New Scientist
ESSENTIAL GUIDE №23

هوشیو

ماشین چگونه می‌آموزد؟
مسیر رسیدن به خودآگاهی
ماشین‌های متفکر و حسی
دانشمندان هوش مصنوعی و ریاضی‌دانان
چشم‌اندازهای آینده از آرمان‌شهر تا انقراض
و بیشتر

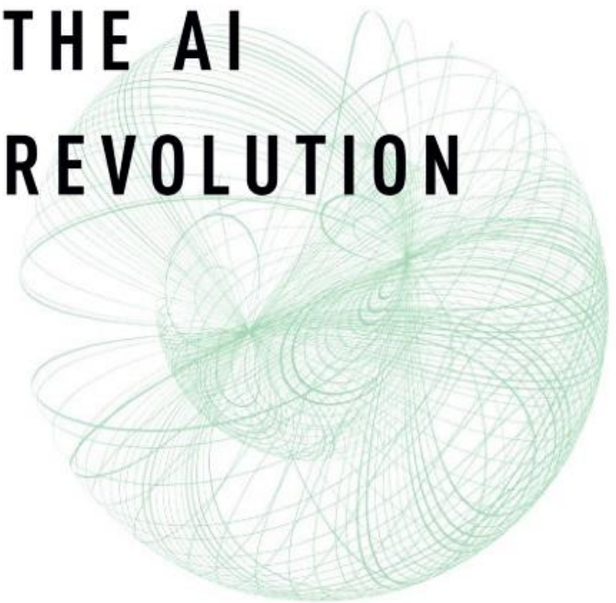
انقلاب هوش مصنوعی

دوران جدید هوش مصنوعی برای بشریت چه معنایی دارد؟

سردبیر:
کریس استوکل-واکر

NEW SCIENTIST ESSENTIAL GUIDE

THE AI REVOLUTION



وقتی نخستین نسخه راهنمای کلیدی هوش مصنوعی را در سال ۲۰۲۰ نوشتیم، این فناوری در دوره‌ای کاملاً متفاوت بود. ما در آن زمان هوش مصنوعی را «شاید انقلابی‌ترین تحول فناورانه یک نسل که با سرعتی حیرت‌آور در حال رخ دادن است» توصیف کردیم؛ اما نمی‌دانستیم چه چیزهایی در راه است.

انتشار ChatGPT در نوامبر ۲۰۲۲ آغازگر عصر هوش مصنوعی مولد بود. این ابزارهای قدرتمند، محتوای نوآورانه‌ای مانند متن، تصویر و موسیقی تولید می‌کنند؛ آن هم بر اساس الگوهای موجود در مجموعه‌های عظیم داده‌های آموزشی. آن‌ها دارند همه آنچه درباره توانایی‌ها و نحوه کارکرد این فناوری می‌دانستیم را دگرگون می‌کنند.

اکنون بیش از هر زمان دیگری، توجه رسانه‌ها، افکار عمومی و نهادهای قانون‌گذاری به هوش مصنوعی جلب شده است. بنابراین برای بیست‌وسومین شماره از این مجموعه راهنماهای کلیدی (Essential Guides)، مناسب به نظر می‌رسد که دوباره بررسی کنیم هوش مصنوعی چگونه ما و جهان اطراف‌مان را در آینده نزدیک تغییر خواهد داد. در صفحات پیش رو، بررسی می‌کنیم این فناوری چه معنایی برای زندگی، کار و برنامه‌های علمی و ریاضی ما خواهد داشت.

ما میزان آگاهی این نسل جدید از هوش مصنوعی نسبت به کارهایی که انجام می‌دهد؛ یا اینکه آیا صرفاً یک ترفند هوشمندانه است را بررسی می‌کنیم و برخی چشم‌اندازهای آینده را ترسیم می‌کنیم؛ اینکه این فناوری، چه به نفع و چه به ضرر، ممکن است ما را به کجا ببرد.

عناوین دیگر مجموعه راهنماهای کلیدی را می‌توانید در www.shop.newscientist.com تهیه کنید؛ همچنین بازخوردهای شما در نشانی essentialguides@newscientist.com پذیرفته می‌شود.

کریس استوکل-واکر
Chris Stokel-Walker

درباره سردبیر

«کریس استوکل-واکر» روزنامه‌نگار و ویراستار آزاد حوزه فناوری و نویسنده کتاب «اپلیکیشن دینامیتی چین و رقابت ابرقدرت‌ها برای رسانه‌های اجتماعی» (TikTok Boom: China's dynamite app and the superpower race for social media) است.

NEW SCIENTIST ESSENTIAL GUIDES
NORTHCLIFFE HOUSE, 2 DERRY STREET,
LONDON, W8 5TT
+44 (0)203 615 6500

NEW SCIENTIST ESSENTIAL GUIDES
ARE PUBLISHED BY NEW SCIENTIST LTD
ISSN 2634-0151
PRINTED IN THE UK BY
PRECISION COLOUR PRINTING LTD AND
DISTRIBUTED BY MARKETFORCE UK LTD
+44 (0)20 3148 3333

COVER: IMAGINIMA/ISTOCK
ABOVE: NAQIEWEI/ISTOCK
CHAPTER BACKGROUNDS:
JUST_SUPER/UNDEFINED/ISTOCK

SERIES EDITOR Thomas Lewton
EDITOR Chris Stokel-Walker
DESIGN Craig Mackie
SUBEDITOR Bethan Ackerley
PRODUCTION AND APP Joanne Keogh
TECHNICAL DEVELOPMENT (APP)
Amardeep Sian
PUBLISHER Roland Agambar
DISPLAY ADVERTISING
+44 (0)203 615 6456
displayads@newscientist.com

ADDITIONAL CONTRIBUTORS
Neil Briscoe, Michael Brooks, Daniel Cossins, Edd Gent, Emma Hart, Jeremy Hsu, Graham Lawton, Thomas Lewton, April Reese, Timothy Revell, Mordechai Rorvig, Amanda Ruggeri, Mary-Ann Russon, Matthew Sparkes, Katharine Sanderson, Chris Stokel-Walker, Alex Wilkins, James Woodford

فصل ۱

لحظه هوش مصنوعی

هوش مصنوعی مدت‌هاست که به‌عنوان فناوری فردا و همیشه در آستانه ورود در نظر گرفته می‌شود. اما از زمان انتشار ChatGPT در سال ۲۰۲۲، همراه با مدل‌های هوش مصنوعی مولد مشابه که به راحتی محتوا و ایده‌های جدید تولید می‌کنند، فردا بالاخره به امروز تبدیل شده است.

