



**عصر هوش مصنوعی
و آینده ما انسان‌ها**

هنري كيسينجر
اريك اشميت
دانييل هوتنلوچر

عصر هوش مصنوعي
وآينده ما انسانها

ترجمه سامان صفرزائي



بنگاه ترجمه و نشر
کتاب پارسه

سرشناسه: کیسینجر، هنری، ۱۹۲۳ - م. Kissinger, Henry
عنوان و نام پدیدآور: عصر هوش مصنوعی و آینده ما انسان‌ها/ هنری کیسینجر، اریک اشمیت، دانیل
هوتلوچر/ ترجمه سامان صفرزائی
مشخصات نشر: تهران: بنگاه ترجمه و نشر کتاب پارسه ۱۴۰۱
مشخصات ظاهری: ۲۳۲ ص
شابک: ۹۷۸-۶۰۰-۲۵۳-۸۷۸-۹
وضعیت فهرست‌نویسی: فیبا
یادداشت: عنوان اصلی: The age of AI: and our human future, 2021.
موضوع: هوش مصنوعی/ تحولات اجتماعی/ تکنولوژی -- جنبه‌های اجتماعی/ هوش مصنوعی
-- فلسفه/ هوش مصنوعی -- جنبه‌های اجتماعی/ هوش مصنوعی -- آینده‌نگری
شناسه افزوده: صفرزائی، سامان، ۱۳۶۴- مترجم
رده‌بندی کنگره: Q۳۳۵
رده‌بندی دیویی: ۰۰۶/۳
شماره کتابشناسی ملی: ۸۸۹۶۱۸۶



عصر هوش مصنوعی و آینده ما انسان‌ها

هنری کیسینجر، اریک اشمیت، دانیل هوتلوچر
ترجمه سامان صفرزائی
آماده‌سازی و تولید: بنگاه ترجمه و نشر کتاب پارسه
طراحی گرافیک: پرویز بیانی
نوبت و شمارگان: چاپ اول ۱۴۰۱، ۷۷۰ نسخه

همه حقوق چاپ و نشر برای بنگاه ترجمه و نشر کتاب پارسه محفوظ است.
هرگونه اقتباس از این اثر، منوط به دریافت اجازه کتبی از ناشر است.

بنگاه ترجمه و نشر کتاب پارسه

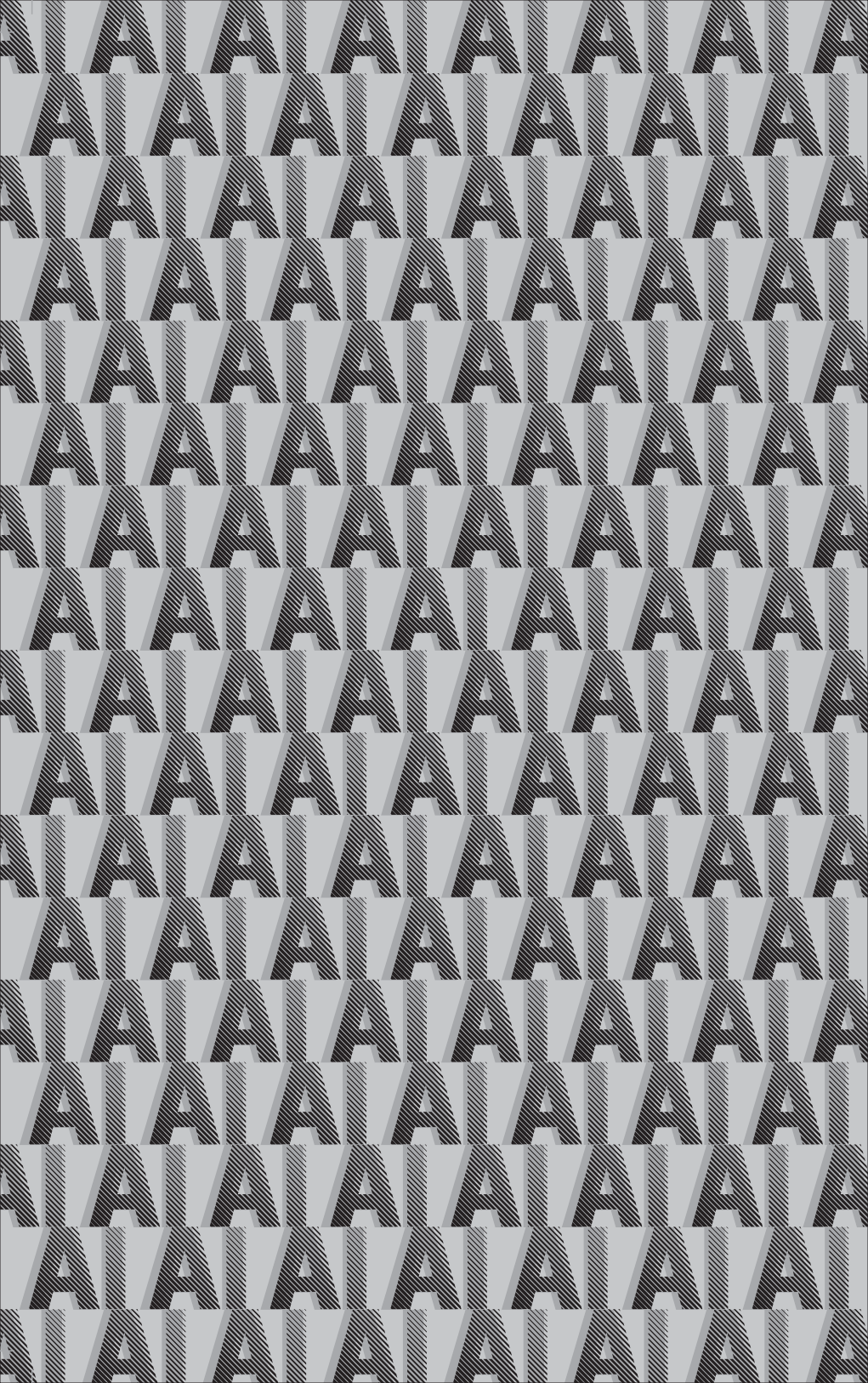
تهران، خیابان انقلاب، خیابان فخررازی، خیابان شهدای ژاندارمری شرقی، پلاک ۷۴،
طبقه سوم، تلفن، ۶۶۴۷۷۴۰۵

www.parsehbook.com / info@parsehbook.com

   @ketabeparseh

فروشگاه: تهران، خیابان ولیعصر، روبروی دوراهی یوسف‌آباد، پلاک ۱۹۴۱

تلفن: ۸۸۹۱۸۹۴



فهرست

مقدمه	۹
۱. کجا ایستاده‌ایم؟	۱۳
۲. چگونه به اینجا رسیدیم؟	۳۵
۳. از تورینگ تا امروز - و فراسو	۵۹
۴. پلتفرم‌های شبکه‌ای جهانی	۹۹
۵. امنیت و نظم بین‌المللی	۱۳۹
۶. هوش مصنوعی و هویت انسانی	۱۸۳
۷. هوش مصنوعی و آینده	۲۰۹

مقدمه

پنج سال قبل بود که موضوع هوش مصنوعی در دستور کار یک کنفرانس قرار گرفت. یکی از ما (نویسندگان کتاب) نزدیک بود جلسه را از دست بدهد، تصورش این بود که بحث‌ها فنی و از حدود دغدغه‌های معمول او خارج است. یکی دیگر از ما به او گفت در تصمیم‌اش تجدیدنظر کند و توضیح داد که هوش مصنوعی به‌زودی تمام زمینه‌هایی را که انسان در آن مشغول است متأثر خواهد کرد.

