی که می‌توان در نظر گرفت از این قرارند:

- گروهی متخصص استخدام کنید تا درباره کاربردهای AI راهنمایی عملی و اجرایی ارائه کند.
- توصیه‌هایی اصولی برای عوامل حکومتی تدوین کنید تا در حین تصمیم‌گیری برای حمایت‌های مالی،

^۱ One Size Fits All

قاعده‌ای سرانگشتی است که برخی آن را در فرآیند محافظت از داده به کار می‌گیرند، گرچه امروزه این نحوه از محافظت از داده به چالش کشیده شده است.

^۲ POC = Production Office Coordinator

^۳ US NIH

^۴ Computer Vision

از آن‌ها استفاده کنند.

- برنامه‌های آزمایشی (پیلوت) را تشویق کنید تا به‌کارگیری AI در خدمات شهروندی شتاب گیرد.
- با استخدام متخصصان یا همکاری با آن‌ها، فرصت‌هایی را که برای افزایش تخصص داخلی در حوزه AI است دریابید.
- با پلتفرم‌هایی چون [Kaggle](#)^۱ همکاری کنید تا موجب رونق رقابت در استفاده از پایگاه‌های داده‌ی عمومی برای کمک به ارتقای خدمات عمومی شوید.
- موانع را بشناسید و در زیرساخت‌ها و آموزش سرمایه‌گذاری کنید تا مجموعه‌های حکومتی تشویق شوند در جهت بهبود به‌اشتراک‌گذاری و تصدی^۲ پایگاه‌های داده‌ی عمومی بکوشند.

۷- در جهت آماده‌شدن برای دوره‌ی گذارِ نیروی کار، اقداماتی انجام دهید.

در اینکه AI پیکربندی دوباره‌ای برای اشتغال ایجاد خواهد کرد، اجماع عمومی وجود دارد. حتی اگر سرعت و ابعاد این تأثیرگذاری هنوز ناشناخته باشد. حکومت‌ها می‌توانند نقشی کلیدی در توانمندسازی مردم برای توسعه‌ی مهارت‌های آینده‌شان ایفا کنند؛ مثلاً از طریق برنامه‌های تقویت سواد دیجیتال در مدارس یا فراهم آوردن مشوق‌هایی برای یادگیری ضمن کار، می‌توانند چنین نقشی ایفا کنند. برای اطمینان از اینکه به این مهارت‌ها اهمیت داده شود، لازم است انعطاف‌پذیری و قابلیت تغییر در نیروی کار، بهبود یابد تا کسب‌وکارها ساده‌تر نیرو استخدام کنند. همچنین موانع اشتغال مانند مدرک و گواهی‌نامه‌های غیرضروری کنار گذاشته شود. به‌موازات آن، کسانی که پیش‌قدم شده‌اند تا کارآفرینی پیشرفت کند و انعطاف‌پذیری تقویت شود، می‌توانند مردم را برای استفاده از فرصت‌های جدید کاری تجهیز کنند. درنهایت، با توجه به تغییر چشم‌انداز اشتغال، مهم است که درباره‌ی شبکه‌های ایمنی اجتماعی^۳ و تکامل مالی و امکاناتی آن‌ها فکر شود.

کنش‌هایی که می‌توان در نظر گرفت در ادامه بیان می‌شوند:

- در بخش‌های دارای اولویت با صنعت همکاری کنید تا دوره‌های آموزشی برای نسل بعد برگزار کند به‌ویژه دوره‌های ضمن خدمت برای شاغلان دوره‌ی گذار.
- با نمایندگان کارفرمایان و کارمندان ارتباط بگیرید تا گواهی‌نامه‌های کاری را بررسی کنید و مطمئن شوید که محدودیت‌های کاری همچنان کارایی دارند.
- گروه‌های متخصص منطقه‌ای راه‌اندازی کنید. این گروه‌ها باید متشکل از این افراد باشند: نمایندگان دانشگاه، صنعت، نیروی کار و کسانی که بتوانند بر اساس روندهای اشتغال محلی، اولویت‌های آموزش شغلی تدوین کنند.

^۱ جامعه‌ای آنلاین از دانشمندان داده و آموزش‌دهندگان ماشین است که گوگل راه‌اندازی کرده است. این پلتفرم اجازه می‌دهد کاربران پایگاه‌های داده را بیابند، منتشر کنند و مدل‌هایی مبتنی بر علم داده در وب بسازند.

^۲ Curation

^۳ Social Safety Nets

- حمایت و مشوق‌هایی برای کارگران (و خانواده‌هایشان) که در معرض تغییرند فراهم کنید تا به کسب و کارها کمک شود که موقعیت‌های شغلی خالی را در بخش‌های اولویت‌دار پر کنند.
- بورسیه یا مشوق‌هایی به بزرگ‌سالانی که مشغول یادگیری هستند اختصاص دهید. (مانند کاهش مالیات)
- از پژوهش‌ها حمایت مالی کنید و طرح‌های آزمایشی (پایلوت) را تشویق کنید. با این دست اقدامات کمک کنید تا یادگیری در حین کار تقویت شود.
- راه را برای استفاده‌ی کارجویان از فرصت‌های یادگیری غیررسمی، هموار کنید. مثلاً برنامه‌های گواهی‌نامه‌ی هدفمند^۱ یا اردوهای آموزشی فناوری^۲ برگزار کنید.
- درباره‌ی رویکردهای جدید به شبکه‌های ایمن اجتماعی پژوهش کنید و این ایده را در عمل بیازمایید.

۴-۳- دیدگاه‌هایی درباره‌ی مسائل حکمرانی هوش مصنوعی

در ادامه و همچنین در مطلبی دیگر، حوزه‌های کلیدی که در آن حکومت می‌تواند با جامعه‌ی شهروندی، مهندسان و کارورزان AI تعامل کند تا رهنمودی درباره‌ی کاربرد و توسعه‌ی AI مسئولانه ارائه دهد بیان شده است. در اینجا اهم مطالب به‌طور خلاصه ارائه می‌شود.

