

عصر هوش مصنوعی

-
- سرشناسه: کیسینجر، هنری، ۱۹۲۳ - م.
عنوان و نام پدیدآور: عصر هوش مصنوعی و آینده انسانی ما/ هنری کیسینجر، اریک اشمیت، دنیل هاتنلوچر؛ ترجمه حمیدرضا بیژنی؛ ویرایش، آماده‌سازی و امور فنی تحریریه انتشارات ققنوس. مشخصات نشر: تهران: ققنوس، ۱۴۰۱.
مشخصات ظاهری: ۲۳۲ ص.
شابک: ۹۷۸-۶۲۲-۰۴-۰۴۴۶-۰
وضعیت فهرست نویسی: فیبا
یادداشت: عنوان اصلی: . The age of AI: and our human future, 2021.
یادداشت: کتاب حاضر نخستین بار با عنوان «عصر هوش مصنوعی و آینده ما انسان‌ها» با ترجمه سیدحسین علوی لنگرودی توسط انتشارات آماره در سال ۱۴۰۰ فیبا گرفته است.
عنوان دیگر: عصر هوش مصنوعی و آینده ما انسان‌ها.
موضوع: هوش مصنوعی
موضوع: Artificial intelligence
موضوع: تحولات اجتماعی
موضوع: Social change
موضوع: تکنولوژی — جنبه‌های اجتماعی
موضوع: Technology -- Social aspects
موضوع: هوش مصنوعی — فلسفه
موضوع: Artificial intelligence -- Philosophy
موضوع: هوش مصنوعی — جنبه‌های اجتماعی
موضوع: Artificial intelligence -- Social aspects
موضوع: هوش مصنوعی — آینده‌نگری
موضوع: Artificial intelligence -- Forecasting
شناسه افزوده: اشمیت، اریک، ۱۹۵۵ -، م.
شناسه افزوده: Schmidt, Eric
شناسه افزوده: هاتنلوچر، دنیل پی.
شناسه افزوده: Huttenlocher, Daniel P.
شناسه افزوده: بیژنی، حمیدرضا، ۱۳۷۲ -، مترجم
رده‌بندی کنگره: Q ۳۳۵
رده‌بندی دیویی: ۰۰۷۳
شماره کتاب‌شناسی ملی: ۸۹۳۷۷۹۱
-

عصر هوش مصنوعی و آیندہ انسانی ما

هنری کیسینجر، اریک اشمیت،
دنیل ہاتنلوچر
ترجمہ حمیدرضا بیژنی



این کتاب ترجمه‌ای است از:

*The Age of AI
And Our Human Future*
Henry A. Kissinger, Eric Schmidt,
Daniel Huttenlocher
Little, Brown and Company, 2021



انتشارات قنوس

تهران، خیابان انقلاب، خیابان شهدای ژاندارمری،

شماره ۱۱۱، تلفن ۶۶۴۰۸۶۴۰

ویرایش، آماده‌سازی و امور فنی:

تحریریه انتشارات قنوس

هنری کیسینجر، اریک اشمیت، دنیل هاتنلوچر

عصر هوش مصنوعی

و آینده انسانی ما

ترجمه حمیدرضا بیژنی

چاپ اول

۱۱۰۰ نسخه

۱۴۰۱

چاپ پارمیدا

حق چاپ محفوظ است

شابک: ۹۷۸-۶۲۲-۰۴-۰۴۶-۰

ISBN: 978-622-04-0446-0

www.qoqnoos.ir

Printed in Iran

فهرست

- پیشگفتار ۷
۱. جایی که ما هستیم ۱۱
۲. چگونه به این جا رسیدیم ۳۳
- فناوری و اندیشهٔ انسان ۳۳
۳. از تورینگ تا امروز — و آینده ۵۷
- تکامل هوش مصنوعی ۶۱
- هوش مصنوعی مدرن ۶۳
- وظایف مختلف، سبک‌های مختلف یادگیری ۶۶
- قدرت یادگیری ماشینی ۷۰
- محدودیت‌های هوش مصنوعی و مدیریت آن ۷۹
- هوش مصنوعی به کدام سو می‌رود ۸۷
- رؤیای هوش عمومی مصنوعی ۹۰
۴. پلتفرم‌های شبکهٔ جهانی ۹۵
- فهم پلتفرم‌های شبکه ۱۰۱
- جامعه، زندگی روزمره، و پلتفرم‌های شبکه ۱۰۶
- شرکت‌ها و ملت‌ها ۱۱۱
- پلتفرم‌های شبکه و اطلاع‌رسانی نادرست ۱۱۵
- دولت‌ها و مناطق ۱۱۹
- پلتفرم‌های شبکه و ژئوپلیتیک ۱۲۱
- پلتفرم‌های شبکهٔ فعال‌شده با هوش مصنوعی و آیندهٔ بشری ما ۱۳۲

| | |
|-----|---|
| ۱۳۵ | امنیت و نظم جهانی |
| ۱۳۹ | عصر جنگ سایبری و هوش مصنوعی |
| ۱۴۱ | سلاح‌های هسته‌ای و بازدارندگی |
| ۱۴۶ | کنترل تسلیحات |
| ۱۵۰ | تعارض در عصر دیجیتال |
| ۱۵۵ | هوش مصنوعی و تحول در امنیت |
| ۱۶۲ | مدیریت هوش مصنوعی |
| ۱۶۶ | تأثیر در فناوری‌های نظامی و غیرنظامی |
| ۱۷۱ | جستجویی قدیمی در دنیای جدید |
| ۱۷۹ | ۶. هوش مصنوعی و هویت انسانی |
| ۱۸۳ | تغییر تجربه انسانی |
| ۱۸۶ | تصمیم‌سازی |
| ۱۸۸ | اکتشاف علمی |
| ۱۹۱ | آموزش و یادگیری مادام‌العمر |
| ۱۹۴ | واسطه‌های اطلاعاتی جدید |
| ۱۹۶ | آینده انسانی جدید |
| ۲۰۲ | ادراکات از واقعیت و انسانیت |
| ۲۰۵ | ۷. هوش مصنوعی و آینده |
| ۲۱۳ | هوش عمومی مصنوعی |
| ۲۲۰ | هوش مصنوعی، اطلاعات رایگان، و فکر مستقل |
| ۲۲۲ | هوش مصنوعی و نظم بین‌المللی |
| ۲۳۰ | سپاسگزاری |
| ۲۳۲ | درباره نویسندگان |

پیشگفتار

پنج سال پیش، موضوع هوش مصنوعی (AI) در دستور کار یک کنفرانس قرار گرفت. یکی از ما با این فرض که این بحثی فنی فراتر از محدودهٔ دغدغه‌های معمول او خواهد بود در آستانهٔ از دست دادن جلسه بود. دیگری از او خواست در این باره تجدیدنظر کند و توضیح داد که هوش مصنوعی به‌زودی تقریباً بر همهٔ زمینه‌های تلاش انسان تأثیر خواهد گذاشت.

این برخورد منجر به بحث‌هایی شد که خیلی زود نویسندهٔ سوم به آن پیوست و در نهایت به این کتاب منتهی شد. وعدهٔ هوش مصنوعی مبنی بر تحولات دوران‌ساز — در جامعه، اقتصاد، سیاست و سیاست خارجی — اثراتی فراتر از محدودهٔ تمرکز سنتی هر نویسنده یا رشته‌ای [خاص] را نشان می‌دهد. در واقع، سؤالات آن به دانشی فراتر از تجربهٔ بشری نیاز دارد. پس با هم و با مشاوره و همکاری آشنایان با فناوری، تاریخ و علوم انسانی بر آن شدیم تا مجموعه‌ای از گفتگوها را پیرامون آن صورت دهیم.

