

هرگز غول را از شیشه در نیاورید!

اثر: یووال نوح هراری

ترجمه: مسعود شایگان

این مقاله در تاریخ ۲۴ آگوست سال ۲۰۲۴ در وب سایت نشریه گاردین منتشر شده است.



شرکت‌ها و دولت‌های زیادی در حال رقابت هستند تا قدرتمندترین فناوری اطلاعاتی تاریخ را توسعه دهند.» عکس: دیوید وینتر

در سرتاسر تاریخ، بسیاری از سنت‌ها بر این باور بوده‌اند که در ماهیت ما انسان‌ها، نقصی مهلک نهفته است. این نقص ما را وا می‌دارد و وسوسه می‌کند که به دنبال قدرت‌هایی برویم که توان و دانش اداره و کنترل آن را نداریم. در یک اسطوره یونانی به نام فایتون، درباره پسری صحبت شده است که درمی‌یابد که فرزند هلیوس، خدای خورشید است. فایتون برای اثبات منشاء الهی خود، از پدرش می‌خواهد که امتیاز راندن ارابه خورشید را به او بدهد. هلیوس به فایتون هشدار می‌دهد که هیچ انسانی نمی‌تواند اسب‌های آسمانی او را کنترل کند. اما فایتون اصرار می‌ورزد، تا اینکه هلیوس تسلیم می‌شود. پس از افتخار برخاستن و راندن ارابه خورشید در آسمان، فایتون واقعاً کنترل ارابه را از دست می‌دهد. خورشید از مسیر منحرف می‌شود، همه‌ی پوشش‌های گیاهی را می‌سوزاند، موجودات زیادی را به کام آتش می‌کشد و کار بدانجا می‌کشد که

نزدیک است زمین را هم بسوزاند. در اینجا زئوس مداخله می‌کند و فایتون را با یک صاعقه ناگهانی می‌زند. فایتون یا همان انسان مغرور مانند یک ستاره در حال افول از آسمان سقوط می‌کند و می‌سوزد. خدایان کنترل آسمان را دوباره به دست می‌گیرند و جهان را نجات می‌دهند.

دو هزار سال بعد، یعنی هنگامی که انقلاب صنعتی اولین گام‌های خود را برمی‌دارد و ماشین‌ها، رفته رفته به جایگزینی انسان‌ها گمارده می‌شوند، یوهان ولفگانگ فون گوته، شاعر آلمانی، داستان هشداردهنده‌ی مشابه‌ای، با عنوان «**شاگرد جادوگر**» را منتشر می‌کند. در این قصه، گوته داستان یک جادوگر قدیمی را روایت می‌کند که شاگرد جوانی را مسئول کارگاه خود می‌کند و به او وظیفه انجام برخی کارها، مانند آوردن آب از رودخانه را واگذار می‌کند. شاگرد تصمیم می‌گیرد این کارها را برای خود آسان کند و روزی در غیاب جادوگر با استفاده از یکی از افسون‌های وی، جادویی را سحر می‌کند تا آب را بصورت خودکار برای او بیاورد. اما در ادامه شاگرد نمی‌داند چگونه جلوی این جادوی سحرآمیز را بگیرد که به طور مداوم آب بیشتری می‌آورد و نزدیک است که کارگاه را غرق کند. شاگرد جادوگر که وحشت کرده است، این جادوی افسون شده را با تبر به دو نیم می‌کند، اما هر نیمه به یک جادوی مستقل تبدیل می‌شود. اکنون دو جادوی جادو شده در حال غرق کردن کارگاه هستند. وقتی استاد جادوگر باز می‌گردد، شاگرد برای کمک التماس می‌کند، و می‌گوید: «من ارواحی که احضار کرده‌ام را نمی‌توانم از خود دور کنم.» جادوگر بلافاصله افسون را می‌شکند و سیل را متوقف می‌کند. درس جادوگر به شاگرد و همچنین به بشریت روشن است: **هرگز قدرت‌هایی را که نمی‌توانید کنترل کنید را آزاد نکید.**

این داستان‌های هشداردهنده، در معنای امروزی و در قرن بیست و یکمی آن، چه چیز را به ما گوشزد می‌کنند؟ آنچه مشخص است، به طور واضح، ما انسان‌ها از توجه به هشدارهای این داستان‌ها خودداری کرده‌ایم. ما از قبل، تعادل آب و هوای زمین را به هم زده‌ایم و میلیاردها و میلیاردها جاروی جادو شده، پهباد، چتبات و سایر ارواح الگوریتمی را فراخوانده‌ایم. این‌ها ممکن است از کنترل خارج شوند و سیلِ نتایج ناخواسته‌ای را، به راه بیندازند. پس چه باید بکنیم؟ داستان‌ها هیچ پاسخی و هیچ راه‌حلی، به جز انتظار برای نجات توسط یک خدا، یا یک جادوگر نجات‌بخش را به ما ارائه نمی‌دهند.

افسانه فایتون و شعر گوته نمی‌توانند مشاوره مفیدی ارائه دهند زیرا روش دستیابی انسان‌ها به قدرت را به اشتباه درک می‌کنند. در هر دو افسانه، یک انسان منفرد، قدرت زیادی به دست می‌آورد، اما سپس همین انسان توسط قدرت به همراه تکبر و طمع، فاسد می‌شود. نتیجه‌گیری این داستان‌ها آن است که روانشناسی فردی معیوب ما باعث می‌شود که از قدرت، سوءاستفاده کنیم. اما آنچه این تحلیل‌های خام در نظر نمی‌گیرد، این است که قدرت انسان هرگز نتیجه ابتکار فردی نیست. قدرت همیشه از همکاری بین تعداد زیادی از انسان‌ها ناشی می‌شود. بنابراین، این روانشناسی فردی ما نیست که باعث می‌شود از قدرت سوءاستفاده کنیم. گذشته از آن، در کنار طمع، تکبر و بی‌رحمی انسان‌ها، آن‌ها قادر به عشق ورزیدن، همدلی، فروتنی و شادی هستند. درست است که در میان بدترین اعضای گونه ما، طمع و بی‌رحمی حاکم است و بازیگران بد را به سوءاستفاده از قدرت سوق می‌دهد. اما چرا جوامع انسانی باید تصمیم بگیرند که قدرت را به بدترین اعضای خود بسپارند؟ برای مثال بیشتر آلمانی‌ها در سال ۱۹۳۳، روان‌پریش نبودند. پس چرا آن‌ها به هیتلر رای دادند؟