فصل ۲

زیر پوست

هوش مصنوعی مولد با مجموعه‌ای از نوآوری‌ها نیرو گرفته است. برخی از این نوآوری‌ها، تجدید حیات ایده‌های قبلی هستند که محققان به دلیل محدودیت‌های سخت‌افزاری قادر به تحقق آن‌ها نبودند. برخی دیگر، تغییرات پارادایم در معماری‌های هوش مصنوعی هستند که جهش‌هایی را در قابلیت‌های آن‌ها ممکن ساخته است.

فصل ۳

هوش مصنوعی مولد چگونه می‌تواند کمک کند؟

شرکت‌های فناوری اغلب هوش مصنوعی را «برهم‌زننده یا اختلال‌گر» توصیف می‌کنند؛ به این معنی که به طور چشمگیری نحوه عملکرد جوامع و اقتصادهای سراسر جهان را تغییر می‌دهد.

فصل ۴

ماشین‌های متفکر و حسی

ما مستعد این هستیم که ماشین‌ها را دارای ویژگی‌هایی شبیه به انسان ببینیم. اما آیا هوش مصنوعی واقعاً می‌تواند استدلال کند، همدلی کند یا تجربه‌ای آگاهانه از جهان داشته باشد؟

حتی اگر ماشین‌ها فقط این ویژگی‌ها را تقلید کنند، ما برای حمایت عاطفی به آنها روی می‌آوریم.

فصل ۵

دانشمندان ماشین

مدل‌های هوش مصنوعی قدرتمند به ما در یافتن راه‌حل‌هایی برای تغییرات اقلیمی و درمان بیماری‌های مختلف کمک می‌کنند. این ماشین‌های هوشمند شیوه علم را تغییر داده‌اند. به‌جای ربات‌هایی که به‌عنوان ابزار یا دستیار عمل می‌کنند، هوش مصنوعی به همکاران خلاقی تبدیل می‌شود که می‌توانند قوانین جدید طبیعت را کشف کرده و قضایای ریاضی را اثبات کنند.

فصل ۶

چشم‌اندازهای آینده: از آرمان‌شهر تا انقراض

برای برخی، لحظه کنونی در هوش ماشینی بسیار انرژی‌بخش است؛ نشانه‌ای از اینکه ما از رسیدن آینده‌ای آرمانی دور نیستیم. برای برخی دیگر، این لحظه نشانه‌ای نگران‌کننده از انقراض به دست یک فناوری بسیار قدرتمند است.

فصل ۱

لحظه هوش مصنوعی

ریشه واژه «هوش مصنوعی» (Artificial Intelligence)، به‌عنوان یک اصطلاح به دهه ۱۹۵۰ میلادی برمی‌گردد. هوش مصنوعی مدت‌هاست که به‌عنوان فناوری فردا و همیشه در آستانه ورود در نظر گرفته می‌شود. اما از زمان انتشار ChatGPT در سال ۲۰۲۲، همراه با مدل‌های هوش مصنوعی مولد مشابه که به راحتی محتوا و ایده‌های جدید تولید می‌کنند، فردا بالاخره به امروز تبدیل شده است.

برخی از کارشناسان این دوره را تابستان هوش مصنوعی می‌نامند و برخی دیگر می‌گویند زمستان دیگری برای هوش مصنوعی در راه است. در مورد چرخه‌های هیجان‌انگیز هوش مصنوعی و فرصت‌های ناشی از این رویداد فناورانه تحول‌آفرین چه باید کرد؟

ورود هوش مصنوعی به جریان اصلی

تحقیقات مربوط به هوش مصنوعی معمولاً با هیاهویی درباره ظرفیت‌ها و خطرات این فناوری همراه است. اما چه آن‌هایی که مرزهای توانایی‌های ممکن را جابه‌جا می‌کنند و چه آن‌هایی که خواستار احتیاط هستند، ظاهراً بر سر یک موضوع توافق دارند. نسل جدید هوش مصنوعی مولد، بیش از هر ابزار هوش مصنوعی پیش از خود، جامعه را تحت‌تأثیر قرار داده است.

این لحظه برای هوش مصنوعی بی‌سابقه است. مدل‌های زبانی قدرتمند به طور ناگهانی جهشی بزرگ در توانایی‌های خود داشته‌اند و اکنون می‌توانند حجم زیادی از نوشته‌های معقول تولید کنند؛ متن‌هایی که اغلب از نوشته انسان قابل‌تشخیص نیستند. آن‌ها می‌توانند به پرسش‌های فنی پیچیده نیز پاسخ دهند، از جمله پرسش‌هایی که معمولاً از وکلا و برنامه‌نویسان پرسیده می‌شود. حتی می‌توانند به آموزش بهتر سایر مدل‌های هوش مصنوعی نیز کمک کنند.

در مرکز این تحولات، مدل‌های زبانی بزرگ و انواع دیگر هوش مصنوعی مولد قرار دارند که می‌توانند در پاسخ به فرمان‌های انسانی، متن و تصویر تولید کنند. از سال ۲۰۲۲ استارتاپ‌هایی که با پشتیبانی غول‌های فناوری فعالیت می‌کنند، روند عرضه این مدل‌های مولد را شتاب بخشیده‌اند و دسترسی میلیون‌ها نفر به چت‌بات‌هایی قانع‌کننده اما اغلب نادقیق را امکان‌پذیر کرده‌اند؛ درحالی‌که اینترنت را با نوشته‌ها و تصاویر تولیدشده توسط هوش مصنوعی پر کرده‌اند، به شکلی که می‌تواند جامعه را دگرگون کند.

