

پیدایش: هوش مصنوعی، امید و سرشت انسانی

"A profound exploration of how we can protect human dignity and values
in an era of autonomous machines." —WALTER ISAACSON

ARTIFICIAL INTELLIGENCE,

HOPE,

AND THE HUMAN SPIRIT

GENESIS

HENRY A. KISSINGER

CRAIG MUNDIE

ERIC SCHMIDT

FOREWORD BY NIALL FERGUSON

Henry A. Kissinger, Craig Mundie, and Eric Schmidt

مولفان: هنری کیسینجر، اریک اشمیت و کریگ ماندی

مترجمان: مهندس علیرضا صالحین

دکتر صادق غضنفری

دکتر پویا آرزومند لنگرودی امیدی



انتشارات علمی صالحین

نام کتاب: پیدایش: هوش مصنوعی، امید و سرشت انسانی

مولفان: هنری کیسینجر، اریک اشمیت و کریگ ماندی

مترجمان: مهندس علیرضا صالحین / دکتر پویا آرزومند لنگرودی امیدی / دکتر صادق غضنفری

ویراستار: پویا آرزومند لنگرودی امیدی

گرافیکست: کتابون علی نوری

سال چاپ: اول ۱۴۰۴

شمارگان: ۱۰۰

شابک: ۹۷۸-۶۲۲-۸۸۹۸۹-۴-۰

حق چاپ برای نشر علمی صالحین محفوظ می‌باشد.

نشانی: تهران، خیابان آزادی، بعد از خیابان شادمهر، ساختمان اداری شادمان، پلاک ۳۹۹، طبقه چهارم، واحد ۲۳

سرشناسه	: کیسینجر، هنری، ۱۹۲۳ - م. Kissinger, Henry
عنوان و نام پدیدآور	: پیدایش: هوش مصنوعی، امید و سرشت انسانی / مولفان هنری کیسینجر، اریک اشمیت، کریگ ماندی؛ ترجمه علیرضا صالحین، پویا آرزومند لنگرودی امیدی، صادق غضنفری؛ ویراستار پویا آرزومند لنگرودی امیدی.
مشخصات نشر	: تهران: علمی صالحین، ۱۴۰۴.
مشخصات ظاهری	: ۱۲۵ ص.
شابک	: ۹۷۸-۶۲۲-۸۸۹۸۹-۴-۰
وضعیت فهرست نویسی	: فیبا
یادداشت	: عنوان اصلی: Genesis : artificial intelligence, hope, and the human spirit, ۲۰۲۴.
یادداشت	: کتاب حاضر با عنوان "پیدایش: هوش مصنوعی، ذات انسان، و امید به آینده" با ترجمه مریم درویدیان توسط انتشارات نشر پازل در سال ۱۴۰۳ ترجمه و منتشر شده است.
عنوان دیگر	: پیدایش: هوش مصنوعی، ذات انسان، و امید به آینده.
موضوع	: هوش مصنوعی Artificial intelligence تحولات اجتماعی Social change هوش مصنوعی -- فلسفه Artificial intelligence -- Philosophy هوش مصنوعی -- جنبه‌های اجتماعی Artificial intelligence -- Social aspects هوش مصنوعی -- آینده‌نگری Artificial intelligence -- Forecasting
شناسه افزوده	: اشمیت، اریک، ۱۹۵۵ - م.
شناسه افزوده	: Schmidt, Eric
شناسه افزوده	: ماندی، کریگ
شناسه افزوده	: Mundie, Craig
شناسه افزوده	: صالحین، علیرضا، ۱۳۶۵ - م. مترجم
شناسه افزوده	: آرزومند لنگرودی، پویا، ۱۳۷۱ - م. مترجم، ویراستار
شناسه افزوده	: غضنفری، صادق، ۱۳۶۹ - م. مترجم
رده بندی کنگره	: Q۳۳۵
رده بندی دیویی	: ۰۰۶/۳
شماره کتبخانی ملی	: ۱۰۲۵۹۳۳۷
اطلاعات رکورد	: فیبا
کتبخناسی	

فهرست مطالب

•	تقدیم‌نامه	۴
•	پیشگفتار — به قلم نایل فرگوسن	۹
•	به یاد هنری کیسینجر	۱۳
•	مقدمه	۱۵
•	بخش اول: در آغاز	۱۵
•	فصل ۱: کشف	۲۶
•	فصل ۲: مغز	۴۲
•	فصل ۳: واقعیت	۴۹
•	بخش دوم: چهار شاخه	۴۹
•	فصل ۴: سیاست	۶۱
•	فصل ۵: امنیت	۷۳
•	فصل ۶: رفاه	۸۷
•	فصل ۷: علم	۹۳
•	بخش سوم: درخت زندگی	۹۳
•	فصل ۸: راهبرد (استراتژی)	۱۰۶
•	نتیجه‌گیری	۱۰۹
•	سپاسگزاری	۱۱۱
•	درباره نویسندگان	۱۱۲
•	آثاری از هنری کیسینجر	۱۱۳
•	تحسین‌های پیش از انتشار برای کتاب «پیدایش»	۱۱۵
•	منابع و ارجاعات	۱۲۳
•	درباره مترجمان	

پیشگفتار به قلم نایل فرگوسن (Niall Ferguson)

هنگامی که هنری کیسینجر در ژوئن ۲۰۱۸ مقاله‌اش با عنوان «چگونه عصر روشنگری به پایان می‌رسد» را در مجله *آتلانتیک* منتشر کرد، بسیاری از اینکه بزرگ‌مرد عرصه سیاست و دیپلماسی، دیدگاهی درباره موضوع هوش مصنوعی دارد، شگفت‌زده شدند. کیسینجر به‌تازگی ۹۵ ساله شده بود. هوش مصنوعی هنوز به آن «موضوع داغی» تبدیل نشده بود که پس از عرضه ChatGPT توسط OpenAI در اواخر سال ۲۰۲۲، به آن بدل شد.

با این حال، به عنوان زندگی‌نامه‌نویس کیسینجر، برای من کاملاً طبیعی بود که موضوع هوش مصنوعی توجه او را به خود جلب کرده باشد. به هر حال، او در سال ۱۹۵۷ با کتابی درباره یک فناوری جدید و دگرگون‌کننده به شهرت رسیده بود. کتاب *سلاح‌های هسته‌ای و سیاست خارجی* آنچنان با تحقیقی موشکافانه نوشته شده بود که حتی تأیید رابرت اوپنهایمر را نیز به دست آورد. اوپنهایمر آن را کتابی «به‌شکلی خارق‌العاده آگاهانه و از این حیث در حوزه تسلیحات هسته‌ای کاملاً بی‌سابقه... در پایبندی به واقعیت‌ها و سواس‌گونه و در استدلال هم پرشور و هم سرسخت» توصیف کرد.

اگرچه کیسینجر در دوران دانشجویی دکتری، خود را غرق در تاریخ دیپلماتیک اروپای اوایل قرن نوزدهم کرده بود، اما در تمام دوران فعالیت حرفه‌ای خود به‌خوبی آگاه بود که الگوهای جاودانه سیاست قدرت‌های بزرگ، به‌طور دوره‌ای در معرض گسست ناشی از تحولات فناوریانه قرار دارند. او نیز مانند بسیاری از اعضای نسل خود که در جنگ جهانی دوم خدمت کردند، نه تنها مرگ و ویرانی انبوهی را که سلاح‌های مدرن می‌توانستند به بار آورند، بلکه پیامدهای هولناک «علم منحرف» رایش سوم هیتلر را نیز، به تعبیر به یادماندنی چرچیل، برای هم‌کیشان یهودی خود به چشم دیده بود.

برخلاف شهرت ناموجهش به عنوان یک جنگ‌طلب، کیسینجر در تمام دوران بزرگسالی خود عمیقاً با این الزام برانگیخته می‌شد که از جنگ جهانی سوم — پیامدی که بسیاری از آن در صورت تبدیل شدن جنگ سرد میان ایالات متحده و اتحاد جماهیر شوروی به یک جنگ تمام‌عیار می‌ترسیدند — جلوگیری کند. او به‌خوبی می‌فهمید که فناوری شکافت هسته‌ای، جنگ جهانی دیگری را به آتشی حتی بزرگ‌تر از جنگ جهانی دوم بدل خواهد کرد. کیسینجر در اوایل کتاب *سلاح‌های هسته‌ای و سیاست خارجی*، اثرات مخرب یک بمب ده‌مگاتی بر نیویورک را تخمین زد و سپس چنین برآورد کرد که یک حمله تمام‌عیار شوروی به پنجاه شهر بزرگ آمریکا بین ۱۵ تا ۲۰ میلیون نفر را کشته و بین ۲۰ تا ۲۵ میلیون نفر را مجروح خواهد کرد. ۵ تا ۱۰ میلیون نفر دیگر بر اثر پیامدهای بارش رادیواکتیو جان خود را از دست می‌دادند و شاید ۷ تا ۱۰ میلیون نفر دیگر نیز بیمار می‌شدند. بازماندگان با «فروپاشی اجتماعی» روبرو می‌شدند. او خاطرنشان کرد که حتی پس از چنین حمله‌ای، ایالات متحده همچنان قادر خواهد بود ویرانی مشابهی را بر اتحاد جماهیر شوروی تحمیل کند. نتیجه واضح بود: «از این پس، تنها نتیجه یک جنگ تمام‌عیار این خواهد بود که هر دو طرف متخاصم بازنده خواهند بود.» کیسینجر در مقاله خود در سال ۱۹۵۷ با عنوان «استراتژی و سازماندهی» استدلال کرد که در چنین درگیری‌ای هیچ برنده‌ای نمی‌تواند وجود داشته باشد، «زیرا حتی طرف ضعیف‌تر نیز ممکن است بتواند چنان درجه‌ای از تخریب را تحمیل کند که هیچ جامعه‌ای توان تحمل آن را نداشته باشد.»

با این حال، آرمان‌گرایی جوانی کیسینجر او را به یک صلح‌طلب تبدیل نکرد. او در کتاب *سلاح‌های هسته‌ای و سیاست خارجی* کاملاً صریح بود که «بعید است بتوان با کاهش تسلیحات هسته‌ای» یا با سیستم‌های بازرسی تسلیحات، از «وحشت جنگ هسته‌ای» جلوگیری کرد. سؤال این نبود که آیا می‌توان به‌کلی از جنگ اجتناب کرد، بلکه این بود که آیا «می‌توان کاربردهایی از قدرت را تصور کرد که فاجعه‌بارتر از جنگ تمام‌عیار ترموهسته‌ای نباشد؟» زیرا اگر چنین امکانی وجود نداشت، پیروزی برای ایالات متحده و متحدانش در جنگ سرد بسیار دشوار می‌شد. کیسینجر در مقاله‌ای با عنوان

«کنترل‌ها، بازرسی‌ها و جنگ محدود» که در مجله ریپورتر منتشر شد، هشدار داد: «فقدان هرگونه درک عمومی از حدود جنگ، چارچوب روانی مقاومت در برابر اقدامات کمونیست‌ها را تضعیف می‌کند. جایی که جنگ معادل خودکشی ملی تلقی شود، تسلیم ممکن است شرّ کوچک‌تر به نظر برسد.»

بر این اساس بود که کیسینجر دکترین خود مبنی بر جنگ هسته‌ای محدود را ارائه داد، همانطور که در «استراتژی و سازماندهی» تشریح شده است:

در پس‌زمینه شوم ویرانی ترموهسته‌ای، هدف جنگ دیگر نمی‌تواند پیروزی نظامی به آن شکلی که ما می‌شناسیم باشد. بلکه باید دستیابی به شرایط سیاسی مشخصی باشد که برای طرف مقابل کاملاً قابل درک باشد. هدف از جنگ محدود، تحمیل تلفات یا قرار دادن دشمن در معرض خطراتی است که با اهداف مورد مناقشه تناسبی ندارد. هرچه هدف میانه‌روتر باشد، احتمالاً جنگ کمتر خشونت‌آمیز خواهد بود.

این امر مستلزم درک روانشناسی طرف مقابل و همچنین توانایی نظامی آن بود.