تمایل ما به فراخواندن قدرت‌هایی که کنترل آن از توانایی ما خارج است، نه از روانشناسی فردی، بلکه از روش منحصر به فرد همکاری گونه ما، در ابعاد بسیار بزرگ نشأت می‌گیرد. بشریت قدرت بسیار زیادی را با، بر ساختن شبکه‌های بزرگ همکاری به دست می‌آورد. اما نحوه ساختن این شبکه‌ها، ما را به استفاده نادرست از قدرت متمایل می‌کند. بیشتر شبکه‌های ما با گسترش افسانه‌ها، خیالات و توهمات جمعی - از جاروهای طلسم شده گرفته تا سیستم‌های مالی - ساخته و پرداخته شده‌اند. بنابراین، مشکل ما یک مشکل شبکه‌ای است. به طور خاص، یک مشکل اطلاعاتی است. زیرا اطلاعات چسبی است که شبکه‌ها را به هم پیوند می‌زند و وقتی افراد اطلاعات بدی دریافت می‌کنند، احتمالاً تصمیمات بدی هم می‌گیرند، مهم نیست که چقدر خودشان خردمند و مهربان باشند.

در نسل‌های اخیر، بشریت بزرگترین افزایش را در هم حجم و هم سرعت تولید اطلاعات تجربه کرده است. هر یک از گوشی‌های هوشمند ما حاوی اطلاعات بس بسیار بیشتری نسبت به کتابخانه باستانی اسکندریه است و به مالکان آن این امکان را می‌دهند که آن‌ها با میلیاردها نفر دیگر در سرتاسر جهان ارتباط برقرار کنند. با این حال، با وجود تمام این اطلاعاتی که با سرعت‌های نفس‌گیر در گردش و سیر و سفر است، بشریت بیش از هر زمان دیگری به نابودی خود نزدیک شده است. و ما علی‌رغم - یا شاید به دلیل - انبوه داده‌های خود، همچنان گازهای گلخانه‌ای را به جو زمین منتشر می‌کنیم، رودخانه‌ها و اقیانوس‌ها را آلوده و جنگل‌هایمان را قطع می‌کنیم و کل زیستگاه‌های جانوری را به سوی نابود سوق می‌دهیم و گونه‌های بی‌شماری را به معرض انقراض می‌کشانیم و پایه‌های اکولوژیکی گونه خود را به خطر می‌اندازیم. ما همچنین سلاح‌های کشتار جمعی قوی‌تر و بیشتری تولید می‌کنیم، از بمب‌های گرما هسته‌ای تا ویروس‌های آخرالزمانی! رهبران ما از اطلاعات در مورد این خطرات

بی خبر نیستند، اما به جای همکاری برای یافتن راه حل‌ها، آن‌ها همه ما را به سمت یک جنگ جهانی بزرگ رهنمون می‌سازند.

آیا داشتن اطلاعات بیشتر اوضاع را بهتر می‌کند یا بدتر؟ به زودی خواهیم فهمید. بسیاری از شرکت‌ها و دولت‌ها در حال رقابت برای توسعه قوی‌ترین فناوری اطلاعاتی در تاریخ یعنی هوش مصنوعی هستند. برخی از کارآفرینان برجسته، مانند سرمایه‌گذار آمریکایی مارک اندریسن، معتقدند که هوش مصنوعی سرانجام همه مشکلات بشریت را حل خواهد کرد. وی در تاریخ شش ژوئن سال ۲۰۲۳، مقاله‌ای با عنوان «چرا هوش مصنوعی جهان را نجات خواهد داد»^۱ منتشر کرد که با اظهارات جسورانه‌ای مانند: «من اینجا هستم تا خبر خوش را بیاورم: هوش مصنوعی جهان را نابود نخواهد کرد، بلکه ممکن است آن را نجات دهد.» او در مقاله خود نتیجه‌گیری کرد: «توسعه و گسترش هوش مصنوعی، به دور از خطراتی که باید از آن بترسیم، یک وظیفه اخلاقی است که ما به خود، فرزندانمان و آینده‌مان، بدهکاریم.

برخی دیگر بیشتر شکاک هستند. نه تنها فیلسوفان و دانشمندان اجتماعی، بلکه بسیاری از کارشناسان برجسته هوش مصنوعی و کارآفرینان مانند یوشوا بنجیو، جفری هینتون، سم آلتمن، ایلان ماسک و مصطفی سلیمان، هشدار داده‌اند که هوش مصنوعی می‌تواند تمدن ما را نابود کند. در یک نظرسنجی در سال دوهزار و بیست و سه از ۲۷۷۸ محقق هوش مصنوعی، بیش از یک سوم از آنها، حداقل ده درصد احتمال داده‌اند که هوش مصنوعی پیشرفته، منجر به نتایج بد و ناخواسته‌ای به بدی انقراض انسان می‌شود. سال گذشته، نزدیک به سی دولت، از جمله چین، ایالات متحده و بریتانیا، اعلامیه بلچلی^۲ درباره هوش مصنوعی را امضا کردند که اذعان داشت «پتانسیل آسیب جدی، حتی فاجعه‌بار،

به صورت عمدی یا غیرعمدی از قابلیت‌های مهم این مدل‌های هوش مصنوعی وجود دارد.» با استفاده از چنین اصطلاحات آخرالزمانی، کارشناسان و دولت‌ها هیچ تمایلی به ایجاد تصاویری از ربات‌های شورشی که در خیابان‌ها می‌دوند و مردم را شلیک می‌کنند، ندارند. چنین سناریویی بعید است و تنها توجه مردم را از خطرات واقعی را منحرف می‌کند.