داستان هوش مصنوعی چرخه‌ای تکرارشونده از موج‌های علاقه و سرمایه‌گذاری است که پس از برآورده‌نشدن انتظارات بزرگ، وارد رکود می‌شود. در دهه ۱۹۵۰، شوروشوق زیادی برای ساخت ماشین‌هایی وجود داشت که بتوانند هوشی در سطح انسان نشان دهند. خود اصطلاح «هوش مصنوعی» در کارگاهی تابستانی در کالج Dartmouth در New Hampshire در سال ۱۹۵۶ ابداع شد. اما آن هدف بلندپروازانه محقق نشد؛ زیرا سخت‌افزار و نرم‌افزار رایانه‌ای به‌سرعت با محدودیت‌های فنی روبه‌رو شدند. نتیجه آن، به‌وجودآمدن زمستان‌های هوش مصنوعی در دهه ۱۹۷۰ و اواخر دهه ۱۹۸۰ بود؛ دورانی که سرمایه‌گذاری و علاقه شرکت‌ها به‌شدت کاهش یافت.

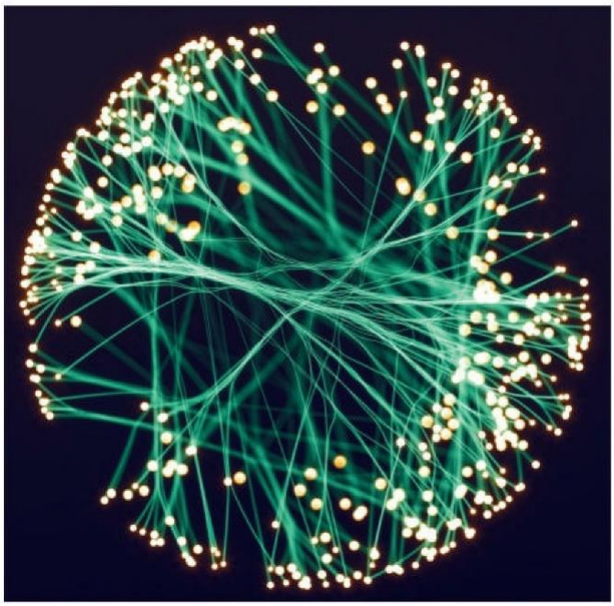
دهه گذشته میلادی را می‌توان به‌نوعی تابستان هوش مصنوعی دانست؛ هم برای پژوهشگرانی که به دنبال بهبود توانایی یادگیری هوش مصنوعی بودند و هم برای شرکت‌هایی که قصد داشتند این فناوری را به کار گیرند. به لطف ترکیبی از پیشرفت‌های عظیم در قدرت پردازش رایانه و دردسترس‌بودن داده‌ها، رویکردی که از ساختار مغز الهام گرفته بود توانست به موفقیت‌های زیادی برسد. قابلیت‌های تشخیص صدا و چهره در گوشی‌های هوشمند معمولی با چنین مدل‌های شبکه عصبی کار می‌کنند. همچنین مدل‌های هوش مصنوعی قدرتمندی که بهترین بازیکنان جهان را در بازی باستانی Go شکست داده‌اند و چالش‌های علمی بسیار دشوار را حل کرده‌اند؛ مانند پیش‌بینی ساختار تقریباً همه پروتئین‌های شناخته‌شده، از همین روش بهره می‌برند.

تحولات پژوهشی در این حوزه معمولاً طی سال‌ها شکل‌گرفته‌اند و ابزارهای جدید هوش مصنوعی برای انجام کارهای تخصصی یا در قالب‌هایی نامرئی در محصولات و خدمات تجاری موجود مانند موتورهای جست‌وجوی اینترنتی، به کار رفته‌اند.

اما طی دو سال گذشته مدل‌های هوش مصنوعی مولد که آن‌ها نیز از شبکه‌های عصبی استفاده می‌کنند، به مرکز توجه صنعت فناوری تبدیل شده‌اند؛ صنعتی که می‌خواهد این مدل‌ها را با سرعت از آزمایشگاه‌های شرکت‌ها بیرون آورده و در اختیار عموم قرار دهد. نتیجه این روند آشفته، گاه چشمگیر و اغلب غیرقابل‌پیش‌بینی بوده است؛ زیرا افراد و سازمان‌ها به آزمایش این مدل‌ها می‌پردازند.

«تیمنیت گبرو» (Timnit Gebru)، بنیان‌گذار Distributed AI Research Institute در کالیفرنیا، می‌گوید: «واقعاً انتظار چنین انفجاری از مدل‌های مولد را نداشتم. هرگز چنین افزایش سریع و گسترده‌ای در تعداد محصولات ندیده‌ام.»

جرقه این انفجار را OpenAI زد؛ زمانی که در ۳۰ نوامبر ۲۰۲۲ نسخه اولیه ChatGPT را منتشر کرد و تنها طی پنج روز یک میلیون کاربر جذب کرد. مایکروسافت، سرمایه‌گذار چندمیلیارددلاری OpenAI، در فوریه ۲۰۲۳ با ارائه یک چت‌بات مجهز به همان فناوری ChatGPT در موتور جست‌وجوی Bing، این روند را ادامه داد که به‌عنوان حرکتی آشکار برای به چالش کشیدن سلطه طولانی‌مدت گوگل در بازار موتورهای جست‌وجو شناخته می‌شود.