در آن زمان، بسیاری از تأمل خونسردانه کیسینجر در مورد جنگ هسته‌ای محدود، رویگردان شدند. برخی از محققان، مانند توماس شلینگ، این ایده را که می‌توان از یک تشدید تنش غیرقابل توقف جلوگیری کرد، زیر سؤال بردند؛ حتی خود کیسینجر نیز بعداً از استدلال خود فاصله گرفت. با این حال، هر دو ابرقدرت به ساخت و استقرار سلاح‌های هسته‌ای تاکتیکی یا میدانی ادامه دادند و دقیقاً از همان منطقی پیروی کردند که کیسینجر در سلاح‌های هسته‌ای و سیاست خارجی تشریح کرده بود. جنگ هسته‌ای محدود شاید در تتوری کارساز نبود، اما برنامه‌ریزان نظامی در هر دو طرف طوری رفتار می‌کردند که گویی در عمل ممکن است کارساز باشد. (در واقع، چنین سلاح‌هایی تا به امروز وجود دارند. دولت روسیه از زمان درگیر شدن در تهاجم به اوکراین، بیش از یک بار تهدید به استفاده از آنها کرده است.) کیسینجر جوان در مورد سلاح‌های هسته‌ای، حتی بیش از آنچه خود می‌دانست، حق داشت.

کیسینجر هرگز از تأمل در باب پیامدهای تغییرات فناورانه در عرصه سیاسی دست نکشید. در یک مقاله فراموش‌شده که او در ژانویه ۱۹۶۸ برای نلسون راکفلر نوشت، کیسینجر به روش‌هایی نگرینست که کامپیوتری‌سازی می‌توانست به مقامات در مدیریت جریان روزافزون اطلاعات تولید شده توسط سازمان‌های دولتی ایالات متحده کمک کند. از دید او، مقامات ارشد در خطر جدی غرق شدن در انبوه داده‌ها بودند. او نوشت: «سیاست‌گذار ارشد آنقدر اطلاعات در اختیار دارد که در شرایط بحرانی، مدیریت آن را غیرممکن می‌یابد.» کیسینجر استدلال می‌کرد که تصمیم‌گیرندگان باید «به‌طور مداوم در مورد نقاط بحرانی احتمالی توجیه شوند»، از جمله نقاط بحرانی بالقوه «حتی زمانی که اولویت بالایی به آنها اختصاص داده نشده باشد.» همچنین باید «مجموعه‌ای از گزینه‌های عملی... که جایگزین‌های اصلی را در پاسخ به شرایط قابل پیش‌بینی با ارزیابی پیامدهای احتمالی داخلی و خارجی هر جایگزین مشخص می‌کند» در اختیار آنها قرار گیرد.

کیسینجر اذعان داشت که دستیابی به چنین پوشش جامعی، نیازمند سرمایه‌گذاری‌های عمده در برنامه‌نویسی، ذخیره‌سازی، بازیابی و گرافیک است. خوشبختانه، «فناوری سخت‌افزاری» برای انجام هر چهار عملکرد اکنون وجود داشت:

ما اکنون می‌توانیم چند صد قلم اطلاعات مربوط به هر فرد در ایالات متحده را روی یک نوار مغناطیسی ۲۴۰۰ فوتی ذخیره کنیم... کامپیوترهای نسل سوم اکنون قادر به انجام عملیات پایه‌ای ماشین در نانو ثانیه، یعنی یک میلیارد ثانیه هستند... سیستم‌های اشتراک زمانی آزمایشی اکنون نشان داده‌اند که قابلیت دسترسی چندگانه برای کامپیوترهای دیجیتال بزرگ مقیاس امکان‌پذیر است تا امکان ورود/خروج اطلاعات هم در ایستگاه‌های اجرایی و هم در ایستگاه‌های

اپراتور در سراسر جهان فراهم شود... [و] به‌زودی نمایشگرهای لامپ پرتوی کاتدی رنگی برای خروجی کامپیوتر در دسترس خواهند بود.

بعدها، پس از اولین سال حضورش در کاخ سفید به عنوان مشاور امنیت ملی ریچارد نیکسون، کیسینجر تلاش کرد تا چنین کامپیوتری را برای استفاده شخصی خود به دست آورد. سازمان سیا این درخواست را رد کرد، احتمالاً به این دلیل که کیسینجر بدون کامپیوتر هم برای جامعه اطلاعاتی کافی بود و بس.

هنری کیسینجر هرگز بازنشسته نشد. او همچنین هرگز از نگرانی درباره آینده بشریت دست برنداشت. چنین مردی به سختی می‌توانست یکی از مهم‌ترین پیشرفت‌های فناوریانه اواخر عمر خود را نادیده بگیرد: توسعه و استقرار هوش مصنوعی مولد. در واقع، وظیفه درک پیامدهای این فناوری نوظهور، بخش قابل توجهی از سال‌های پایانی عمر کیسینجر را به خود اختصاص داد.

پیدایش، آخرین کتاب کیسینجر، با همکاری دو فناور برجسته، کریگ ماندی و اریک اشمیت، نوشته شد و مهر خوش‌بینی ذاتی آن نوآوران را بر خود دارد. نویسندگان مشتاقانه منتظر «تکامل انسان فناور (*Homo technicus*) — گونه‌ای از انسان که ممکن است در این عصر جدید، در هم‌زیستی با فناوری ماشینی زندگی کند» هستند. آنها استدلال می‌کنند که هوش مصنوعی به‌زودی می‌تواند برای «ایجاد یک سطح پایه جدید از ثروت و رفاه بشری... که دست‌کم تنش‌های کار، طبقه و درگیری را که پیش از این بشریت را از هم گسسته است، کاهش داده یا حتی از بین ببرد» به کار گرفته شود. پذیرش هوش مصنوعی حتی ممکن است به «برابری‌های عمیق... در سراسر نژاد، جنسیت، ملیت، محل تولد و پیشینه خانوادگی» منجر شود.

با این وجود، سهم مسن‌ترین نویسنده در مجموعه‌ای از هشدارها که ایده محوری کتاب را تشکیل می‌دهند، قابل تشخیص است. نویسندگان مشاهده می‌کنند: «ظهور هوش مصنوعی، یک مسئله بقای بشری است... یک هوش مصنوعی که به‌طور نامناسب کنترل شود... می‌تواند دانش را به شیوه‌ای مخرب انباشت کند...». در اینجا، پرسش اصلی کیسینجر از مقاله سال ۲۰۱۸ خود در *آتلانتیک* با عنوان «چگونه عصر روشنگری به پایان می‌رسد»، بازنویسی شده برای پیدایش اما بلافاصله قابل تشخیص است:

ظرفیت عینی [هوش مصنوعی] برای رسیدن به نتایج جدید و دقیق در مورد جهان ما از طریق روش‌های غیرانسانی، نه‌تنها اتکای ما به روش علمی را که به‌طور مداوم برای پنج قرن دنبال شده است، مختل می‌کند، بلکه ادعای انسان مبنی بر درک انحصاری یا منحصربه‌فرد از واقعیت را نیز به چالش می‌کشد. این به چه معناست؟ آیا عصر هوش مصنوعی نه‌تنها در پیشبرد بشریت ناکام خواهد ماند، بلکه در عوض بازگشتی به پذیرش پیشامدرن اقتدار توضیح‌ناپذیر را تسریع خواهد کرد؟ به طور خلاصه: آیا ما، آیا ممکن است ما، در آستانه یک وارونگی بزرگ در شناخت انسان باشیم — یک روشنگری تاریک؟

در بخشی که برای این خواننده قدرتمندترین قسمت کتاب بود، نویسندگان به یک مسابقه تسلیحاتی هوش مصنوعی عمیقاً نگران‌کننده می‌اندیشند. نویسندگان می‌نویسند: «اگر... هر جامعه انسانی بخواهد موقعیت یک‌جانبه خود را به حداکثر برساند، آنگاه شرایط برای یک رقابت روانی بین نیروهای نظامی و آژانس‌های اطلاعاتی رقیب فراهم می‌شود، رقابتی که بشریت هرگز با نمونه آن روبرو نشده است. امروز، در سال‌ها، ماه‌ها، هفته‌ها و روزهای منتهی به ظهور اولین ابرهوش، یک معضل امنیتی با ماهیت وجودی در انتظار ماست.»

اگر ما هم‌اکنون شاهد «رقابتی برای رسیدن به یک هوش واحد، کامل و بی‌چون‌وچرا مسلط» هستیم، پس پیامدهای محتمل چه خواهند بود؟ نویسندگان، به شمارش من، شش سناریو را پیش‌بینی می‌کنند که هیچ‌کدام جذاب نیستند:

۱. بشریت کنترل یک مسابقه وجودی بین بازیگران متعددی که در یک معضل امنیتی گرفتار شده‌اند را از دست خواهد داد.
 ۲. بشریت از اعمال هژمونی مطلق توسط یک پیروز که توسط سازوکارهای مهار و توازی که به‌طور سنتی برای تضمین حداقل امنیت برای دیگران لازم است، مهار نشده، رنج خواهد برد.
 ۳. تنها یک هوش مصنوعی برتر وجود نخواهد داشت، بلکه نمودهای متعددی از هوش برتر در جهان وجود خواهد داشت.
 ۴. شرکت‌هایی که هوش مصنوعی را در اختیار دارند و توسعه می‌دهند، ممکن است قدرت فراگیر اجتماعی، اقتصادی، نظامی و سیاسی به دست آورند.
 ۵. هوش مصنوعی ممکن است بیشترین ارتباط و گسترده‌ترین و بادوام‌ترین بیان خود را نه در ساختارهای ملی، بلکه در ساختارهای مذهبی بیابد.
 ۶. انتشار کنترل‌نشده و متن‌باز این فناوری جدید می‌تواند به ظهور گروه‌ها یا قبایل کوچکتری با ظرفیت هوش مصنوعی پایین‌تر اما همچنان قابل توجه منجر شود.
- کیسینجر عمیقاً نگران سناریوهایی از این دست بود و تلاش او برای جلوگیری از آنها با نوشتن این کتاب به پایان نرسید. بر کسی پوشیده نیست که آخرین تلاش زندگی او — که توان باقیمانده‌اش را در ماه‌های پس از صدمین سالگرد تولدش تحلیل برد — آغاز فرآیند مذاکرات محدودسازی تسلیحات هوش مصنوعی بین ایالات متحده و چین بود، دقیقاً به امید جلوگیری از چنین پیامدهای ویران‌شهری.
- نتیجه‌گیری پیلایش به شیوه‌ای آشکارا کیسینجری است:
- آنچه برخی لنگری برای استوار نگه داشتن خود در طوفان می‌بینند، دیگران افساری می‌دانند که ما را عقب نگه می‌دارد. آنچه برخی به عنوان گام‌های ضروری به سوی اوج پتانسیل انسانی ستایش می‌کنند، دیگران شتابی سرسام‌آور به سوی یک مفاک می‌بینند.
- در این مورد، واگرایی‌های غریزی عاطفی — و خطوط ذهنی که توسط همه طرف‌ها ترسیم می‌شود — وضعیتی غیرقابل پیش‌بینی و انفجاری ایجاد خواهد کرد. مواضع به طور فزاینده آشکار «برندگان» و «بازندگان» بالقوه، فشار این شرایط را تشدید خواهد کرد. ترسوها توسعه خود را کند کرده و توسعه دیگران را خرابکاری خواهند کرد. بیش از حد مطمئن‌ها قدرت خود را پنهان کرده و در خفا، کار خود را تسریع خواهند بخشید. خط زمانی بحران‌های آینده فراتر از تجربه پیشین بشری شتاب خواهد گرفت؛ به سرعت، ما غرق خواهیم شد، و مشخص نیست که آیا یا چگونه زنده خواهیم ماند.
- پاسخ معمول یک فناوری به چنین پیش‌بینی‌های نگران‌کننده‌ای این است که مزایای ملموس هوش مصنوعی را به ما یادآوری کند، که در حوزه علوم پزشکی از هم‌اکنون بسیار آشکار است. من مخالف نیستم. به نظر من، آلفا فولد — مدلی مبتنی بر شبکه عصبی که ساختارهای سه‌بعدی پروتئین را پیش‌بینی می‌کند — پیشرفتی بسیار مهم‌تر از ChatGPT بود. با این حال، علوم پزشکی در قرن بیستم پیشرفت‌های قابل مقایسه‌ای داشت. جنگ‌های جهانی و هولوکاست با این وجود رخ دادند، حتی در حالی که آنتی‌بیوتیک‌ها، واکسن‌های جدید و بی‌شمار درمان دیگر کشف و به طور گسترده در دسترس قرار گرفتند.