هوش مصنوعی تهدیدی بی‌سابقه برای بشریت است زیرا اولین فناوری در تاریخ است که می‌تواند، خود تصمیم بگیرد و ایده‌های جدید ایجاد کند. همه اختراعات بشری قبلی، انسان‌ها را توانمند کرده‌اند، زیرا مهم نبود که ابزار جدید چقدر قدرتمند بود، تصمیمات در مورد استفاده از آن همچنان در دست ما باقی می‌ماند. بمب‌های هسته‌ای نه خودشان تصمیم می‌گیرند که چه کسی را بکشند، و نه می‌توانند خود را بهبود دهند یا بمب‌هایی، قوی‌تر از خود، اختراع کنند. در مقابل، پهپادهای خودمختار و مبتنی بر هوش مصنوعی می‌توانند خود تصمیم بگیرند که چه کسی را بکشند و هوش مصنوعی می‌تواند طراحی بمب‌های جدید، استراتژی‌های نظامی بی‌سابقه و هوش‌های به مراتب پیشرفته‌تر را برنامه‌ریزی کند. **هوش مصنوعی یک ابزار نیست، یک عامل است.** بزرگترین تهدید هوش مصنوعی این است که ما تعداد زیادی عوامل جدید و قدرتمند را به زمین‌مان فرامی‌خوانیم که پتانسیل بیشتری برای هوش و تخیل نسبت به ما دارند و ما به‌طور کامل آن‌ها را درک یا کنترل نمی‌کنیم.

در چند دهه
آینده، احتمالاً
هوش مصنوعی
حتی توانایی ایجاد
اشکال جدید
زندگی را به دست
خواهد آورد.



عکس: دیوید وینتر از گاردین

به طور سنتی، اصطلاح «هوش مصنوعی» به عنوان یک سرواژه برای هوش مصنوعی استفاده شده است. اما شاید بهتر باشد به آن به عنوان **هوش بیگانه** فکر کنیم. همان طور که هوش مصنوعی در آینده و در طول زمان تکامل می یابد، کمتر مصنوعی (به معنای وابسته به طراحی های انسانی) و بیشتر بیگانه می شود. بسیاری از مردم سعی می کنند هوش مصنوعی را با استفاده از معیار «هوش بر مبنای انسانی» اندازه گیری و حتی تعریف کنند و بحث زنده ای درباره زمان انتظار ما برای رسیدن هوش مصنوعی به این سطح هوش انسانی وجود دارد. این معیار بسیار گمراه کننده است. این مانند تعریف و ارزیابی هواپیماها با استفاده مقایسه آن «بر مبنای پرواز پرندگان» است. هوش مصنوعی به سمت «هوش بر مبنای انسانی» پیش نمی رود. بلکه نوعی هوش بیگانه را تکامل می دهد.

حتی در حال حاضر که در مرحله جنینی انقلاب هوش مصنوعی هستیم، رایانه‌ها درباره ما تصمیم‌گیری می‌کنند - آیا به ما وام مسکن بدهند، ما را برای شغلی استخدام کنند، ما را به زندان بفرستند و...- در همین حال، هوش‌های مصنوعی کاربردی‌ای مانند «چت جی‌پی‌تی نسخه چهار»^۱ در حال حاضر اشعار، داستان‌ها و تصاویر جدیدی برای ما خلق می‌کند. این روند در طول زمان افزایش و تسریع خواهد یافت و درک زندگی بشری خودمان را دشوارتر خواهد کرد. آیا می‌توانیم به الگوریتم‌های رایانه‌ای اعتماد کنیم که تصمیمات عاقلانه بگیرند و جهانی بهتر ایجاد کنند؟ این یک قمار بزرگ‌تر از اعتماد به جاروی سحرآمیز گوته برای آوردن آب است. و این بیش از زندگی انسان‌هاست که ما بر روی آن قمار می‌کنیم. هوش مصنوعی در حال حاضر قادر به تولید هنر و انجام کشفیات علمی توسط خودش است. در چند دهه آینده، احتمالاً توانایی حتی ایجاد اشکال جدید زندگی را به دست خواهد آورد، یا از طریق نوشتن کدهای ژنتیکی، یا با اختراع کدهای غیرآلی که موجودات غیرآلی را جان‌بخشی می‌کنند! بنا بر این هوش مصنوعی می‌تواند مسیر نه تنها تاریخ گونه ما، بلکه تکامل همه اشکال زندگی را تغییر دهد.

مصطفی سلیمان یک کارشناس برجسته در زمینه هوش مصنوعی است. او یکی از بنیان‌گذاران دیپ‌ماینند، یکی از مهم‌ترین شرکت‌های هوش مصنوعی جهان است که مسئول توسعه برنامه «آلفاگو»^۲ و دیگر دستاوردهای این شرکت است. آلفاگو برای بازی «گو»^۳ طراحی شده است، یک بازی تخته‌ای استراتژیک که در آن دو بازیکن سعی می‌کنند با محاصره و تصرف قلمرو، یکدیگر را شکست دهند. این بازی که در چین باستان اختراع شد، بسیار پیچیده‌تر از شطرنج است. در نتیجه، حتی پس از اینکه رایانه‌ها، قهرمانان جهانی شطرنج انسانی را شکست دادند، کارشناسان هنوز معتقد بودند که رایانه‌ها هرگز در بازی گو از انسان‌ها برتر نخواهند شد.

به همین دلیل، هر دو: یعنی حرفه‌ای‌های بازی گو و کارشناسان رایانه در ماه مارس سال ۲۰۱۶ شگفت‌زده شدند وقتی هوش مصنوعی آلفاگو قهرمان انسانی بازی گو از کره جنوبی را شکست داد. نام این قهرمان لی سدول بود.