IMAGINIMA/ISTOCK

این اقدام باعث شد گوگل نیز در مارس ۲۰۲۳ چت‌بات خود به نام Bard (که بعدها به Gemini تغییر کرد) را معرفی کند. گوگل همچنین ۳۰۰ میلیون دلار در Anthropic، استارتاپی که توسط کارکنان سابق OpenAI تأسیس شد، سرمایه‌گذاری کرده است. شرکتی که چت‌بات خود به نام Claude را از مارس ۲۰۲۳ در اختیار گروه محدودی از افراد و شرکای تجاری قرار داد و سپس به صورت عمومی عرضه کرد. شرکت‌های بزرگ فناوری چین، مانند Baidu و Alibaba نیز وارد رقابت شده‌اند و چت‌بات‌های هوش مصنوعی را در موتورهای جست‌وجو و سایر خدمات خود ادغام کرده‌اند.

این مدل‌های مولد اکنون بر حوزه‌هایی مانند آموزش تأثیر گذاشته‌اند؛ به‌طوری که برخی مدارس ChatGPT را ممنوع کرده‌اند، چون می‌تواند مقاله‌های کامل و غیرقابل‌تشخیص نسبت به نوشته دانش‌آموز تولید کند. توسعه‌دهندگان نرم‌افزار نشان داده‌اند که ChatGPT می‌تواند خطاهای کدنویسی را پیدا و اصلاح کند و حتی برنامه‌ها را از صفر بنویسد. مشاوران املاک از ChatGPT برای تولید آگهی‌های فروش و پست‌های شبکه‌های اجتماعی استفاده کرده‌اند و شرکت‌های حقوقی چت‌بات‌های هوش مصنوعی را برای تهیه پیش‌نویس قراردادهای حقوقی به کار گرفته‌اند. حتی آزمایشگاه‌های پژوهشی دولت آمریکا در حال بررسی این هستند که فناوری OpenAI چگونه می‌تواند با سرعت مقالات منتشرشده را غربال کند تا به طراحی آزمایش‌های علمی جدید کمک کند.

طبق گزارش تحلیلگران Goldman Sachs، حدود ۳۰۰ میلیون شغل تمام‌وقت ممکن است حداقل تا حدی تحت‌تأثیر خودکارسازی ناشی از هوش مصنوعی مولد قرار گیرند؛ اما همان‌طور که آن‌ها نوشته‌اند، این موضوع به این بستگی دارد که «هوش مصنوعی مولد تا چه حد به توانایی‌های وعده‌داده‌شده خود عمل کند.» گزاره‌ای آشنا که پیش‌تر نیز در چرخه‌های رونق و رکود هوش مصنوعی شنیده شده است.

آنچه روشن است، این است که خطرات واقعی هوش مصنوعی مولد نیز با سرعتی سرسام‌آور رخ می‌دهند. ChatGPT و دیگر چت‌بات‌ها اغلب اشتباهات را واقعی ارائه می‌دهند و رویدادها یا مقالات کاملاً ساختگی را ذکر می‌کنند. استفاده از ChatGPT همچنین منجر به رسوایی‌های نقض حریم خصوصی شده است؛ از جمله افشای داده‌های محرمانه شرکت‌ها و حتی دسترسی کاربران ChatGPT به تاریخچه چت و اطلاعات پرداخت دیگران.

هنرمندان و عکاسان نیز نگرانی‌های بیشتری مطرح کرده‌اند؛ از جمله اینکه آثار تولیدشده توسط هوش مصنوعی ممکن است معیشت آنان را تهدید کند، درحالی‌که برخی شرکت‌ها مدل‌های مولد را با استفاده از آثار همان هنرمندان آموزش می‌دهند؛ بدون اینکه قوانین حق نشر را رعایت کنند. تصاویر تولیدشده توسط هوش مصنوعی می‌تواند به انتشار گسترده اطلاعات نادرست منجر شود؛ همان‌طور که تصاویر جعلی از بازداشت دونالد ترامپ و تصویری از پاپ فرانسیس با کت سفید براق، در اینترنت وایرال شد. شمار زیادی از مردم فریب خوردند و باور کردند واقعی است.

گبرو بسیاری از این خطرات احتمالی را پیش‌بینی کرده بود؛ زمانی که او و همکارانش در سال ۲۰۲۰ مقاله‌ای بنیادین درباره خطرات مدل‌های زبانی بزرگ نوشتند و او یکی از رهبران تیم اخلاق هوش مصنوعی گوگل بود. گبرو می‌گوید که پس از آنکه مدیران گوگل از او خواستند مقاله را پس بگیرد، مجبور به ترک شرکت شد؛ هرچند گوگل استعفای او را «خروج داوطلبانه» توصیف کرده است. گبرو می‌گوید: «این وضعیت شبیه یک چرخه هیجان‌زده دیگر است، اما تفاوت این است که اکنون محصولات واقعی وجود دارند که آسیب‌زا هستند.»

شبکه عصبی چیست؟

در قلب تحقیقات فعلی هوش مصنوعی، برنامه‌های کامپیوتری الهام‌گرفته از مغز انسان به نام شبکه‌های عصبی قرار دارند که هر کدام از تعداد زیادی نورون دیجیتال متصل تشکیل شده‌اند.

ما از طریق حواس خود ورودی دریافت می‌کنیم، اما مدل‌های هوش مصنوعی ورودی را به‌عنوان داده، مانند تصاویر، صدا یا متن، دریافت می‌کنند. در پاسخ به یک ورودی، مغز ما با توالی‌های آشفته و پیچیده‌ای از شلیک سیناپس‌ها بین نورون‌ها فوران می‌کند، درحالی‌که شبکه‌های عصبی مصنوعی داده‌ها را به‌صورت ساختاریافته‌تری پردازش می‌کنند.