مشکل اصلی پیشرفت فناوریانه در طول عمر هنری کیسینجر خود را نشان داد. شکافت هسته‌ای در سال ۱۹۳۸ در برلین توسط دو شیمیدان آلمانی، اوتو هان و فریتز اشتراسمان، کشف شد. این پدیده در سال ۱۹۳۹ توسط فیزیکدانان

اتریشی‌الاصل، لیزه مایتنر و برادرزاده‌اش اوتو رابرت فریش، به صورت نظری توضیح داده شد (و نام‌گذاری شد). امکان یک واکنش زنجیره‌ای هسته‌ای که منجر به «تولید انبوه انرژی و عناصر رادیواکتیو، و متأسفانه شاید همچنین بمب‌های اتمی» شود، بینش فیزیکدان مجارستانی لئو زیلارد بود. این امکان که چنین واکنش زنجیره‌ای ممکن است در یک راکتور هسته‌ای برای تولید گرما نیز به کار گرفته شود، در همان زمان تشخیص داده شد. با این حال، تنها کمی بیش از پنج سال طول کشید تا اولین بمب اتمی ساخته شود، در حالی که تا سال ۱۹۵۱ اولین نیروگاه هسته‌ای افتتاح نشد. از خود پرسید: بشر در هشتاد سال گذشته کدام یک را بیشتر ساخته است: کلاهک‌های هسته‌ای یا نیروگاه‌های هسته‌ای؟ امروزه تقریباً ۱۲،۵۰۰ کلاهک هسته‌ای در جهان وجود دارد و این تعداد در حال حاضر رو به افزایش است زیرا چین به سرعت به زرادخانه هسته‌ای خود می‌افزاید. در مقابل، ۴۳۶ راکتور هسته‌ای در حال کار هستند. تولید برق هسته‌ای در ارقام مطلق در سال ۲۰۰۶ به اوج خود رسید و سهم آن از کل تولید برق جهان از ۱۵٫۵٪ در سال ۱۹۹۶ به ۸٫۶٪ در سال ۲۰۲۲ کاهش یافته است، که بخشی از آن نتیجه واکنش‌های سیاسی بیش از حد به تعداد کمی از حوادث هسته‌ای است که تأثیرات آنها بر سلامت انسان و محیط زیست در مقایسه با اثرات انتشار دی‌اکسید کربن از سوخت‌های فسیلی ناچیز بود.

درس زندگی هنری کیسینجر روشن است. پیشرفت‌های فناوریانه می‌توانند هم پیامدهای خوش‌خیم و هم بدخیم داشته باشند، بسته به اینکه ما به‌طور جمعی چگونه تصمیم به بهره‌برداری از آنها بگیریم. هوش مصنوعی البته از جهات بسیاری با شکافت هسته‌ای متفاوت است. اما این یک خطای فاحش خواهد بود که فرض کنیم ما این فناوری جدید را بیشتر برای اهداف سازنده به کار خواهیم برد تا اهداف بالقوه مخرب.

همین نوع بینش بود، که از تجربه تاریخی و همچنین نشأت گرفته بود، که هنری کیسینجر را بر آن داشت تا بخش بزرگی از زندگی خود را وقف مطالعه نظم جهانی و اجتناب از جنگ جهانی کند. همین بینش بود که باعث شد او با چنین اشتیاقی — و نگرانی — به پیشرفت‌های اخیر در هوش مصنوعی واکنش نشان دهد. و به همین دلیل است که این اثر پس از مرگ، به اندازه هر آنچه که او در طول زندگی طولانی و پربار خود نوشت، اهمیت دارد.

آکسفورد

ژوئیه ۲۰۲۴

به یاد هنری کیسینجر

در ۲۹ نوامبر ۲۰۲۳، دکتر هنری کیسینجر در سن یکصد سالگی درگذشت. او که الهام‌بخش تمام کسانی بود که می‌شناختندش، تا آخرین لحظات بر روی این مجلد، بیست و دومین کتابش، کار می‌کرد. در دیدارهای مکرر ما در آخرین سال زندگی‌اش، او با قاطعیت بر اهمیت موضوع کتاب و نیاز فوری به انتشار پیام آن پافشاری می‌کرد. ما، به عنوان دو همکار نویسنده‌اش، از آخرین افرادی بودیم که در روزهای پیش از مرگش او را دیدیم و با او سخن گفتیم. اکنون با به سرانجام رساندن این پروژه از جانب او و به درخواستش، کوشیده‌ایم تا اصالت اندیشه و طنین صدای او را در موضوعی که برای آینده بشریت از اهمیت بالایی برخوردار است، حفظ کنیم. به پایان رساندن آنچه او آغاز کرده بود، تا این آخرین اثر ادبی او با او نمیرد بلکه در جهان بدون او زنده بماند، تنها سهم کوچکی در بزرگداشت یاد اوست.

او پس از آنکه آن همه برای ساختن دنیای ما تلاش کرد، ساعات پایانی عمر خود را صرف تلاشی حیاتی برای نجات آن نمود. در واقع، آخرین کلام مکتوب او درخواستی است از تمام بشریت برای ادامه پروژه عظیم تضمین آینده گونه ما. در سال‌های میانی قرن پیشین، دکتر کیسینجر یکی از معماران اصلی تلاش فلسفی و دیپلماتیک برای محافظت از بشریت در برابر نابودی اتمی بود — مواجهه قرن بیستم با واقعیت‌های تلخ خطر وجودی (existential risk). این مدافع شجاع در برابر آن خطر، درست در زمانی از میان ما می‌رود که خطری جدید از راه می‌رسد. زندگی او درست زمانی به پایان می‌رسد که شکلی جدید از حیات آغاز می‌شود. در حالی که با سپیده‌دم عصر هوش مصنوعی روبرو هستیم، آگاهیم که چه تعداد اندکی جز دکتر کیسینجر — شاگرد قرن نوزدهم، استاد قرن بیستم، و پیشگوی قرن بیست و یکم — به اندازه او در موقعیتی مناسب برای بنیان نهادن آینده ما بوده‌اند.

دکتر کیسینجر پیش و بیش از هر چیز، یک فیلسوف تاریخ بود. از دل تحقیقات عمیق او در باب تراژدی، تلاشی مادام‌العمر برای اثبات این موضوع پدیدار شد که آرمان‌گرایی قلبی می‌تواند با واقع‌گرایی ذهنی سازگار باشد و آن را تعالی بخشد. به تعبیر نویسنده فرانسوی رومن رولان، می‌توان همزمان هم «بدینی خرد» و هم «خوش‌بینی اراده» را داشت. در جایی که خوش‌بین آرزوی کنترل قطعی انسان بر امور را دارد، بدین وضعیت ما را محصول نیروهایی فراتر از کنترل ما می‌بیند: قوانین طبیعت و چرخه‌های تاریخ.

بی‌تردید، او به‌خوبی می‌دانست که چگونه یک آرمان‌گرایی پرشور می‌تواند توسط ایدئولوگ‌هایی که در ریختن خون بی‌رحم و در اعمال زور بی‌پروا هستند، مورد سوءاستفاده قرار گیرد. فاشیسم، کمونیسم، توتالیتاریسم، بنیادگرایی مذهبی ستیزه‌جو — همگی ادعای دستیابی به آرمانی‌ترین اهدافی را داشته‌اند که در تاریخ ما دنبال شده است. او که ابتدا قربانی و سپس مبارزی نظامی و دیپلماتیک علیه چنین افراط‌گری‌های غیرانسانی بود، بر آن شد تا به بازسازی دوباره جهان، بر بنیانی از نظم بدون شرمساری و امنیت بدون احساس گناه، یاری رساند. دکتر کیسینجر از طریق مدیریت فعال خود در امور بین‌الملل، کشورخوانده‌اش — و جهان — را از میان تحولات پرآشوب عبور داد و خود را بر خاک سخت واقعیت‌های تاریخی و منافع ملی استوار ساخت.

دکتر کیسینجر با همان هوشمندی که در باب ضرورت گزینشی واقع‌گرایی داشت، یک آرمان‌گرا نیز بود — و به گفته زندگی‌نامه‌نویسش نایل فرگوسن، به «نقش آزادی، انتخاب و عاملیت انسان در شکل دادن به جهان» احترام می‌گذاشت. او در نظر و عمل، اعتقاد خود را نشان داد که انسان‌ها طوری زندگی نمی‌کنند و نمی‌توانند زندگی کنند که گویی آینده امری اجتناب‌ناپذیر است. هنری کیسینجر ۲۷ ساله در رساله ارشد خود در هاروارد با عنوان *معنای تاریخ*، با همان بحث فلسفی دست و پنجه نرم می‌کرد که اکنون آخرین اثر او را جان می‌بخشد: «هر تصویری که درباره ضرورت رویدادها

داشته باشیم، [...] و هرگونه که پس از وقوع، کنش‌ها را توضیح دهیم، تحقق آنها با اعتقاد درونی به انتخاب همراه بوده است.»

برای او مسلم نبود که بشریت از طرح‌های غیرانسانی برآمده از آتش کوره خود جان سالم به در خواهد برد. او که با چشم‌انداز هولناک فاجعه هسته‌ای روبرو — و زیر بار آن — بود، نه تسلیم جبرگرایی شد و نه به پیشگویی‌های شوم تن داد. درست است، ترس‌های وجودی می‌توانستند به نیهیلیسم منجر شوند، اما همچنین می‌توانستند در بهترین‌های ما، نیروی لازم برای به چالش کشیدن شر و دفاع از آنچه باید برای آینده‌گونه ما حفظ می‌شد را بدمند. در اوایل دهه ۱۹۵۰، به عنوان یک استاد جوان در هاروارد، او در مجموعه‌ای از جلسات شرکت کرد که در آن دانشمندان و دانشگاهیان برجسته‌ای مانند خودش گرد هم آمدند تا در مورد پیامدهای بالقوه جنگ هسته‌ای و اقدامات لازم برای جلوگیری از آن بحث و گفتگو کنند. از آن جلسات، دکتربین‌هایی پدید آمد که از آن زمان تاکنون جهان ما را از بدترین ترس‌های شرکت‌کنندگان در امان نگه داشته است.

دهه‌ها بعد، در گفتگو با ما دو نفر، او اغلب درباره آن جلسات — ساختار، هدف و اهمیت گذشته‌نگریشان — سخن می‌گفت. دیدگاه او تا پایان یکسان باقی ماند، دیدگاهی که هم تسلیم سرنوشت نبود و هم به هیچ چشم‌انداز آرمان‌شهری باور نداشت. همین توازن در مورد هوش مصنوعی نیز مانند زمینه هسته‌ای برقرار است: گروه‌های کوچک افراد متعهد می‌توانند با ورود به صحنه و تجلی بخشیدن به ارزش‌هایشان، تاریخ را تغییر دهند. با این حال، و فارغ از نبوغ دانشمندی که در حال ساخت هوش‌های جدید بودند، او معتقد بود که آموزش آنها به خودی خود برای تضمین حداقل ایمنی و امنیت لازم در عملکرد این ابزارهای جدید کافی نخواهد بود.

بدین منظور، میراث او در زمینه هوش مصنوعی تنها به کاوش‌های صرفاً فلسفی و علمی محدود نمی‌شود، بلکه شامل پیشنهادها و عملی نیز هست. نیم قرن پس از پرواز مخفیانه‌اش به پکن و گشایش متعاقب روابط بین ایالات متحده آمریکا و جمهوری خلق چین، دکتر کیسینجر یک سفر دیگر به پایتخت چین داشت. او به دعوت فوری و مشخص رئیس‌جمهور شی جین پینگ رفت تا به عنوان موضوع اصلی، درباره خطراتی که بشریت با هوش مصنوعی با آن روبروست، گفتگو کند. این آخرین سفر خارجی او و آخرین مأموریت دیپلماتیکش بود.

اگر در گذشته، دکتر کیسینجر مطالعه و عمل کشورداری را به سطح یک هنر ارتقا داد، امروز جستجوی او برای یافتن پاسخ، هوش مصنوعی را به موضوعی فراتر از علم بدل کرده است. او با یکی از ما، و با پروفسور دانیل هاتن‌لاکر از MIT، کتاب *عصر هوش مصنوعی: و آینده انسانی* ما را نوشت که در سال ۲۰۲۱ منتشر شد و پیش‌بینی کرد که ظهور هوش مصنوعی عصری نوین در تاریخ ایجاد خواهد کرد، که در تأثیرگذاری خود بر تغییر عمیق و غیرمنتظره تفکر انسانی، شبیه به عصر روشنگری قرن هجدهم است. با این حال، در این عصر جدید، بشریت به جای حرکت از پرسش‌های مطرح شده توسط انسان‌ها، با پاسخ‌هایی روبرو می‌شود که هوش مصنوعی به پرسش‌هایی که هیچ انسانی هرگز نپرسیده، ارائه می‌دهد. در حالی که هوش مصنوعی قلمروهای دانش بشری را تسخیر می‌کند، دکتر کیسینجر به دنبال آن بود که بر منابع خرد انسانی تکیه و از آن بهره‌برداری کند.