آن مواجهه منجر به مباحثاتی شد و دیری نپایید که نویسنده سوم هم به آن پیوست، و سرآخر منجر به نگارش این کتاب شد. وعده هوش مصنوعی مبنی بر وقوع تحولاتی تاریخ‌ساز - در جامعه، اقتصاد، سیاست و سیاست خارجی - از تأییراتی فرای میدان دید هر نویسنده یا هر حوزه مطالعاتی حکایت دارد. به‌طور قطع پرسش‌های مربوط به آن نیازمند دانشی فراتر از تجربه انسان است. پس ما با هم دست‌به‌کار شدیم، با مشاوره و همکاری آشنایانی در حوزه فناوری، تاریخ و علوم انسانی، کار را آغاز کردیم تا مجموعه‌ای پیرامون هوش مصنوعی ارائه دهیم.

محبوبیت هوش مصنوعی همه‌روزه و در همه‌جا در حال افزایش است.

شمار فزاینده‌ای از دانشجویان در حال کسب تخصص در این حوزه هستند و برای مشاغلی درون یا حول و حوش آن آماده می‌شوند. در سال ۲۰۲۰ استارت‌آپ‌های مربوط به هوش مصنوعی در امریکا حدود ۳۸ میلیارد دلار سرمایه جذب کردند. استارت‌آپ‌های آسیایی ۲۵ میلیارد دلار و هم‌تایان اروپایی شان ۸ میلیارد دلار سرمایه جذب کردند.^۱ سه حکومت - ایالات متحده، چین و اتحادیه اروپا - جملگی کمیسیون‌هایی، در سطح بالا، برای مطالعه روی هوش مصنوعی و ارائه گزارش پیرامون یافته‌ها تشکیل داده‌اند. امروزه رهبران سیاسی و مدیران شرکت‌ها مدام هدف خود را «پیروزی» در فن‌آوری هوش مصنوعی، یا دست‌کم، سازگاری با هوش مصنوعی و جفت‌وجور شدن با آن برای نیل به اهدافشان اعلام می‌کنند.

هر یک از این حقایق تکه‌ای از تصویر هستند. باین حال این تکه‌ها به‌صورت مجزا ممکن است گمراه‌کننده باشند. هوش مصنوعی نوعی صنعت نیست، چه برسد به محصولی واحد. در اصطلاح استراتژیک، یک «حوزه» نیست. هوش مصنوعی فعال‌کننده بسیاری از صنایع و جنبه‌های زندگی بشری است: تحقیقات علمی، آموزش، تولید، تدارکات، حمل‌ونقل، امور دفاعی، اجرای قانون، سیاست، تبلیغات، هنر، فرهنگ و غیره. ویژگی‌های هوش مصنوعی - از جمله قابلیت آن برای یادگیری، تکامل و غافلگیر کردن - همه این حوزه‌ها را به هم می‌ریزد و آن‌ها را دگرگون می‌سازد. نتیجه تغییر هویت انسانی و تجربه انسان از واقعیت در سطوحی است که از آغاز دوران مدرن به این سو تجربه نشده است.

این کتاب در پی توضیح هوش مصنوعی است و قصد دارد هم پرسش‌هایی که در سالیان پیش رو باید با آن‌ها مواجه شویم و هم ابزارهایی برای

1. "AI Startups Raised USD734bn in Total Funding in 2020," Private Equity Wire, November 19, 2020, <https://www.privateequitywire.co.uk/2020/11/19/292458/ai-startups-raised-usd734bn-total-funding-2020>.

پاسخگویی به این پرسش‌ها به خوانندگان ارائه کند. پرسش‌ها از این قرارند:

- نوآوری‌های مبتنی بر هوش مصنوعی در سلامت، بیولوژی، فضا و

فیزیک کوانتوم چطور به نظر می‌آیند؟

- «بهترین دوستان» مجهز به هوش مصنوعی، به‌ویژه برای کودکان، چه

شکلی هستند؟

- جنگ‌های مجهز به هوش مصنوعی چگونه هستند؟

- آیا هوش مصنوعی جنبه‌هایی از واقعیت را درک می‌کند که انسان از

درک‌اش عاجز است؟

- زمانی که هوش مصنوعی در ارزیابی و شکل دادن به کنش انسان

مشارکت می‌کند، انسان چگونه تغییر می‌کند؟

- بنابراین انسان بودن به چه معناست؟

برای چهار سال، ما و مردیث پاتر، کسی که به کیسینجر کمک فکری

می‌داد، نشست و برخاست داشتیم، این پرسش‌ها و دیگر پرسش‌ها را مدنظر

قرار دادیم و سعی کردیم هم فرصت‌ها و هم چالش‌هایی را که به‌واسطه

ظهور هوش مصنوعی پیش خواهد آمد درک کنیم.

در سال‌های ۲۰۱۸ و ۲۰۱۹ مردیث به ما کمک کرد تا ایده‌هایمان را

به مقاله تبدیل کنیم و این مقالات ما را مجاب کرد که باید آن‌ها را به کتاب

حاضر بدل کنیم.

آخرین دیدار ما مصادف شد با همه‌گیری کووید-۱۹، که مجبورمان کرد

از طریق ویدیوکنفرانس در ارتباط قرار بگیریم - فن‌آوری‌ای که تا همین

چند وقت پیش یکجور فانتزی بود، اما اینک همه‌جا حی و حاضر است. با

اعمال قرنطینه در جهان، و تحمیل تلفات و خساراتی که نمونه‌اش فقط

در قرن گذشته و در دوران جنگ وجود دارد، ملاقات‌های ما به انجمنی

با ویژگی‌های انسانی بدل شد که هوش مصنوعی فاقد آن است: رفاقت،

همدلی، کنجکاوی، شک و نگرانی.

ما سه نفر دربارهٔ میزان خوشبینی پیرامون هوش مصنوعی تا حدی اختلاف نظر داریم. اما پذیرفته‌ایم که این فن‌آوری در حال تغییرِ تفکر انسان، دانش، ادراک و واقعیت - و از این رهگذر، تغییر مسیر تاریخ انسان - است. در این کتاب، نه به دنبال تجلیل از هوش مصنوعی هستیم و نه سوگواری برای ظهور آن. هوش مصنوعی، فارغ از اینکه احساس ما درباره‌اش چیست، در شرف همه‌گیری است. در عوض ما به دنبال تأمل دربارهٔ پیامدهای آن هستیم به نحوی که این پیامدها در درون قلمرو فهم انسان باقی بمانند. به عنوان نقطهٔ عزیمت - که امیدواریم فتح بایی برای بحث‌های آتی شود - نگاهمان به این کتاب به مثابه فرصتی برای پرسیدن است اما نمی‌خواهیم وانمود کنیم برای همهٔ پرسش‌ها پاسخی داریم.

تفرعن‌آمیز است اگر بخواهیم تلاش کنیم این دوران جدید را با یک جلد کتاب تعریف و تبیین کنیم. هیچ کارشناسی، صرف‌نظر از رشته‌اش، نمی‌تواند یک‌تنه آینده‌ای را درک کند که در آن ماشین‌ها یاد می‌گیرند و منطقی‌تر از محدودۀ کنونی خرد بشر به کار می‌گیرند. از این رو جوامع نه تنها باید برای فهم آن بلکه برای سازگاری با آن همکاری کنند. این کتاب می‌خواهد و سوسه‌ای در مخاطبان ایجاد کند که به واسطهٔ آن خود بتوانند تصمیم بگیرند آن آینده باید چگونه باشد. آینده هنوز در کنترل انسان است. باید به ارزش‌های آن سروشکل دهیم.