بنا بر دیدگاه گوگل، AI می‌تواند منافع بزرگی برای اقتصاد و اجتماع به ارمغان بیاورد. بنابراین از تصمیم‌گیری منصفانه‌تر، ایمن‌تر، دربرگیرنده‌تر و آگاهانه، پشتیبانی می‌کند. اما بیشتر این وعده‌ها محقق نخواهند شد مگر اینکه تلاش و مراقبت عظیمی صورت گیرد. بخشی از این تلاش‌ها شامل ملاحظات درباره‌ی چگونگی راهبری استفاده و توسعه‌ی AI است. تلاش برای پاسخ به این سؤال که چه میزانی از نظارت اخلاقی و حقوقی، توسط چه کسی و در چه زمانی لازم است؟

تاکنون دانشگاه‌ها، شرکت‌ها و مجموعه‌های فنی در جلوگیری از استفاده‌ی نامطلوب از AI تا حد زیادی موفق عمل کرده‌اند. همچنین بسیاری از کاربردها در محدوده‌ی چارچوب‌های موجود حاکمیتی، به‌خوبی پیش می‌روند. اما این به معنای عدم نیاز به اقدامات حکومت‌ها نیست. برعکس، این مقاله از حکومت‌ها و جامعه‌ی مدنی درخواست می‌کند که مشارکتی اساسی در بحث حکمرانی AI داشته باشند.

گوگل روی پنج حوزه برای وضوح‌بخشی به انتظارات درباره‌ی کاربرد AI تأکید می‌کند. البته حکومت‌ها، جامعه‌ی مدنی و شاغلان AI نقشی اساسی در این باره ایفا می‌کنند: ۱- استانداردهای تفسیرپذیری، ۲- رویکردهای ارزیابی انصاف، ۳- ملاحظات ایمنی، ۴- الزامات همکاری انسانی و AI و ۵- چارچوب‌های التزام کلی^۳. این موارد استفاده‌ی مسئولانه از AI را تضمین خواهند کرد.

¹ Targeted Certification Programs

² Technology “Boot Camps”

³ General Liability Frameworks

در این گزارش، گوگل دیدگاه خود را درباره این موضوعات مطرح می‌کند. گوگل بیان می‌کند جواب همه سؤالات را نمی‌داند؛ بلکه بالعکس، ضروری است که ذی‌نفعان سیاستی از سراسر جهان وارد بحث شوند. همان‌طور که فناوری AI تحول می‌یابد و تجارب ما با آن رشد می‌کند، این انتظار را داریم که جامعه جهانی به‌عنوان یک کل به یادگیری ادامه دهد و ظرایف بیشتر نمود پیدا کنند. این ظرایف شامل فهمی کامل‌تر از سبک‌سنگین کردن‌ها^۱ و پیامدهای ناخواسته‌ی بالقوه می‌شود که منشأ بسیاری از انتخاب‌های سخت همین‌ها هستند.

تابه‌حال بیشتر بحث‌های جاری درباره حکمرانی AI بین سیاست‌گذاران سطح بالا بوده است. نوشته حاضر تلاش دارد بحث را به‌سوی مواجهه عملی‌تر سوق دهد. تدوین قواعد زمین بازی AI نیاز به زمان دارد تا ملاحظات عمیق و آگاهانه از اولویت‌ها و دیدگاه‌های اجتماعی و اقتصادی، در آن گنجانده شود و با امکانات و الزامات فناوری همگام گردد. در ادامه پنج حوزه‌ی کلیدی حکمرانی AI و پیشنهادها در این زمینه ارائه می‌شود.

۴-۳-۱- حوزه‌های کلیدی برای وضوح‌بخشی و اقدامات پیشنهادی

- استانداردهای تفسیرپذیری
 - مجموعه‌ای از بهترین توضیحات از اقدامات به همراه حاشیه‌ای بر ویژگی‌های ستودنی آن‌ها تدوین کنید تا الهام بخشی عملی صورت گیرد.
 - راهنمایی‌هایی برای موارد استفاده‌ی احتمالی ارائه کنید تا صنعت بتواند تصمیم بگیرد که چگونه بین منافع استفاده از سامانه‌های AI پیچیده و محدودیت‌هایی که استانداردهای مختلف تفسیرپذیری تحمیل می‌کند، تعادل برقرار کند.
 - استانداردهای حداقلی قابل قبول در بخش‌های مختلف صنعت و زمینه‌های کاربرد را توصیف کنید.
- ارزیابی انصاف
 - چارچوب‌هایی را برای ایجاد تعادل بین اهداف متضاد و همچنین تعریف انصاف، تدوین کنید.
 - عامل‌های متضاد را در برخی از موقعیت‌های فرضی اولویت‌بندی کنید، حتی اگر این در فرهنگ‌ها و مناطق جغرافیایی مختلف با هم متفاوت باشد.
- ملاحظات ایمنی
 - روندهای کاری پایه^۲ و استانداردهای ارائه‌ی مدارک را برای کاربردهای خاص بیان کنید تا نشان دهد تلاش مناسب و کافی در انجام بازرسی ایمنی انجام شده است.
 - گواهی‌های ایمنی را امتیازبندی کنید تا نشان دهد یک خدمت، آزمون‌های خاصی را که برای کاربردهای حساس نیاز است گذرانده است.

¹ Trade-Offs

² Basic Workflows

- همکاری انسانی و AI
 - زمینه‌هایی را مشخص کنید که در آنها تصمیم‌گیری نباید تماماً به‌طور خودکار با سامانه‌ی هوشمند انجام شود، بلکه نیاز است عامل انسانی در چرخه‌ی تصمیم‌گیری حضور داشته باشد.
 - رویکردهای مختلف به توانمندسازی داوری و نظارت انسان بر سامانه‌های AI را ارزیابی کنید.
- چارچوب‌های التزام کلی
 - ضعف‌های احتمالی قواعد فعلی التزام حقوقی را ارزیابی کنید و قوانین مکمل برای کاربردهای با ریسک بالا تدوین کنید.
 - چارچوب‌های کاری ایمن و کلاهک‌های التزام حقوقی^۱ را بازبینی کنید تا در زمینه‌هایی که نگرانی وجود دارد این قوانین تأثیر مخرب در نوآوری مفید برای اجتماع نداشته باشد.
 - بدیل‌های بیمه برای فضاهایی که در آن قواعد التزام حقوقی ناکافی یا غیرعملی وجود دارد، در نظر بگیرید.