یک شبکه عصبی را می‌توان با عبور حجم عظیمی از داده‌ها از طریق آن و سپس بررسی خروجی آن در مقایسه با یک نتیجه موردانتظار شناخته‌شده، برای انجام یک کار آموزش داد. به‌عنوان‌مثال، برای ساخت یک هوش مصنوعی که می‌تواند دست خط را رونویسی کند، می‌توانید بارهاوبارها نمونه‌هایی از نوشته را به یک شبکه عصبی نشان دهید و خروجی آن را با متن واقعی مقایسه کنید. سپس مقادیر اتصالات در شبکه به‌روزرسانی می‌شوند تا خروجی به نتیجه مطلوب نزدیک‌تر شود. این کار را برای میلیون‌ها یا میلیاردها قطعه داده آموزشی انجام دهید و در نهایت مدلی هوش مصنوعی‌ای خواهید داشت که می‌تواند دستخطی که قبلاً هرگز ندیده است را رونویسی کند.

همین رویکرد، ویژگی تشخیص چهره در تلفن‌های هوشمند و مدل‌های هوش مصنوعی که تصاویر واقع‌گرایانه از پیام‌های متنی ایجاد می‌کنند و همچنین مدل‌های زبانی بزرگ را تقویت می‌کند. تقریباً تمام تحقیقات فعلی هوش مصنوعی مبتنی بر شبکه‌های عصبی یا زیرمجموعه‌های تخصصی آن‌ها است که از ترفندها و روش‌های مختلفی برای دستیابی به عملکرد بهتر استفاده می‌کنند.

«هیجان و هیاهوی انتقادی، هر دو می‌تواند از وظیفه مدیریت خطرات واقعی ناشی از هوش مصنوعی مولد منحرف شوند.»

فناوری هوش مصنوعی مولد بر پایه یک دهه پژوهش بنا شده است؛ پژوهش‌هایی که توانایی‌های هوش مصنوعی را در تشخیص تصویر، دسته‌بندی مقالات بر اساس موضوع و تبدیل گفتار به متن به طور چشمگیری بهبود داده‌اند. «آروین نارایانان» (Arvind Narayanan) از دانشگاه پرینستون می‌گوید: «با معکوس کردن این روند، اکنون می‌توان تصاویر مصنوعی بر اساس یک توضیح، مقاله‌هایی درباره یک موضوع مشخص یا نسخه صوتی متن‌های نوشتاری تولید کرد.» او می‌گوید: «هوش مصنوعی مولد واقعاً امکان انجام کارهای جدید زیادی را فراهم کرده است.» بااین‌حال، ارزیابی این فناوری دشوار است.

مدل‌های زبانی بزرگ شاهکارهای مهندسی هستند که از حجم عظیمی از توان پردازشی در مرکز داده‌های شرکت‌هایی مانند مایکروسافت و گوگل استفاده می‌کنند. آن‌ها به مقدار عظیمی از داده‌های آموزشی نیاز دارند؛ داده‌هایی که شرکت‌ها اغلب آن‌ها را از مخازن اطلاعات عمومی در اینترنت، مانند ویکی‌پدیا، جمع‌آوری می‌کنند. این فناوری همچنین به تعداد زیادی نیروی انسانی متکی است تا در طول فرایند آموزش، بازخورد ارائه دهند و هوش مصنوعی را در مسیر درست هدایت کنند.

اما مدل‌های هوش مصنوعی قدرتمندی که توسط شرکت‌های بزرگ فناوری منتشر می‌شوند، معمولاً سیستم‌های بسته‌ای هستند که دسترسی عمومی یا دسترسی توسعه‌دهندگان بیرونی را محدود می‌کنند. سیستم‌های بسته می‌توانند در کنترل ریسک‌ها و آسیب‌های احتمالی ناشی از اینکه هر کسی بتواند این هوش‌های مصنوعی را دانلود و استفاده کند مؤثر باشند، اما درعین‌حال قدرت را در دست سازمان‌هایی متمرکز می‌کنند که آن‌ها را توسعه داده‌اند؛ بدون اینکه به مردمی که زندگی‌شان تحت‌تأثیر این مدل‌های هوش مصنوعی قرار می‌گیرد، اجازه مشارکت بدهند.

«ایرین سولایمان» (Irene Solaiman)، مدیر سیاست‌گذاری Hugging Face، شرکتی که ابزارهایی برای اشتراک‌گذاری کد و داده‌های هوش مصنوعی توسعه می‌دهد می‌گوید: «نگران‌کننده‌ترین مسئله در روند بسته‌سازی این است که تعداد بسیار کمی از مدل‌ها خارج از چند سازمان توسعه‌دهنده در دسترس خواهند بود.»

این روند در تغییر رویکرد OpenAI نیز آشکار است؛ سازمانی که باوجود آغاز به کار به‌عنوان یک نهاد غیرانتفاعی برای توسعه متن‌باز هوش مصنوعی، اکنون به سمت رویکردی مالکیتی و بسته حرکت کرده است. زمانی که OpenAI فناوری زیربنایی ChatGPT را به GPT-4 ارتقا داد، دلیل عدم افشای نحوه کار این مدل را «چشم‌انداز رقابتی و پیامدهای ایمنی مدل‌های بزرگ‌مقیاس مانند GPT-4 اعلام کرد.»

این رویکرد، ارزیابی توانایی‌ها و محدودیت‌های هوش‌های مصنوعی مولد را برای افراد بیرون از سازمان دشوار می‌کند و ممکن است به دامن‌زدن به هیجان و بزرگ‌نمایی منجر شود. «لی وینسل» (Lee Vinsel)، تاریخ‌دان فناوری از Virginia Techمی‌گوید: «حباب‌های فناوری انرژی احساسی زیادی ایجاد می‌کنند، هم هیجان و هم ترس؛ اما محیط‌های اطلاعاتی بدی هستند.»