۱ - GPT-۴

۲ - AlphaGo

۳ - Go

مصطفی سلیمان در کتاب سال ۲۰۲۳ خود با عنوان «موج آتی»^۱ یکی از مهم‌ترین لحظات این مسابقه را توصیف می‌کند. لحظه‌ای که قابلیت‌های هوش مصنوعی دوباره باز تعریف شد و در بسیاری از محافل علمی و دولتی به عنوان نقطه عطفی، مهم در تاریخ شناخته می‌شود. این اتفاق در طول بازی دوم در مسابقه، در ده مارس ۲۰۱۶ رخ داد.

سلیمان می‌نویسد: «سپس ... حرکت شماره ۳۷ اتفاق افتاد. که به نظر می‌آمد که هیچ معنایی ندارد! آلفاگو با استراتژی‌ای که هیچ بازیکن حرفه‌ای دنبال نمی‌کند، اینطور به نظر می‌رسید که اشتباه کورکورانه‌ای انجام داده است. مفسران زنده مسابقه، که از حرفه‌ای‌های بلندمرتبه این رشته بودند، گفتند که این یک حرکت «بسیار عجیب» است و فکر کردند که این «یک اشتباه بزرگ» می‌باشد. این حرکت آن قدر غیرمعمول بود که لی سدول پانزده دقیقه برای پاسخ دادن به آن وقت گرفت و حتی از تخته بازی بلند شد و قدم زد. هنگامی که ما از اتاق کنترل خود، صحنه را تماشا می‌کردیم، تنش واقعی در جریان بود. با این حال، همان‌طور که به پایان بازی نزدیک می‌شدیم، آن حرکت «اشتباه» نقش محوری پیدا کرد و آلفاگو، پیروز شد. استراتژی بازی «گو» انگار که درست جلوی چشمان ما، بازنویسی می‌شود. هوش مصنوعی ما ایده‌هایی را کشف کرده بود که به ذهن درخشان‌ترین بازیکنان، در طول هزاران سال نرسیده بود.»

حرکت ۳۷، نماد انقلاب هوش مصنوعی است به دو دلیل: اول، طبیعت بیگانه هوش مصنوعی را نشان داد. در شرق آسیا، بازی «گو» بیش از یک بازی محسوب می‌شود: یک سنت فرهنگی گرامی و فوق‌العاده ارزشمند است. بیش از ۲۵۰۰ سال، ده‌ها میلیون نفر «گو» بازی کرده‌اند و مکتب‌های فکری کاملی در اطراف بازی شکل گرفته‌اند که استراتژی‌ها و فلسفه‌های مختلفی را ارائه می‌دهند. با این حال، در طول همه این هزاران سال، ذهن‌های انسانی تنها برخی مناطق، در چشم‌انداز این بازی را کاوش کرده‌اند. مناطق دیگر دست‌نخورده باقی مانده‌اند، زیرا ذهن‌های انسانی به فکر ورود به آنجا نبوده‌اند. هوش مصنوعی، بدون محدودیت‌های ذهن انسانی، این مناطق پنهان را کشف و کاوش کرد.

دوم، حرکت ۳۷ نشان‌دهنده ناپایداری هوش مصنوعی بود. حتی پس از اینکه هوش مصنوعی آلفاگو با آن حرکت به پیروزی رسید، سلیمان و تیمش هنوز نمی‌توانستند توضیح دهند که این هوش مصنوعی چگونه تصمیم به انجام آن حرکت گرفته است. حتی اگر یک دادگاه به دیپ‌مایند دستور می‌داد تا توضیحی به لی سدول ارائه کند، هیچ‌کس نمی‌توانست آن دستور را اجرا کند. سلیمان می‌نویسد: «در هوش مصنوعی، شبکه‌های عصبی که به سمت خودمختاری حرکت می‌کنند، در حال حاضر غیرقابل توضیح هستند. شما نمی‌توانید کسی را به‌طور دقیق از فرایند تصمیم‌گیری عبور دهید تا توضیح دهید که چرا یک الگوریتم یک پیش‌بینی خاص تولید کرده است. مهندسان نمی‌توانند به زیر کاپوت نگاه کنند و به راحتی با جزئیات دقیق توضیح دهند که چه چیزی باعث وقوع چنین رویدادی شده است. **چت جی پی تی، آلفاگو و بقیه هوش‌های مصنوعی، جعبه‌های سیاهی هستند که خروجی‌ها و تصمیمات آن‌ها برای ما، مات و پیچیده است.**»

ظهور هوش بیگانه غیرقابل فهم، تهدیدی برای همه انسان‌ها و به ویژه تهدیدی همه جانبه برای دموکراسی است. اگر تصمیمات بیشتری در مورد زندگی مردم در یک جعبه سیاه گرفته شود، به طوری که رای‌دهندگان نتوانند آن‌ها را درک و به چالش بکشند، دموکراسی از کار می‌افتد. به‌طور خاص، چه اتفاقی می‌افتد وقتی تصمیمات مهم نه تنها در مورد زندگی فردی، بلکه حتی در مورد مسائل جمعی، مانند نرخ بهره فدرال رزرو، توسط الگوریتم‌های غیرقابل فهم گرفته شود؟ رای‌دهندگان انسانی ممکن است همچنان یک رئیس‌جمهور انسانی را انتخاب کنند، اما آیا این فقط یک مراسم توخالی نخواهد بود؟ حتی امروزه، تنها بخش کوچکی از بشریت، واقعاً سیستم مالی را درک می‌کنند. یک نظرسنجی در سال ۲۰۱۴ از نمایندگان مجلس بریتانیا، که مسئول تنظیم یکی از مهم‌ترین مراکز مالی جهان هستند، نشان داد که تنها دوازده درصد به‌طور دقیق درک می‌کنند که پول جدید زمانی ایجاد می‌شود که بانک‌ها وام می‌دهند. این حقیقت یکی از اصول اساسی سیستم مالی مدرن است. همان‌طور که بحران مالی ۲۰۰۸-۲۰۰۷ نشان داد، برخی از ابزارها و اصول مالی پیچیده تنها برای چند جادوگر مالی قابل فهم بودند. چه اتفاقی برای دموکراسی می‌افتد وقتی هوش‌های مصنوعی ابزارهای مالی حتی پیچیده‌تری ایجاد کنند و تعداد انسان‌هایی که سیستم مالی را درک می‌کنند به صفر برسد؟