۴-۴ - نقش آفرینی گوگل در به‌اشتراک‌گذاری مسئولانه‌ی داده‌های باز^۲

داده‌های دسترس‌پذیر می‌توانند به مردم و سازمان‌ها کمک زیادی کنند. می‌توان با بازگذاشتن دسترسی به پایگاه‌های داده، فراهم کردن دسترسی به API^۳ها، داده‌های محصول تجمیع‌شده^۴ و ابزارهای توسعه‌دهنده^۵، در این کار مشارکت کرد. می‌توان همچنین در بازگذاشتن دسترسی به داده، سبک‌سنگین کرد؛ یعنی بین امور حساسیت‌برانگیز و منافع بالقوه‌ی اشتراک‌گذاری داده، توازن ایجاد کرد.

۲ مارس ۲۰۱۹ به‌عنوان روز داده‌های باز نام‌گذاری شده که تجلیلی است برای در اختیار همگان قراردادن داده‌ها. جوامعی در سراسر جهان رویدادهایی برگزار کردند و در گوگل نیز برنامه‌ای برگزار شد تا دیدگاه‌های مختلف درباره اهمیت داده‌های باز مطرح شود. داده‌های دسترس‌پذیر می‌تواند آدم‌ها و سازمان‌ها را کمک زیادی کند. در ادامه در این رابطه توضیح داده می‌شود.

^۱ Liability Caps

منظور از کلاهک التزام حقوقی قراردادی است که مقداری که شخص ممکن است به جهت اهمال یا نقض عهد مورد پیگیری حقوقی واقع شود را محدود می‌کند.

^۲ Open Data

^۳ Application Programming Interface

^۴ Aggregated Product Data

^۵ Developing Tools

۴-۴-۱ - بازکردن مسئولانه‌ی پایگاه‌های داده

همین‌طور که افراد بیشتری از ML در بسترهایی باز مانند TensorFlow بهره می‌برند، به اشتراک‌گذاری پایگاه‌های داده بیش‌ازپیش اهمیت می‌یابد. گوگل بیش از ۵۰ پایگاه‌های داده‌ی باز برای توسعه‌دهندگان و محققان، در دسترس قرار داده است. این‌ها شامل موارد زیر هستند:

- [YouTube 8M](#)
- [HDR+ Burst Photography dataset](#)
- [Pixel features like Portrait Mode](#)
- [Open Images](#)
- [Open Images Extended](#)

صرف این‌که داده‌ها باز هستند به این معنا نیست که سودمندند. اولاً یک پایگاه داده باید سالم باشد (حاوی داده‌های مختل نباشد). تا درک مبتنی بر آن براساس مثال‌های دقیق و ساختاریافته باشد. سالم‌سازی پایگاه‌های بزرگ داده، کار کوچکی نیست. قبل از بازکردن داده‌ها، ساعت‌ها وقت گذاشته می‌شود تا داده‌ها استانداردسازی شوند و کیفیت آن‌ها تأیید شود. ثانیاً پایگاه داده باید در فرمت قابل‌خواندن برای ماشین به اشتراک گذاشته شود (مثلاً JSON و نه PDF) تا برای دیگران نیز به راحتی قابل استفاده باشد. نهایتاً بررسی می‌شود که آیا پایگاه داده برای محتوای موردنظر نمایانگر است یا نه. حتی اگر داده سودمند و نمایانگر برخی موقعیت‌ها باشد، ممکن است برای هر کاربردی مناسب نباشد. برای مثال، اگر یک پایگاه داده، بیشتر شامل عکس حیوانات شمال آمریکا باشد، می‌تواند در طبقه‌بندی یک گوزن به کار بیاید و نه یک زرافه. ابزارهایی چون [Facets](#) می‌توانند به شما کمک کنند تا ساختار یک پایگاه داده را تحلیل کنید و بهترین راه‌ها برای استفاده از آن را ارزیابی کنید. همچنین از طریق واسطه‌هایی مانند [Crowdsourcing application](#) می‌توان پایگاه‌های داده‌ای که نمایانگرترند ساخت. برای راهنمایی استفاده‌ی دیگران از پایگاه داده‌ی شما، یک کارت منتشر کنید که مالکیت، ترکیب و موارد استفاده پیشنهادی را بیان کند. (این یک مثال از انتشار [Open Images Extended](#) است.)

۴-۴-۲ - یافتنی و سودمندکردن داده‌ها

این کافی نیست که صرفاً داده‌ای خوب را در دسترس قرار دهیم، بلکه لازم است آن را پیداشدنی^۱ کنیم. [Kaggle](#) اجتماعی از دانشمندان داده و علاقه‌مندان به ML است که کمک می‌کند تا کاربران پایگاه‌های داده‌ی بزرگ را ذخیره و در آن کندوکاو کنند. اما پژوهشگران، توسعه‌دهندگان، روزنامه‌نگاران و دیگرانی که پیگیر داده^۲ اند اغلب تلاش دارند داده‌ها را به‌طور پخش‌شده در هزاران محل مختلف قرار دهند. ابزار جستجوی پایگاه داده کمک می‌کند تا حدی داده طوری توصیف شده باشد که موتورهای جستجو بتوانند آن را پیدا کنند و منابع داده‌ها را در هر هاستی که هست بیابند.