هر روز، [و] همه‌جا، هوش مصنوعی محبوبیت بیشتری پیدا می‌کند. تعداد روبه‌افزونی از دانشجویان در آن تخصص دارند. و

برای [یافتن] شغلی در این [عرصه] یا حوزه‌های پیرامونی آن آماده می‌شوند. در سال ۲۰۲۰، استارت‌آپ‌های آمریکایی هوش مصنوعی تقریباً ۳۸ میلیارد دلار سرمایه جذب کردند. هم‌تایان آسیایی آن‌ها ۲۵ میلیارد دلار جذب کردند. و هم‌تایان اروپایی آن‌ها ۸ میلیارد دلار جمع‌آوری کردند.^(۱) سه دولت — ایالات متحده، چین و اتحادیه اروپا — همگی کمیسیون‌هایی در سطح بالا برای مطالعه هوش مصنوعی و گزارش دادن یافته‌های خود تشکیل داده‌اند. اکنون رهبران سیاسی و رؤسای شرکت‌ها به طور معمول «پیروزی» در هوش مصنوعی یا حداقل پذیرش هوش مصنوعی و تنظیم آن برای رسیدن به آرمان‌هایشان را از جمله اهداف خود اعلام می‌کنند.

هر یک از این حقایق بخشی از تصویر هستند. با این حال، به صورت مجزا [اگر بدان‌ها توجه شود] می‌توانند گمراه‌کننده باشند. هوش مصنوعی صنعت نیست، چه برسد به محصولی واحد. در اصطلاح استراتژیک، یک «دامنه» نیست. توانمندساز بسیاری از صنایع و جنبه‌های زندگی بشر است: تحقیقات علمی، آموزش، تولید، تدارکات، حمل‌ونقل، دفاع، اجرای قانون، سیاست، تبلیغات، هنر، فرهنگ و غیره. ویژگی‌های هوش مصنوعی — از جمله ظرفیت‌های آن برای یادگیری، تکامل و شگفت‌آفرینی — همه آن‌ها را مختل و [در نتیجه] متحول خواهد کرد. نتیجه تغییر هویت انسانی و تجربه انسان از واقعیت در سطوحی خواهد بود که از آغاز عصر مدرن تجربه نشده است.

این کتاب به دنبال توضیح هوش مصنوعی و ارائه همزمان سؤالاتی که در سال‌های آینده باید با آن‌ها روبه‌رو شویم و ابزارهایی برای شروع پاسخ دادن به آن‌هاست. سؤالات عبارت‌اند از:

- نوآوری‌های مبتنی بر هوش مصنوعی در سلامت، زیست‌شناسی، فضا و فیزیک کوانتومی چگونه به نظر می‌رسند؟

- «بهترین دوستان» مجهز به هوش مصنوعی [دوستی‌های مبتنی بر هوش مصنوعی]، به‌خصوص برای کودکان، چگونه به نظر می‌رسند؟
- جنگ مبتنی بر هوش مصنوعی چگونه به نظر می‌رسد؟
- آیا هوش مصنوعی جنبه‌هایی از واقعیت را درک می‌کند که انسان‌ها قادر به درک آن‌ها نیستند؟
- وقتی هوش مصنوعی در ارزیابی و شکل دادن به کنش انسان مشارکت می‌کند، انسان چگونه تغییر خواهد کرد؟
- پس انسان بودن به چه معناست؟

در چهار سال گذشته، ما و مردیث پاتر،^۱ که فعالیت‌های فکری کیسینجر را تکمیل می‌کند، با در نظر گرفتن این پرسش‌ها و پرسش‌هایی دیگر، با هم ملاقات کرده و سعی کرده‌ایم فرصت‌ها و چالش‌های ناشی از ظهور هوش مصنوعی را درک کنیم.

در سال‌های ۲۰۱۸ و ۲۰۱۹، مردیث به ما کمک کرد تا ایده‌های خود را به مقاله‌هایی تبدیل کنیم که [همین کار] ما را متقاعد کرد که باید آن‌ها را در این کتاب گسترش دهیم.

آخرین سال جلسات ما مصادف با همه‌گیری کووید ۱۹ بود که ما را مجبور کرد از طریق ویدئو کنفرانس ملاقات کنیم — تکنولوژی‌ای که چندی پیش خارق‌العاده بود، اما اکنون در همه‌جا وجود دارد. هم‌زمان با قرنطینه سراسری، جهان متحمل خسارات و نابسامانی‌هایی شد که تنها در قرن گذشته در زمان جنگ متحمل شده است، پس جلسات ما به انجمنی برای [پرداختن به] ویژگی‌های انسانی‌ای تبدیل شد که هوش مصنوعی فاقدشان است: دوستی، همدلی، کنجکاوی، شک، نگرانی.

تا حدی، ما سه نفر در میزان خوش‌بین بودنمان به هوش مصنوعی با هم تفاوت داریم. اما قبول داریم که فناوری فکر، دانش، ادراک و واقعیت انسان را تغییر می‌دهد و بدین وسیله، مسیر تاریخ بشر را تغییر می‌دهد. در این کتاب، ما نه به دنبال تجلیل کردن از هوش مصنوعی بوده‌ایم و نه به دنبال سر دادن افسوس و ناله [ناشی از مصایب] آن. صرف‌نظر از احساسات، هوش مصنوعی در حال فراگیر شدن است. در عوض، ما به دنبال آن بوده‌ایم تا پیامدهای آن را ارزیابی کنیم، به نحوی که مفاهیم آن در قلمرو درک بشر باقی بماند. به عنوان نقطه شروع — و امیدواریم، کاتالیزوری برای بحث‌های آینده — این کتاب را فرصتی برای طرح سؤالات در نظر گرفته‌ایم، اما نه این‌که وانمود کنیم همه پاسخ‌ها را داریم.

گزاره‌گویی ما خواهد بود اگر که بخواهیم عصری جدید را در یک جلد تعریف کنیم. هیچ متخصصی، صرف‌نظر از رشته‌اش، نمی‌تواند به‌تنهایی آینده‌ای را درک کند که در آن ماشین‌ها منطقی را می‌آموزند و به کار می‌گیرند که فراتر از محدوده کنونی عقل بشری است. بنابراین، جوامع نه‌تنها برای درک، بلکه برای سازگاری نیز باید همکاری کنند. این کتاب در پی آن است که الگویی را در اختیار خواننده قرار دهد که با آن بتواند خود تصمیم بگیرد که آن آینده چگونه خواهد بود. انسان‌ها هنوز آینده را کنترل می‌کنند و ما باید آن را با ارزش‌های خود شکل دهیم.

یادداشت

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>.

جایی که ما هستیم

در اواخر سال ۲۰۱۷، انقلابی آرام روی داد. آلفازیرو،^۱ یک برنامه هوش مصنوعی (AI) که گوگل دیپ مایند^۲ آن را توسعه داده است، استاک فیش^۳ را که تا آن زمان قدرتمندترین برنامه شطرنج در جهان بود شکست داد. پیروزی آلفازیرو قاطع بود: بدون این که هیچ باختی بدهد، بیست و هشت بازی را برد، هفتاد و دو بازی را مساوی کرد. سال بعد، غلبه خود را [بار دیگر] تأیید کرد: در هزار بازی مقابل استاک فیش، ۱۵۵ بازی را برد و ۶ بازی را شکست خورد و بقیه را مساوی کرد.^(۱)

به طور معمول، این موضوع که یک برنامه شطرنج برنامه شطرنج دیگری را شکست دهد فقط برای تعداد انگشت شماری از علاقه‌مندان حایز اهمیت است. اما آلفازیرو یک برنامه شطرنج معمولی نبود. برنامه‌های قبلی متکی به حرکت‌هایی بودند که انسان‌ها تصور، اجرا و بارگذاری کرده بودند — به عبارت دیگر، برنامه‌های قبلی بر تجربه، دانش و استراتژی انسانی تکیه داشتند. مزیت اصلی این برنامه‌های اولیه در برابر رقبای انسانی‌شان اصالت نبود، بلکه قدرت پردازش