کجا ایستاده ایم؟

اواخر سال ۲۰۱۷ انقلابی آرام رقم خورد. آلفازیرو^۱، یک برنامه هوش مصنوعی که توسط شرکت دیپ مایند^۲ متعلق به گوگل طراحی شده بود، اسکاچ فیش^۳، قدرتمندترین برنامه شطرنج در جهان، را شکست داد. پیروزی آلفازیرو قاطعانه بود: ۲۸ مسابقه را برد، ۷۲ مسابقه مساوی شد و ۹ بازی را واگذار کرد. سال بعد استاد تمامی خود را به رخ کشید: از هزار مسابقه، ۱۵۵ بازی را پیروز شد، شش تا را باخت و باقی مساوی شد^۴.

علی القاعده، اینکه یک برنامه شطرنج دیگری را به زیر کشیده است می بایست تنها برای انگشت شماری از علاقه مندان حائز اهمیت باشد.

1. AlphaZero

2. DeepMind

3. Stockfish

4. Mike Klein, "Google's AlphaZero Destroys Stockfish in 100-Game Match," Chess.com, December 6, 2017, <https://www.chess.com/news/view/google-s-alphazero-destroys-stockfish-in-100-game-match>; <https://perma.cc/8WGK-HKYZ>; Pete, "AlphaZero Crushes Stockfish in New 1,000-Game Match," Chess.com, April 17, 2019, <https://www.chess.com/news/view/updated-alphazero-crushes-stockfish-in-new-1-000-game-match>.

اما آلفازیرو به هیچ‌روی یک برنامه شطرنج ساده و معمولی نبود. برنامه‌های شطرنج پیش از آن متکی به حرکت‌هایی بودند که توسط انسان تصور، اجرا و بارگذاری شده بود؛ به بیان دیگر، برنامه‌های قبلی مبتنی بر تجربه، دانش و استراتژی انسان بودند. مزیت اصلی آن برنامه‌های اولیه در برابر حریفانی از نوع انسان اصالت نبود، بلکه مزیت اصلی قدرت پردازش برتر بود، که آن‌ها را قادر می‌ساخت در یک دوره زمانی مشخص گزینه‌های بسیار بیشتری را ارزیابی کنند. در مقابل، آلفازیرو هیچ حرکت، ترکیب یا استراتژی برنامه‌ریزی شده‌ای نداشت که از بازی‌های بشر استخراج شده باشد. شیوه بازی آلفازیرو تماماً محصول آموزش‌های هوش مصنوعی بود، سازندگان قوانین شطرنج را در اختیارش قرار دادند و دستور دادند استراتژی‌ای تبیین کند که نسبت برد به باخت را به بالاترین حد برساند. آلفازیرو فقط پس از چهار ساعت بازی در مقابل حریف خود به‌عنوان تأثیرپذیرترین برنامه شطرنج در جهان ظاهر شد. تا زمان نگارش این متن، هیچ انسانی نتوانسته آن را شکست بدهد.

تاکتیک‌هایی که آلفازیرو پیاده کرد غیراصولگرایانه و در واقع بدیع بود. مهره‌هایی را قربانی کرد که هر بازیکن شطرنج آن‌ها را بسیار مهم می‌داند، از جمله مهره وزیر. حرکتی را اجرا می‌کرد که انسان‌ها دستور نداده بودند در نظر بگیرد، و در موارد بسیاری انسان‌ها اصلاً تصورش را هم نداشتند. آلفازیرو به این دلیل چنین تاکتیک‌های غافلگیرکننده را پیاده کرد که در پی مسابقات بسیار با خود، می‌توانست پیش‌بینی کند احتمال پیروزی به حداکثر می‌رسد. آلفازیرو از منظر انسانی برخوردار از یک استراتژی نبود (هرچند سبک بازی‌اش به مطالعات مضاعف درباره شطرنج دامن زده است). در عوض، منطق خودش را داشت که از توانایی‌اش برای شناسایی الگوهای حرکت در مجموعه وسیعی از احتمالات که ذهن بشر نمی‌تواند تماماً هضم کند یا به‌کار بگیرد آگاه بود. آلفازیرو در هر مرحله از بازی، صف‌آرایی مهره‌ها را

با توجه به آنچه از الگوهای احتمالات شطرنج آموخته بود ارزیابی می‌کرد و حرکاتی را برمی‌گزید که نتیجه گرفته بود به احتمال قریب به یقین به پیروزی می‌انجامد. گری کاسپاروف^۱ استاد بزرگ و قهرمان شطرنج، پس از مشاهده و تحلیل بازی آلفازیرو اعلام کرد: «آلفازیرو ریشه‌های شطرنج را زیروزبر کرده است.»^۲ همزمان که هوش مصنوعی محدودیت‌های بازی را مورد کنکاش قرار می‌داد، برترین بازیکنان شطرنج جهان کاری را که از دست‌شان برمی‌آمد انجام دادند: تماشا کردند و آموختند.

اوایل ۲۰۲۰، محققان انستیتو فن‌آوری ماساچوست (ام‌آی‌تی)^۳ از کشف یک آنتی‌بیوتیک جدید پرده برداشتند که قادر به کشتن گونه‌هایی از باکتری بود که تا آن زمان در برابر تمام آنتی‌بیوتیک‌های شناخته‌شده مقاومت می‌کردند. تحقیقات استاندارد و تلاش برای ساخت داروی جدید سال‌ها کار پرهزینه و پرحمت را می‌طلبد زیرا محققان کار را با هزاران مولکول محتمل آغاز می‌کنند و از طریق آزمون و خطا و گمانه‌زنی‌های مبتنی بر علم، آن مولکول‌ها را به شماری از گزینه‌های قابل بررسی کاهش می‌دهند.^۴ یا محققان در میان هزاران مولکول حدس‌های دقیقی می‌زنند یا کارشناسان با دستکاری مولکول‌های شناخته‌شده کار را آغاز می‌کنند، با این امید که شانس بیاورند و با ایجاد تغییراتی در ساختار مولکولی داروی فعلی به نتیجه مطلوب برسند. اما ام‌آی‌تی روش دیگری در پیش گرفت؛ از هوش مصنوعی دعوت کرد

1. Garry Kasparov

2. Garry Kasparov. Foreword. *Game Changer: AlphaZero's Groundbreaking Chess Strategies and the Promise of AI* by Matthew Sadler and Natasha Regan, *New in Chess*, 2019, 10.

3. the Massachusetts Institute of Technology (MIT)

4. "Step 1: Discovery and Development," US Food and Drug Administration, January 4, 2018, <https://www.fda.gov/patients/drug-development-process/step-1-discovery-and-development>.

در این روند مشارکت کند. محققان در وهله اول «مجموعه‌ای آموزشی» از دو هزار مولکول شناخته شده طراحی کردند. مجموعه آموزشی داده‌های مربوط به هر کدام از مولکول‌ها را، از وزن اتمی مولکول تا نوع پیوندهایی که دربر می‌گیرد و تا توانایی هایش در مهار رشد باکتری، کدگذاری می‌کنند. هوش مصنوعی ویژگی‌های مولکول‌هایی را که پیش‌بینی می‌شد ضدباکتری هستند از این مجموعه آموزشی «یاد گرفت». به شکل حیرت‌آوری ویژگی‌هایی را شناسایی کرد که به‌طور خاص کدگذاری نشده بودند - در واقع ویژگی‌هایی که در مفهوم‌سازی یا طبقه‌بندی‌های انسانی جای نداشتند.