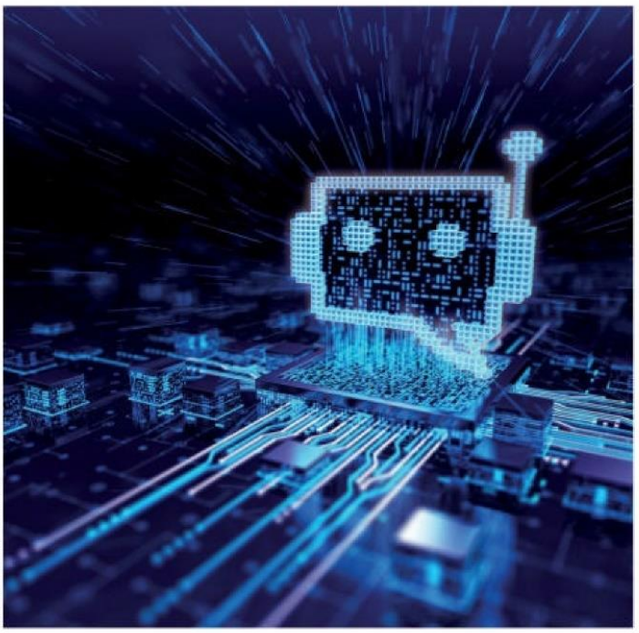
بسیاری از حباب‌های فناوری شامل هر دو عامل هیجان (Hype) و چیزی هستند که وینسل آن را «هیاهوی انتقادی» (Criti-Hype) می‌نامد؛ یعنی انتقاداتی که با قبول ادعاهای اغراق‌آمیز شرکت‌ها و تبدیل آن به سناریوهای پرخطر، خود به بزرگ‌نمایی فناوری کمک می‌کند. این موضوع را می‌توان در واکنش‌ها به ChatGPT به‌خوبی دید.

بیانیه مأموریت OpenAI عنوان می‌کند که این شرکت متعهد است مزایای «هوش جامع مصنوعی»، (Artificial General Intelligence) به‌عنوان مدل‌هایی که می‌توانند در هر وظیفه فکری از انسان بهتر عمل کنند را گسترش. ولی ChatGPT از آن هدف فاصله بسیاری دارد. بااین‌حال، در مارس ۲۰۲۳، پژوهشگران هوش مصنوعی مانند «یوشوا بنجیو» (Yoshua Bengio) و چهره‌های فناوری مانند «ایلان ماسک» نامه‌ای سرگشاده را امضا کردند که در آن از آزمایشگاه‌های تحقیقاتی خواسته شده بود آزمایش‌های هوش مصنوعی غول‌پیکر را متوقف کنند، درحالی‌که هوش مصنوعی را «ذهن‌های غیرانسانی که ممکن است در نهایت از ما بیشتر شوند، ما را شکست دهند، ما را منسوخ کنند یا جایگزینمان شوند» خطاب کرده بودند.

متخصصانی که با New Scientist مصاحبه کرده‌اند هشدار می‌دهند که هم هیجان و هم هیاهوی انتقادی می‌توانند توجه را از کار ضروری مدیریت ریسک‌های واقعی هوش مصنوعی مولد منحرف کنند. به گفته «دارون عجم‌اوغلو» (Daron Acemoglu)، اقتصاددان MIT: «GPT-4 می‌تواند بسیاری از کارها را خودکار کند، در مقیاس وسیع اطلاعات نادرست تولید کند، سلطه چند شرکت بزرگ فناوری را تثبیت کند و حتی به دموکراسی‌ها آسیب بزند. این کارها را می‌تواند انجام دهد بدون اینکه حتی به هوش جامع مصنوعی نزدیک شده باشد.»

عجم‌اوغلو می‌گوید اکنون «نقطه عطفی» برای نهادهای قانون‌گذاری است تا اطمینان حاصل کنند این فناوری‌ها به کارمندان کمک می‌کنند، به شهروندان قدرت می‌دهند و همچنین «قدرت‌گیری اربابان تکنولوژی که این فناوری را کنترل می‌کنند»، محدود شود.

اوایل سال ۲۰۲۴، قانون‌گذاران اتحادیه اروپا «قانون هوش مصنوعی» را تصویب کردند که نخستین استانداردهای گسترده جهانی برای تنظیم‌گری این فناوری را ایجاد کرد. این قانون مدل‌های هوش مصنوعی پرریسک را تنظیم‌گری می‌کند و بحث‌هایی نیز درباره گنجاندن ChatGPT و مدل‌های مولد مشابه با کاربردهای عمومی در دسته «پرخطر» در جریان است.



BLACKJACK3D/ISTOCK

چرا ChatGPT این قدر خوب است؟

چت‌بات OpenAI توسط نوعی شبکه عصبی به نام «ترنسفورمر» (Transformer,) پشتیبانی می‌شود که در سال ۲۰۱۷ توسط گوگل اختراع شد. ترنسفورمر با تلاش برای توجه به مهم‌ترین بخش‌های ورودی داده، درحالی‌که بقیه را به‌صورت روشمند ارزیابی می‌کند، یک شبکه عصبی استاندارد را بهبود می‌بخشد. به طور خلاصه، این کار زمینه ایجاد می‌کند.

«ساموئل بومن» (Samuel Bowman) از دانشگاه نیویورک می‌گوید: مدل‌های هوش مصنوعی مانند ChatGPT چشمگیر هستند، اما نتایج آنها تا حد زیادی در مقیاس کوچک و مجموعه‌های وسیعی از داده‌های آموزشی است. «هیچ نوآوری بزرگ و عمیقی وجود ندارد، هیچ ایده بزرگی پشت ترانسفورماتورها وجود ندارد؛ فقط یک روش خاص برای کنار هم قرار دادن چیزهایی است که اتفاقاً به‌خوبی با سخت‌افزار و قابلیت اطمینان مطابقت دارد تا آزمایش‌ها واقعاً در مقیاس بزرگ انجام شود.»

رویکردهای دیگری نیز وجود دارند که می‌توانند پیشرفت‌های جزئی ایجاد کنند؛ اما در حال حاضر هیچ انگیزه قوی‌ای برای تیم‌های تحقیقاتی وجود ندارد که به جایگاه دیگری نگاه کنند، زیرا ترنسفورمرها به طور قابل‌اعتمادی کار می‌کنند و نتایج خوبی را به همراه دارند.