در این مجلد، ما به همراه دکتر کیسینجر تأثیر هوش مصنوعی را بر هشت حوزه مختلف فعالیت و تفکر انسانی بررسی می‌کنیم که با پاسخ فلسفی خود او به جستجوی مداوم برای یک راهبرد کارآمد جهت ایجاد توازن بین منافع و خطرات، به اوج می‌رسد. او در راستای این هدف، چشم‌اندازهای هم‌زیستی انسان با هوش مصنوعی و در زمان مقتضی، هم‌فرگشتی انسان و هوش مصنوعی را کاوش می‌کند. او با گشودن مفهومی امکان‌آستی بین این دو گونه — یکی ارگانیک، دیگری

مصنوعی — نیاز به یک انتخاب را نیز آشکار می‌سازد: خلق جهانی که در آن هوش مصنوعی بیشتر شبیه ما شود، یا جهانی که در آن ما بیشتر شبیه هوش مصنوعی شویم.

از زمان انتشار اولین کتابش در مورد عصر هوش مصنوعی، دکتر کیسینجر به طور فزاینده‌ای به محدودیتی در سودمندی خرد در روزهای پایانی چنین عصری پی برد. برای ما انسان‌ها، توضیحاتی فراتر از درک خودمان — یا ساخته خودمان — می‌توانند کاملاً گیج‌کننده به نظر برسند؛ گزینه ما این است که فرض کنیم آنها از توضیحات علمی ما کمتر پیشرفته و ابتدایی‌تر هستند، یک گام به عقب و نه به جلو. اما این فرضی خطرناک است.

اگر به گفته آرتور سی. کلارک، «هر فناوری به اندازه کافی پیشرفته از جادو قابل تشخیص نیست»، و اگر معجزات از ریاضیات مهندسی شوند، آینده باید غیرقابل توضیح، گیج‌کننده و حتی جادویی باشد. در طول دهه‌هایی که ما دکتر کیسینجر را می‌شناختیم، او سخاوتمندانه از آموخته‌های خود در کشورداری — آن قلمرو پیچیده امور انسانی — به ما آموخت که اگرچه خرد پارادایم غالبی بوده که انسان‌ها از طریق آن بر جهان خود مسلط شده‌اند، اما نمی‌تواند پارادایمی باشد که با آن بر خود مسلط شویم.

بنابراین، در حرکت رو به جلو، نباید انتظار داشته باشیم که صرفاً به خرد — سوخت تاریخی بزرگترین پیشرفت‌های بشری ما — تکیه کنیم. اما همچنین نمی‌توانیم آن را به طور کامل رها کنیم. تحقیقات نهایی او در مورد آینده ما، نه بی‌شابهت به میانه‌روی او بین آرمان‌گرایی و واقع‌گرایی، به تعادلی بین آن کیفیت تجربی حقیقت و چیزی دیگر دست می‌یابد — چیزی که از نظر فلسفی فراتر از خرد است اما از نظر زمانی از آن عقب‌تر است. درست همانطور که سیاست‌گذاری خارجی توان تحمل افراط در هیچ یک از این دو قطب را ندارد، چارچوب ما برای آینده نیز چنین توانی را ندارد.

بنابراین، هوش مصنوعی یک چالش منحصر به فرد است که نیازمند تفکری است که ممکن است در ابتدا غیرمنطقی یا اغراق‌آمیز به نظر برسد — و در واقع، سناریوهای توصیف شده در این کتاب تکان‌دهنده هستند. اما او با پند دادن به ما و دیگران مبنی بر اینکه خود او تنها شاگردی فروتن در مکتب بشریت و آخرین و بالقوه نهایی‌ترین مخلوق آن است، این نکته را بر ما مسلم ساخت که بزرگترین خطری که از سوی هوش مصنوعی متوجه ماست، این است که خیلی زود، یا خیلی کامل، اعلام کنیم که آن را می‌فهمیم.

عمق خرد و درک او از مردم، خصوصیات نیستند که انتظار داشته باشیم دوباره با آنها روبرو شویم. ما هیچ کس دیگری را نمی‌شناسیم که در سن ۹۳ سالگی بتواند بر یک حوزه کاملاً جدید و پیشتر ناشناخته از دانش فنی تسلط یابد. با کنجکاوی سیری‌ناپذیر و سرزندگی ذهنی‌اش، همراه با تعهد به کار و احساس رسالت، هیچ درد جسمی یا روحی هرگز برای فرونشاندن شور او برای پیشرفت کافی نبود. صرف نظر از ناتوانی‌های پیری، او هر روز با اراده‌ای شکست‌ناپذیر برای به پیش بردن جهان برمی‌خاست. قدرت بی‌بدیل او، شاید، از انضباط بی‌مانندش نشأت می‌گرفت — انضباطی که در جوانی با ظلم سخت شده، با خدمت در جنگ صیقل یافته، و برای دهه‌ها با کشمکش‌های زندگی عمومی آزموده شده بود.

ما تنها دو نفر از بسیاری هستیم که زندگی‌شان توسط این مرد خارق‌العاده شکل گرفته است. دلتنگ او خواهیم بود — بی‌تردید، به شیوه‌هایی بیش از آنچه اکنون می‌توانیم تصور کنیم. او که در آستانه عدم قطعیتی بزرگ ما را ترک می‌کند، اکنون بیش از هر زمان دیگری به او نیاز است. به همین دلیل است که برای این کتاب، هیچ عنوانی مناسب‌تر از *پیدایش* به نظر نمی‌رسید — یک سرآغاز جدید برای او و برای همه انسان‌ها. چه بشریت موفق شود و چه شکست بخورد، او

دیگر برای مشاهده نتیجه نهایی تلاش‌هایش حضور نخواهد داشت. دست‌کم اکنون خرد او را برای راهنمایی خود در اختیار داریم.

اریک اشمیت و کریگ ماندی

مقدمه مولفان

تنها چند سال پیش، هوش مصنوعی (AI) گوشه کوچکی از مباحث عمومی را به خود اختصاص داده بود. امروز، به دنبال پیشرفت‌های سریع، هوش مصنوعی موضوعی در صفحه اول خبرگزاری‌ها در همه جا و مسئله‌ای در ذهن رهبران علم، تجارت، روزنامه‌نگاری، خدمات عمومی، آموزش و سیاست در سراسر جهان است.

از دید ما، هم عموم مردم و هم بسیاری از کارشناسان این حوزه همچنان از جنبه‌های مهم این عصر جدید هوش مصنوعی غافل هستند. اشکال جدید هوش مصنوعی و واکنش‌های انسان به آنها می‌تواند چیزی کمتر از رابطه انسان با واقعیت و حقیقت، کاوش دانش و همچنین تکامل فیزیکی بشریت، نحوه دیپلماسی و نظام بین‌الملل را دگرگون نکنند. این‌ها از جمله مسائل حیاتی دهه‌های آینده هستند و باید دغدغه‌های راهنمای رهبران در هر عرصه‌ای باشند.

آخرین قابلیت‌های هوش مصنوعی، هرچند چشمگیر، در نگاه به گذشته ضعیف به نظر خواهند رسید زیرا قدرت آن با سرعتی فزاینده افزایش می‌یابد. قدرت‌هایی که ما هنوز تصور نکرده‌ایم، قرار است در زندگی روزمره ما نفوذ کنند. سیستم‌های آینده، پیشرفت‌های عظیم و عمدتاً مفیدی را تسهیل خواهند کرد، سلامت ما را بهبود بخشیده و در عین حال ثروت تولید می‌کنند.

اما این قابلیت‌ها با خطرات فنی و انسانی همراه هستند، که برخی از آنها شناخته شده و برخی ناشناخته‌اند. فناوری‌های امروزی از هم‌اکنون به شیوه‌هایی عمل می‌کنند که مخترعان‌شان پیش‌بینی نکرده بودند، و این الگو احتمالاً ادامه خواهد یافت. هر مسیر تحقیقاتی پرثمری که دانشمندان ما طی کنند — و بیش از یک مسیر وجود خواهد داشت — می‌تواند شاخه‌های جدیدی از قدرت‌های پیش‌بینی‌نشده را به بار آورد که ممکن است برای انسان‌ها قابل درک یا مفید باشند یا نباشند.

به نظر می‌رسد هوش مصنوعی مقیاس‌های زمانی انسان را فشرده می‌کند. اشیاء در آینده نزدیک‌تر از آن چیزی هستند که به نظر می‌رسند. به عنوان تنها یک مثال، ماشین‌هایی با قابلیت تعریف اهداف خود، چندان دور نیستند. اگر آمیدی به همگام شدن با خطرات مربوطه داشته باشیم، باید در کوتاه‌ترین بازه زمانی قابل تصور، پاسخ داده و عمل کنیم. با آگاهی از اهمیت موضوع و فوریت وظیفه‌ای که در پیش است، در اینجا تنها به چند جنبه از وجوه متعدد آن اشاره می‌کنیم.

با فراگیر شدن مشارکت‌های انسان و ماشین، انسان‌ها باید ماهیت مناسب این روابط را تعیین کنند. پاسخ‌ها ممکن است از منطق امنیت و کارایی گرفته شوند، از طریق مطالعه تاریخ استنتاج شوند، یا از امر الهی تشخیص داده شوند. افراد، ملت‌ها، فرهنگ‌ها و ادیان باید حدود اختیارات هوش مصنوعی بر حقیقت راه، در صورت وجود، تعیین کنند. آنها باید تصمیم بگیرند که آیا به هوش مصنوعی اجازه دهند تا به واسطه‌ای بین انسان و واقعیت تبدیل شود. در این راستا، آنها باید بین حفظ نقش سنتی فعالیت انسانی (در حالی که احتمالاً رهبری در کشف دانش جدید را به هوش مصنوعی واگذار می‌کنند) از یک سو، و رها کردن ذهن انسان با محدودیت‌های بیولوژیکی‌اش به نفع یک مشارکت بالقوه مهندسی‌شده با هوش مصنوعی در مرزهای فکری از سوی دیگر، انتخاب کنند. آیا ما اهداف خود را انتخاب می‌کنیم و هوش مصنوعی را برای دستیابی به آنها به کار می‌گیریم، یا اجازه می‌دهیم هوش‌های مصنوعی خودشان به انتخاب برخی از اهداف کمک کنند؟ به فوریت تمام، بشریت باید تعریفی مدرن و پایدار از کرامت انسانی ارائه دهد که بتواند جهت‌گیری فلسفی برای تصمیمات ما در سال‌های آینده فراهم کند.

از دید ما، ظهور هوش مصنوعی، مسئله بقای انسان است. همانطور که بعداً در این کتاب توضیح می‌دهیم، توانایی‌های آینده هوش مصنوعی، که با سرعت‌های غیرانسانی عمل می‌کنند، مقررات سنتی را بی‌فایده خواهند ساخت. ما به شکلی اساساً جدید از کنترل نیاز خواهیم داشت. برای جامعه علمی جهانی، وظیفه فوری، یافتن اقدامات فنی برای نهادینه کردن

پادمان‌های ذاتی در هر سیستم هوش مصنوعی است. به نوبه خود، ملت‌ها و سازمان‌های بین‌المللی، پس از رسیدن به یک اجماع، باید ساختارهای سیاسی جدیدی برای نظارت، اجرا و واکنش به بحران ایجاد کنند. این امر مستلزم حل نه یک، بلکه دو «مسئله هم‌راستاسازی» (alignment problems) است: هم‌راستاسازی فنی ارزش‌ها و نیت انسانی با اقدامات هوش مصنوعی، و هم‌راستاسازی دیپلماتیک انسان‌ها با هم‌نوعان خود.

دکتر هنری کیسینجر، یکی از نویسندگان این کتاب، دو همکار خود را در مورد مسئله دوم از نزدیک راهنمایی کرد، در حالی که آنها — به عنوان رهبران ارشد سابق مایکروسافت و گوگل او را در مورد مسئله اول هدایت می‌کردند. کریگ ماندی رابط اصلی سیاست‌گذاری فناوری مایکروسافت با واشنگتن و دولت‌های خارجی در سراسر جهان بود و هم‌زمان بر تحقیقات مایکروسافت نظارت داشت، پیش از آنکه اخیراً به رهبری سازمان تحقیقاتی OpenAI مشاوره دهد. اریک اشمیت به مدت یک دهه گوگل را رهبری کرد و سپس دهه بعد را صرف ایجاد شبکه‌ای از استعدادها و ایده‌ها در تلاقی فناوری، بشردوستی و امنیت با هدف حفاظت از بشریت کرد.

فوریت مسائلی که ما با هم با آنها روبرو شده‌ایم به حدی است که، به جای منتظر ماندن برای یک بحران، ضروری می‌دانیم که جامعه ما، و در واقع گونه ما، به طور پیشگیرانه به آنها بپردازد. و، اگرچه آرمان ایمنی انسان یکی از مؤلفه‌های ضروری یک پاسخ موفق به هوش مصنوعی است، اما نمی‌تواند به تمام سؤالاتی که هوش مصنوعی مطرح می‌کند پاسخ دهد — زیرا، در عصر هوش مصنوعی، بشریت تغییر خواهد کرد. سؤال این است که آیا و تا چه حد ما انسان‌ها انتخاب خواهیم کرد که بر چگونگی وقوع آن تغییر، همچنان اعمال قدرت کنیم.