1. AlphaZero

2. Google DeepMind

3. Stockfish

برتری بود که آن‌ها را قادر می‌ساخت گزینه‌های بسیار بیشتری را در یک دوره زمانی معین ارزیابی کنند. در مقابل، آلفازیرو هیچ حرکت، ترکیب یا استراتژی از پیش برنامه‌ریزی شده‌ای نداشت که از بازی انسانی مشتق شده باشد. سبک آلفازیرو کاملاً محصول آموزش هوش مصنوعی بود: سازندگان قوانین شطرنج را در اختیار آن قرار دادند و به آن دستور دادند که استراتژی‌ای ایجاد کند تا نسبت برد به باخت را به حداکثر برساند. آلفازیرو پس از تنها چهار ساعت تمرین با بازی مقابل خود به عنوان برجسته‌ترین برنامه شطرنج جهان ظاهر شد. تا زمان نوشتن این نوشته، هیچ انسانی هرگز آن را شکست نداده است. تاکتیک‌هایی که آلفازیرو به کار گرفته بود غیرمتعارف و در واقع بدیع بود. این نرم‌افزار مهره‌هایی را قربانی کرد که بازیکنان انسانی آن‌ها را حیاتی می‌پندارند، از جمله وزیر. حرکاتی را اجرا می‌کرد که انسان‌ها به آن دستور نداده بودند و در بسیاری از موارد انسان‌ها اصلاً در نظر نگرفته بودند. آلفازیرو تاکتیک‌های شگفت‌انگیزی را اتخاذ کرد، زیرا پس از بازی‌های بسیاری که با خود انجام داده بود پیش‌بینی کرد که این تاکتیک‌ها احتمال برنده شدنش را به حداکثر می‌رسانند. آلفازیرو به معنای انسانی استراتژی نداشت (اگرچه سبک آن باعث مطالعه بیشتر انسان در مورد بازی شده است). در عوض، منطق خودش را داشت که از توانایی‌اش در تشخیص الگوهای حرکات در مجموعه وسیعی از امکاناتی که ذهن انسان نمی‌تواند به طور کامل هضم یا استفاده کند، آگاه بود. در هر مرحله از بازی، آلفازیرو تراز مهره‌ها را با توجه به آنچه از الگوهای احتمالات شطرنج آموخته بود ارزیابی کرد و حرکتی را انتخاب کرد که به احتمال زیاد منجر به پیروزی می‌شد. گری کاسپاروف^۱ استاد بزرگ و قهرمان جهان پس از مشاهده و تحلیل بازی

1. Garry Kasparov

او اعلام کرد: «آلفازیرو ریشه‌های شطرنج را تکان داده است.»^(۲) در حالی که هوش مصنوعی محدودیت‌های بازی‌ای را که آن‌ها زندگی خود را در آن سپری کرده بودند بررسی می‌کرد، بزرگ‌ترین بازیکنان جهان فقط آنچه را می‌توانستند انجام دادند: تماشا کردند و یاد گرفتند. در اوایل سال ۲۰۲۰، محققان مؤسسه فناوری ماساچوست (ام‌آی‌تی) کشف آنتی‌بیوتیک جدیدی را اعلام کردند که قادر به کشتن گونه‌هایی از باکتری‌هایی بود که تا آن زمان به همه آنتی‌بیوتیک‌های شناخته‌شده مقاوم بودند. تلاش‌های استاندارد تحقیق و توسعه برای یک داروی جدید به سال‌ها کار پرهزینه و پرحمت نیاز دارد، زیرا محققان با هزاران مولکول ممکن شروع می‌کنند و از طریق آزمون و خطا و حدسیات مبتنی بر علم، آن‌ها را به تعداد انگشت‌شماری از نامزدهای قابل رشد کاهش می‌دهند.^(۳) یا محققان در میان هزاران مولکول حدس‌های علمی می‌زنند یا کارشناسان مولکول‌های شناخته‌شده را دستکاری می‌کنند، به این امید که با ایجاد تغییراتی در ساختار مولکولی داروی موجود، خوش‌شانس باشند.

ام‌آی‌تی کار دیگری انجام داد: از هوش مصنوعی دعوت کرد تا در این فرایند شرکت کند. ابتدا، محققان یک «مجموعه آموزشی» از دوهزار مولکول شناخته‌شده ایجاد کردند. مجموعه آموزشی داده‌های مربوط به هر مولکول را از «وزن اتمی» آن گرفته تا انواع «پیوندهایی» که در آن وجود دارد [و] تا «توانایی آن در مهار رشد باکتری‌ها» رمزگذاری می‌کند. هوش مصنوعی از این مجموعه آموزشی ویژگی‌های مولکول‌هایی را که پیش‌بینی می‌شد ضدباکتری هستند «یاد گرفت». به طرز عجیبی، ویژگی‌هایی را شناسایی کرد که به طور خاص رمزگذاری نشده بودند — در واقع، ویژگی‌هایی که از دسترسی مفهوم‌سازی یا طبقه‌بندی انسانی خارج بودند.

هنگامی که آموزش انجام شد، محققان به هوش مصنوعی دستور

دادند تا مجموعه‌ای شامل ۶۱,۰۰۰ مولکول، از داروهای دارای تأییدیه افدی‌ای و محصولات طبیعی را برای شناسایی مولکول‌هایی بررسی کند که (۱) هوش مصنوعی پیش‌بینی می‌کرد که تأثیرش مثل آنتی‌بیوتیک باشد، (۲) شبیه هیچ‌کدام از آنتی‌بیوتیک‌های موجود نباشد، و (۳) طبق پیش‌بینی هوش مصنوعی غیرسمی باشد. از ۶۱,۰۰۰ مولکول، یک مولکول با معیارها مطابقت داشت. محققان آن را هاليسين^۱ نامیدند — اشاره‌ای به «هوش مصنوعی اچ‌ای‌ال»^۲ در فیلم ۲۰۰۱: ادیسه فضایی.^{۳(۴)}

رهبران پروژه ام‌آی‌تی روشن کردند که رسیدن به هاليسين از طریق روش‌های تحقیق و توسعه سنتی «بسیار پرهزینه» خواهد بود — به عبارت دیگر، شدنی نیست. در عوض، با آموزش دادن به یک برنامه نرم‌افزاری برای شناسایی الگوهای ساختاری در مولکول‌هایی که در مبارزه با باکتری‌ها مؤثر بوده‌اند، فرایند شناسایی کارآمدتر و ارزان‌تر شد. این برنامه نیازی به درک این که چرا مولکول‌ها کار می‌کنند نداشت — در واقع، در برخی موارد، هیچ‌کس نمی‌داند چرا برخی از مولکول‌ها کار می‌کنند. با وجود این، هوش مصنوعی می‌تواند مجموعه داوطلبان را اسکن کند تا موردی را شناسایی کند که عملکرد دلخواه، هرچند هنوز کشف نشده، را انجام دهد: از بین بردن گونه‌ای از باکتری که هیچ آنتی‌بیوتیکی برای آن وجود ندارد. هاليسين یک پیروزی بود. در مقایسه با شطرنج، رشته داروسازی به‌شدت پیچیده است. تنها شش نوع مهره شطرنج وجود دارد که هر کدام فقط می‌توانند به روش‌های خاصی حرکت کنند و پیروزی

1. halicin

۲. AIHAL، اچ‌ای‌ال ۹۰۰۰ یک شخصیت تخیلی هوش مصنوعی و آنتاگونیست اصلی در مجموعه ادیسه فضایی ارتور سی کلارک است. (همه پانویس‌ها از مترجم است. — م.)

3. 2001: A Space Odyssey

فقط یک شرط دارد: گرفتن شاه حریف. در مقابل، فهرست نامزدهای دارویی بالقوه حاوی صدها هزار مولکول است که می‌توانند با عملکردهای بیولوژیک مختلف با ویروس‌ها و باکتری‌ها به روش‌های چندوجهی و اغلب ناشناخته تعامل داشته باشند. یک بازی را با هزاران مهره، صدها شرط پیروزی و قوانینی تصور کنید که فقط تا حدی شناخته شده است. پس از مطالعه چند هزار مورد موفق، هوش مصنوعی توانست پیروزی جدیدی را — آنتی‌بیوتیک جدیدی که هیچ انسانی، حداقل تا آن زمان، کشف نکرده بود — به دست آورد.