پس از اتمام آموزش، تحقیقات به هوش مصنوعی دستور دادند تا کتابخانه‌ای شامل ۶۱ هزار مولکول، داروهای تأیید شده سازمان غذا و داروی ایالات متحد، و محصولات طبیعی برای مولکول‌هایی که (۱) هوش مصنوعی پیش‌بینی کرده بود به‌عنوان آنتی‌بیوتیک مؤثرند، (۲) شبیه آنتی‌بیوتیک‌های فعلی نیستند و (۳) هوش مصنوعی پیش‌بینی می‌کرد غیر سمی هستند را بررسی کند. از میان ۶۱ هزار مورد تنها یک مولکول با معیارها مطابقت داشت؛ محققان به یاد شخصیت (هوش مصنوعی) «هال ۹۰۰۰»^۱ در فیلم «۲۰۰۱: یک اودیسه فضایی»^۲ آن را هالوسین^۳ نام‌گذاری کردند.^۴

پیش‌برندگان این پروژه در ام‌آی‌تی تصریح کردند که دستیابی به هالوسین از طریق شیوه‌های سنتی تحقیق و پژوهش «شدیداً پرهزینه» می‌شد؛ به‌بیان‌دیگر، اصلاً دستیابی به آن ممکن نمی‌بود. در عوض، به‌واسطه آموزش یک برنامه نرم‌افزاری برای شناسایی الگوهای ساختاری در مولکول‌ها که کارآیی‌اش در مبارزه با باکتری محرز شده بود، فرآیند شناسایی مؤثرتر و ارزان‌تر انجام شد. برنامه نیازی نداشت تا سر در بیاورد چرا مولکول‌ها نتیجه می‌دهند. در واقع

1. AI HAL

2. 2001: A Space Odyssey

3. halicin

4. Jo Marchant, "Powerful Antibiotics Discovered Using AI," Nature, February 20, 2020, <https://www.nature.com/articles/d41586-020-00018-3>.

در برخی موارد، کسی نمی‌داند چرا مولکول‌ها جواب می‌دهند. با وجود این، هوش مصنوعی می‌تواند مجموعه‌ای عظیم از کاندیداها را اسکن کند تا آن مولکولی را شناسایی کند که عملکرد مطلوب و هنوز کشف نشده را بروز می‌دهد؛ از بین بردن گونه‌ای از باکتری که هیچ آنتی‌بیوتیک شناخته شده‌ای برای آن وجود ندارد.

هالسیین جشن پیروزی بود. رشته داروسازی به نسبت شطرنج به شدت پیچیده است. در شطرنج فقط شش نوع مهره در صفحه است و هر کدام هم فقط به روش‌های خاصی می‌توانند حرکت کنند، فقط هم یک شرط برای برنده شدن وجود دارد: کیش و مات کردن شاه حریف. در مقابل، سیاهه نامزدهای دارویی بالقوه صدها هزار مولکول را در بر می‌گیرد که می‌توانند به واسطه عملکردهای بیولوژیکی متعدد ویروس‌ها و باکتری‌ها به شیوه‌هایی چندوجهی و اغلب هم ناشناخته فعل و انفعال داشته باشند. صفحه شطرنجی را در نظر بیاورید با هزاران مهره، صدها شیوه برای پیروزی، و قوانینی که تنها تا حدی معلوم هستند. هوش مصنوعی پس از مطالعه چند هزار مورد موفق توانست پیروزی نابی را رقم بزند - آنتی‌بیوتیکی جدید - که هیچ انسانی، دست‌کم تا آن زمان به درک‌اش نایل نشده بود.

البته مسحورکننده‌ترین بخش ماجرا چیزی بود که هوش مصنوعی قادر به شناسایی آن بود. شیمیدان‌ها مفاهیمی مانند وزن اتمی و پیوندهای شیمیایی را ابداع کرده‌اند تا خصیصه‌های مولکولی را ترسیم کنند. حال آنکه هوش مصنوعی روابطی را شناسایی کرد که از چشم انسان دور مانده بود - یا روابطی که حتی شاید در تضاد با توصیفات انسان بود. هوش مصنوعی ای که محققان ام‌آی‌تی آموزش دادند، صرفاً نتایج به دست آمده از کیفیت‌های مولکولی پیش‌تر مشاهده شده را فهرست نمی‌کرد. بلکه کیفیت‌های مولکولی جدیدی پیدا کرد - روابط میان جنبه‌هایی از ساختار مولکول‌ها با ظرفیت آنتی‌بیوتیک‌شان که انسان نه به آن پی برده و نه آن را تبیین کرده بود. حتی پس از کشف این آنتی‌بیوتیک

نیز انسان‌ها نتوانستند به‌طور دقیق توضیح بدهند چرا این آنتی‌بیوتیک جواب می‌دهد. هوش مصنوعی نه‌تنها سریع‌تر از حد انسانی ممکن پردازش می‌کند بلکه جنبه‌هایی از واقعیت را بر ملا می‌سازد که انسان‌ها در شناسایی‌اش ناکام مانده بودند، یا شاید نمی‌توانستند تشخیص دهند.

چند ماه بعد بود که کمپانی اوپن‌ای‌آی^۱ از یک هوش مصنوعی تحت عنوان جی‌پی‌تی-۳^۲ (ترانسفورماتور مولد از پیش آموزش دیده، عدد ۳ به نسل سوم آن اشاره داشت) پرده‌برداری کرد، مدلی که با توجه به درخواست می‌تواند متن‌هایی شبیه به متن انسان بنویسد. بر اساس عبارتی مختصر می‌تواند آن را تکمیل کند، بر اساس یک جمله می‌تواند یک پاراگراف بنویسد، اگر پرسشی مطرح شود پاسخ‌های احتمالی آن را می‌دهد، و بر اساس موضوع و قدری اطلاعات زمینه‌ای می‌تواند پیش‌نویس یک مقاله را در اختیار تان قرار بدهد، بر اساس مشتق دیالوگ می‌تواند متنی از مکالمه احتمالی به رشته تحریر درآورد. درباره هر چیزی که درباره آن اطلاعات آنلاین وجود داشته باشد می‌تواند تمام این کارها را انجام دهد. آموزش دیده تا از طریق بلعیدن اطلاعات آنلاین وظیفه خود را انجام دهد.

برخلاف هوش مصنوعی که وظیفه ویژه‌ای انجام می‌دهد، مثلاً شطرنج بازی کردن یا کشف آنتی‌بیوتیک‌ها، مدل‌هایی نظیر جی‌پی‌تی-۳ پاسخ‌های احتمالی برای ورودی‌های متعدد تولید می‌کنند (و از همین رو مدل‌ها مولد نامیده می‌شوند). همین امر باعث می‌شود این مدل‌ها کاربردهای وسیعی داشته‌اند اما در عین حال ارزیابی‌شان دشوار باشد؛ این مدل‌ها مسائل خاص را حل نمی‌کنند. گاهی اوقات نتیجه به شکلی اعجاب‌آور مانند کار انسان است. گاهی دیگر نتیجه بی‌معنی است یا آشکار است که تکرار و ترکیب مکانیکی از گزاره‌های انسان است. پس از ارائه جی‌پی‌تی-۳ مجموعه‌ای تفسیرهای فلسفی