چیزی که داده را سودمند می‌کند نحوه‌ی تحلیل شدنش است. اگرچه امروزه تعداد بیشتری داده‌ی باز در اختیار است، دانشمندان علم داده وقت زیادی صرف می‌کنند تا داده‌ها را از منابع چندگانه تحلیل کنند. برای

^۱ Findable

^۲ Data-Seeker

رفع این مشکل، Data Commons راه‌اندازی شده است. گراف دانشی^۱ از منابع داده‌ای است که به کاربران این امکان را می‌دهد که پایگاه‌های داده‌ی مطلوبشان را -فارغ از اینکه چه منبع و فرمتی دارند- طوری استفاده کنند که گویا همه‌ی آن‌ها از یک پایگاه داده محلی واحد اخذ شده‌اند. هر کس می‌تواند در پایگاه‌های داده مشارکت داشته باشد یا با استفاده از این زیرساخت، برنامه‌هایی بسازد. کسانی که از این پلتفرم استفاده می‌کنند زمان کمتری را صرف مهندسی کردن داده‌ها صرف می‌کنند و بیشتر وقتشان را صرف تولید بینش^۲ می‌کنند. یک استفاده جالب از Data Commons این است: در یک درس علم داده که توسط Josh Hug و Fernando Perez تدریس شد، دانش‌آموزان از داده Census، CDC و Bureau of Labor Statistics استفاده کردند تا سطح چاقی در سطح شهرهای آمریکا را به دیگر عامل‌های اقتصادی و سلامتی ربط دهند. به‌طور معمول، این تحلیل روزها یا هفته‌ها به طول می‌انجامد. اما دانش‌آموزان با استفاده از Data Commons قادر شدند تا مدل‌های با کیفیت بالا را در کمتر از یک ساعت بسازند. گوگل اظهار امیدواری کرد که در صورت علاقه‌مند بودن آموزگاران و دانش‌آموزان، با آن‌ها در datacommons.org همکاری کند.

۴-۳-۴- ایجاد تعادل در سبک‌سنگین کردن‌ها

در دسترس قرار دادن داده‌ها با سبک‌سنگین کردن‌هایی همراه است. در این راستا گوگل درصدد است برای منافع بالقوه‌ی به‌اشتراک‌گذاری داده و مضرات امور حساسیت‌برانگیز آن، دست به سبک‌سنگین کردن بزند. یک ملاحظه این است که داده‌های باز لازمی پیشرفت سامانه‌های AI است و ملاحظه‌ی دیگر اینکه بازبودن گسترده‌ی داده می‌تواند موجب سرعت‌بخشی به استفاده‌ی شود که الزاماً با اصول اخلاق AI گوگل هم‌راستا نیست. در همین راستا، گوگل اخیراً دسترسی به داده‌های گفتاری ساختگی را صرفاً برای پژوهشگرانی که در چالش ASVspooof سال ۲۰۱۹ شرکت کردند^۳ باز گذاشته است تا مطمئن باشد که داده می‌تواند در جهت توسعه‌ی ابزارهای حذف دیپ‌فیک^۴ (صوت و ویدئوهای ساختگی) به کار آید و درعین حال سوءاستفاده از آن محدود باشد.

باز گذاشتن بیش‌ازحد داده همچنین می‌تواند کاربر یا اطلاعات مالک را در معرض خطر قرار دهد، موجب خدشه‌دار شدن حریم خصوصی شود یا امنیت پلتفرم را تهدید کند. گوگل به گروه‌های توسعه‌دهنده‌ی دیگر اجازه می‌دهد که روی سرویس‌هایی چون Maps و Gmail و بیشتر از طریق API‌ها محصولاتشان را بسازند اما داده‌های کاربران امن بماند. آن‌ها همچنین داده‌های تجمیع‌شده‌ی محصولات را در جایی چون Search

^۱ Knowledge Graph

^۲ Insight

^۳ چالش ASVspooof چالشی برای تمییز داده‌های ساختگی (مثل صدایی که با کامپیوتر شبیه‌سازی شده) از داده‌های واقعی (مثل صدای واقعی ضبط شده از یک شخص) بود.

^۴ Deepfakes

Trends منتشر می‌کنند تا اطلاعات علایق عمومی، در حالی که حریم خصوصی شخصی حفظ می‌شود، به اشتراک گذاشته شود.

در حالی که استفاده از داده‌های حساسیت‌برانگیز مثلاً برای پیش‌بینی شرایط یا اتفاقات پزشکی مفید است، بسیار مهم است که ایمن‌بان‌هایی تعبیه شود تا حین آموزش مدل‌های ML، حریم خصوصی فردی از بین نرود. انجام تحقیق و پژوهش‌های مختلف، مسیرهای جدید و نویدبخشی برای یادگیری از داده‌های حساسیت‌برانگیز فراهم آورده است. یک مورد یادگیری متحد (Federated Learning) است که تکنیکی است برای آموزش مدل‌های ML جامع^۱ بدون اینکه حتی داده از دستگاه شخص خارج شود که این مورد اخیراً منبع باز شد: TensorFlow Federated. مورد دیگر، حریم خصوصی افتراقی (Differential Privacy) است که می‌تواند تضمین محکمی بدهد که جزئیات داده‌های آموزش به‌طور نامناسب در معرض مدل‌های ML قرار نمی‌گیرد. به‌علاوه، پژوهشگران استفاده از پایگاه‌های داده‌ی آموزش کوچک را بدون آشکارسازی نمونه^۲ تجربه می‌کنند. برای توضیح بیشتر این موضوع در تحقیق ردیابی سرطان پروستات و یا ترجمه‌ی گوگل به لینک‌های مرتبط مراجعه کنید.

در نهایت نویسنده‌ی این مقاله اظهار امیدواری می‌کند که این تلاش‌ها کمک کند تا دسترسی به داده‌ی باز سالم سودمند و حافظ حریم خصوصی پیشرفت کند. او همچنین سازمان‌های دیگر را به مشارکت در این راستا دعوت می‌کند. آن‌ها می‌توانند با بازکردن پایگاه‌های داده‌شان از این مزایا بهره‌مند شوند: تسریع در قابل‌استفاده‌کردن داده‌ها با سالم‌سازی آن‌ها پیش از انتشار، استفاده از استانداردهای فراداده‌ی schema.org برای افزایش قابلیت پیداشدن، ارتقای شفافیت از طریق کارت‌های داده^۳ و سبک‌سنگین کردن بین حریم خصوصی کاربر و امکان سوءاستفاده. در پایان نویسنده مقاله را با این جمله خاتمه می‌دهد: این پیام تقدیم به همه‌ی کسانی که در روز داده‌ی باز دور هم جمع شدند: منتظریم ببینیم چه می‌سازید.