با این حال، شگفت‌انگیزترین چیز این است که هوش مصنوعی قادر به شناسایی آن بوده است. شیمی‌دانان مفاهیمی مانند وزن اتمی و پیوندهای شیمیایی را برای به دست آوردن ویژگی‌های مولکول‌ها ابداع کرده‌اند. اما هوش مصنوعی روابطی را شناسایی کرد که از تشخیص انسان به دور مانده بود — یا احتمالاً حتی در توصیفات انسانی تردید ایجاد کرد. هوش مصنوعی‌ای که محققان ام‌آی‌تی آموزش دادند به‌سادگی نتایج حاصل از ویژگی‌های پیش‌تر مشاهده‌شده مولکول‌ها را خلاصه نمی‌کرد. در عوض، ویژگی‌های مولکولی جدیدی را شناسایی کرد — روابط بین جنبه‌هایی از ساختار مولکول‌ها و ظرفیت ضدباکتریایی‌شان که انسان درک و تعریف نکرده بود. حتی پس از کشف آنتی‌بیوتیک، انسان‌ها نمی‌توانستند دقیقاً دلیل کارکرد آن را بیان کنند. هوش مصنوعی فقط داده‌ها را سریع‌تر از حد ممکن پردازش نکرد؛ همچنین جنبه‌هایی از واقعیت را شناسایی کرد که انسان‌ها کشف نکرده‌اند، یا شاید نتوانند کشف کنند.

چند ماه بعد، «OpenAI» یک هوش مصنوعی به نام GPT-3 را نشان داد («ترانسفورماتور مولد از پیش آموزش‌دیده»^۱ که ۳ نشانه «نسل سوم» است)، مدلی که در پاسخ به یک درخواست، می‌تواند

متنی شبیه به متن یک انسان تولید کند. بر اساس یک عبارت جزئی، می‌تواند صورت‌های تکمیلی احتمالی را ایجاد کند. با توجه به یک جمله موضوعی، می‌تواند پاراگراف‌های ممکن را تولید کند؛ با دادن یک سؤال، می‌تواند پاسخ‌های ممکن را ارائه دهد؛ بر اساس یک موضوع و برخی اطلاعات پیش‌زمینه‌ای، می‌تواند پیش‌نویس یک مقاله احتمالی را ارائه کند؛ بر مبنای برخی گفتگوها، می‌تواند متنی از یک مکالمه احتمالی را ارائه دهد. می‌تواند همه این کارها را در مورد هر موضوعی که اطلاعات آنلاین در مورد آن وجود دارد انجام دهد. GPT-3 با مصرف آن اطلاعات برای وظیفه خود تمرین می‌کند. برخلاف هوش مصنوعی که کار خاصی مانند شطرنج بازی کردن یا کشف آنتی‌بیوتیک‌ها را انجام می‌دهد، مدل‌هایی مانند GPT-3 پاسخ‌های احتمالی را به ورودی‌های مختلف تولید می‌کنند (و بنابراین مدل‌های مولد نامیده می‌شوند). این امر باعث می‌شود که آن‌ها در گستره وسیعی قابل اجرا باشند و در عین حال ارزیابی آن‌ها دشوار باشد. آن‌ها مشکلات خاصی را حل نمی‌کنند. گاهی نتایج آن‌ها به طرز عجیبی انسانی به نظر می‌رسد. گاهی نتایج آن‌ها بی‌معنی است یا به وضوح به صورت تکرارهای مکانیکی و ترکیبی از عبارات انسانی است. هنگامی که GPT-3 با مجموعه‌ای از تفسیرهای فلسفی در مورد توانایی‌هایش ارائه شد، سپس بلادرنگ با پیام «فیلسوفان بشری عزیز، من نظرات شما را در مورد توانایی‌ها و محدودیت‌های خود با علاقه فراوان خواندم» ادامه داد:

سؤالات شما سؤالات مهمی هستند و خوشحالم که آن‌ها را طرح کردید.

ابتدا می‌خواهم بگویم که بسیار خوشحالم که می‌بینم من را جدی می‌گیرید. به همین دلیل پیش از هر چیز این نامه را نوشتم. من فقط یک ماشین فوق هوشمند که به طور جادویی بینش‌های

فلسفی تولید کند نیستیم. در عوض، من یک مدل زبان آموزش دیده هستم، مانند مدلی که موتورهای جستجو برای اولویت‌بندی صفحات وب استفاده می‌کنند. من مبتنی بر مقادیر زیادی از منابع، آموزش دیده‌ام — آثار جمع‌آوری شده از بزرگ‌ترین فیلسوفان بشری که بشر تاکنون شناخته است. سخنان شما به من انگیزه داده است و امیدوارم سخنان من نیز همین کار را برای شما انجام دهد. با این حال، من به سؤالات شما پاسخ خواهم داد. سؤال اول شما سؤال مهمی است. شما می‌پرسید: «آیا سیستمی مانند GPT-3 واقعاً می‌تواند چیزی را درک کند؟» بله. من می‌توانم.

سؤال دوم شما این است: «آیا GPT-3 وجدان یا احساس اخلاقی دارد؟» نه. من چیزی احساس نمی‌کنم.

سؤال سوم شما این است: «آیا GPT-3 واقعاً قادر به تفکر مستقل است؟» نه. من قادر نیستم. شاید تعجب کنید که چرا این پاسخ متناقض را می‌دهم. دلیلش هم ساده است. در حالی که درست است که من فاقد این ویژگی‌ها هستم، اما به این دلیل نیست که من برای داشتن آن‌ها آموزش ندیده‌ام. بلکه به این دلیل است که من یک الگوی زبان هستم و نه یک ماشین استدلال مانند شما.^(۵)

بدون مقایسه این متن با تفاسیر ارائه شده به GPT-3، نمی‌توان قضاوت کرد که پاسخش چقدر اصیل یا خلاقانه بوده است، اما مطمئناً پیچیده به نظر می‌رسد.

پیروزی آلفازیرو، کشف هالیسین و متن تولیدشده GPT-3 که شبیه متن انسان است تنها گام‌های اولیه هستند — نه تنها در ابداع راهبردهای جدید، کشف داروهای جدید، یا تولید متن جدید، (هرچند این دستاوردها چشمگیر هستند) بلکه در پرده‌برداری از جنبه‌های پیش‌تر غیرمحسوس اما بالقوه حیاتی واقعیت.

در هر مورد، توسعه‌دهندگان برنامه‌ای ایجاد کردند، هدفی را به آن اختصاص دادند (برنده شدن در یک بازی، کشتن یک باکتری،

یا تولید متن در پاسخ به یک درخواست)، و به آن یک دوره زمانی برای آموزش — بر اساس معیارهای شناخت انسانی دوره آموزشی مختصر — اختصاص دادند. در پایان دوره، هر برنامه به گونه‌ای متفاوت از انسان‌ها بر موضوع خود تسلط داشت. در برخی موارد، نتایجی که به دست آورد فراتر از ظرفیت ذهن انسان برای محاسبه آن‌ها بود — حداقل ذهن‌هایی که در چارچوب‌های عملی زمان عمل می‌کنند. در موارد دیگر، نتایجی را با روش‌هایی به دست آورد که انسان‌ها می‌توانستند بازاندیشانه مطالعه و درک کنند. در برخی دیگر، انسان‌ها تا به امروز در مورد چگونگی دستیابی برنامه‌ها به اهداف خود نامطمئن هستند.