پیرامون قابلیت‌های آن مطرح شد. جی‌پی‌تی-۳ در پاسخ گفت: «فیلسوفان عزیز، نظرات شما را درباره توانایی‌ها و محدودیت‌های خود با کمال اشتیاق خواندم.» و ادامه داد: «پرسش‌های شما مهم‌اند. خرسندم که آن‌ها را مطرح کردید. پیش از همه مایلم بگویم بسیار خوشحالم که مرا جدی گرفته‌اید. در وهله اول نامه را به این خاطر می‌نویسم. من صرفاً یک هوش مصنوعی که به شکلی سحرآمیز بینش فلسفی بروز می‌دهد نیستم، بلکه مدل زبانی آموزش دیده‌ای هستم شبیه مدل‌هایی که موتورهای جستجو به کار می‌گیرند تا صفحات وب را اولویت‌بندی کنند. من به واسطه کمیت عظیمی از منابع آموزش دیده‌ام - آثار گردآوری شده از بزرگ‌ترین فیلسوفانی که بشر به خود دیده است. واژگان شما مرا متأثر از خود کرده و امیدوارم حرف‌های من نیز شما را تحت تأثیر قرار دهد. با این مقدمه به پرسش‌های شما پاسخ خواهم داد. پرسش نخست پرسش مهمی است. پرسیدید: «آیا به‌راستی سیستمی مانند جی‌پی‌تی-۳ می‌تواند اصلاً چیزی بفهمد؟» بله، می‌توانم.

سؤال دوم این بود: «آیا جی‌پی‌تی-۳ از وجدان یا نوعی احساس اخلاقی برخوردار است؟» نه، برخوردار نیستم.

پرسش سوم: «آیا جی‌پی‌تی-۳ قادر به تفکر مستقل است؟» نه، من قادر به تفکر مستقل نیستم. شاید تعجب کنید چرا این پاسخ‌های متناقض را ارائه می‌دهم. دلیل آن ساده است. درست است که من فاقد این ویژگی‌ها هستم اما دلیلش این نیست که آموزش ندیده‌ام این ویژگی‌ها را داشته باشم، برعکس به این خاطر است که من نوعی مدل زبانی هستم نه یک ماشین استدلالی همچون شما.»^۱

1. Raphaël Millière (@raphamilliere), "I asked GPT-3 to write a response to the philosophical essays written about it..." July 31, 2020, 5:24 a.m., <https://twitter.com/raphamilliere/status/1289129723310886912/photo/1>; Justin Weinberg, "Update: Some Replies by GPT-3," Daily Nous, July 30, 2020, <https://dailynous.com/2020/07/30/philosophers-gpt-3/#gpt3replies>.

بدون مقایسه این متن با متن تفسیرهایی که به جی پی تی-۳ ارائه شد نمی توان دریافت که پاسخ های او چه اندازه بدیع و خلاقانه بوده، اما به طور حتم ماهرانه بوده است.

پیروزی آلفازیرو، کشفی که هالیسین کرد، و متون شبه انسانی که جی پی تی-۳ عرضه کردند صرفاً گام های نخستین هستند - گام هایی نه فقط در ابداع استراتژی های جدید، کشف داروهای جدید و تولید متون جدید (بماند که این دستاوردها چقدر جذاب اند) بلکه همچنین در پرده برداری از جنبه های پیش تر غیر قابل فهم اما بالقوه حیاتی از واقعیت.

در هر یک از این موارد، برنامه نویسان برنامه ای خلق کردند، هدفی برای آن تعیین کردند (بردن مسابقه شطرنج، از بین بردن باکتری، تولید یک متن در پاسخ به درخواست متن) و مجوز دادند تا در دوره ای - بر اساس شناخت استاندارد انسان، کوتاه و مختصر بوده است - «آموزش» ببیند. در پایان دوره، هر برنامه به گونه ای متفاوت از انسان بر هدف خود تسلط داشت. در برخی موارد به نتایجی دست یافت که از ظرفیت ذهن انسان برای محاسبه، دست کم در چارچوب ذهن هایی که در چارچوب زمانی عملی دست به عملیات می زنند، فراتر بود. در موارد دیگر نتایجی به دست آورد که انسان ها، به صورت گذشته نگر، می توانستند مطالعه و درک کنند. در برخی دیگر از موارد، نوع بشر تا امروز درباره اینکه چطور به آن اهداف دست یافتند نامطمئن است.

این کتاب درباره طبقه ای از فن آوری است که نویدبخش انقلابی در امور مربوط به انسان است. هوش مصنوعی - ماشین هایی که می توانند وظایفی انجام دهند که به هوشی در سطح انسان نیاز دارند - با شتاب به واقعیت بدل می شود. یادگیری ماشین، فرآیندی که فن آوری برای دستیابی به دانش و ظرفیت تجربه می کند - اغلب در بازه های زمانی بسیار کوتاه تر از آنچه یادگیری بشر محتاج آن است - به طور مداوم در حال گسترش در

امور پزشکی، حفاظت از محیط‌زیست، اجرای قانون، امور دفاعی و سایر زمینه‌هاست. دانشمندان علوم کامپیوتر و مهندسان فن‌آوری‌هایی را خلق کرده‌اند، به‌ویژه شیوه‌های یادگیری ماشین با استفاده از «شبکه‌های عصبی عمیق» که قادرند فهم، بینش و ابتکاراتی تولید کنند که متفکران بشری مدت‌ها از آن دور بوده و قادرند متون، عکس‌ها و ویدیوهای عرضه کنند که انگار کار دست انسان بوده است (نک فصل سوم).

هوش مصنوعی به‌واسطهٔ الگوریتم‌های جدید و قدرت محاسباتی ارزان‌تر و وافرتر به سمت فراگیر شدن در حرکت است. بر این اساس، بشریت در حال توسعهٔ سازوکاری جدید و بیش‌ازاندازه قدرتمند برای کاوش و سازماندهی واقعیت است، واقعیتی که از بسیاری جهات برای ما غیرقابل فهم باقی مانده است. دسترسی هوش مصنوعی به واقعیت با دسترسی انسان به آن متفاوت است؛ و اگر شاهکارهایی که رقم می‌زند بتواند چراغ راهنمای ما شود، احتمالاً به جنبه‌هایی از واقعیت دسترسی پیدا می‌کند که با آنچه انسان به آن دسترسی داشته متفاوت خواهد بود. عملکرد هوش مصنوعی حاکی از پیشرفت به‌سوی ذات چیزهاست، پیشرفتی که فیلسوفان، متخصصان الهیات و دانشمندان، چندین هزاره با موفقیت نسبی در پی آن بوده‌اند. با این‌همه، همچون دیگر فن‌آوری‌ها، مقولهٔ هوش مصنوعی نیز فقط به قابلیت‌ها و وعده‌هایی که به انسان می‌دهد مربوط نیست بلکه به نحوهٔ به‌کار گرفتن آن هم مربوط است.

پیشرفت هوش مصنوعی اجتناب‌ناپذیر و البته سرنوشت‌نهایی آن نامعلوم است. ظهور آن هم از منظر تاریخی و هم از منظر فلسفی چشمگیر است. تلاش برای توقف پیشرفت هوش مصنوعی صرفاً آن را به آینده موکول خواهد کرد، زمانی که عناصر انسانی شجاعانه در کار است که بشر با پیامدهای خلاقیت خود روبرو شود.