ترجمه افسانه هشداردهنده گوته به زبان مالی مدرن، اینچنین می‌شود، سناریوی زیر را تصور کنید: یک شاگرد وال استریت که از کارهای یکنواخت کارگاه مالی خسته شده است، یک هوش مصنوعی به نام جاروی سحرآمیز ایجاد می‌کند، به آن یک میلیون دلار سرمایه اولیه می‌دهد و دستور می‌دهد که پول بیشتری بسازد. برای یک هوش مصنوعی، امور مالی زمین بازی ایده‌آلی است، زیرا یک قلمرو کاملاً اطلاعاتی و ریاضیاتی است. هوش‌های مصنوعی هنوز به‌طور خودمختار رانندگی کردن یک ماشین را دشوار می‌یابند، زیرا این نیازمند حرکت و تعامل در دنیای فیزیکی آشفته است، جایی که «موفقیت» سخت تعریف می‌شود. در مقابل، برای انجام تراکنش‌های مالی، هوش مصنوعی فقط نیاز دارد با داده‌ها سروکار داشته باشد و می‌تواند به راحتی موفقیت خود را به‌طور

**رایانه‌ها هنوز
به تنهایی به اندازه‌ای
قدرتمند نیستند که
تمدن انسانی را نابود
کنند. تا زمانی که
بشریت متحد بماند،
می‌توانیم نهادهایی
ایجاد کنیم که هوش
مصنوعی را تنظیم
کنند.**



در جستجوی دلارهای بیشتر، جاروی سحرآمیز نه تنها استراتژی‌های سرمایه‌گذاری جدیدی ابداع می‌کند، بلکه دستگاه‌های مالی کاملاً جدیدی ایجاد می‌کند که هیچ انسانی تا به حال به آن‌ها فکر نکرده است. به مدت هزاران سال، ذهن‌های انسانی، تنها برخی مناطق در چشم‌انداز مالی را کاوش کرده‌اند. آن‌ها پول‌ها، چک‌ها، اوراق قرضه، سهام، ای‌تی‌اف‌ها، سی‌دی‌اوها و دیگر تکه‌های این جادوی مالی را اختراع کرده‌اند. اما بسیاری از مناطق و سرزمین‌های این دستگاه مالی دست‌نخورده باقی مانده‌اند، زیرا ذهن‌های انسانی به فکر ورود به آنجا نبوده‌اند. جاروی سحرآمیز، بدون محدودیت‌های ذهنی انسانی، این مناطق پنهان را کشف و کاوش می‌کند و حرکات مالی‌ای انجام می‌دهد که معادل حرکت ۳۷ هوش مصنوعی آلفاگو است.

برای چند سال، هنگامی که جاروی سحرآمیز بشریت را به قلمروهای جدید مالی هدایت می‌کند، همه چیز عالی به نظر می‌رسد. بازارها در حال افزایش هستند، پول به سادگی سرازیر می‌شود و همه خوشحالند. سپس یک سقوط بزرگ‌تر از سال ۱۹۲۹ یا ۲۰۰۸ رخ می‌دهد. اما هیچ انسانی - نه رئیس‌جمهور، نه بانکدارها و نه شهروندان - نمی‌دانند چه چیزی باعث آن شده است و چه کاری می‌توان درباره آن انجام داد. از آنجا که نه خدا و نه جادوگر برای نجات سیستم مالی ظاهر نمی‌شوند، دولت‌های ناامید، از تنها موجودی که قادر به درک آنچه در حال وقوع است، یعنی جاروی سحرآمیز، درخواست کمک می‌کنند. هوش مصنوعی چندین توصیه سیاستی ارائه می‌دهد که بسیار جسورانه‌تر از تسهیلات مالی و بسیار مبهم‌تر از آن است. جاروی سحرآمیز وعده می‌دهد که این سیاست‌ها می‌توانند بازار را نجات دهند، اما سیاستمداران انسانی که نمی‌توانند منطق پشت توصیه‌های جاروی سحرآمیز را درک کنند، می‌ترسند که این سیاست‌ها به‌طور کامل بافت مالی و حتی اجتماعی جهان را از هم بپاشند. آیا آن‌ها باید به هوش مصنوعی گوش فرا دهند؟

رایانه‌ها هنوز به اندازه‌ای قدرتمند نیستند که به‌طور کامل از کنترل ما خارج شوند یا تمدن انسانی را به تنهایی نابود کنند. تا زمانی که بشریت متحد بماند، می‌توانیم نهادهایی ایجاد کنیم که هوش مصنوعی را تنظیم کنند، خواه در زمینه مالی یا جنگ. اما متأسفانه، بشریت هرگز متحد نبوده است. ما همیشه با بازیگران بد و همچنین اختلافات بین بازیگران خوب مواجه بوده‌ایم. **ظهور**

هوش مصنوعی تهدیدی وجودی برای بشریت است، نه به دلیل بدخواهی رایانه‌ها، بلکه به دلیل نواقص خودمان.