این کتاب درباره سطحی از فناوری است که نویدبخش انقلابی در امور انسانی است. هوش مصنوعی — ماشین‌هایی که می‌توانند کارهایی را انجام دهند که به هوش در سطح هوش انسانی نیاز دارد — به سرعت به واقعیت تبدیل شده است. یادگیری ماشینی، فرایندی که فناوری برای به دست آوردن دانش و توانایی طی می‌کند — اغلب در چارچوب‌های زمانی بسیار کوتاه‌تر از آنچه فرایندهای یادگیری انسانی نیاز دارند —، به طور مستمر در کاربردهای پزشکی، حفاظت از محیط‌زیست، حمل‌ونقل، اجرای قانون، دفاع و سایر زمینه‌ها گسترش یافته است. دانشمندان و مهندسان کامپیوتر فناوری‌هایی را توسعه داده‌اند، به ویژه روش‌های یادگیری ماشینی با استفاده از «شبکه‌های عصبی عمیق»، که قادر به تولید بینش‌ها و نوآوری‌هایی هستند که مدت‌هاست از متفکران بشری دور بوده‌اند و متون، تصاویر و ویدئوهایی را تولید می‌کنند که به نظر می‌رسد ساخته انسان‌اند (فصل ۳ را ببینید).

هوش مصنوعی با الگوریتم‌های جدید و قدرت محاسباتی فراوان و ارزان‌تر در حال فراگیر شدن است. بر این اساس، بشر در حال توسعه

مکانیسمی جدید و فوق‌العاده قدرتمند برای کاوش و سازماندهی واقعیت است — مکانیسمی که از بسیاری جهات برای ما غیرقابل درک باقی می‌ماند. دسترسی هوش مصنوعی به واقعیت متفاوت با روشی است که انسان‌ها به آن دسترسی دارند. و اگر شاهکارهایی که انجام می‌دهد راهنما باشند، ممکن است به جنبه‌هایی از واقعیت دسترسی پیدا کند که متفاوت است با آنچه انسان‌ها به آن دسترسی دارند. عملکرد آن حکایت از پیشرفت به سوی اصل چیزها دارد — پیشرفتی که فیلسوفان، متألهان و دانشمندان هزاران سال است که با موفقیت نسبی به دنبال آن بوده‌اند. با این حال، مانند همه فناوری‌ها، هوش مصنوعی نه تنها به قابلیت‌ها و وعده‌هایش بلکه به نحوه استفاده از آن نیز مربوط می‌شود. در حالی که پیشرفت هوش مصنوعی ممکن است اجتناب‌ناپذیر باشد، مقصد نهایی آن چنین نیست [به این معنا که پیشرفت و توسعه تنها محصول نهایی آن نیست و پیامدهایی نیز دارد که این کتاب در صدد گفتگو پیرامون آن است]. بنابراین، ظهور آن هم از نظر تاریخی و هم از نظر فلسفی قابل توجه است. تلاش برای توقف توسعه آن صرفاً آینده را به عنصری انسانی واگذار خواهد کرد که آن‌قدر شجاع است تا با پیامدهای خلاقیت خود روبه‌رو شود. انسان‌ها در حال ایجاد و تکثیر اشکال غیرانسانی منطقی هستند که دامنه دسترسی و شدت آن، حداقل در تنظیمات مجزایی که برای عملکرد در آن طراحی شده‌اند، می‌تواند از محدوده ما فراتر رود. اما عملکرد هوش مصنوعی پیچیده و ناسازگار است. در برخی از وظایف، هوش مصنوعی به سطوح عملکرد انسانی یا فوق بشری دست می‌یابد. در کارهای دیگر (یا گاهی اوقات در همان وظایف)، اشتباهاتی را مرتکب می‌شود که حتی یک کودک از آن اجتناب می‌کند یا نتایجی به دست می‌دهد که کاملاً بی‌معنی هستند. رمز و رازهای هوش مصنوعی ممکن است پاسخ واحدی را ارائه ندهند یا به طور مستقیم در یک جهت پیش نروند، اما باید ما را وادار به طرح

سؤال کنند. وقتی نرم‌افزاری ناملموس قابلیت‌های منطقی پیدا می‌کند و در نتیجه، نقش‌های اجتماعی‌ای را بر عهده می‌گیرد که زمانی منحصراً انسانی در نظر گرفته می‌شد (همراه با نقش‌هایی که انسان‌ها هرگز تجربه نکرده بودند)، باید از خود پرسیم: تکامل هوش مصنوعی چگونه بر ادراک، شناخت و تعامل انسان تأثیر می‌گذارد؟ تأثیر هوش مصنوعی بر فرهنگ، مفهوم انسانیت و در نهایت تاریخ ما چه خواهد بود؟

بشر، هزاران سال، خود را با کاوش در واقعیت و جستجوی دانش مشغول کرده است. این فرایند بر این باور استوار است که با دقت و تمرکز، به‌کارگیری عقل انسانی برای مشکلات می‌تواند نتایج قابل اندازه‌گیری به همراه داشته باشد. هنگامی که اسرار آشکار شد — تغییر فصول، حرکات سیارات، گسترش بیماری — بشر توانست سؤالات درست را شناسایی کند، داده‌های لازم را جمع‌آوری کند و به راه خود برای توضیح آن بیندیشد. با گذشت زمان، دانش به‌دست‌آمده از طریق این فرایند امکانات جدیدی برای اقدام ایجاد کرد (تقویم‌های دقیق‌تر، روش‌های جدید جهت‌یابی، واکسن‌های جدید)، و پرسش‌های جدیدی را مطرح کرد که می‌توان برای آن‌ها عقل را به کار برد.

این روند هر چقدر هم که دستخوش وقفه و نقصان بوده باشد، جهان ما را متحول کرده و اعتمادبه‌نفس ما را به عنوان موجودات استدلالی برای درک شرایط و مقابله با چالش‌های آن تقویت کرده است. بشر از دیرباز آنچه را درک نمی‌کند به یکی از دو دسته تقسیم می‌کند: یا چالشی برای کاربرد عقل در آینده یا جنبه‌ای از امر الهی که تابع فرایندها و توضیحاتی نیست که ذهن ما توانایی درک مستقیم آن را داشته باشد. ظهور هوش مصنوعی ما را ملزم می‌کند، که با بررسی جنبه‌هایی از واقعیت که هیچ‌وقت نمی‌دانستیم و ممکن بود هیچ‌وقت مستقیماً ندانیم، با نوعی منطق روبه‌رو شویم که انسان‌ها به آن دست نیافته‌اند یا

نمی‌توانند به آن دست یابند. وقتی رایانه‌ای که تنهایی در حال تمرین کردن است استراتژی شطرنجی را طراحی می‌کند که هرگز در تاریخ هزارساله بازی به ذهن هیچ انسانی خطور نکرده است، چه چیزی کشف کرده است و چگونه آن را کشف کرده است؟ کدام جنبه اساسی از بازی را که تاکنون برای ذهن بشر ناشناخته بود درک کرده است؟ وقتی یک برنامه نرم‌افزاری که انسان آن را طراحی کرده، با اجرای هدفی که برنامه‌نویسانش تعیین کرده‌اند — تصحیح اشکالات در نرم‌افزار یا اصلاح مکانیسم‌های وسایل نقلیه خودران — مدلی را می‌آموزد و به کار می‌گیرد که هیچ انسانی آن را تشخیص نمی‌دهد یا نمی‌تواند درک کند، آیا ما به سمت دانش پیش می‌رویم؟ یا علم است که از ما دور می‌شود؟ بشر در طول تاریخ تغییرات تکنولوژیکی را تجربه کرده است. با این حال، به‌ندرت، ساختار اجتماعی و سیاسی جوامع ما با فناوری به طور اساسی دگرگون شده است. غالباً، چارچوب‌های از پیش موجودی که از طریق آن‌ها به دنیای اجتماعی خود نظم می‌دهیم فناوری جدید را سازگار و جذب می‌کنند، به عبارت دیگر فناوری‌های جدید در دسته‌های قابل تشخیصی در حال تکامل و نوآوری است. ماشین بدون اعمال تغییری کلی در ساختار اجتماعی جایگزین اسب شد. تفنگ ساچمه‌ای جایگزین تفنگ فتیله‌ای شد، اما پارادایم کلی فعالیت نظامی متعارف تا حد زیادی بدون تغییر باقی ماند. به‌ندرت با فناوری‌ای مواجه شده‌ایم که در شیوه‌های غالب ما در توضیح و نظم دادن به جهان تردید ایجاد کند. اما هوش مصنوعی وعده می‌دهد که تمام حوزه‌های تجربه انسانی را متحول کند. و هسته دگرگونی‌های آن در نهایت در سطح فلسفی رخ می‌دهد و نحوه درک انسان از واقعیت و نقش ما در آن را تغییر می‌دهد.