بخش اول : در آغاز

فصل ۱ : کشف

کشف شاید هیجان‌انگیزترین ظرفیت گونه انسان باشد. ما که با کنجکاوی رانده و با شگفتی به وجد می‌آییم، خلأهایی را که درک می‌کنیم پر کرده و پرسش‌هایی را که مطرح می‌کنیم به پاسخ تبدیل می‌نماییم. کاوش آنچنان با تعریف ما از خود عجین است که علی‌رغم خطرات و ناکامی‌های آشکار، بی‌وقفه در مسیرهای متعدد آن به پیش رفته‌ایم.

در طول تاریخ، کاوش انسان، به‌ویژه در محیط فیزیکی ما، داستانی از شجاعت عظیم در مواجهه با خطر شدید بوده است. تیم‌های انسانی که برای چنین تلاش‌هایی عازم می‌شدند، اغلب در رقابتی با میرایی خود بودند. در اوایل قرن شانزدهم، سفر فردیناند ماژلان به دور کره زمین، یک ادیسه سه‌ساله مملو از خشونت، گرسنگی و مرگ بود. سفر ماژلان اولین سفری بود که با موفقیت مسیری به دور کره زمین را طی کرد. در این فرآیند، او رکورد استقامت در دریاهای آزاد را شکست، ابعاد سیاره را مشخص کرد، و — با توجه به اینکه این اقدام در چارچوب استعمار اروپا صورت گرفت — راه را برای تبادلات اجتماعی و اقتصادی در مقیاس بین‌المللی گشود.

بیشتر ملوانان ماژلان می‌دانستند که در حال به جان خریدن فاجعه هستند. اگرچه دیگر به طور گسترده باور بر این نبود که زمین صاف است، اما گرد بودن آن به عنوان یک واقعیت اثبات نشده بود و بسیاری از خدمه او ممکن است از افتادن از لبه دنیا ترسیده باشند.

هم او و هم خدمه‌اش آگاه بودند که بهای سنگینی برای هرگونه فرض اشتباه خواهند پرداخت. و در واقع اشتباه کردند و بها پرداختند: آنها طول سفر خود، مقادیر غذای مورد نیاز برای زنده ماندن، خطرات مسمومیت‌های دسته‌جمعی، و خطر کشتی‌های آسیب‌دیده یا غیرعملیاتی را دست‌کم گرفته بودند. از ناوگان اصلی متشکل از پنج کشتی که تقریباً ۲۷۰ خدمه را حمل می‌کرد، تنها یک کشتی با هجده بازمانده شیخ‌گون به بندر اسپانیا بازگشت. ناخدایشان در میان آنها نبود، او در طول راه پس از اصابت یک تیر زهرآلود به پایش، جان باخته بود.^۱

در ۴۰۰ سال پس از دوران ماژلان، هر گوشه از جهان نقشه‌برداری شد — به جز قطب جنوب، سرزمینی به ویرانی یک سیاره بیگانه. کاوشگر انگلیسی-ایرلندی، ارنست شکلتون، بیش از هر کس دیگری قبل از خود به رسیدن به قطب جنوب آن سرزمین نزدیک شد: انتهای دنیا. در سال ۱۹۰۹، با خدمه‌ای بی‌تجربه و بدون حمایت هیچ دولتی، بلکه تنها با چند وام خصوصی و کمک‌های فردی، شکلتون و مردانش رکورد طولانی‌ترین سفر این‌چنینی را ثبت کرده و راه را برای کاوشگران پس از خود هموار ساختند.

اگرچه شکلتون نتوانست ادعا کند که اولین کسی است که به خود قطب جنوب رسیده، اما احترام نسل‌های بعدی را به دست آورد. او این کار را با اولویت دادن به ارزش‌های انسانی بر جاه‌طلبی‌های اکتشافی انجام داد. یک سال پس از آغاز سفر، اعضای تیم، که روزانه به مدت ده ساعت طاقت‌فرسا به سورت‌مه‌ها بسته شده بودند، هر روز تنها چند مایل پیشروی می‌کردند. اگرچه آنها غذای کافی برای رسیدن به قطب را داشتند، اما آذوقه باقیمانده برای بازگشتشان به کشتی کافی نبود. همانجا، در فاصله ۹۷ مایلی پیروزی، شکلتون، به جای به خطر انداختن جان مردانش، تصمیم به بازگشت گرفت. او در دفتر خاطراتش اعلام کرد: «ما بهترین تلاش خود را کردیم.»^۲ در بازگشت، شکلتون تنها بیسکویت روزانه خود را به فرانک وایلد، یکی از اعضای بیمار خدمه، تعارف کرد. وایلد در خاطراتش نوشت: «تمام پولی که تاکنون ضرب شده نمی‌توانست آن بیسکویت را بخرد، و یاد آن فداکاری هرگز از ذهنم پاک نخواهد شد.»^۳

شکلتون که از شکست دلسرد نشده بود، بعدها سفرهای دیگری به قطب جنوب انجام داد. برای دهه‌ها، داستانی (که بعداً رد شد) نقل می‌شد که او آگهی زیر را در روزنامه تایمز لندن منتشر کرده بود:

مردانی برای سفری پرخطر تحت تعقیبند. دستمزد کم، سرمای تلخ، ماه‌های طولانی تاریکی مطلق، خطر دائمی، بازگشت سالم بعید. افتخار و شهرت در صورت موفقیت.^۴

این آگهی ممکن است جعلی بوده باشد، اما فداکاری‌ای که به تصویر می‌کشد، جعلی نبود. اینها واقعیت‌های اکتشاف تنها یک قرن پیش بود؛ پیشرفت ما در مرزهای دانش به تعداد روح‌های شجاعی محدود بود که چنین احتمالات ناچیزی را می‌پذیرفتند.

شاید به دلیل شناخت این خطرات، برخی دولت‌ها صلاح دیدند که از سفرهای اکتشافی حمایت کرده و به آنها پاداش دهند، و این ventures به بخشی از بازی رقابت بین‌المللی تبدیل شدند. به عنوان مثال، سفر ماژلان با سیاست تعریف می‌شد. ماژلان که نتوانسته بود حمایت مالی پادشاه پرتغال — حاکم خودش — را جلب کند، فرار کرده و در عوض تحت حمایت سخاوتمندانه‌تر پادشاهی اسپانیا به دریا زد. پس از مرگ او، خدمه یک اسپانیایی به نام خوان سباستین الکانو را برای فرماندهی انتخاب کردند. در سفر بازگشت، الکانو که ناامید و تقریباً بدون غذا و آذوقه بود، تلاش کرد در جزایر کیپ ورد — مستعمره پرتغال در سواحل غرب آفریقا — توقف کند و سیزده نفر از خدمه را برای مذاکره با فرماندار مستعمره به ساحل فرستاد. اما درخواست آنها با ردی تحقیرآمیز مواجه شد.

الکانو با قلبی شکسته اما اکنون با عزمی راسخ‌تر برای نشان دادن حماقت پرتغال و دستاورد اسپانیا، دستور داد لنگر را بکشند و به سمت خانه ادامه دهند. (او اعضای مأموریت پیشرو خود را در سرزمینی غیردوستانه رها کرد). سرانجام با به پایان رساندن سفر — و تحقق بخشیدن به چشم‌انداز ماژلان — الکانو به پادشاه اسپانیا که اکنون امپراتور مقدس روم و قدرتمندترین حاکم جهان بود، نوشت:

اعلیحضرت بهتر از هر کس دیگری خواهند دانست که آنچه ما باید بیش از همه ارج نهیم و به آن بچسبیم این است که ما تمام گردی جهان را کشف کرده و پیموده‌ایم، که با رفتن به غرب، از شرق بازگشته‌ایم.^۵

سرانجام، انسان‌های بی‌باک و حامیان سیاسی آنها، پروژه اکتشاف را کاملاً فراتر از کره خاکی ما بردند. ما نه تنها به مطالعه سیارات — که از هزاران سال پیش تمرینی جاری بود — بلکه به احساس کردن سطوح آنها در زیر پاهایمان پرداختیم. پس از جنگ جهانی دوم، «مسابقه فضایی» بین ایالات متحده و اتحاد جماهیر شوروی، مکملی برای جنگ سرد ژئوپلیتیک، دو ابرقدرت را به رقابتی برای فرستادن انسان‌ها به جایی که هرگز پیش از این نرفته بودند، کشاند. خود فضانوردان را می‌توان پیشگام در نظر گرفت، اما هیچ‌یک از سفرهای فضایی که مسکو و واشنگتن انجام دادند، محصول یک قمار فردی بر سر زندگی و ثروت نبود. بلکه، هر یک مأموریتی هماهنگ‌شده دیپلماتیک و نظامی بود که با سرمایه‌گذاری‌های عظیم پول، زمان و آزادی عمل برای آزمایش، قدرت می‌گرفت. در هفتاد سال گذشته، بیش از ۶۰۰ فضانورد آسمان‌ها را شکافته و فضا را گشته‌اند، که برخی از آنها فراتر رفته و به دور ماه چرخیده، در مدار آن قرار گرفته یا بر روی آن راه رفته‌اند. ۶ بدین ترتیب، رقابت آمریکا و شوروی ما را همزمان به آستانه نابودی هسته‌ای و به سوی ستارگان بالا برد.

بیش از یک قرن پیش از عصر اکتشافات — نامی که به دوران مدرن اولیه ماژلان، واسکو دا گاما، آمریکا و سپوچی و بسیاری دیگر داده شد — جاه‌طلبی‌های چین در دریاهای آزاد از نظر گستره و مقیاس بی‌رقیب بود. «ناوگان گنج» ژنگ هه، دریاسالار بزرگ دوران مینگ، با کوچک شمردن منابع و حمایتی که دولت‌های غربی به کاوشگران ارائه می‌دادند، متشکل از ده‌ها و گاهی حتی صدها کشتی پیشرفته بود که ده‌ها هزار ملوان، سرباز، دیپلمات و بازرگان را حمل می‌کردند. هر یک از سفرهای این دریاسالار چینی دو سال به طول می‌انجامید. کشتی‌های او پس از پیمودن آب‌های اقیانوس آرام

در نزدیکی سواحل جنوب شرقی چین، به سمت غرب به خلیج بنگال، اقیانوس هند، دریای عرب، دریای سرخ و سرانجام سواحل سواحیلی شرق آفریقا ادامه دادند، در کاوش‌هایی که تقریباً سه دهه از ۱۴۰۵ تا ۱۴۳۳ به طول انجامید. ۷ سفرهای ژنگ هه، که از نظر منشأ و انگیزه به برنامه‌های فضایی مدرن آمریکا و شوروی نزدیک‌تر بودند تا به معادل‌های غربی در زمان خود، چندان پرشی به سوی ناشناخته‌ها نبودند، بلکه محصولات غنی کشورداری امپراتوری بودند. اما در همین جا نیز مشکلی نهفته است. همان چیزهایی که برای تضمین موفقیت یک سفر، حتی در یک دولت با منابع خوب، لازم است، می‌تواند در درازمدت موجب نابودی آن شود. سیاست‌ها تغییر می‌کنند. اولویت‌ها عوض می‌شوند. صبر انسان به راحتی تمام می‌شود. هزینه‌های هنگفتی که برای «ناوگان گنج» صرف شد، به قدری زیاد بود که جناح‌هایی در دربار مینگ شروع به سرزنش امپراتور برای تأمین مالی این مأموریت‌ها کردند. شکست‌های سیاسی و بلایای طبیعی، ناآرامی‌ها را تشدید کرد. در نهایت، دولت چین تصمیم گرفت بهترین کشتی‌های خود را به همراه بسیاری از سوابق سفرهای ژنگ هه، نابود کرده یا عمدتاً به حال خود رها کند، مبادا یک رویاپرداز مشابه دوباره برخیزد و رهبری کشور را مجذوب و اغوا کند. کشتی‌ها پوسیدند و کشتی‌هایی مانند آنها تا ۴۰۰ سال بعد دیده نشدند.

به همین ترتیب، پس از پیروزی ایالات متحده در مسابقه فضایی، بدون وجود رقیبی برای انگیزه بخشیدن به تلاش‌های ملی‌اش، حمایت در واکنش برای اکتشافات فضایی کاهش یافت و بودجه سازمان ملی هوانوردی و فضایی (ناسا) قطع شد. در طول پنج دهه، ظرفیت آمریکا برای پروازهای فضایی سرنشین‌دار به طور پیوسته کاهش یافت. از تنها ملتی در زمین که قادر به فرود آوردن انسان بر روی ماه بود، ایالات متحده به جایی رسید که تنها توانایی حمل انسان به «مدار پایین زمین» را داشت، و سپس، در نهایت، توانایی بردن هر کسی به مدار را به کلی از دست داد.