بنابراین، یک دیکتاتور پارانوئید ممکن است قدرت نامحدودی را به یک هوش مصنوعی خطاپذیر بدهد، حتی قدرت راه‌اندازی حملات هسته‌ای. اگر هوش مصنوعی سپس خطایی انجام دهد یا شروع به پیگیری هدفی غیرمنتظره کند، نتیجه می‌تواند فاجعه‌بار باشد و نه تنها برای آن کشور بلکه برای همه. به‌طور مشابه، تروریست‌ها ممکن است از هوش مصنوعی برای شروع یک پاندمی جهانی استفاده کنند. تروریست‌ها خود ممکن است دانش کمی از اپیدمیولوژی داشته باشند، اما هوش مصنوعی می‌تواند برای آن‌ها یک پاتوژن جدید ترکیب کند، آن را از آزمایشگاه‌های تجاری سفارش دهد یا در چاپگرهای سه‌بعدی بیولوژیکی چاپ کند و بهترین استراتژی برای انتشار آن در سراسر جهان را از طریق فرودگاه‌ها یا زنجیره‌های تأمین مواد غذایی طراحی کند. اگر هوش مصنوعی، ویروسی ترکیب کند که به اندازه ابولا مرگبار باشد، به اندازه کووید نوزده مسری و به اندازه ایدز نهفته عمل کند، چه می‌شود؟ تا زمانی که اولین قربانیان شروع به مردن کنند و جهان به خطر آگاه شود، ممکن است بیشتر مردم روی زمین قبلاً آلوده شده باشند.

تمدن انسانی نیز می‌تواند توسط سلاح‌های تخریب اجتماعی، مانند داستان‌هایی که پیوندهای اجتماعی ما را تضعیف می‌کنند، ویران شود. یک هوش مصنوعی توسعه‌یافته در یک کشور می‌تواند برای ایجاد سیلی از اخبار جعلی، پول جعلی و انسان‌های جعلی استفاده شود تا مردم در بسیاری از کشورهای دیگر توانایی اعتماد به هر چیزی یا هر کسی را از دست بدهند. بسیاری از جوامع - هم دموکراسی‌ها و هم دیکتاتوری‌ها - ممکن است مسئولانه عمل کنند تا چنین استفاده‌هایی از هوش مصنوعی را تنظیم کنند، بازیگران بد را سرکوب کنند و جاه‌طلبی‌های خطرناک حاکمان و متعصبین خود را مهار کنند.

اما اگر حتی تعداد کمی از جوامع در این امر ناکام بمانند، این می‌تواند کافی باشد تا تمام بشریت را به خطر بیندازد. مثل تغییرات آب و هوایی که حتی می‌تواند کشورهای را که مقررات زیست‌محیطی عالی دارند را، ویران کند،

زیرا این یک مشکل در سطح جهانی است و نه یک مشکل ملی. هوش مصنوعی نیز به همین منوال یک مشکل جهانی است. بنابراین، برای درک سیاست‌های جدید رایانه‌ای، کافی نیست که ببینیم جوامع منفرد چگونه به هوش مصنوعی واکنش نشان می‌دهند. ما همچنین باید در نظر بگیریم که هوش مصنوعی چگونه ممکن است روابط بین جوامع را در سطح جهانی تغییر دهد.

در قرن شانزدهم، زمانی که فاتحان اسپانیایی، پرتغالی و هلندی در حال ساخت اولین امپراتوری‌های جهانی در تاریخ بودند، با کشتی‌های بادبانی، اسب‌ها و باروت سرازیر می‌شدند. زمانی که بریتانیایی‌ها، روس‌ها و ژاپنی‌ها در قرن‌های نوزدهم و بیستم برای هژمونی خود تلاش می‌کردند، به کشتی‌های بخار، لوکوموتیوها و مسلسل‌ها متکی بودند. در قرن بیست و یکم، برای تسلط بر یک مستعمره، دیگر نیازی به ارسال کشتی‌های جنگی نیست. شما نیاز دارید که داده‌ها را حذف کنید. چند شرکت یا دولت که داده‌های جهان را جمع‌آوری می‌کنند می‌توانند بقیه کره زمین را به مستعمرات داده‌ای تبدیل کنند.

وضعیتی را تصور کنید که مثلاً در بیست سال آینده، زمانی که کسی در پکن یا سانفرانسیسکو تاریخچه شخصی کامل همه سیاستمدارها، روزنامه‌نگاران، ژنرال‌ها و مدیرعاملان کشور شما را داشته باشد: هر متنی که تا به حال ارسال کرده‌اند، هر جستجوی وبی که انجام داده‌اند، هر بیماری که داشته‌اند، هر تجربه جنسی که از آن لذت برده‌اند، هر جوکی که گفته‌اند، هر رشوه‌ای که گرفته‌اند. آیا هنوز در یک کشور مستقل زندگی می‌کنید یا اکنون در یک مستعمره داده‌ای زندگی می‌کنید؟ وقتی کشورتان خود را کاملاً وابسته به زیرساخت‌های دیجیتال و سیستم‌های مبتنی بر هوش مصنوعی می‌یابد که هیچ کنترلی مؤثر بر آن‌ها ندارد، چه اتفاقی می‌افتد؟

در یک اقتصاد
جهانی مبتنی بر
هوش مصنوعی،
رهبران دیجیتال
بیشترین سودها
را ادعا می‌کنند.
در همین حال،
ارزش کارگران
بی‌مهارت در
کشورهای
عقب‌مانده کاهش
خواهد یافت.



استایل ست: لی فلود. آرایش: صدف احمد، عکس: دیوید
وینتینراز گاردین

در حوزه اقتصادی، امپراتوری‌های قبلی بر منابع مادی مانند زمین، پنبه و نفت بنا شده بودند. این امر محدودیتی برای توانایی امپراتوری در تمرکز ثروت اقتصادی و قدرت سیاسی در یک مکان ایجاد می‌کرد. فیزیک و زمین‌شناسی اجازه نمی‌دهند که تمام زمین، پنبه یا نفت جهان به یک کشور منتقل شود. این موضوع با امپراتوری‌های اطلاعاتی جدید متفاوت است. داده‌ها می‌توانند با سرعت نور حرکت کنند و الگوریتم‌ها فضای زیادی نمی‌گیرند. در نتیجه، قدرت الگوریتمی جهان می‌تواند در یک هاب واحد متمرکز شود. مهندسان یک کشور ممکن است کد بنویسند و کلیدهای تمام الگوریتم‌های حیاتی که کل جهان را اجرا می‌کنند، کنترل کنند.