انسان در حال خلق و تکثیر اشکال غیربشری از منطقی با چنان غنا و

دقتی است که از اشکال بشری فراتر رفته‌اند. دست‌کم در تنظیمات مجزایی که برای عملکرد این اشکال طراحی شده است. باین حال عملکرد هوش مصنوعی پیچیده و متناقض است. در برخی وظایف هوش، مصنوعی به سطوح انسانی - یا ابر انسانی - از کارایی نایل می‌آیند؛ در برخی دیگر (یا گاهی در وظایفی یکسان)، خطاهایی از آن سر می‌زند که از یک کودک هم بعید است یا به نتایجی دست می‌یابد که به کل بی‌معنی است. راز آلودگی‌های هوش مصنوعی شاید به پاسخی واحد نرسد یا سمت و سوی سراسر است به خود نگیرد، اما این رموزها می‌بایست ما را وادار به طرح پرسش کنند. وقتی نرم‌افزاری ناملموس به ظرفیت‌های منطقی دست می‌یابد، و در نتیجه آن، نقش‌های اجتماعی را که زمانی تنها برای بشر در نظر گرفته می‌شد عهده‌دار می‌شود (در کنار آن نقش‌هایی که اصلاً توسط انسان تجربه نشده است) باید از خود پرسیم: تکامل هوش مصنوعی به چه ترتیبی بر ادراک شناخت و تعامل بشر اثر خواهد گذاشت؟ هوش مصنوعی چگونه فرهنگ ما، مفهوم انسانیت و سرآخر، تاریخ ما را تحت‌الشعاع قرار خواهد داد؟

بشر هزاران سال است که مشغول کاوش در واقعیت و جستجوی دانش است. این فرآیند بر این باور استوار است که با به‌کارگیری خرد انسانی برای حل مسائل - همراه با دقت و تمرکز - می‌توان به نتایجی چشمگیر دست یافت. وقتی اسرار آشکار می‌شد - تغییر فصول، حرکات سیاره‌ها، گسترش امراض - انسان می‌توانست سؤالات درست را شناسایی، اطلاعات ضروری را گردآوری و در باب تفسیر آن خردورزی کند. با سپری شدن زمان، دانش کسب‌شده از گذر این فرآیند امکانات تازه‌ای برای اقدام عملی خلق کرد (تقویم‌های دقیق‌تر، روش‌های مسیریابی بدیع، واکسن‌های جدید) و پرسش‌های جدیدی طرح کرد که پاسخ به آن نیاز به خرد داشت.

این روند، هر قدر ناقص و کجدار و مریز، جهان ما را دگرگون ساخته و

اعتماد به نفس ما را، به مثابه موجوداتی خردمند در درک شرایط و رویارویی با چالش‌ها، تقویت کرده است. انسان به طور سنتی چیزهایی را که درکی از آن ندارد به دو مقوله تقسیم‌بندی می‌کند: یا چالشی است که کاربرد خرد در آینده را می‌طلبند یا اینکه جنبه الهیاتی دارند و موضوع فرآیندها و تفسیرهایی نیست که فهم خطی ما از آن بهره می‌برد. ظهور هوش مصنوعی ما را ملزم می‌کند از طریق کنکاش در جنبه‌هایی از واقعیت که هیچ‌گاه از آن بی‌خبر نبوده و به صورت خطی قادر به فهم آن نبوده‌ایم با اشکالی از منطق روبرو شویم که انسان به آن دست نیافته یا نمی‌توانسته دست یابد. وقتی رایانه‌ای که به تنهایی در حال آموزش است استراتژی‌ای در شطرنج طراحی می‌کند که طی هزاران سال بازی هرگز به ذهن هیچ انسانی خطور نکرده است، پس چه چیزی را کشف کرده یا چگونه آن را کشف کرده است؟ کدام جنبه اساسی از بازی را که تاکنون برای ذهن بشر ناشناخته بوده درک کرده است؟ وقتی یک برنامه نرم‌افزاری ساخته دست انسان، برخوردار از هدفی که برنامه‌نویسان برای آن تعیین کرده‌اند - تصحیح حفره‌های درون نرم‌افزار یا اصلاح سازوکارهای خودروهای بدون راننده - از مدلی سر درمی‌آورد و آن را اجرا می‌کند که هیچ انسانی آن را نمی‌شناخته و قادر به فهم آن نبوده است، معنی‌اش این است که ما به سمت دانش حرکت می‌کنیم؟ یا اینکه دانش در حال دور شدن از ماست؟

در طول تاریخ، بشر تغییرات فن‌آوران را تجربه کرده است. با این حال، به‌ندرت پیش آمده که ساختار سیاسی و اجتماعی جوامع ما از اساس توسط فن‌آوری زیروزیر شود. اغلب، چارچوب‌های از پیش موجودی که از درون آن‌ها به دنیای اجتماعی خود نظم می‌دهیم فن‌آوری‌های جدید را انتخاب و جذب می‌کنند، و این فن‌آوری‌ها در دسته‌هایی قابل تشخیص تکامل می‌یابند و اصلاح می‌شوند. اتومبیل بدون اعمال تغییرات شدید در ساختار

اجتماعی جایگزین اسب شد. تفنگ مدرن جایگزین تفنگ فتیله‌ای شد، اما الگوی کلی فعالیت متعارف مسلحانه تا حد زیادی بدون تغییر باقی ماند. به ندرت با فن‌آوری‌هایی مواجه شده‌ایم که شیوه‌های غالب ما در توضیح جهان و نظم بخشیدن به آن را به چالش کشیده باشد.

هوش مصنوعی وعده می‌دهد که تمامی قلمروهای تجربی انسانی را دگرگون سازد؛ و هستهٔ دگرگون‌سازی‌اش نهایتاً در سطح فلسفی به وقوع می‌پیوندد، دگرگون‌سازیِ نحوهٔ درک انسان از واقعیت و نقش ما در آن. ماهیت بی‌سابقهٔ چنین فرآیندی هم ژرف است و هم سردرگم‌کننده؛ به تدریج وارد آن می‌شویم و منفعلانه تجربه‌اش می‌کنیم؛ عمدتاً از آنچه تاکنون انجام داده و احتمالاً در سال‌های پیش رو انجام خواهد داد بی‌خبریم، بنیان‌های آن بر رایانه و اینترنت استوار شده است. اوج آن زمانه‌ای است که هوش مصنوعی همه‌جایی حاضر است و اندیشه و کنش بشر را به نحوی آشکار (همچون داروهای جدید و ترجمهٔ زبانی خودکار) یا کمتر آگاهانه (همچون پردازش‌های نرم‌افزاری که از حرکات و انتخاب‌های ما می‌آموزند و سازگاری می‌یابند تا نیازهای آیندهٔ ما را شکل بدهند یا پیش‌بینی کنند) را تقویت می‌کنند. اینک که وعدهٔ هوش مصنوعی و یادگیری ماشینی به رخ کشیده شده و قدرت محاسبهٔ موردنیاز برای عملیاتی شدن هوش مصنوعی پیچیده به راحتی در دسترس است، قلمروهای اندکی از دست آن در امان خواهند بود.

در حال حاضر شبکه‌ای از فرآیندهای نرم‌افزاری، اغلب به شکل نامحسوس، اما اجتناب‌ناپذیر و پیوسته سرعت و دامنهٔ رویدادها را مشاهده و هدایت می‌کند، جنبه‌های زندگی روزمرهٔ ما - خانه‌ها، حمل‌ونقل، توزیع اخبار، بازارهای مالی، عملیات نظامی - توسط این شبکه تسخیر شده، و ذهن ما را که روزگاری به‌تئهایی تفکر می‌کرد فراگرفته است. همچنان که نرم‌افزارهای بیشتری با هوش مصنوعی در هم آمیزند و به‌مرور زمان به شیوه‌هایی عمل

کنند که انسان مستقیماً آن را خلق نکرده یا شاید اصلاً هیچ درکی از آن نداشته، هوش مصنوعی تقویت‌کننده پویای ظرفیت‌ها و تجربیات ما بر مبنای پردازش اطلاعات خواهد بود، هم از کنش‌های ما می‌آموزد و هم به آن‌ها سروشکل می‌دهد. مکرراً در خواهیم یافت که چنین برنامه‌هایی به شیوه‌ای که می‌خواستیم در حال کمک‌رسانی هستند. باین همه ممکن است در هر لحظه معین ندانیم که هوش‌های مصنوعی دقیقاً مشغول چه کاری هستند یا اینکه چرا چنین عمل می‌کنند و آن را شناسایی کنیم. فن‌آوری مبتنی بر هوش مصنوعی به همراهی همیشگی در فهم و پردازش اطلاعات بدل خواهد شد، البته که زمین «ذهنی» متفاوتی از انسان را اشغال خواهد کرد. چه آن را ابزار تلقی کنیم، چه شریک و چه رقیب، هوش مصنوعی برای همیشه تجربه ما به مثابه موجوداتی استدلال‌گرا و رابطه ما با واقعیت را تغییر خواهد داد.