فصل ۵- تحلیل و ارزیابی

۵-۱- مقدمه

در این فصل، فصول پیشین به‌عنوان یک کل، با دیدی تحلیلی، بررسی و ارزیابی می‌شود تا اولاً فهم بهتری از مطالب ارائه‌شده به‌دست آید و ثانیاً جنبه‌های مثبت و منفی آن تا حدودی روشن شود. در ادامه ابتدا خلاصه

^۱ Global

^۲ Zero-Shot

^۳ Data Cards

ای از گزارش‌های مختلف گوگل بیان و پیشنهادهایی برای الگوبرداری از آن‌ها در کشورمان ارائه می‌شود. در بندهای دیگر نیز مبانی و ابعاد مختلف گزارش ارائه‌شده در فصول پیشین، تحلیل و ارزیابی می‌شود.

۵-۲- خلاصه گزارش؛ الگوبرداری

گزارش‌های مختلفی که واحد AI گوگل درباره بحث‌های هنجاری ارائه کرده است، شامل هفت باید و چهار نباید اخلاقی می‌شود که بر اساس چهار ارزش انصاف، تفسیرپذیری، حریم خصوصی و امنیت ارائه شده‌اند. در هرکدام از این ارزش‌ها کنش‌های انضمامی نیز پیشنهاد شده است تا معلوم شود در هر حوزه باید دقیقاً دست به چه اقداماتی زد. در سطح سیاست‌گذاری نیز توصیه‌ی گوگل به سیاست‌گذاران و حکومت‌ها در حوزه‌ی AI بیان می‌شود و نهایتاً موضع گوگل در به‌اشتراک‌گذاری مسئولانه داده‌های باز تبیین می‌گردد. گوگل دو گزارش را در فاصله شش‌ماهه در خصوص عملکرد خود در حوزه بررسی و تدوین بایدها، نبایدها و ارزش‌های اخلاقی مربوط به شرکت خود، ارائه می‌کند.

با الگوبرداری از ساختار گزارش‌هایی که گوگل ارائه کرده است می‌توان چند پیشنهاد برای توسعه‌ی AI در کشور داشت:

- لازم است نهادهای سیاست‌گذار در حوزه AI وارد شوند و سند پیشرفت در این حوزه را طوری تدوین کنند که اصول اخلاقی پایه، بایدها و نبایدها و نهایتاً اقدامات عملیاتی برای تحقق آن اصول و ارزش‌ها به‌نحو روشنی بیان شود. غفلت از ابعاد اخلاقی AI در تدوین سند پیشرفت AI و توجه صرف به ابعاد فنی می‌تواند نتایج جبران‌ناپذیری از نظر اخلاقی و اجتماعی به بار بیاورد.
- کار دیگری که نهادهای حاکمیتی در ایران باید انجام دهند تدوین استانداردهای اخلاقی فناوری است که هر فناوری AI لازم باشد آن استاندارد را دریافت کند. این استاندارد می‌تواند در شاخص‌های مختلفی چون حریم خصوصی، انصاف، امنیت و... محصول فناورانه را ارزیابی کند.
- شرکت‌هایی که در حوزه‌ی AI فعالیت می‌کنند ملزم به تدوین اصول و چارچوب هنجاری روشن شوند. این اصول و چارچوب هم مبنای اخلاقی و سیاستی این شرکت‌ها را روشن می‌کند هم تعامل آن‌ها با نهادهای دولتی و حکومتی را وضوح می‌بخشد. این اصول و چارچوب می‌تواند در عین تبعیت از ساختاری واحد که قوانین کشور الزام‌بخشی می‌کند، متناسب با زمینه‌های کاری شرکت‌های مختلف، از تنوع برخوردار باشد.
- یکی از دلایل موفقیت گوگل را می‌توان توجه به حوزه عمومی در سیاست‌گذاری دانست. این امر خود را در آگاهی‌سازی عمومی توسط برنامه‌های ترویجی مانند کنفرانس و سخنرانی‌های عمومی، درمیان‌گذاشتن دغدغه‌های اخلاقی با فضای فکری جامعه، استفاده از متخصصان علوم انسانی و اخلاق برای تدوین نقشه‌راه پیشرفت AI و غیره نشان می‌دهد. این می‌تواند به‌عنوان نکته‌ای مثبت در تدوین سیاست‌های توسعه در حوزه AI در کشور ما به کار آید: این سیاست‌ها باید در نسبت با جامعه‌ی وسیع مردم و همچنین در مشورت با متخصصانی از رشته‌های فنی مانند کامپیوتر و برق و رشته‌های انسانی همچون حقوق، جامعه‌شناسی و فلسفه تدوین شوند. سیاستی که در نسبت با عموم مردم و همچنین در مشورت با متخصصان مختلف تدوین شده است، در عمل نیز مورد استقبال قرار می‌گیرد و ضمانت اجرایی بالاتری دارد.