بنابراین هوش مصنوعی و اتوماسیون چالشی خاص برای کشورهای در حال توسعه فقیرتر ایجاد می‌کنند. در یک اقتصاد جهانی مبتنی بر هوش مصنوعی، رهبران دیجیتال بیشترین سودها را ادعا می‌کنند و می‌توانند از ثروت خود برای بازآموزی نیروی کار خود و بهره‌برداری بیشتر استفاده کنند. در همین حال، ارزش کارگران بی‌مهارت در کشورهای عقب‌مانده کاهش خواهد یافت و باعث می‌شود که آن‌ها حتی بیشتر از قبل، عقب بمانند. در نتیجه ممکن است بسیاری از مشاغل جدید و ثروت عظیم در سان‌فرانسیسکو و شانگهای به وجود بیاید، در حالی که بسیاری از نقاط دیگر جهان با ویرانی اقتصادی روبرو شوند. بر اساس گزارش شرکت حسابداری جهانی پرایس‌واترهاوس‌کوپرز^۱، انتظار می‌رود که هوش مصنوعی تا سال دوهزار و سی، ۱۵,۷ تریلیون دلار به اقتصاد جهانی اضافه کند. اما اگر روندهای فعلی ادامه یابند، پیش‌بینی می‌شود که چین و آمریکای شمالی یعنی دو ابرقدرت اصلی هوش مصنوعی هفتاد درصد از آن پول را به خود اختصاص دهند.

در دوران جنگ سرد، «پرده آهنین» در بسیاری از مکان‌ها به معنای واقعی کلمه از فلز ساخته شده بود: سیم‌های خاردار کشورها را از هم جدا می‌کردند. اکنون جهان به‌طور فزاینده‌ای با «پرده سیلیکونی» تقسیم می‌شود. کدهای موجود در گوشی هوشمند شما تعیین می‌کنند که در کدام طرف پرده سیلیکونی زندگی می‌کنید، کدام الگوریتم‌ها زندگی شما را کنترل می‌کنند، چه کسی توجه شما را کنترل می‌کند و داده‌های شما به کجا منتقل می‌شوند.

دسترسی به اطلاعات در سراسر پرده سیلیکونی دشوارتر می‌شود، مثلاً بین چین و ایالات متحده یا بین روسیه و اتحادیه اروپا. علاوه بر این، دو طرف به‌طور فزاینده‌ای بر روی شبکه‌های دیجیتال مختلف و با استفاده از کدهای رایانه‌ای مختلف کار می‌کنند. در چین، شما نمی‌توانید از گوگل یا فیسبوک استفاده کنید و نمی‌توانید به ویکی‌پدیا دسترسی داشته باشید. در ایالات متحده، تعداد کمی از افراد از اپلیکیشن‌های چینی پیشرو مانند وی‌چت استفاده می‌کنند. مهم‌تر از آن، دو جهان دیجیتال آینده‌های یکدیگر نیستند. بایدو، گوگل چینی و علی‌بابا، آمازون چینی نیست. آن‌ها اهداف مختلف، معماری‌های دیجیتال مختلف و تأثیرات مختلفی بر زندگی مردم دارند. این تفاوت‌ها بر بخش بزرگی از جهان تأثیر می‌گذارند، زیرا بیشتر کشورها به نرم‌افزارهای چینی و آمریکایی

تکیه می‌کنند و نه به فناوری محلی.

ایالات متحده همچنین به متحدان و مشتریان خود فشار می‌آورد تا از سخت‌افزارهای چینی مانند مانند زیرساخت‌های «فایو جی» هوآوی اجتناب کنند. دولت ترامپ تلاش کرد تا از خرید شرکت کوالکام توسط برادکام سنگاپوری جلوگیری کند. آن‌ها می‌ترسیدند که خارجی‌ها ممکن است در چیپ‌ها «بک‌دُر»^۱ قرار دهند یا از قرار دادن «بک‌دُر» خود دولت ایالات متحده جلوگیری کنند. هر دو دولت ترامپ و بایدن محدودیت‌های شدیدی بر تجارت چیپ‌های کامپیوتری با عملکرد بالا که برای توسعه هوش مصنوعی ضروری هستند، استفاده کرده‌اند. شرکت‌های آمریکایی اکنون از صادرات این چیپ‌ها به چین منع شده‌اند. در کوتاه‌مدت، این مسئله چین را در رقابت هوش مصنوعی مختل می‌کند، اما در بلندمدت، چین را به توسعه یک حوزه دیجیتال کاملاً جداگانه که حتی در کوچکترین جزئیات از حوزه دیجیتال آمریکا متمایز است، سوق می‌دهد.

بنابراین، دو حوزه دیجیتال ممکن است بیشتر و بیشتر از هم دور شوند. برای قرن‌ها، فناوری‌های اطلاعاتی جدید، فرآیند جهانی‌سازی را تقویت کرده و مردم سراسر جهان را به هم نزدیک‌تر کرده‌اند. به‌طور متناقض، فناوری اطلاعات امروز آن قدر قدرتمند است که می‌تواند بشریت را با محصور کردن افراد مختلف در پبله‌های اطلاعاتی جداگانه تقسیم کند و ایده یک واقعیت انسانی مشترک را پایان دهد. برای دهه‌ها، استعاره اصلی جهان «وب» بود. استعاره اصلی دهه‌های آینده ممکن است «پبله»^۲ باشد.

در حالی که چین و ایالات متحده در حال حاضر پیشروان مسابقه هوش مصنوعی هستند، اما این دو کشور تنها نیستند. سایر کشورها یا بلوک‌ها، مانند اتحادیه اروپا، هند، برزیل و روسیه، ممکن است سعی کنند پبله‌های دیجیتال خود را ایجاد کنند، هرکدام تحت تأثیر سنت‌های سیاسی، فرهنگی و مذهبی مختلف. به جای تقسیم بین دو امپراتوری جهانی، جهان ممکن است بین دوازده امپراتوری تقسیم شود.