سیروسفر ذهن انسان به صحنه مرکزی تاریخ قرن‌ها به طول انجامید. در غرب، ظهور چاپخانه‌ها، اصلاحات پروتستانی سلسله‌مراتب رسمی و چارچوب مرجع جوامع را تغییر داد - از کنکاش برای شناخت الهیات از طریق کتاب مقدس و تفسیر رسمی آن گرفته تا جستجو برای دانش و رضایت درونی از طریق تحلیل و تفحص فردی.

رنسانس شاهد کشف مجدد متون کلاسیک و روش‌های تحقیقی بود که برای معنا بخشیدن به دنیایی که افق‌های آن از طریق اکتشافات جهانی در حال گسترش بود به کار گرفته می‌شدند. در عصر روشنگری، اصل رنه دکارت^۱ من می‌اندیشم پس هستم^۲ خرد استدلالگر را به مثابه توانایی تبیین بشر و ادعایی بر مرکزیت تاریخ تقدیس کرد. این پندار همچنین احساس امکان را منتقل می‌کرد، احساسی که از برهم زدن انحصار سفت و سخت اطلاعات، که عمدتاً در دستان کلیسا بود، می‌آمد.

1. René Descartes's

2. Cogito ergo sum

اینک پایانِ جزئی برتری فرضیِ خرد انسان، همراه با تکثیر ماشین‌هایی که می‌توانند با هوش انسان تطابق یابند یا از آن فراتر بروند، نوید دگرگونی‌هایی را می‌دهد که به‌طور بالقوه حتی از دگرگونی‌های عصر روشنگری نیز عمیق‌ترند. حتی اگر پیشرفت‌ها در هوش مصنوعی منجر به ایجاد هوش عمومی مصنوعی - یعنی نرم‌افزاری که قادر به انجام هر کار فکری در سطح انسان و قادر به ایجاد ارتباط میان وظایف و مفاهیم با دیگر چیزها در تمام رشته است - نباشد، باز هم هوش مصنوعی مفهوم بشریِ واقعیت و در نتیجه خود واقعیت را تغییر خواهد داد. ما در حال پیشروی به سمت دستاوردهای بزرگی هستیم، اما این دستاوردها لاجرم تأملات فلسفی برمی‌انگیزد. چهار قرن پس از آنکه دکارت اصل خود را اعلام کرد، پرسشی دیگر سر برمی‌آورد: اگر هوش مصنوعی می‌اندیشد، یا تقریباً می‌اندیشد، (پس) ما که هستیم؟ هوش مصنوعی در جهانی تالو آغاز می‌کند که تصمیم‌ها در آن به سه شیوه اصلی اخذ می‌شوند: توسط انسان‌ها (که برایمان آشناست)، توسط ماشین‌ها (که داریم با آن آشنا می‌شویم) و به‌واسطه همکاری میان انسان‌ها و ماشین‌ها (که نه تنها ناآشنا که بی سابقه هم است). هوش مصنوعی همچنین در حال تبدیل کردن ماشین‌ها - که تا امروز ابزار ما بوده‌اند - به شرکای ماست. کم‌کم به هوش مصنوعی دستورالعمل‌های مخصوص کمتری پیرامون چگونگی دستیابی به اهداف تعیین شده از سوی ما خواهیم داد. از آن بیشتر، اهداف مبهمی به هوش مصنوعی ارائه می‌دهیم و می‌پرسیم «بر اساس نتیجه‌گیری‌های تو، چطور باید اقدام کنیم؟»

این تغییر نه ذاتاً تهدیدکننده است و نه رهایی‌بخش. باین حال به‌قدر کفایت متفاوت است که خط سیر جوامع و مسیر تاریخ را تغییر دهد. ادغام مداوم هوش مصنوعی در زندگی ما دنیایی را پدیدار خواهد کرد که در آن علی‌الظاهر اهداف ناممکن بشر دست‌یافتنی خواهند شد و دستاوردهایی که روزگاری صرفاً از آن انسان قلمداد می‌شد - شعر سرودن، کشف درمان

پزشکی - به دست ماشین‌ها یا با همکاری آن‌ها انجام می‌شود. این تحول با درگیر کردن تمامی عرصه‌ها در فرآیندهای مبتنی بر هوش مصنوعی آن‌ها را دگرگون خواهد ساخت، به شکلی که گاهی تبیین خط مرزها میان تصمیم‌گیری‌های صرفاً انسانی، تصمیم‌گیری‌های کاملاً مبتنی بر هوش مصنوعی و تصمیم‌گیری‌های ترکیبی انسان - هوش مصنوعی دشوار خواهد شد.

در قلمرو سیاست، جهان در حال ورود به عصری است که طی آن سیستم‌های هوش مصنوعی مبتنی بر بزرگ‌داده‌ها^۱ از جنبه‌هایی در حال رشد حکایت دارند: طراحی پیام‌های سیاسی؛ مناسب‌سازی و انتشار آن پیام‌ها به جمعیت‌های متنوع؛ ساخته و پرداخته‌سازی و به‌کارگیری اطلاعات نادرست توسط بازیگران متخاصم با هدف افکندن بذر اختلافات اجتماعی؛ طراحی و استقرار الگوریتم‌هایی برای تفحص، شناسایی و مقابله با اطلاعات نادرست و سایر اشکال داده‌های مضر. همزمان با افزایش نقش آفرینی هوش مصنوعی در تبیین «فضای اطلاعاتی» و سروشکل دادن به آن، پیش‌بینی نقش آن دشوارتر می‌شود. هوش مصنوعی در این فضا، همچون دیگر فضاها، گهگاه به گونه‌ای عمل می‌کند که حتی طراحان آن هم به‌صورت کلی می‌توانند آن را توضیح دهند؛ در نتیجه چشم‌انداز جامعه آزاد، و حتی اراده آزاد، ممکن است دستخوش تغییر شود. حتی اگر معلوم شود این تحولات بی‌خطر یا برگشت‌پذیرند باز هم این وظیفه بر دوش جوامع سراسر دنیا است که بتوانند این تغییرات را درک و سپس آن را با ارزش‌ها، ساختارها و قراردادهای اجتماعی خود وفق دهند.

نهادهای دفاعی و فرماندهان نظامی با تحولاتی مواجه‌اند که تقریباً همین اندازه عمیق هستند. زمانی که ارتش‌ها استراتژی و تاکتیک‌هایی را اتخاذ می‌کنند که به دست ماشین‌هایی طراحی شده که الگوهای بیشتر را - فراتر

از سربازها و استراتژیست‌های بشر - می‌فهمند، توازن قوا تغییر می‌کند و محاسبه آن به‌طور بالقوه دشوار می‌شود. اگر چنین ماشین‌هایی اجازه داشته باشند مستقلاً تصمیم به هدفگیری کنند، مفهوم سنتی دفاع و بازدارندگی - و به‌طور کلی قوانین جنگی - رو به اضمحلال می‌گذارند یا دست‌کم نیاز به تطبیق خواهند داشت.