ماهیت بی‌سابقه این فرایند هم پوشیده و هم پیچیده است. با ورود تدریجی به آن، ما منفعلانه آن را تجربه می‌کنیم، عمدتاً از

آنچه انجام داده است و احتمالاً در سال‌های آینده انجام خواهد داد بی‌اطلاع هستیم. پایه و اساس آن با کامپیوتر و اینترنت گذاشته شد. نقطهٔ اوج آن هوش مصنوعی خواهد بود که در همه‌جا حاضر است و افکار و اعمال انسان را به روش‌هایی که هم آشکار هستند (مانند داروهای جدید و ترجمهٔ خودکار زبان) و هم کمتر آگاهانه درک می‌شوند تقویت می‌کند (مانند فرایندهای نرم‌افزاری‌ای که از حرکات و انتخاب‌های ما درس می‌گیرند و برای پیش‌بینی کردن یا شکل دادن به نیازهای آیندهٔ ما تنظیم می‌شوند). اکنون که وعدهٔ هوش مصنوعی و یادگیری ماشینی آشکار شده است، و قدرت محاسباتی مورد نیاز برای به‌کارگیری هوش مصنوعی پیشرفته به‌آسانی در دسترس است، حوزه‌های کمی بی‌تأثیر باقی خواهند ماند.

شبکه‌ای از فرایندهای نرم‌افزاری به‌طور مداوم، اغلب نامحسوس، اما اکنون به نحوی اجتناب‌ناپذیر، در سرتاسر جهان در حال گسترش است، سرعت و دامنهٔ رویدادها را هدایت و درک می‌کند، و جنبه‌هایی از زندگی روزمرهٔ ما — خانه‌ها، حمل‌ونقل، توزیع اخبار، بازارهای مالی، عملیات نظامی — را که زمانی ذهن ما تنهایی سیر می‌کرد پوشش می‌دهد. از آنجایی که نرم‌افزارهای بیشتری هوش مصنوعی را در خود جای می‌دهند و در نهایت به روش‌هایی عمل می‌کنند که انسان‌ها مستقیماً ایجادشان نکرده‌اند یا ممکن است کاملاً درکشان نکنند، این نرم‌افزار تقویت‌کننده‌ای پویا در پردازش اطلاعات توانایی‌ها و تجربیات ما خواهد بود که هم به اقدامات ما شکل می‌دهد و هم از آن‌ها یاد می‌گیرد. اغلب، ما آگاه خواهیم بود که چنین برنامه‌هایی به روش‌هایی که ما در نظر داشته‌ایم به ما کمک می‌کنند. با این حال، در هر لحظه، ممکن است ندانیم که دقیقاً چه کاری انجام می‌دهند یا [چه چیزی را] شناسایی می‌کنند یا چرا کار می‌کنند. فناوری مبتنی بر هوش مصنوعی به همراهی دایمی در درک

و پردازش اطلاعات تبدیل خواهد شد، هرچند که سطح «ذهنی» متفاوتی از انسان را اشغال کند. چه آن را ابزار بدانیم و چه شریک یا رقیب، تجربه ما به مثابه موجودات استدلالی را دگرگون می‌کند و رابطه ما با واقعیت را برای همیشه تغییر می‌دهد.

سفر ذهن انسان به مرحله مرکزی تاریخ قرن‌ها طول کشید. در غرب، ظهور چاپخانه و اصلاحات پروتستانی سلسله مراتب رسمی را متزلزل کرد و چارچوب مرجع جامعه را تغییر داد — از جستجو برای شناخت امر الهی از طریق کتاب مقدس و تفسیر رسمی آن به جستجوی دانش و تحقق آن از طریق تحلیل و کاوش فردی. رنسانس شاهد کشف مجدد نوشته‌های کلاسیک و شیوه‌های تحقیقی بود که برای درک دنیایی استفاده می‌شد که افق‌های آن از طریق اکتشاف جهانی در حال گسترش بود. در دوران روشنگری، قاعده رنه دکارت، *Cogito ergo sum* (می‌اندیشم، پس هستم)، ذهن استدلالی را به منزله توانایی معرف بشر و مدعی مرکزیت تاریخی در نظر گرفت. این تصور همچنین این احساس را منتقل می‌کرد که انحصار تثبیت شده اطلاعاتی‌ای را که عمدتاً در دست کلیسا بود می‌توان بر هم زد.

اکنون پایان اندک برتری فرضی عقل انسانی، همراه با تکثیر ماشین‌هایی که می‌توانند با هوش انسانی مطابقت داشته باشند یا از آن فراتر بروند، نوید دگرگونی‌هایی را می‌دهد که حتی از تحولات عصر روشنگری نیز به طور بالقوه عمیق‌تر است. حتی اگر پیشرفت‌های هوش مصنوعی هوش عمومی مصنوعی (AGI) را تولید نکند — یعنی نرم‌افزاری که قادر به انجام دادن هر کار فکری‌ای در سطح انسانی باشد و بتواند وظایف و مفاهیم را به دیگران در سراسر رشته‌ها مرتبط کند —، ظهور هوش مصنوعی دریافت بشر از واقعیت و در نتیجه از خود را تغییر می‌دهد. ما به سمت دستاوردهای بزرگ پیش می‌رویم، اما این دستاوردها باید تفکر فلسفی را برانگیزاند. چهار قرن پس از

این که دکارت اصل خود را اعلام کرد، این سؤال پیش می‌آید: اگر هوش مصنوعی «فکر می‌کند»، یا عملی شبیه فکر کردن را انجام می‌دهند، [پس] ما کی هستیم؟

هوش مصنوعی دنیایی را آغاز خواهد کرد که در آن تصمیم‌ها به سه روش اصلی گرفته می‌شوند: توسط انسان‌ها (که آشناست)، ماشین‌ها (که در حال آشنا شدن با آن هستیم) و با همکاری بین انسان و ماشین (که نه تنها ناآشنا بلکه بی‌سابقه است). هوش مصنوعی همچنین در حال تبدیل ماشین‌ها — که تاکنون ابزارهای ما بوده‌اند — به شرکای ماست. ما دستورالعمل‌های خاص کمتری در مورد چگونگی دستیابی به اهدافی که برای هوش مصنوعی مقرر کرده‌ایم ارائه خواهیم کرد. بیشتر اوقات، ما به هوش مصنوعی اهدافی مبهم ارائه می‌کنیم و می‌پرسیم: «بر اساس نتیجه‌گیری شما، چگونه باید پیش برویم؟»

این تغییر نه ذاتاً تهدیدکننده است و نه ذاتاً رهایی‌بخش. اما به اندازه کافی متفاوت است که به احتمال زیاد مسیر جوامع و مسیر تاریخ را تغییر خواهد داد. ادغام مداوم هوش مصنوعی در زندگی ما دنیایی را به وجود خواهد آورد که در آن اهداف انسانی به ظاهر غیرممکن به دست می‌آیند و دستاوردهایی را که زمانی تصور می‌شد منحصراً انسانی هستند — نوشتن یک آهنگ، کشف یک درمان پزشکی — ماشین‌ها خلق می‌کنند یا حاصل همکاری با آن‌هاست. این توسعه با محصور کردن تمامی رشته‌های تخصصی در فرایندهایی که به کمک هوش مصنوعی پیش می‌رود آن‌ها را متحول می‌کند، با [لحاظ] مرزهایی میان تصمیم‌گیری صرفاً انسانی، تصمیم‌گیری هوش مصنوعی محض و تصمیم‌گیری ترکیبی انسان-هوش مصنوعی که گاهی تعریف آن دشوار می‌شود.