مؤسسه AI Now در نیویورک و سایر متخصصان اخلاق هوش مصنوعی، از جمله گبرو، پیشنهاد می‌کنند بار مسئولیت بر دوش شرکت‌های بزرگ فناوری قرار گیرد و آن‌ها مجبور شوند ثابت کنند که مدل هوش مصنوعی‌شان آسیب‌رسان نیست؛ به‌جای اینکه از نهادهای نظارتی انتظار برود پس از وقوع آسیب‌ها به آن رسیدگی کنند.

«سارا مایرز وست» (Sarah Myers West)، مدیر AI Now، می‌گوید: «برخی از بازیگران صنعت از اولین کسانی بودند که گفتند ما به قانون‌گذاری نیاز داریم. اما ای‌کاش سؤال این بود که چطور مطمئن هستید کاری که انجام می‌دهید از ابتدا قانونی است؟»

آنچه اتفاق می‌افتد تا حد زیادی به این بستگی دارد که این فناوری‌ها چگونه استفاده و تنظیم شوند. عجم‌اوغلو می‌گوید: «مهم‌ترین درس تاریخ این است که ما به‌عنوان جامعه، انتخاب‌های بسیار بیشتری درباره نحوه توسعه و عرضه فناوری‌ها نسبت به آنچه آینده‌پژوهان فناوری به ما می‌گویند، داریم.»

«سم آلتمن»، مدیرعامل OpenAI سناریوهای بسیار افراطی‌تری را درباره آینده‌ای با مدل‌های هوش مصنوعی قدرتمند که در مجموع از انسان‌ها بهتر عمل می‌کنند مطرح کرده است. او از «بهترین حالت» گفته که در آن مدل‌های هوش مصنوعی می‌توانند «همه جنبه‌های واقعیت را بهبود دهند و به ما اجازه دهند زندگی بهتری داشته باشیم» و همچنین هشدار داده که «حالت بد» می‌تواند «به معنای پایان همه چیز باشد». اما او تأکید کرده که توسعه کنونی هوش مصنوعی هنوز فاصله بسیار زیادی از هوش جامع مصنوعی دارد.

گبرو و همکارانش بیانیه‌ای منتشر کردند که هشدار می‌دهد: «خطرناک است که خودمان را با رؤیای آرمان‌شهری یا آخرالزمانی مبتنی بر هوش مصنوعی مشغول کنیم؛ رؤیایی که آینده‌ای شکوفا یا احتمالاً فاجعه‌آمیز را وعده می‌دهد. رقابت کنونی برای انجام آزمایش‌های هوش مصنوعی هرچه بزرگ‌تر، یک مسیر اجتناب‌ناپذیر نیست که تنها انتخاب ما سرعت حرکت باشد؛ بلکه مجموعه‌ای از تصمیم‌هاست که با انگیزه مالی هدایت می‌شود. اقدامات و انتخاب‌های شرکت‌ها باید توسط مقرراتی شکل گیرد که از حقوق و منافع مردم محافظت کند.»

به گفته «ساشا لوچونی» (Sasha Luccioni)، پژوهشگر هوش مصنوعی در Hugging Face: «اگر حباب انتظارات تجاری پیرامون هوش مصنوعی مولد به سطحی ناپایدار برسد و در نهایت بترکد، این موضوع می‌تواند توسعه آینده هوش مصنوعی را نیز تضعیف کند» بااین‌حال، رونق فعلی لزوماً به زمستان دیگری برای هوش مصنوعی منجر نمی‌شود. یکی از دلایل این است که برخلاف چرخه‌های گذشته، بسیاری از سازمان‌ها همچنان مسیرهای دیگری از پژوهش هوش مصنوعی را دنبال می‌کنند و همه چیز را روی مدل‌های مولد سرمایه‌گذاری نکرده‌اند.

سازمان‌هایی مانند Hugging Face خواستار فرهنگ باز در پژوهش و توسعه هوش مصنوعی هستند؛ فرهنگی که بتواند از تشدید غیرقابل‌کنترل هیجان و آثار واقعی اجتماعی جلوگیری کند. لوچونی با برگزارکنندگان NeurIPS، یکی از بزرگ‌ترین رویدادهای پژوهشی هوش مصنوعی همکاری می‌کند تا یک کد اخلاق کنفرانسی را ایجاد کند؛ کدی که پژوهشگران را ملزم می‌کند داده‌های آموزشی خود را افشا کنند، دسترسی به مدل‌هایشان را فراهم کنند و کار خود را شفاف نشان دهند، به‌جای اینکه آن را به‌عنوان فناوری مالکیتی پنهان کنند.

«نیما باسکارینو» (Nima Boscarino)، مهندس اخلاق در Hugging Face، می‌گوید پژوهشگران هوش مصنوعی باید واضحاً توضیح دهند که مدل‌ها چه کارهایی می‌توانند و چه کارهایی نمی‌توانند انجام دهند، بین توسعه محصول و پژوهش علمی تمایز قائل شوند و با جوامعی که بیشترین تأثیر را می‌پذیرند همکاری نزدیک داشته باشند تا ویژگی‌ها و ملاحظات ایمنی مرتبط با آن‌ها را بیاموزند. باسکارینو همچنین بر نیاز به ارزیابی عملکرد هوش مصنوعی در قبال افراد با هویت‌های گوناگون تأکید می‌کند.

پژوهش بر روی هوش مصنوعی مولد، اگر به این شیوه انجام شود، می‌تواند شکلی پایدارتر و پایا از توسعه فناوری‌های مفید را برای آینده تضمین کند. باسکارینو می‌افزاید: «این‌ها دوران هیجان‌انگیزی در حوزه اخلاق هوش مصنوعی هستند و امیدوارم بخش گسترده‌تر یادگیری ماشین، مسیر کاملاً مخالف آنچه OpenAI انجام داده را در پیش بگیرد.»