شهرت آمریکا تنها با تلاش‌های کاوشگران خصوصی نجات یافت که به رهبری شرکت هوافضای اسپیس‌ایکس، جاه‌طلبی‌های کیهانی را در غرب احیا کردند. تلاش‌های آن شرکت اکنون برخی را به این نقطه تاکنون غیرقابل تصور رسانده است که نزدیکترین همسایگان آسمانی ما را نه به عنوان مقصدهایی موقت، بلکه به عنوان خانه‌هایی دائمی ببینند. یک قرن پیش، شکل‌تون به ایجاد حضور انسانی در قطب جنوب زمین کمک کرد. امروز، دهانه شکل‌تون — فرورفتگی‌ای که به افتخار او در قطب جنوب، نه قطب جنوبگان، بلکه ماه، نام‌گذاری شده — محل برنامه‌ریزی شده برای پایگاه بعدی بشریت است.

وجود حامیان جایگزین برای ادامه اکتشافات حیاتی بوده است. در اروپای ماژلان، اگر یک پادشاه از حمایت خودداری می‌کرد، یک کاوشگر می‌توانست از دیگری بودجه جمع‌آوری کند. تا قرن بیستم، ارنست شکل‌تون که آرزوی تصاحب قطب جنوب را برای امپراتوری رو به زوال بریتانیا داشت که با شروع جنگ جهانی اول در سال ۱۹۱۴ حواشی پرت شده بود و نمی‌توانست بودجه لازم را از پادشاهی تأمین کند، عمدتاً به اهداکنندگان خصوصی روی آورد. ظهور شرکت‌های خصوصی و سودآور — که امکان تجمع سرمایه‌گذاری و ریسک را فراهم می‌کنند — امکانات بیشتری را گشود. باید پرسید که اگر ژنگ هه و جانشینانش چنین گزینه‌هایی داشتند، چه بر سرشان می‌آمد.

ورود هوش مصنوعی

برای طولانی‌ترین دوره در تاریخ غرب، کاوش واقعیت بر موجودیت‌های جغرافیایی متمرکز بود — سیاره ما و نزدیک‌ترین همسایگان آسمانی‌مان. همانطور که انسان‌ها به تدریج بر محیط فیزیکی بی‌واسطه خود — در خشکی، در دریا و در آسمان‌های بالا — تسلط یافتند، تنها مسئله زمان بود تا غریزه بی‌قرار انسانی ما برای کشف، دامنه خود را از فضای اطرافمان به ایده‌های درونمان گسترش دهد. امروز، ما در مرز کاوش، نه فیزیکی، بلکه فکری ایستاده‌ایم.

توسعه هوش مصنوعی، عصر جدیدی از اکتشاف را آغاز کرده است. در جایی که هوش مصنوعی در سیستم‌های فیزیکی ادغام می‌شود، حسگرهای رباتیک نقش‌هایی را بر عهده می‌گیرند که پیش از این توسط انسان‌ها انجام می‌شد، و بدین ترتیب کشف انسانی را از خطر فیزیکی برای کاوشگر رها می‌سازند و در نتیجه، صفوف توسعه‌دهندگان و سرمایه‌گذاران مشتاق را چند برابر می‌کنند.

هوش مصنوعی همچنین ترسی را تجربه نمی‌کند، و بنابراین از گستره وسیع واقعیت هراسی ندارد. شرم نیز احساس نمی‌کند، و بنابراین بدون تردید شکست می‌خورد — اما یک هوش مصنوعی آنچنان سریع می‌تواند خود را بازتنظیم کند که از طریق بداهه‌پردازی و آزمایش مداوم، می‌تواند نرخ‌های بالای شکست را بدون ایجاد وقفه برای توسعه‌دهندگان و سرمایه‌گذاران مذکور، پذیرا باشد.

امروزه، کشف با هوش مصنوعی پروژه‌ای است که تقریباً به طور انحصاری توسط شرکت‌های خصوصی و کارآفرینان رهبری می‌شود و دولت‌ها به عنوان حامیان مکمل ظاهر می‌شوند. با این حال، حتی بدون اقدامات تکمیلی دولت، رشد و گسترش هوش مصنوعی احتمالاً با تغذیه از فراوانی منابع متنوع سرمایه، بدون وقفه ادامه خواهد یافت. درست است، در مراحل هنوز ابتدایی توسعه امروزی، ممکن است مقادیر قابل توجهی از سرمایه انسانی و حمایت اجتماعی همچنان مورد نیاز باشد؛ اما تداوم کاوش با هوش مصنوعی در آینده ممکن است دیگر یک بار مالی و سیاسی بر جوامعی که آن را به کار می‌گیرند، نباشد. در غیاب تحولات پیش‌بینی‌نشده و برخلاف اعصار پیشین کاوش — که همگی پیش از رسیدن به پتانسیل کامل خود پایان یافتند — می‌توانیم انتظار داشته باشیم که کشف از طریق و همراه با هوش مصنوعی بدون کاهش ادامه یابد.

با این وجود، اگرچه هوش مصنوعی تا حدی از قید و بندهای پیشین بر سر راه کشف رها شده است، اما نمی‌تواند از همه آنها بگریزد — به‌ویژه با آشکارتر شدن تأثیراتش. میزان تحمل ریسک در میان جوامع دموکراتیک و آینده نامشخص بازی‌های بین‌المللی همچنان یک عامل مجهول (X factor) مهم در حوزه هوش مصنوعی خواهد بود. شاید داستان به شکل یک «مسابقه هوش مصنوعی» درآید. شاید به چیزی معادل نابودی کامل «ناوگان گنج» ژنگ‌هه توسط دولت مینگ منجر شود. یا شاید، دولت‌ها پیشرفت را در مسیری میانه هدایت کنند.

ذهن علامه

در نگاه به گذشته، بدیهی به نظر می‌رسد که قلمرو کشف محکوم به گسترش فراتر از حوزه فیزیکی — جولانگاه دریاسالاران، فضانوردان و ماجراجویان — بود و کاشفان آن نیز متنوع‌تر می‌شدند. و در واقع، از همان اوایل، تاریخ شاهد ظهور نهایی نوعی جدید — یا اگر نه جدید، پس قاطعانه متفاوت — از کاشف انسانی بود: علامه (polymath).

افراد علامه که به دلیل توانایی‌شان در تسلط بر حوزه‌های متعدد دانش استثنایی هستند، که هر یک به طور معمول می‌توانست کار یک عمر را به خود اختصاص دهد، در طول تاریخ شاید به حدس، تنها به چند صد نفر برسند. چه وقف هنر شده باشند چه علوم، یا هر دو، همگی سرشار از شور انقلابی کردن، یا از ابتدا خلق کردن، کل رشته‌های مطالعاتی بوده‌اند. آنها نه چندان با شجاعت قلبی، بلکه با قدرت محض ذهن به پیش رانده شده و بی‌باکانه به اعماق دانش و تخیل بشری قدم گذاشته‌اند: سرزمینی حتی پهناورتر از آنچه یک کاوشگر معمولی دنیای فیزیکی با آن روبروست.

گاهی اوقات، به این افراد منحصربه‌فرد با ظرفیت‌های شگفت‌انگیزشان برای رمزگشایی از جهان ما، هم با هیبت و هم با سوءظن به عنوان جادوگر، یا به عنوان واسطه‌های احتمالی با خالق الهی جهان نگریسته می‌شد — شهرتی که اغلب می‌توانست آنها را با مقامات مذهبی یا سیاسی در تضاد قرار دهد. در مواقع دیگر، به دلیل استعدادهایشان گرمای داشته شده، تشویق به ادامه مطالعات خود تحت حمایت مستقیم همین مقامات می‌شدند و برای تلاش‌هایشان پاداش می‌یافتند.

در عصر طلایی اسلام، علامه‌ها به دنبال راه‌هایی برای پیشگامی در علم در خدمت ایمان بودند. ابن هیثم، از بصره در عراق امروزی، مفهوم روش علمی را پیشنهاد کرد — پنج قرن پیش از آنکه معاصران رنسانس در غرب ادعای آن را مطرح کنند. ابن هیثم که به همان اندازه در هندسه، نجوم، اپتیک و روانشناسی تجربی تبحر داشت، دانش عمیقی نیز در مهندسی هیدرولیک داشت.

همین دانش اخیر او را با مذهب در تضاد قرار داد. او که ادعای توانایی تنظیم طغیان رود نیل را داشت، پدیده‌ای طبیعی که در آن زمان هنوز معتقد بودند امری ماوراء طبیعی است، برای ملاقات با خلیفه به بغداد دعوت شد. در آنجا مشخص شد که پروژه‌های مهندسی پیشنهادی او با الهیات اسلامی در تضاد است. به عنوان مجازات برای ادعاهای جسورانه و تفکر انقلابی‌اش، او مجبور به پنهان شدن تا زمان مرگ خلیفه شد.

علامه‌های دیگر مانند محمد بن موسی خوارزمی — یک ایرانی از قلمرو ترکمنستان امروزی — موفقیت بیشتری در حوزه‌هایی از دانش یافتند که صراحتاً در خدمت اربابان الهیاتی خود بودند. خوارزمی به عنوان منجم دربار و رئیس کتابخانه بیت‌الحکمه در بغداد منصوب شد. ۹ نجوم تحت حمایت خلفای عباسی شکوفا شد، که حامیان سخاوتمند مردانی چون او برای کمک‌های مستقیمشان به ایمان اسلامی شدند. به عنوان مثال، مختصات جغرافیایی مکان‌های مقدس، و مهمتر از همه جهت مکه — دانشی که برای مقاصد نماز اسلامی الزامی است — به لطف دسترسی بهبود یافته نجوم قرون وسطی به موقعیت دقیق ستارگان، بسیار دقیق‌تر قابل محاسبه شده بود.

معاصران ابن هیثم و خوارزمی راه‌های خلاقانه‌تری برای حفظ روحیه کشف از طریق تعمیق اتحاد بین علم و دین یافتند و پناهگاهی را تضمین کردند — برای کسانی که به قول ابن رشد (همچنین شناخته شده به نام اوروئس)، یکی دیگر از علامه‌های بزرگ مسلمان همان دوران، دارای «وحدت عقل» بودند — در عصری که نه به خاطر دستاوردهای عقلی‌اش، مانند روشنگری اروپا، بلکه به خاطر دینداری پرشورش نام‌گذاری شده بود. ۱۰

هزاران کیلومتر دورتر از بغداد — و از یکدیگر — علامه‌های چینی و هندی خود را نه با قدرت الهی، بلکه با ساختارهای حکومتی همسو کردند و بدین ترتیب به نزدیکی متمایزی به سیاست دست یافتند که هم با امتیاز و هم با تلاش به دست آمده بود. در قرن دوازدهم، هماغاندر — «دانای تمام دانش در زمان خود» — به عنوان مشاور پادشاه کوماراپالا در گجرات امروزی خدمت می‌کرد. قرن‌ها بعد، اکبر کبیر — امپراتور جوان مغول — هم خود دولت را اداره می‌کرد و هم به موفقیت‌های علمی در معماری، مهندسی و ادبیات دست می‌یافت.

در مورد نوابخ چینی، علاوه بر مهارتشان در امور فکری، آنها نیز عمیقاً در خدمت دربار به عنوان مشاور، مقامات دانشمند، و مدیران عالی‌رتبه مشغول بودند. دولت به همان اندازه که حامی آنها بود، مشتری آنها نیز به شمار می‌رفت. در هر ساعتی از روز، ممکن بود این متفکران را در حال نظارت بر تدوین علم تقویم در وظایف خود به عنوان مقامات ارشد دیوان نجوم بیابید؛ یک ساعت بعد، ممکن بود به کابینه مرکزی امپراتور در مورد بهترین راه برای افزایش بازده محصولات کشاورزی مشاوره دهند. به آنها مأموریت ساخت ماشین‌های جنگی بزرگ داده می‌شد، به عنوان سفیر در مأموریت‌های دیپلماتیک به پادشاهی‌های همسایه اعزام می‌شدند، و وظیفه مشاوره به امپراتور در مورد مسائل سیاست اقتصادی را بر عهده داشتند.