- در مجموع برخی پیشنهادهای گوگل به نهادهای سیاست‌گذار و حاکمیتی می‌تواند مورد استفاده نهادهای مربوطه در کشور واقع شود. به‌طور مثال، کمک به ارتقاء فهم عمومی از AI، آگاه‌کردن واحدهای صنعتی از کارکردهای AI برای هر واحد، به‌اشتراک‌گذاری و استفاده مسئولانه از داده و فعالیت در جهت گذار نیروی کار، مواردی است که در کشور ما نیز لازم است پیگیری شود. همچنین، پرداختن به ارزش‌هایی چون انصاف و ایمنی، و تنظیم چارچوب التزام حقوقی یا تلاش برای ایجاد توازن در استفاده از نیروی انسانی و ماشینی در انجام وظایف به‌خصوص، مواردی است که به زمینه‌ی خاص بستگی ندارد و خوب است آن‌ها را در کشورمان عملی کنیم.
- لازم است همچنین به تدوین اصول اخلاقی و سیاست‌هایی دست زد که متفاوت با اصول اخلاقی و سیاست‌های پیشنهادی گوگل است. این امر از دو جهت حائز اهمیت است. اول این‌که بافتار فرهنگی ما می‌تواند ارزش‌ها و اولویت‌های متفاوتی را برای توسعه اخلاقی AI پیشنهاد کند. دوم این‌که در مواردی که گوگل فایده‌ی AI یا شرکت خود را در اولویت قرار داده است می‌توان با دیدی وسیع‌تر، ارزش‌های مهم‌تری را در اولویت قرار داد. به‌طور مثال، در توصیه هفتم از بخش «سیاست‌گذاران چگونه می‌توانند کمک کنند» که مربوط به حوزه کاربری‌کاری می‌شود، توصیه‌های گوگل بیشتر ناظر به حمایت از AI و شرکت‌های فعال در این زمینه است تا بتوانند بدون دغدغه نیروی کار خود را جذب کنند، درحالی‌که سیاست‌گذاری در کشوری دیگر با شرایط متفاوت کاری، می‌تواند سیاست‌های متفاوتی را اقتضا کند. مثال دیگر در مورد تشویق گوگل به داده‌های باز است. گوگل در راستای تملک بیشتر داده‌ها، مدل‌ها و الگوریتم‌ها، پلتفرم‌هایی طراحی می‌کند که در آن، داده‌ها و مدل‌های مختلف قرار گیرند. این پلتفرم‌ها نهایتاً به پیشرفت فناوری و در اختیار عموم قرار گرفتن آن می‌انجامد، اما نکته اینجاست که پیشرفت این فناوری‌ها به افزایش قدرت گوگل و تبدیل شدن این شرکت به قدرت واحد (مونوپولی) می‌انجامد. تقویت نیروهای مردمی و اقدام در جهت مردمی‌سازی قدرت شرکت‌های AI در تضاد با محور قرار گرفتن گوگل است. دولت‌ها و حکومت‌های محلی وظیفه دارند برای حفظ حقوق شهروندان خود، در جهت دموکراتیک کردن قدرت شرکت‌هایی چون گوگل برنامه‌ریزی کنند.

۵-۳- اراده‌گرایی یا جبرگرایی

در مباحث سنتی فلسفه فناوری بحثی در این باره انجام می‌شود که آیا توسعه‌ی فناوری جبری است یا توسط اراده‌ی انسانی هدایت می‌شود؟ در این بحث، به نظر می‌رسد در گزارش‌های گوگل رویکرد میانه‌ای اتخاذ می‌شود. از سویی، زمینه‌هایی که توسعه‌ی تکنیکی AI فراهم می‌آورد، امری است که تا حدودی خارج از اراده انسانی است و به ملاحظات فنی مثل توسعه‌ی سخت‌افزارها یا الگوریتم‌های AI بستگی دارد از سویی دیگر می‌توان تأثیر اراده انسانی را در زمینه‌هایی که AI در آن توسعه می‌یابد پیگیری کرد. در بخشی از فصل چهارم چنین آمد که:

گرچه محققان AI می‌توانند زمینه‌ای به لحاظ تکنیکی ایجاد کنند، اما تأثیرات AI به تمایلاتی که صنعت و جامعه از خود نشان می‌دهند و همچنین به رهنمودها و حدود مرزهایی که حکومت‌ها برای کاربرد آن قرار می‌دهند بستگی دارد. به همین خاطر سیاست‌گذاران در این جهت از اهمیتی بسزا برخوردارند تا چشم‌انداز را مشخص کنند و چارچوبی بنا گذارند که پیشرفت AI را پشتیبانی کند.

توصیه‌ی گوگل چنین است که AI در زمینه‌هایی چون پزشکی، سرعت‌بخشی به اکتشافات علمی و فهم زبان که برای جامعه نفع دارند، توسعه یابد. ضمن این‌که طبق مطالب بیان‌شده، ML می‌تواند طوری اتفاق بیفتد که حریم خصوصی افراد آسیب نبیند، یا طوری تنظیم شود که عاری از سوگیری غیرمنصفانه باشد. این موارد نشان از این دارند که توسعه‌ی AI می‌تواند (و باید) متناسب با اراده‌ی خیر باشد. در همین راستا است که گوگل به توسعه‌دهندگان، سیاست‌گذاران و دیگر صنایع توصیه می‌کند که چارچوبی مناسب برای پیشرفت AI بنا گذارند. نتیجه این‌که در بحث جبرگرایی-اراده‌گرایی، گزارش حاضر رویکردی میانه اتخاذ می‌کند و زمینه‌های اعمال اراده انسانی را برجسته می‌کند.

۴-۵ - ذات‌گرایی یا ابزارگرایی

بحث دیگر در فضای فلسفه فناوری سنتی به این مربوط می‌شود که آیا فناوری ابزاری است که بتوان از آن به‌انحاء مختلف استفاده کرد یا فناوری، ذاتی به‌خصوص دارد که فارغ از نحوه‌ی استفاده از آن، مؤثر است. در این‌باره، بخش‌های عمده‌ای از گزارش گوگل به سمت ابزارگرایی تمایل دارد گویا AI ابزاری است که می‌تواند هم فایده هم ضرر داشته باشد اما باید آن را در جهتی خوب هدایت کرد تا به‌مثابه ابزاری مفید مورد استفاده قرار گیرد. به‌طور مثال، چنین بیان می‌شود که باید AI را در زمینه‌هایی که منفعت عمومی دارند، مثلاً در حوزه علم یا درمان، به کار برد و در حوزه‌های مخرب همچون ساخت سلاح از آن استفاده نکرد. البته در برخی از مباحث ایراد شده، خصوصاً در جایی که بحث از ML است، گویی نوعی ذات یا هویت می‌تواند به آنچه از یادگیری (ترینینگ) با داده به‌دست می‌آید نسبت داد. مثلاً می‌توان مدل یادگیری را منصفانه و دیگری را دارای سوگیری دانست. بر این اساس، چنین توصیه می‌شود که داده‌ها باید طوری متنوع شوند و همچنین مدل یادگیری باید با فرآیندی حاصل آید که نتیجه‌ای منصفانه به بار بیاورد. خلاصه این‌که، در بحث ذات‌گرایی-ابزارگرایی در اغلب موارد نگاهی ابزارگرایانه بر گزارش‌های گوگل حاکم است، گرچه در مباحثی که حول ML می‌شود نوعی ذات‌گرایی نیز می‌توان ردیابی کرد.