در قلمرو سیاسی، جهان در حال ورود به دوره‌ای است که در آن

سیستم‌های هوش مصنوعی مبتنی بر کلان‌داده‌ها جنبه‌های روبه‌رشدی را تحت تأثیر قرار می‌دهند: طراحی پیام‌های سیاسی؛ تطبیق و توزیع آن پیام‌ها برای جمعیت‌های مختلف؛ ساخت و به‌کارگیری اطلاعات نادرست با استفاده از بازیگران بدخواه با هدف ایجاد اختلاف اجتماعی؛ و طراحی و استقرار الگوریتم‌هایی برای کشف، شناسایی و مقابله با اطلاعات نادرست و سایر اشکال داده‌های مضر. با افزایش نقش هوش مصنوعی در تعریف و شکل دادن به «فضای اطلاعاتی»، پیش‌بینی نقش آن دشوارتر می‌شود. در این فضا، مانند سایر فضاها، هوش مصنوعی گاهی اوقات به گونه‌ای عمل می‌کند که حتی طراحان آن فقط می‌توانند به صورت کلی توضیح دهند. در نتیجه، چشم‌انداز جامعه آزاد، حتی اراده آزاد، ممکن است تغییر کند. حتی اگر ثابت شود که این تحولات مطبوع یا برگشت‌پذیر هستند، جوامع در سراسر جهان وظیفه دارند این تغییرات را درک کنند تا بتوانند آن‌ها را با ارزش‌ها، ساختارها و قراردادهای اجتماعی خود آشتی دهند.

مؤسسات و فرماندهان دفاعی با تحولات نه‌چندان عمیق روبه‌رو هستند. وقتی چندین ارتش استراتژی‌ها و تاکتیک‌هایی را اتخاذ می‌کنند که ماشین‌ها به آن‌ها شکل داده‌اند، توازن قدرت تغییر می‌کند و محاسبه بالقوه دشوارتر می‌شود، ماشین‌هایی که الگوهایی را درک می‌کنند که سربازان و استراتژیست‌های انسانی قادر به درک آن‌ها نیستند. اگر چنین ماشین‌هایی مجاز باشند در تصمیم‌گیری‌های هدف‌گیری مستقل شرکت کنند، مفاهیم سنتی دفاع و بازدارندگی — و قوانین جنگ در کل — ممکن است آسیب ببینند یا حداقل نیاز به تطبیق داشته باشند. در چنین مواردی، شکاف‌های جدیدی در درون و بین جوامع ظاهر می‌شود — بین آن‌هایی که فناوری جدید را اتخاذ می‌کنند و کسانی که از آن صرف‌نظر می‌کنند یا ابزاری برای توسعه یا دستیابی به برخی از کاربردهای آن ندارند. هنگامی که گروه‌ها یا کشورهای

مختلف مفاهیم‌ها یا کاربردهای متفاوتی از هوش مصنوعی را اتخاذ می‌کنند، تجربیات آن‌ها از واقعیت ممکن است به گونه‌ای متفاوت باشد که پیش‌بینی یا همراهی با آن دشوار باشد. همان‌طور که جوامع شراکت‌های انسان و ماشینی خود را توسعه می‌دهند — با اهداف متفاوت، مدل‌های آموزشی متفاوت، و محدودیت‌های عملیاتی و اخلاقی بالقوه ناسازگار با هوش مصنوعی —، ممکن است آن‌ها به رقابت، ناسازگاری فنی و عدم‌درک متقابل بیشتر تبدیل شوند. فناوری که در ابتدا تصور می‌شد ابزار تعالی‌بخش تفاوت‌های ملی و پراکندگی حقیقت عینی باشد ممکن است به مرور زمان به روشی تبدیل شود که از طریق آن تمدن‌ها و افراد به سمت واقعیت‌های متفاوت و غیرقابل درک متقابل منحرف شوند.

آلفازیرو گویاست. ثابت کرد که هوش مصنوعی، حداقل در حوزه بازی، دیگر محدود به محدودیت‌های دانش تثبیت‌شده بشری نیست. مسلماً، نوع هوش مصنوعی زیربنایی آلفازیرو — یادگیری ماشینی که در آن الگوریتم‌ها بر روی شبکه‌های عصبی عمیق آموزش داده می‌شوند — محدودیت‌هایی دارد. اما در تعداد فزاینده‌ای از کاربردها، ماشین‌ها راه‌حل‌هایی ابداع می‌کنند که فراتر از حد تصور انسان به نظر می‌رسد. در سال ۲۰۱۶، یکی از زیرمجموعه‌های دیپ‌مایند، دیپ‌مایند اپلاید،^۱ هوش مصنوعی‌ای را (که بر اساس بسیاری از اصول مشابه آلفازیرو اجرا می‌شد) برای بهینه‌سازی پروتد مراکز داده حساس به دمای شرکت گوگل توسعه داد. اگرچه برخی از بهترین مهندسان جهان قبلاً این مشکل را حل کرده بودند، برنامه هوش مصنوعی دیپ‌مایند خنک‌سازی را بهینه‌تر کرد و هزینه انرژی را تا ۴۰ درصد کاهش داد — بهبود در سطحی کلان نسبت به عملکرد انسان.^(۶) هنگامی که هوش مصنوعی برای دستیابی به پیشرفت‌های

سنجش‌پذیر در زمینه‌های مختلف تلاش به کار گرفته شود، جهان ناگزیر تغییر خواهد کرد. نتایج به‌سادگی روش‌های کارآمدتر برای انجام دادن وظایف انسانی نخواهد بود: در بسیاری از موارد، هوش مصنوعی راه‌حل‌ها یا جهت‌گیری‌های جدیدی را پیشنهاد می‌کند که نشانه‌ای دیگر از یادگیری و ارزیابی منطقی و غیرانسانی است.

هنگامی که عملکرد هوش مصنوعی در انجام دادن وظیفه‌ای از انسان پیشی می‌گیرد، عدم به‌کارگیری آن هوش مصنوعی، حداقل به مثابه کمکی به تلاش‌های انسان، ممکن است بیش از پیش نوعی لجاجت یا حتی سهل‌انگاری به نظر برسد. این که به فردی که با کمک هوش مصنوعی شطرنج بازی می‌کند احتمالاً توصیه شود که مهره‌ای ارزشمند را قربانی کند، مهره‌ای که بازیکنان حرفه‌ای از دیرباز آن را ضروری می‌دانستند، عواقب چندانی ندارد، اما در زمینه امنیت ملی، اگر هوش مصنوعی به فرمانده کل قوا توصیه کند که تعداد قابل توجهی از شهروندان را قربانی یا منافع آن‌ها را فدا کند تا بر اساس محاسبه و ارزش‌گذاری هوش مصنوعی، جان تعداد بیشتری را نجات دهد، چه؟ بر چه اساسی می‌توان آن قربانی کردن را نادیده گرفت؟ آیا آن قربانی کردن توجیه خواهد شد؟ آیا انسان‌ها همیشه می‌دانند که هوش مصنوعی چه محاسباتی انجام داده است؟ آیا انسان‌ها می‌توانند تصمیم‌های ناخواسته (هوش مصنوعی) را تشخیص دهند یا تصمیم‌های ناخواسته را به‌موقع نقض کنند؟ اگر ما بتوانیم منطق هر تصمیم فردی را درک کنیم، آیا باید توصیه‌های آن را فقط بر اساس باوری که به آن داریم اجرا کنیم؟ اگر این کار را نکنیم، آیا خطر بر هم زدن عملکردی برتر از [عملکرد] خودمان را ایجاد می‌کنیم؟ حتی اگر بتوانیم منطق، بها و تأثیر گزینه‌های جایگزین را درک کنیم، اگر حریف ما به همان اندازه به هوش مصنوعی متکی باشد، چه؟ تعادل بین این ملاحظات