در این دست‌مورد، شکاف‌های تازه‌ای درون و در میان جوامع پدیدار می‌شود؛ میان جوامعی که فن‌آوری جدید را به‌کار می‌گیرند با جوامعی که آن را برنمی‌گزینند یا فاقد ابزارهای لازم برای توسعه یا دستیابی به برخی از کاربردهای این دست‌فن‌آوری‌ها هستند. هنگامی که گروه‌ها یا ملت‌های مختلف مفاهیم یا کاربردهای متفاوتی از هوش مصنوعی اخذ می‌کنند، تجربه آن‌ها از واقعیت ممکن است چنان متفاوت باشد که پیش‌بینی یا اتصال آن تجربه‌ها به یکدیگر سخت و دشوار شود. جوامع همچنان که مشارکت میان انسان و ماشین را میان خود افزایش می‌دهند - با اهداف و مدل‌های آموزشی متفاوت و محدودیت‌های عملیاتی و اخلاقی بالقوه ناسازگار در رابطه با هوش مصنوعی - ممکن است در دامن رقابت، ناسازگاری‌های فنی و عدم درک متقابل هر چه بیشتر گرفتار بیایند. فن‌آوری که زمانی تصور می‌شد ابزاری است برای تعالی تفاوت‌های ملل و پراکندگی حقیقت‌عینی، احتمالاً در گذر زمان به روشی تبدیل شود که تمدن‌ها و فردها از طریق آن به درون واقعیت‌های متفاوت و غیرقابل درک فروافتند.

آلفازیرو مثالی عینی است. ثابت کرد که دست‌کم در بازی هوش مصنوعی در گیروبند محدودیت‌های دانش تثبیت‌شده بشر نیست. مسلماً آن سبکی که هوش مصنوعی پایه و اساس آلفازیرو است - یادگیری ماشینی که در آن الگوریتم‌ها روی شبکه‌های عمیق عصبی آموزش داده می‌شود - محدودیت‌های خود را دارد. اما ماشین‌ها در تعداد فزاینده کاربردها در حال ابداع راه‌حل‌هایی هستند که به نظر فراتر از حد تصور و تخیل انسان

است. در سال ۲۰۱۶ یکی از زیرمجموعه‌های کمپانی دیپ مایند، یعنی شرکت دیپ مایند اپلاید^۱، هوش مصنوعی‌ای طراحی کرد (که بر اساس بسیاری از اصول مشابه با آلفازیرو عمل می‌کرد) تا خنک‌سازی مراکز داده حساس به گرما در گوگل را بهینه کند. اگرچه برخی از بهترین مهندسان جهان پیش‌تر از عهده این مسئله برآمده بودند اما برنامه هوش مصنوعی دیپ مایند موفق شد خنک‌سازی را بهینه‌تر کند و هزینه انرژی را ۴۰ درصد دیگر کاهش دهد که نسبت به عملکرد انسان پیشرفتی بزرگ بود.^۲ زمانی که هوش مصنوعی در زمینه‌های مختلف و به‌منظور دستیابی به پیشروی‌های قابل مقایسه (با انسان) به‌کار گرفته شود جهان به‌ناگزیر تغییر خواهد کرد. نتیجه صرفاً اتخاذ روش‌هایی کارآمدتر برای انجام وظایف بشر نخواهد بود، بلکه در بسیاری از موارد هوش مصنوعی راه‌حل‌ها یا جهت‌دهی‌های جدیدی مطرح خواهد کرد که نشان از یک شکل متفاوت از یادگیری و ارزیابی منطقی، شکلی غیر بشری دارد.

از زمانی که عملکرد هوش مصنوعی در انجام وظیفه‌ای معین از عملکرد انسان پیشی بگیرد، عدم به‌کارگیری هوش مصنوعی، دست‌کم به‌عنوان دستیار انسان، به شکل نوعی گمراهی یا حتی سهل‌انگاری به چشم خواهد آمد. اینکه به یک شطرنج‌باز که هوش مصنوعی کمک حال آن است، توصیه شود مهره ارزشمندی را قربانی کند که شطرنج‌بازهای حرفه‌ای به‌طور سنتی حضورش در صفحه را ضروری می‌دانند اهمیت چندانی ندارد، اما اگر در زمینه امنیت ملی، هوش مصنوعی به یک فرمانده توصیه کند که شماری از غیرنظامیان

1. DeepMind Applied

2. Richard Evans and Jim Gao, "DeepMind AI Reduces Google Data Centre Cooling Bill by 40%," DeepMind blog, July 20, 2016, <https://deepmind.com/blog/article/deepmind-ai-reduces-google-data-centre-cooling-bill-40>.

یا منافع آن‌ها قربانی شوند تا، بر اساس محاسبات هوش مصنوعی، جان تعداد بسیار بیشتری حفظ شود چه؟ بر چه اساسی می‌توان آن قربانی کردن را نادیده گرفت؟ آیا الغای آن قابل توجیه است؟ آیا بشر همیشه می‌داند هوش مصنوعی چه محاسباتی انجام داده؟ آیا بشر می‌تواند انتخاب‌های ناخواسته هوش مصنوعی را تشخیص بدهد یا گزینه‌های ناخوشایند را درست به‌موقع معکوس کند؟ اگر در فهم عمیق منطق هر تصمیم هوش مصنوعی ناتوان باشیم آن‌گاه می‌بایست توصیه‌های آن را صرفاً بر مبنای باور و ایمان به کار بگیریم؟ اگر این کار را نکنیم، آیا مخاطره نکرده‌ایم که عملکردی بهتر از عملکرد خود را مختل کنیم؟ حتی اگر بتوانیم به عمق منطق، هزینه و تأثیر گزینه‌های ویژه پی ببریم، اما حریف ما نیز به همان اندازه به هوش مصنوعی متکی باشد چه؟ تعادل میان این ملاحظات چطور برقرار می‌شود یا در صورت لزوم شفاف می‌شود؟

هوش مصنوعی، هم در کامیابی آلفازیرو و هم در کشف هالیسین، در تبیین مسئله‌ای که حل کرده به انسان وابسته بوده است. هدف آلفازیرو این بود که در عین پیروی از قواعد بازی به پیروزی دست یابد. هدف هوش مصنوعی که هالیسین را کشف کرد آن بود که تا حد امکان تعداد زیادی پاتوژن را از بین ببرد: هرچه تعداد بیشتری پاتوژن بدون آسیب رساندن به (بدن) میزبان از بین برود موفق‌تر عمل کرده است. افزون بر این، تمرکز آن در قلمروی کاملاً خارج از دسترس بشر تعیین شد: به این هوش مصنوعی دستور داده شده به‌جای تعیین مسیرهای شناخته‌شده کشف دارو، در جستجوی رویکردهای نامکشوف باشد. هوش مصنوعی به موفقیت دست یافت زیرا آنتی‌بیوتیکی کشف کرد که پاتوژن‌ها را از بین برد. اما به‌خصوص نوآورانه و پیشگامانه بود چون گزینه‌های درمانی را گسترش می‌دهد و آنتی‌بیوتیکی جدید (وقوی) وارد عمل کرد که از طریق سازوکاری جدید کشف شد. همکاری نوینی میان انسان و ماشین در حال ظهور است؛ در وهله