۵-۵ - نگاه آرمان‌شهرانه یا ویران‌شهرانه

در بحث‌های اجتماعی و اخلاقی AI معمولاً دو نوع نگاه آرمان‌شهرانه یا ویران‌شهرانه وجود دارد. طبق نگاه آرمان‌شهرانه، AI برای بشریت رفاه، سلامتی و ثروت به بار می‌آورد، اوقات فراغت انسان‌ها را گسترش می‌دهد و از طریق به‌خدمت‌درآوردن ماشین‌ها، قدرت انسان را در برآورده کردن نیازهایش گسترش می‌دهد. در مقابل، نگاه ویران‌شهرانه، AI را زمینه‌ساز دورانی می‌داند که ماشین‌ها بر انسان‌ها مسلط می‌شوند تا انسان را برده خود کنند و حیات او را به خطر بیندازد. طبق این نگاه، از عواقب گریزناپذیر AI، بی‌کاری گسترده و ایجاد ناعدالتی شدید میان صاحبان فناوری AI و دیگر مردم است.

هیچ‌یک از این دو نگاه افراطی خوش‌بینانه یا بدبینانه بر گزارش‌های اخلاقی و سیاست‌گذارانه‌ی گوگل حاکم نیستند. در بخشی از فصل چهارم، چنین بیان می‌شود که «هم‌اکنون، اطمینان عمومی درباره AI توسط

تصاویر علمی-تخیلی مخدوش شده است .. و باعث شده که خطرات مهم‌تر ولی ناواضحی چون مخاطره سوگیری و استفاده بدخواهانه از AI کمتر مورد توجه واقع شود» در این راستا به سیاست‌گذاران چنین توصیه می‌شود که «یک تصویر متوازن و مبتنی بر واقعیت از فرصت‌ها و چالش‌های AI ارائه کنند». این جملات و فضای کلی گزارش‌های گوگل را می‌توان نمایانگر نگاهی دانست که توصیه‌های اخلاقی و سیاست‌گذارانه را وابسته به واقعیت در دست و نه گمانه‌زنی‌های دوردستانه، می‌داند. روندهای گزارش شده در حوزه کار/بی‌کاری، سوگیری مدل‌های ML و استفاده در حوزه‌های خصمانه مواردی هستند که برای دانشمندان اخلاق و سیاست‌گذاران گوگل ایجاد نگرانی کرده‌اند، چون در این حوزه‌ها می‌توان به روشنی آثار سوء عدم استفاده اخلاقی از AI را ردیابی کرد. در مقابل، استفاده خوب از AI در حوزه‌هایی چون علم و درمان و یا استفاده از داده‌ها جهت آموزش (ترنینگ) بی‌آنکه به حریم خصوصی افراد آسیب رساند، مواردی هستند که با ارزش‌های اخلاقی در تناسب اند. همه‌ی این موارد مبتنی بر بازخوردهایی است که از محیط واقعی گرفته شده‌اند. از این رو می‌توان نگاه حاکم بر گزارش‌های اخلاقی و سیاست‌گذارانه‌ی گوگل را واقع‌گرایانه ارزیابی کرد.

حال سؤالی که می‌توان پرسید این است که اگر در زمینه‌ای بازخورد واقعی قابل اعتنا وجود نداشته باشد، نگاه گوگل به‌طور پیش‌فرض مثبت یا منفی است؟ برخی شواهد حاکی از آن است که نگاه مثبت بر نگاه منفی غلبه دارد و این طبیعی است که در شرکتی که در حوزه‌ی فناوری کار می‌کند نگاه مثبت نسبت به توسعه‌ی فناوری مبنا باشد. در بخشی از فصل دوم چنین آمده است که «سامانه‌های کامپیوتری یا سامانه‌های تصمیم‌گیری خودکار این ظرفیت را دارند که منصفانه‌تر و دربرگیرنده‌تر از قواعد دلخواه و بی‌مبنا یا از قضاوت‌های انسانی عمل کنند». در بخشی دیگر از همین فصل چنین می‌خوانیم: «مدل‌های ML حتی می‌توانند در جهت شناسایی برخی از سوگیری‌ها و موانع دربرگیری به کار آید. مواردی که ممکن است در انسان‌ها خودآگاه یا ناخودآگاه، وجود داشته و در طول تاریخ رشد کرده و ریشه دوانده باشد. تا بدین وسیله تغییری مثبت به ارمغان آورده شود». با توجه به این که نتایج انضمامی چنین سامانه‌هایی هنوز به بار ننشسته است، قضاوت مثبت نسبت به چنین سامانه‌های خودکاری را می‌توان، نه ناشی از بازخوردهای واقعی که ناشی از نگاه مثبت توسعه‌دهندگان AI در گوگل دانست.

در مجموع می‌توان چنین گفت که سیاست‌گذاران گوگل تلاش دارند نگاهشان نسبت به توسعه AI مبتنی بر بازخورد واقعی باشد؛ درحالی که، در حوزه‌هایی که بازخورد قابل اعتنا هنوز وجود ندارد نگاه کلی مثبت نسبت به توسعه فناوری در این شرکت وجود دارد.

۵-۶- اخلاق هنجاری: نتیجه‌گرا

نظریات مختلفی چون وظیفه‌گرایی، فضیلت‌گرایی و نتیجه‌گرایی در اخلاق هنجاری وجود دارد. هر یک از این نظریات منبع اخلاق را وابسته به امری همچون وظیفه یا قانون اخلاق، فضیلت عامل اخلاقی یا نتیجه عمل می‌داند. اگر مباحث هنجاری (اخلاقی و سیاست‌گذارانه) گوگل در حوزه AI را بررسی کنیم، به نظر می‌رسد با اخلاق نتیجه‌گرایانه هم‌خوانی بیشتری دارد. طبق نتیجه‌گرایی، اخلاقی بودن یک عمل به نتایج خیر و شری که از آن منتج می‌شود، وابسته است. در گزارش‌های گوگل نیز می‌توان این را ردیابی کرد که خوب یا بد دانستن

عملی را به خیر عمومی آن برای اجتماع وابسته می‌داند. اگر امری خیرش به‌نحو معناداری بیش از شرش باشد، کار خوبی است و باید پیگیری شود؛ وگرنه بد است و باید واگذاشته شود.