اما آنها استقلال نداشتند و تنها تا جایی تحمل می‌شدند که امپراتور به خدمات فکری آنها نیاز داشت. در دنیای چین، سیاست‌های جناحی به همان اندازه مانعی برای نبوغ بود که محدودیت‌های روحانیون در جهان اسلام. نوابی که در هر نسل یک بار ظهور می‌کردند، در معرض رحمت همان سیستمی بودند که در وهله اول استعدادهای آنها را شناسایی کرده بود، و علم همچنان فرمانبردار پسر آسمان بود. شن کو — علامه‌ای از سلسله سونگ — توسط یک افسر نظامی حسود

برکنار شد و در نهایت از چشم امپراتور افتاد و در نتیجه یک رقابت سیاسی با — به طرز عجیبی — تنها علامه بزرگ دیگر آن سلسله، به انزوا کشانده شد.^{۱۱}

علامه‌های تنها در دوران باستان و قرون وسطی در خاورمیانه، هند و چین شکوفا شدند. اما تا پس از عصر کاوش نبود که تحقیقات مفهومی نظام‌مند ابتدا در اروپا و بعداً در ایالات متحده، در آنچه اکنون عصر خرد یا عصر روشنگری می‌نامیم، آغاز شد. چهار قرن بعدی — که ما را به آغاز عصر هوش مصنوعی می‌رساند — که با رنسانس قرن پانزدهم و شانزدهم پیش‌زمینه و تسهیل شده بود، دوره‌ای اساساً متفاوت برای کشف فکری بود.

در دوران پیش از روشنگری، علامه‌ها چاره‌ای جز خدمت به یک قدرت برتر نداشتند، خواه آن قدرت یک امپراتور باشد یا یک خلیفه. در مقابل، بسیاری از چهره‌های برجسته روشنگری اروپا قادر به پیگیری بینش‌های خود نه به عنوان وسیله‌ای برای اهداف سیاسی یا الهیاتی، بلکه به عنوان اهدافی در خودشان بودند. لئون باتیستا آلبرتی، علامه ایتالیایی، با افتخار در مورد انسان رنسانس ادعا کرد: «انسان اگر بخواهد می‌تواند همه کارها را انجام دهد.»

با این حال، اگرچه هوش برای کاوش ضروری است، اما کافی نیست. کاوشگران علاوه بر حفظ اشتها برای ریسک، باید توسط منابع مناسب، محیط مناسب و همکاران مناسب نیز حمایت شوند. در دوران روشنگری، کاوشگران می‌توانستند به هر سه دسترسی پیدا کنند. دولت‌ها و شرکت‌ها، عمدتاً به دلیل علاقه‌شان به ترجمه نظریه علمی به کاربردهای نظامی و تجاری، حامیان و شرکای فعالی برای علامه‌های یورو-آتلانتیک باقی ماندند، در حالی که، در بیشتر موارد، آنها را آزاد گذاشتند تا انرژی‌ها و مهارت‌های خود را آنطور که صلاح می‌دانند هدایت کنند. حتی زمانی که تلاش‌هایی برای به خدمت گرفتن، سرکوب، یا مداخله به طرق دیگر صورت می‌گرفت، اروپا به اندازه‌ای پراکنده بود که به متفکرانی با ایده‌های نامطلوب در یک مکان اجازه می‌داد تا در مکانی دیگر خانه‌ای بیابند. بنابراین، فرانسوی، فرانسوا-ماری آروئه، که بیشتر با نام مستعار ولتر شناخته می‌شود، زمان قابل توجهی را در خارج از فرانسه گذراند، در حالی که میخائیل لومونوسوف روسی، پس از تصمیم‌گیری در سن نوزده سالگی برای «تحصیل علوم»، از شرایط محدودکننده روستای زادگاهش در شمال دور تمام راه را تا مسکو پیاده طی کرد، جایی که قبل از تحصیل در کیف و سپس در آلمان در دانشگاه ماربورگ و سپس فرایبرگ، تحصیلات ابتدایی را کسب کرد.^{۱۲}

یک نتیجه برجسته این بود که پیشرفت بشر اکنون با تمرکز و پیوند جدیدی از متفکران همفکر در نزدیکی فیزیکی یا ذهنی به پیش‌رانده می‌شد، و هوش با بالاترین کیفیت را به رقابت و همکاری وامی‌داشت. پیش از این، داستان نوابغ انسانی داستانی تنها بود که در استبداد مکان و زمان محبوس شده بود. علامه‌ها که اغلب در انزوا کار می‌کردند و به ندرت شخص دیگری در مجاورت مکانی یا زمانی خود برای همکاری و همدردی داشتند، تنها می‌توانستند مرز را تا جایی که ظرفیت‌های خودشان اجازه می‌داد به پیش برانند. علاوه بر این، این ارتباط محدود با چند پیشگام فکری دیگر در داخل و بین دولت‌ها، تقریباً تکرار تلاش توسط مخترعانی را که از تحقیقات یکدیگر اطلاع مستقیمی نداشتند و نمی‌توانستند بر روی کار یکدیگر بنا کنند، تضمین می‌کرد.

به تدریج، برای کسانی که به اندازه کافی خوش‌شانس بودند تا به مواد دقیق، به‌موقع و با ترجمه وفادارانه دسترسی داشته باشند، اختراعات پیچیده می‌توانست با تلاش گروهی — نه تنها به طور همزمان بلکه در طول نسل‌ها — کنار هم چیده شود. تا زمان روشنگری، علامه‌ها قادر به پل زدن نه تنها بین رشته‌ها، بلکه بین حوزه‌های جداگانه‌ای از درک بودند که هرگز پیش از این آشتی داده نشده یا در یک کل منسجم ادغام نشده بودند. دیگر علم فارسی یا علم چینی وجود نداشت؛ فقط علم وجود داشت.

این قابلیت برای ادغام دانش از حوزه‌های متنوع به تولید پیشرفت‌های سریع علامه‌ای کمک کرد و با گذشت زمان بهترین تلاش تا آن زمان برای «هوش جمعی» را به اثبات رساند. به عنوان مثال، در قرن بیستم، پروژه منهن در جنگ جهانی دوم شاهد چنین تراکم نامتناسبی از توانایی ذهنی بود، زیرا با همکاری یکدیگر، درخشان‌ترین ذهن‌های آن دوران، نسل‌ها فیزیک نظری را در کمتر از پنج سال به کاربردی ویرانگر ترجمه کردند — شاهکاری که برای پیشینیان آنها غیرقابل تصور بود. مؤسساتی مانند مؤسسه مطالعات پیشرفته در دانشگاه پرینستون و شرکت رند در کالیفرنیا به طور مشابه به پناهگاه‌هایی برای ذهن‌های با استعداد تبدیل شدند.

طبیعتاً برخی علامه‌ها که زیر بار نبوغ خود بودند، همچنان ترجیح می‌دادند به تنهایی کار کنند. یکی از آنها مخترع صرب-آمریکایی، نیکولا تسلا بود:

ذهن در انزوا و تنهایی بی‌وقفه، تیزتر و نافذتر است. برای فکر کردن به آزمایشگاه بزرگی نیاز نیست. اصالت در انزوای آزاد از تأثیرات بیرونی که بر ما می‌کوبند تا ذهن خلاق را فلج کنند، شکوفا می‌شود. تنها باشید، این راز اختراع است؛ تنها باشید، آن زمانی است که ایده‌ها متولد می‌شوند. به همین دلیل است که بسیاری از معجزات زمینی در محیط‌های فروتنانه سرچشمه گرفته‌اند.^{۱۳}

اما تسلا استثنا بود، نه قاعده. قرن بیستم یک انفجار کامبرین از علم کاربردی را به وجود آورد و بشریت را با سرعت و مقیاسی بی‌مقایسه فراتر از تکامل‌های پیشین به جلو پرتاب کرد. گروه‌هایی از علامه‌ها که با ترکیب ظرفیت‌های ذهنی خود قدرت گرفته بودند، اکنون به ابزارهای مدرن نیز مجهز بودند. تأثیر انباشته چنین شتاب شدیدی به ما کمک کرده است تا بر تعداد زیادی از محدودیت‌های انسانی غلبه کنیم. ارتباطات دیجیتال و جستجوی اینترنتی، که خود محصول گروه‌های علامه‌ای هستند، امکان بزرگ شدن چنین گروه‌هایی و گردآوری دانشی بسیار فراتر از توانایی‌های پیشین انسان را فراهم کرده‌اند.

البته، محدودیتی نیز وجود داشته است. مهم نیست که چقدر طراحی وسایلی را که قرار است ما را به مرزهای جدید ببرند بهینه کنیم، و مهم نیست که چقدر نبوغ را در عملیات کاری سازماندهی کنیم، محدودیت‌های بیولوژیکی و کاستی‌های انسانی همچنان ظرفیت‌های ما را محدود می‌کنند. زمان ما بر روی زمین محدود است. ما به خواب نیاز داریم. به راحتی خسته می‌شویم. به استراحت و مهلت نیاز داریم. حتی زمانی که سر کار هستیم، بیشتر انسان‌ها تنها می‌توانند بر روی یک کار در یک زمان تمرکز کنند.

به عنوان مثال، در اواخر قرن نوزدهم و اوایل قرن بیستم، توماس ادیسون — خود یک علامه و رقیب تسلا — را در نظر بگیرید که هزاران آزمایش او برای ایجاد یک «لامپ الکتریکی رشته‌ای عملی» سه سال طول کشید تا به موفقیت برسد. درست است، ادیسون تا حدی با کمک به بهبود تلفن، اختراع الکساندر گراهام بل، حواسش پرت شده بود. اما حتی با کمک دستیاران بسیار، تلاش او برای یک لامپ به تلاشی فوق‌العاده نیاز داشت. هنوز هم امروزه، تحقیق و توسعه بسیاری از فناوری‌های پیشرفته گران، طاقت‌فرسا، طولانی و از نظر سیاسی و روانی برای سرمایه‌گذاری دلهره‌آور باقی مانده است.

از آنجا که نتایج بسیار نامشخص هستند، کاوش در مرزهای فیزیکی باقیمانده — فضا، اعماق دریا، پوسته داخلی زمین — همچنان پروژه تنها موفق‌ترین شرکت‌ها و ثروتمندترین دولت‌هاست.

البته، دقیقاً همین احتمالات نامساعد است که باعث می‌شود کشف هنوز معنادار و حتی معجزه‌آسا به نظر برسد. فیزیکدان جان فون نویمان، که بعدها به عنوان یکی از آخرین علامه‌های بزرگ شناخته شد، توسط فاینشال تایمز به عنوان «شخص قرن» نام گرفت زیرا مظهر اعتماد به نفس مشخصه قرن بیستم در قدرت ذهن برای «مهار و رام کردن

دنیای فیزیکی» بود. ۱۴ در واقع، فون نویمان خود را با انرژی شدیدی وقف مسائل محوری مانند نظریه ریاضی و بمب اتمی کرده بود، اما مهمتر از همه، وقف ایجاد کامپیوتر: هم آخرین پیشرفت بزرگ قرن و هم از جمله آخرین اختراعاتی که انسان‌ها تا به حال مجبور به تصور و تولید در انزوا بوده‌اند.

با وجود عقل‌های برجسته‌ای مانند فون نویمان، بشریت ممکن است به حدود بالایی توانایی هوش انسانی تقویت‌نشده برای گسترش افق‌های فکری ما رسیده باشد. علامه بودن به ویژه نادر است زیرا معمولاً تسلط بر اصول اولیه یک رشته آنقدر طول می‌کشد که تا زمانی که هر علامه بالقوه‌ای این کار را انجام دهد، دیگر زمانی برای رشته دیگری نخواهد داشت و به طور بالقوه توانایی تفکر خلاق را نیز از دست داده است. در حال حاضر، به نظر می‌رسد نوآوری امروزه به طور فزاینده‌ای بیشتر از تیم‌هایی از افراد ناشی می‌شود تا از هر نابغه واحدی که مجهز به بینش درخشان میان‌رشته‌ای باشد. با این حال، ادغام دانش به دست آمده از چندین ذهن فردی، فرآیندی دشوار باقی می‌ماند. حتی، یا شاید به ویژه، در میان روشنفکران برجسته، تعداد زیاد همکاران می‌تواند مانع ارتباط هم‌افزا شود.

هوش مصنوعی — در مقابل — علامه نهایی خواهد بود. در کاوش مرز دانش بشری، قادر است توده‌های اطلاعات را با سرعتی سهمگین پردازش و بازنمایی‌هایی از آنها تولید کند. الگوها را در بی‌شمار ابعاد و رشته‌ها به طور همزمان ارزیابی می‌کند و ارتباطی بی‌سابقه ایجاد می‌کند. کارایی آن به آن اجازه می‌دهد تا از محدودیت‌های کشف انسانی فراتر رود، تا جایی که حتی انتظار می‌رود در ادغام بسیاری از فعالیت‌های فکری در یک «وحدت دانش» جدید، به قول جامعه‌شناس زیستی آمریکایی ای. او. ویلسون، موفق شود. ۱۵

درست همانطور که دستاوردهای علامه‌های روشنگری به اتصال اطلاعات بستگی داشت، پیشرفت‌های اخیر در یادگیری ماشین تنها با مقیاس عظیم داده‌ها — هوش جمعی — ممکن شده است که نه تنها توسط هوش مصنوعی امروزی فعال شده، بلکه به راحتی در دسترس قرار گرفته است.