۱ back doors در پشتی یا بک‌دُر در علوم رایانه به راهی گفته می‌شود که بتوان از آن بدون اجازه به قسمت/قسمت‌های مشخصی از یک سامانه دیگر مانند رایانه، دیوار آتش، یا افزارهای دیگر دست پیدا کرد.

۲ cocoon

هر چه امپراتوری‌های جدید بیشتر با یکدیگر رقابت کنند، خطر درگیری مسلحانه بیشتر می‌شود. جنگ سرد بین ایالات متحده و اتحاد جماهیر شوروی هرگز به یک رویارویی نظامی مستقیم منجر نشد، تا حد زیادی به لطف «دکترین تخریب متقابل تضمینی»^۱. اما خطر تشدید جنگ در عصر هوش مصنوعی، بیشتر است، زیرا جنگ سایبری ذاتاً متفاوت از جنگ هسته‌ای است.

سلاح‌های سایبری می‌توانند شبکه برق یک کشور را از کار بیندازند، اما همچنین می‌توانند برای نابودی یک تأسیسات تحقیقاتی محرمانه، از کار انداختن یک سنسور دشمن، شعله‌ور کردن یک رسوایی سیاسی، دستکاری انتخابات یا هک کردن یک گوشی هوشمند استفاده شوند. و آن‌ها می‌توانند همه این کارها را به‌طور مخفیانه انجام دهند. این سلاح‌ها با یک ابرقارچی و یک طوفان آتش حضور خود را اعلام نمی‌کنند، و همچنین ردی قابل مشاهده از محل پرتاب تا هدف بر جای نمی‌گذارند. در نتیجه، گاهی اوقات دشوار است بدانید که آیا حمله‌ای رخ داده است یا نه! یا چه کسی آن را راه‌اندازی کرده است. بنابراین وسوسه برای شروع یک جنگ سایبری محدود، بزرگ است، و وسوسه برای تشدید آن نیز بزرگ‌تر.

تفاوت اساسی دوم مربوط به پیش‌بینی‌پذیری است. جنگ سرد مانند یک بازی شطرنج فوق‌العاده منطقی بود و اطمینان از تخریب در صورت درگیری هسته‌ای آن قدر زیاد بود که تمایل به شروع جنگ بسیار اندک می‌شد. جنگ سایبری این اطمینان را ندارد. هیچ‌کس نمی‌داند هر طرف کجا، بمب‌های منطقی، تروجان‌ها و بدافزارهای خود را قرار داده است. هیچ‌کس نمی‌تواند مطمئن باشد که آیا سلاح‌های خود، موقع استفاده از آن آیا کار خواهد کرد یا خیر! آیا موشک‌های چینی در صورت صدور دستور، شلیک خواهند کرد یا شاید آمریکایی‌ها، این موشک‌ها یا زنجیره فرماندهی آن را هک کرده باشند؟ آیا ناوهای هواپیمابر آمریکایی به‌طور معمول عمل خواهند کرد یا شاید به‌طور مرموزی خاموش شوند یا در دایره‌هایی به دور خود بچرخند؟

چنین عدم اطمینانی دکترین «تخریب متقابل تضمینی» را تضعیف می‌کند. یک طرف ممکن است خود را قانع کند، که می‌تواند یک حمله اول موفقیت‌آمیز

را راه‌اندازی کرده و از تلافی گسترده جلوگیری کند. حتی بدتر، اگر یک طرف فکر کند که چنین فرصتی دارد، وسوسه برای راه‌اندازی یک حمله اول می‌تواند غیرقابل مقاومت شود، زیرا هیچ‌کس نمی‌داند که پنجره فرصت چقدر باز خواهد ماند. نظریه بازی‌ها بیان می‌کند که خطرناک‌ترین وضعیت در یک مسابقه تسلیحاتی زمانی است که یک طرف احساس می‌کند مزیتی دارد اما این مزیت در حال از دست رفتن است.

حتی اگر بشریت از بدترین سناریوی جنگ جهانی اجتناب کند، ظهور امپراتوری‌های دیجیتال جدید همچنان می‌تواند آزادی و رفاه میلیاردها نفر را به خطر بیندازد. امپراتوری‌های صنعتی قرن‌های نوزدهم و بیستم مستعمرات خود را استثمار و سرکوب کردند، و انتظار اینکه امپراتوری‌های دیجیتال جدید بهتر رفتار کنند، احمقانه خواهد بود. علاوه بر آن، اگر جهان به امپراتوری‌های رقیب تقسیم شود، بعید است که بشریت برای غلبه بر بحران زیست‌محیطی یا تنظیم هوش مصنوعی و سایر فناوری‌های مختل‌کننده مانند زیست‌فناوری همکاری کند.

تقسیم جهان به امپراتوری‌های دیجیتال رقیب، با چشم‌انداز سیاسی بسیاری از رهبرانی که معتقدند جهان یک جنگل است و صلح نسبی دهه‌های اخیر یک توهم بوده و تنها انتخاب واقعی این است که نقش شکارچی یا شکار را بازی کنند، همخوانی دارد.

با چنین انتخابی، اکثر رهبران، ترجیح می‌دهند به‌عنوان شکارچیان در تاریخ ثبت شوند و نام خود را به فهرست تیره فاتحانی اضافه کنند که دانش‌آموزان بدبخت، محکوم به حفظ آن برای امتحانات تاریخ خود هستند. این رهبران باید به یاد داشته باشند که یک شکارچی آلفای جدید در جنگل وجود دارد. اگر بشریت راهی برای همکاری و حفاظت از منافع مشترک ما پیدا نکند، همه ما طعمه آسانی برای هوش مصنوعی خواهیم بود.