در نظریات اخلاقی نتیجه‌گرا، سبک‌سنگین کردن فایده‌ی خیر و شر عمل، اهمیت می‌یابد. در همین راستا در مباحث پیشین می‌توان عمل سبک‌سنگین کردن را به‌عنوان تکنیکی برای فهم خوبی یا بدی عمل دانست. این سبک‌سنگین کردن البته از منظر ارزش‌هایی چون انصاف (یا عدم سوگیری نسبت به کاربران مختلف)، حریم خصوصی و امنیت اتفاق می‌افتد که در حوزه AI خود را بیشتر نشان داده‌اند و برای مردم و جوامع حساسیت‌برانگیز شده‌اند.

توجه شرکتی چون گوگل که سود اقتصادی برایش اهمیت کلیدی دارد، به این ارزش‌ها و نتایج اخلاقی آن‌ها را می‌توان از این جهت ارزیابی کرد که در صورت عدم توجه به این امور، توجه و اعتماد عمومی نسبت به خدمات این شرکت از بین می‌رود و به تبع آن، سود اقتصادی شرکت نیز خدشه‌دار می‌شود. به نظر می‌رسد گوگل دریافته است که سود اقتصادی‌اش در پرداختن به نحوه‌ی از توسعه‌ی AI است که منافع آن به‌طور ملموسی از مخاطراتش بیشتر باشد.

۵-۷- مسئله‌های باز در اخلاق AI

در گزارش‌های گوگل رویکردی متواضعانه نسبت به مسئله‌های باز در اخلاق AI به چشم می‌خورد. منظور از رویکرد متواضعانه این است که اولاً، این اذعان وجود دارد که خیلی از مسائل اخلاقی در این حوزه هنوز بی‌جواب‌اند و راه‌حل مناسب برای آن‌ها مستلزم پژوهش و تأمل بیشتر است. ثانیاً، راه‌حل نهایی برای خیلی از مسئله‌ها متصور نیست، بلکه صحبت از بهترین راه‌حل فعلی در میان است. یعنی همواره بهترین راه‌حل تجویز می‌شود اما راه برای راه‌حل‌های بهتر همواره باز است. گوگل پیشنهادهای سیاست‌گذارانه‌اش را نیز نهایی نمی‌داند؛ در این راستا، در فصل چهارم این گزارش چنین بیان شد: «گوگل جواب همه سؤالات را نمی‌داند؛ بلکه بالعکس، ضروری است که ذی‌نفعان سیاستی از سراسر جهان وارد بحث شوند. همان‌طور که فناوری AI تحول می‌یابد و تجارب ما با آن رشد می‌کند، ما این انتظار را داریم که جامعه جهانی به‌عنوان یک کل به یادگیری ادامه دهد و ظرایف بیشتر نمود پیدا کنند».

یکی از مصادیق مسئله‌های باز که در متن گزارش‌های گوگل هم چندین بار به آن برخورد می‌کنیم، مسئله حریم خصوصی است که در تضاد با استفاده از داده‌های مختلف برای آموزش (ترینینگ) مدل‌های ML قرار دارد. این‌که چگونه می‌توان هم از داده‌ها استفاده کرد هم حریم خصوصی را حفظ کرد مسئله‌ای باز است که راه‌حل‌های بهتر و بهتر می‌طلبد. بهترین راه‌حل فعلی گوگل استفاده از داده‌ها در داخل دستگاه کاربر برای یادگیری است به‌طوری‌که داده‌ها به خارج از دستگاه منتقل نشوند. راه‌حل دیگر، ارائه‌ی نتایج کلی یک نمونه‌ی آماری بزرگ از کاربران و عدم ارائه‌ی داده‌های شخصی است. عملیاتی‌بودن و ابعاد اخلاقی هرکدام از این راه‌حل‌ها نیازمند پژوهش و تأمل بیشتر است.

مسئله باز دیگری که در گزارش بیان می‌شود به بحث امنیت در ML مربوط می‌شود. تهدیدهایی مانند سمی‌سازی داده‌های آموزشی، بازیابی داده‌های آموزشی حساس، دزدیدن مدل، استفاده از یک شبکه عصبی

که بتواند یک سامانه و شبکه دومی را که تلاش دارد قلب را ردیابی کند، سر کار بگذارد یا به اصطلاح احمق کند و مثال‌های خصمانه دیگر. در این رابطه در فصل دوم چنین آمد: «هم‌اکنون بهترین دفاع علیه مثال‌های خصمانه، هنوز به اندازه کافی برای استفاده در محیط‌های تولید قابل اعتماد نیستند. در این زمینه پژوهش‌های زیادی در حال انجام است. اما .. هنوز دفاع مؤثری وجود ندارد».

منابع

منابع فصل اول:

[1] <https://ai.google/responsibilities/>

[2] <https://ai.google/principles/>

منابع فصل دوم:

[3] <https://ai.google/responsibilities/>

[4] <https://ai.google/responsibilities/responsible-ai-practices/>

منابع فصل سوم:

[5] <https://www.blog.google/technology/ai/google-ai-principles-updates-six-months/>

[6] <https://www.blog.google/technology/ai/responsible-ai-principles/>

منابع فصل چهارم:

[7] <https://ai.google/static/documents/responsible-development-of-ai.pdf>

[8] <https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf>

[9] <https://www.blog.google/technology/ai/sharing-open-data/>

حوزه فضای مجازی به اندازه انقلاب اسلامی اهمیت دارد. این فضا مثل یک رودخانه پر از آب و خروشان است که می آید و دائماً هم بر آب آن افزوده و خروشان تر می شود. اگر ما بر این رودخانه تدبیر کنیم و برنامه داشته باشیم، زهکشی کنیم و هدایت کنیم این رودخانه را تا به سد بریزد، می شود فرصت. اگر رهاش کنیم و برنامه ای برای آن نداشته باشیم می شود یک تهدید.



csri.majazi.ir