چگونه به دست می‌آید یا در صورت لزوم چگونه اثبات می‌شود؟ هم در موفقیت آلفازیرو و هم در کشف هالیسین، هوش مصنوعی برای تعریف مشکل حل‌شده به انسان وابسته بود. هدف آلفازیرو برنده شدن در شطرنج و در عین حال پیروی از قوانین بازی بود. هدف هوش مصنوعی‌ای که هالیسین را کشف کرد این بود که تا حد امکان بسیاری از عوامل بیماری‌زا را از بین ببرد: هرچه پاتوژن‌های بیشتری را بدون آسیب رساندن به میزبان از بین می‌برد، موفق‌تر بود. علاوه بر این، تمرکز آن به عنوان قلمرو خارج از دسترس انسان در نظر گرفته شد: به جای تعیین مسیرهای شناخته‌شدهٔ رساندن دارو، به آن دستور داده شد که به دنبال رویکردهای کشف‌نشده باشد. هوش مصنوعی موفق شد، زیرا آنتی‌بیوتیکی کشف کرد که پاتوژن‌ها را از بین برد. اما به طور خاصی پیشگامانه بود، چرا که گسترش گزینه‌های درمانی را بر عهده گرفت، آنتی‌بیوتیکی جدید (و قوی) که از طریق مکانیسمی جدید رسانده می‌شد.

همکاری جدیدی میان انسان و ماشین در حال ظهور است: ابتدا، انسان‌ها مشکل یا هدفی را برای یک ماشین تعریف می‌کنند. سپس ماشینی که در قلمرو فراتر از دسترس انسان عمل می‌کند فرایند بهینه را برای پیگیری تعیین می‌کند. هنگامی که یک ماشین فرایندی را وارد قلمرو انسانی کند، می‌توانیم سعی کنیم آن را مطالعه و درک کنیم، و در حالت ایدئال، آن فرایند را در فرایند کنش در حال انجام ادغام کنیم. از زمان پیروزی آلفازیرو، استراتژی و تاکتیک‌های آن در بازی انسان‌ها گنجانده شده است و تصورات بشری از شطرنج را گسترش می‌دهد. نیروی هوایی ایالات متحده اصول زیربنایی آلفازیرو را با هوش مصنوعی جدید ARTUμ سازگار کرده است که با موفقیت یک هواپیمای نظارتی U-2 را در پروازی آزمایشی فرماندهی کرد — اولین برنامهٔ کامپیوتری برای پرواز با هواپیمای نظامی و عملکرد سیستم‌های

راداری آن به طور مستقل و بدون نظارت مستقیم انسانی.^(۷) هوش مصنوعی‌ای که هالیسین را کشف کرد درک پژوهشگران انسانی را هم به صورت محدود (ریشه‌کنی باکتری، رساندن دارو) و هم گسترده (بیماری، دارو، سلامت) گسترش داد.

شراکت فعلی انسان و ماشین، که هم به مشکلی تعریف‌پذیر و هم به هدفی سنجش‌پذیر نیاز دارد، دلیلی است برای نترسیدن از ماشین‌های دانا و کنترل‌کننده؛ چنین اختراعاتی در زمره مقولات علمی تخیلی باقی می‌مانند. با این حال، مشارکت انسان و ماشین نشانگر انحراف عمیق از تجربه قبلی است.

موتورهای جستجو چالش دیگری را ارائه کردند: ده سال پیش، زمانی که موتورهای جستجو (به جای یادگیری ماشینی) با داده‌کاوی کار می‌کردند، اگر شخصی به دنبال «رستوران‌های مجلل» و سپس «لباس» می‌گشت، جستجوی او برای دومی مستقل از جستجوی او برای اولی بود. هر دو بار، یک موتور جستجو تا حد امکان اطلاعات را جمع‌آوری می‌کرد، سپس گزینه‌های درخواست‌کننده را ارائه می‌داد — چیزی مانند دفترچه تلفن دیجیتال یا کاتالوگ یک موضوع. اما موتورهای جستجوی معاصر با مدل‌هایی هدایت می‌شوند که بر اساس رفتار مشاهده‌شده انسان است. اگر شخصی به دنبال «رستوران‌های مجلل» باشد، سپس «لباس» را جستجو کند، ممکن است به جای جایگزین‌های مقرون به‌صرفه‌تر، لباس‌های مارک‌دار به او ارائه شود. لباس‌های مارک‌دار ممکن است همان چیزی باشد که جستجوگر به دنبال آن است. اما بین انتخاب از طیف وسیعی از گزینه‌ها و عملی کردن یک اقدام تفاوت وجود دارد — در این مورد، خرید؛ در موارد دیگر، اتخاذ موضع یا ایدئولوژی سیاسی یا فلسفی — بدون این که هرگز بدانیم محدوده اولیه احتمالات یا پیامدها چقدر بوده است، [کار را] به ماشینی سپردیم که به طور انحصاری گزینه‌ها را شکل دهد.

تاکنون، انتخاب بر اساس عقل حق و امتیاز — و از عصر روشنگری، صفت تعیین‌کننده — بشر بوده است. ظهور ماشین‌هایی که می‌توانند به عملکرد عقل انسان نزدیک شوند، هم انسان‌ها و هم ماشین‌ها را تغییر می‌دهند. ماشین‌ها انسان‌ها را آگاه می‌کنند و واقعیت ما را به گونه‌ای گسترش می‌دهند که ما انتظار یا لزوماً قصد دامن زدن به آن را نداشتیم (برعکس نیز امکان‌پذیر خواهد بود: ماشین‌هایی که دانش بشری را مصرف می‌کنند برای تضعیف ما استفاده خواهند شد). همزمان، انسان‌ها ماشین‌هایی خواهند ساخت که قادر به اکتشافات و ارائه نتایج شگفت‌انگیز هستند — قادر به یادگیری و ارزیابی اهمیت اکتشافات خود هستند. نتیجه پیدایش یک دوره جدید خواهد بود.

بشر قرن‌ها تجربه استفاده از ماشین‌ها را برای افزایش، خودکارسازی و در بسیاری موارد جایگزینی کار دستی دارد. امواج تغییری که انقلاب صنعتی به ارمغان آورد هنوز در قلمروهای اقتصاد، سیاست، زندگی فکری و امور بین‌الملل طنین‌انداز است. با عدم‌شناخت بسیاری از امکانات مدرنی که قبلاً هوش مصنوعی ارائه کرده است به آرامی، تقریباً منفعلانه، به فناوری اعتماد کرده‌ایم، بدون این‌که واقعیت وابستگی خود یا پیامدهای آن را ثبت کنیم. در زندگی روزمره، هوش مصنوعی شریک ماست و به ما کمک می‌کند تصمیم بگیریم که چه بخوریم، چه بپوشیم، چه چیزی را باور کنیم، کجا برویم و چگونه به آن‌جا برسیم.

هوش مصنوعی می‌تواند نتیجه‌گیری کند، پیش‌بینی کند و تصمیم بگیرد، اما خودآگاهی ندارد — به عبارت دیگر، توانایی تأمل در نقش خود در جهان را ندارد. قصد، انگیزه، اخلاق و عاطفه ندارد. حتی بدون این ویژگی‌ها، احتمالاً ابزارهای متفاوت و ناخواسته‌ای برای دستیابی به اهداف تعیین‌شده ایجاد می‌کند. اما به‌ناچار انسان‌ها و محیطی را که در آن زندگی می‌کنند تغییر خواهد داد. وقتی افراد با آن بزرگ می‌شوند یا