برای گسترش بیشتر این قیاس، شاید جای تعجب نباشد که آخرین پیشرفت در تکنیک‌های هوش مصنوعی، به جای تکیه بر یک برنامه بزرگ واحد برای انجام کار به تنهایی، نتیجه‌گیری‌های چندین برنامه کوچکتر را در آنچه به عنوان «ترکیبی از متخصصان» (mixture of experts) شناخته می‌شود، ترکیب کرده است. ما پیش‌بینی می‌کنیم که این آخرین نمونه از قدرت تقویت‌شده علامه‌ها در گروه‌ها نخواهد بود.

پروژه کلی کاوش، تا به امروز، با کمیت و کیفیت انسان‌ها در مرز محدود شده است. ما فقط چند هزار پیشگام فیزیکی و تعداد بسیار کمتری علامه داشته‌ایم. بنابراین، هوش مصنوعی آماده است تا انقلابی در کاوش فیزیکی و فکری ایجاد کند. هوش مصنوعی، همانطور که اشاره کردیم، فاقد ترس و شرم است و بنابراین بی‌احساس به سوی مرز، طبق دستور، می‌شتابد. علاوه بر این، کاوش هوش مصنوعی در واقعیت، که به همان اندازه برای کاوش کیلومترها از فضای بیرونی و نانومترها از زیست‌شناسی انسان مجهز است، به طور قابل توجهی توسط تجربه ذهنی یا کار فیزیکی، توسط قدرت مغز انسان یا حواس انسانی محدود نمی‌شود. همچنین، وقتی صحبت از واقعیت فیزیکی می‌شود، کاوش ماشینی از ما نمی‌خواهد که جان خود را فدای این هدف کنیم؛ برعکس، ممکن است زمان بسیار کمتری از ما را بطلبد — یعنی فقط زمانی که توسط انسان‌ها برای آن فراهم می‌شود.

در آینده، محدودیت اصلی هر جامعه معین ممکن است دیگر تعداد علامه‌های با استعدادی نباشد که می‌تواند برای فراهم کردن موتور کوچک و گاهی ناسازگار پیشرفت علمی گرد هم آورد. دیگر پتانسیل بشریت با تعداد کل ماژلان‌ها یا تسلاهایی که تولید می‌کنیم محدود نخواهد شد. و قوی‌ترین ملت جهان نیز ممکن است دیگر آن ملتی نباشد که بیشترین آلبرت اینشتین‌ها و جی. رابرت اوپنهاইمرها را دارد، تا زمانی که آن ملت بتواند هوش مصنوعی را ایجاد کرده و سپس از

آن تا حداکثر پتانسیل خود استفاده کند. و این امکان یک تغییر پارادایم در استاندارد اصلی اندازه‌گیری قدرت ملی را مطرح می‌کند، که در طول قرن‌ها از قلمرو، به منابع، به سرمایه، به سرمایه انسانی — و اکنون، شاید، به سرمایه محاسباتی — حرکت کرده است.

علاوه بر این، یک ماشین یادگیرنده به خوبی می‌تواند به یک ماشین خود-بهبودبخش تبدیل شود. در نهایت، آیا ممکن است آخرین اختراع علامه‌ای — یعنی محاسبات، که قدرت ذهن انسان را به روشی اساساً متفاوت از هر ماشین قبلی تقویت کرد و دهه‌ها بعد پیشرفت‌ها در هوش مصنوعی را تسهیل نمود — به خاطر جایگزین کردن مخترعان خود به یاد آورده شود؟

عصر سوم کشف

از دیدگاه هوش مصنوعی، دانش انباشته بشریت مانند مجمع‌الجزایری از جزایر آتشفشانی است که بر فراز اقیانوسی بی‌کران پراکنده شده‌اند. در این طرح خیالی، مرکز جغرافیایی هر جزیره تحت سلطه یک قله آتشفشانی است: همانطور که نگاه بیننده به سمت پایین به سوی دریا می‌لغزد، قطعیت محو می‌شود و به شیب‌های پایین‌تری از اطمینان کاهش می‌یابد تا به ساحل برسد. ۱۶

با فرض اینکه، برای اهداف این تمرین خیالی، مقدار کافی آب از اقیانوس‌های زمین تخلیه شده باشد، آنگاه بلافاصله یک توپولوژی وسیع از زمین‌های زیر آب که تا کنون تقریباً برای چشم انسان نامرئی بوده است، دیده می‌شود؛ دیگر جزایر بلافاصله به عنوان خشکی‌های آزاد شناور در اقیانوس ظاهر نمی‌شوند، بلکه صرفاً به عنوان برآمدگی‌های سنگی نمایان کوه‌ها یا آتشفشان‌های غول‌پیکر زیر آب — که از پایه غوطه‌ور خود در کف اقیانوس بالا آمده و فقط به اندازه‌ای بلند هستند که سطح را بشکافند — به نظر می‌رسند.

اگر در این طرح، هر جزیره به عنوان نماینده یک رشته از درک بشری در نظر گرفته شود، آبی که هر یک را از دیگری جدا می‌کند، نماینده ارتباطات ناقصی است که هنوز باید برای پیشبرد درک ما از جهان به عنوان یک کل بالقوه منسجم، کشف شوند. اگرچه ممکن است به طور موقت در نقشه‌برداری از واقعیت بی‌واسطه خود احساس امنیت کنیم، اما تصور کمی از آنچه در زیر ما، یا فراتر از ما نهفته است، داریم. هوش مصنوعی می‌تواند این را تغییر دهد.

قلمرو فیزیک را به عنوان مثال علمی quintessential در نظر بگیرید. اگر اسحاق نیوتن قوانین جهان‌های آسمانی و زمینی را هماهنگ کرد، و اگر مایکل فارادی و جیمز کلرک ماکسول همین کار را برای الکتریسیته، مغناطیس و اپتیک انجام دادند، جستجو هنوز برای یک «نظریه وحدت بزرگ» که دو نظریه جداگانه و ناسازگار را که برای توضیح وجود ما از دو سر مخالف واقعیت رقابت می‌کنند، آشتی دهد، ادامه دارد. اینها نظریه کیهانی (نسبیت عام) و نظریه زیراتمی (مکانیک کوانتومی) هستند.

هوش مصنوعی ممکن است سرانجام به حوزه‌های به ظاهر نامتجانس درک، نظم و ساختار بیخشد و در این فرآیند (مانند مجمع‌الجزایری با همان زیرساخت) ارتباطی بین حوزه‌هایی مانند ژنتیک، زبان‌شناسی، کیهان‌شناسی و روانشناسی را آشکار سازد. هوش مصنوعی حتی ممکن است به آشتی دادن شکاف بین و میان مکاتب فکری یا نظام‌های اعتقادی به ظاهر ناسازگار کمک کند.

در بسیاری از رشته‌ها، ما قبلاً طیف وسیعی از حقایق ممکن را شناسایی کرده‌ایم، اگرچه بسیاری از آنها احتمال صدق واقعی پایینی دارند. در مجمع‌الجزایر درک بشری، اینها نقاطی در امتداد خط ساحلی هستند: نه چهل، اما لزوماً هم دانش نیستند. هوش مصنوعی با هدایت به یک حوزه تحقیق در امتداد آن ساحل، می‌تواند با دقتی فوق‌العاده پرثمرترین راه‌ها

را برای کاوش بیشتر قضاوت کند. با انتخاب، آزمایش، معکوس کردن، و انتخاب مجدد در توالی سریع، می‌تواند اثرات میلیون‌ها انتخاب ممکن را ارزیابی کند.

این همان روشی است (همانطور که در فصل ۵ بیشتر بررسی می‌کنیم) که آزمایشگاه دیپ‌ماینند گوگل نه تنها توانست بر بازی تخته‌ای باستانی چینی «گو» آنطور که انسان‌ها می‌شناختند مسلط شود، بلکه تا آنجا که ماشین دانش خود را به انسان‌ها نشان داد، به دانش خود ما از بازی نیز افزود. در مقایسه با برنامه‌های شطرنج‌باز قبلی، که اغلب به محاسبات brute-force تکیه می‌کردند، آلفاگو، که قبلاً خود را بر روی سی میلیون حرکت قبلی «آموزش» داده بود، توانایی ماشین‌ها را برای استدلال انتزاعی به نمایش گذاشت. ۱۷

به این معنا، آموزش ماشین شبیه به «آموزش» ذهن یک دانشجوی دکتری فلسفه بود: در مورد دوم، یک فرآیند تدریجی ساختن ظرفیت تفکر و استدلال از طریق سال‌ها مطالعه فشرده. مانند دانشجویی که از آن سال‌ها بیرون می‌آید تا به سؤالات در دفاع از رساله خود پاسخ دهد، سیستم دیپ‌ماینند خود را «آموزش» داد تا از آموخته‌هایی که قبلاً در معرض آن قرار گرفته بود فراتر رود و — از درون مایه‌سازی انتزاعی‌تر و سطح بالاتر خود — حرکات شطرنجی را تولید کند که استنباط می‌کرد بالاترین احتمال پیروزی را دارند.

در واقع، در برخی مواقع، مدل هوش مصنوعی در انتخاب حرکاتی موفق شد که هرگز پیش از این توسط یک انسان در ۴۰۰۰ سال بازی امتحان نشده بود — که تنها به این دلیل ممکن بود که، در حالی که ذهن انسان به نظر می‌رسد به دستکاری تنها چهار متغیر مستقل در یک زمان محدود است، هوش مصنوعی می‌تواند قضاوت‌های احتمالی بی‌شماری را از ابعاد بی‌شمار به طور همزمان به کار گیرد. ۱۸. بنابراین هوش مصنوعی به ایده‌های اصلی دسترسی پیدا کرد و برای اولین بار، آنها را به درون مرزهای تجربه انسانی آورد.

کاربر انسانی که از یک مدل هوش مصنوعی پرس‌وجو می‌کند — به عنوان مثال، با تایپ یک سؤال در ChatGPT — از آن نمی‌خواهد که صرفاً یک نقطه اطلاعاتی را ارزیابی کند، همانطور که موتورهای جستجوی معمولی انجام می‌دهند، بلکه می‌خواهد چندین نقطه اطلاعاتی را ترکیب کرده و بر آن اساس نتیجه‌گیری ارائه دهد. با حرکت همزمان در چندین جهت و چندین بعد، بازنمایی‌هایی از اطلاعات را در فضای چندبعدی (high-dimensional space) تولید می‌کند، که شامل روابطی در داخل و میان بی‌شمار رشته‌ها و زیررشته‌ها است؛ و از آن بازنمایی‌های پیچیده شبکه‌ای، نتیجه‌گیری‌های خود را استخراج می‌کند.

اینجا، در پاسخ به سؤالات ما، هدیه ظرفیت و سرعت به ظاهر فوق بشری «مدل‌های زبان بزرگ» هوش مصنوعی نهفته است، که بر روی مقادیر عظیمی از داده‌ها از پیش آموزش دیده‌اند. از آنجا که دقت پاسخ‌ها سطح اطمینان ما را به حقایق مختلفی که ما انسان‌ها از آنها حمایت می‌کنیم تعیین می‌کند، چنین مدل‌هایی زمین‌شناسی عمیق را با جزئیات فزاینده‌ای تولید می‌کنند.

هوش مصنوعی احتمالاً دانش جدید را نه تنها با سرعت زیاد، بلکه به روش‌هایی که طیفی از کاوش‌های اضافی را باز می‌گذارد، انباشته خواهد کرد. در بازی شطرنج خود، آلفاگو تمایل داشت به سمت راه‌حل‌هایی با گشودگی خاصی گرایش پیدا کند. بنابراین برخی از مدل‌های هوش مصنوعی ممکن است در آموزش خود، تعصبی را نسبت به حوزه‌هایی که پتانسیل بالایی برای گزینه‌های متعدد ارائه می‌دهند، جذب کرده باشند، و بدین ترتیب کاوش سریع و انعطاف‌پذیر به جلو را ممکن سازند.

ممکن است برای انسان‌ها دشوار باشد که با این شیوه جدید کاوش هوش مصنوعی سازگار شوند. جدی‌ترین چالش شامل این خواهد بود که آیا و چگونه چنین کاوشی، درک ما از واقعیت و هدف انسانی ما را منعکس می‌کند — یا با آن