مصاحبه

ChatGPT واقعاً چقدر باهوش است؟

هوش مصنوعی مولد توجه عموم را جلب کرده، اما واقعاً چقدر باهوش است و ظهور آن برای ما چه معنایی دارد؟ کمتر کسی بهتر از دانشمند سیستم‌های پیچیده (Complexity Scientist) «ملانی میچل» می‌تواند به این پرسش‌ها پاسخ دهد. او در سال ۲۰۲۳ با New Scientist درباره موج توجهی که به هوش مصنوعی شده، چالش‌های ارزیابی اینکه GPT-4 واقعاً چقدر هوشمند است و اینکه چرا هوش مصنوعی مدام ما را وادار به بازاندیشی درباره هوش می‌کند، صحبت کرد.

در حال حاضر موجی از علاقه به هوش مصنوعی وجود دارد. چرا این اتفاق الان رخ می‌دهد؟

اول اینکه این سیستم‌ها اکنون در دسترس عموم قرار دارند. هر کسی می‌تواند به‌راحتی با ChatGPT بازی کند، بنابراین مردم در حال کشف این سیستم‌ها و توانایی‌های آن‌ها هستند. به‌طورکلی، ما شاهد دوره‌ای از پیشرفت شگرف در توانایی‌های زبانی هستیم. در حدود پنج سال گذشته، ظهور مدل‌های زبانی بزرگ را دیده‌ایم که بر روی مقادیر عظیمی از زبان تولیدشده توسط انسان آموزش داده شده‌اند و قادر به تولید متن‌های روان و مشابه انسان هستند. روان‌بودن آن‌ها نیز ظاهری از هوش شبه‌انسانی ایجاد می‌کند. این امر تخیلات مردم را تحریک کرده و این حس به وجود آمد که آنچه در فیلم‌ها و داستان‌های علمی-تخیلی درباره هوش مصنوعی دیده‌ایم بالاخره اینجاست. فکر می‌کنم مردم هم حس شگفتی و هم تا حدی ترس درباره کارهایی که این مدل‌های هوشمند ممکن است انجام دهند را تجربه می‌کنند.

به «هوش انسان‌مانند» (Human-Like Intelligence) اشاره کردید. هوش مدل‌های مولد امروزی، مثل آن‌هایی که متن تولید می‌کنند، دقیقاً چقدر است و چگونه آن را ارزیابی می‌کنیم؟

این موضوع محل بحث فراوان است. دلیل اول این امر این است که همه اصطلاحاتی که به آن‌ها می‌اندیشیم؛ مانند هوش، فهم، خودآگاهی، تعریف‌های روشنی ندارند. دلیل دوم این است که این سامانه‌های هوش مصنوعی بسیار متفاوت از ذهن‌های انسانی کار می‌کنند. اخیراً دیدیم که GPT-4 توانست از سد آزمون وکالت (Bar Exam) عبور کند؛ آزمونی استاندارد که افراد باید آن را بگذرانند تا بتوانند در آمریکا وکالت کنند. اگر یک انسان در این آزمون که شامل سؤالات چندگزینه‌ای و نوشتن یک مقاله حقوقی است، عملکرد خوبی داشته باشد، ما فرض می‌کنیم او هوش عمومی بالایی دارد. اما چه کسی می‌تواند بگوید این‌گونه آزمون‌ها راه مناسبی برای ارزیابی هوش هوش مصنوعی هستند؟

درباره ملانی میچل (Melanie Mitchel)

ملانی میچل استاد مؤسسه Santa Fe در نیومکزیکو است که در آنجا به مطالعه انتزاع مفهومی و ساختن قیاس‌ها در سامانه‌های هوش مصنوعی می‌پردازد و نویسنده کتاب «هوش مصنوعی: راهنمایی برای انسان‌های اندیشمند» (Artificial Intelligence: A guide for thinking humans.) است.

«هوش انسانی مختص به جایگاه تکاملی ماست و ممکن است آن‌قدر که دوست داریم فکر کنیم عمومی نباشد.»

آیا فکر می‌کنید تلاش‌های ما برای درک توانایی‌های هوش مصنوعی ما را مجبور می‌کند تعریفمان از هوش و فهم را عمیق‌تر کنیم؟

چنین چیزی در تمام تاریخ هوش مصنوعی صادق بوده است. در دهه‌های ۱۹۷۰ و ۱۹۸۰ بسیاری می‌گفتند بازی شطرنج در سطح یک استاد بزرگ مستلزم هوش عمومی در سطح انسان است. سپس Deep Blue آمد، ابررایانه‌ای که «گَری کاسپارف» استاد بزرگ شطرنج را شکست داد و گفتند که این برتری ناشی از نیروی خام محاسباتی است که بهترین حرکت‌ها را جست‌وجو می‌کند و اکنون دوباره به همین نقطه رسیده‌ایم. می‌توان گفت ما مرتباً خط اهداف را جابه‌جا می‌کنیم. اما من می‌گویم در یک برداشت مثبت‌تر، هوش مصنوعی همچنان درک ما از مفهوم اینکه «هوش چیست؟» یا «فهمیدن یعنی چه؟» را به چالش می‌کشد.

مسئله این است که می‌دانیم چندین تجلی مختلف از هوش وجود دارد. هوش انسانی هست که بسیار متفاوت از مثلاً هوش هشت‌پا است و باز هم متفاوت از قدرت‌های یک هوش مصنوعی مولد است. برخی از ما از عبارت «هوش‌های متنوع» (Diverse Intelligences) را به‌صورت جمع استفاده کرده‌ایم تا تأکید کنیم هوش یک چیز یگانه نیست. چگونه این هوش‌های متفاوت را توصیف کنیم؟ آیا ویژگی‌های مشترکی دارند؟ یا کاملاً متفاوت‌اند؟ این‌ها سؤالاتی هستند که می‌خواهیم با آن‌ها دست‌وپنجه نرم کنیم.

آیا چیزی در توانایی‌های مدل‌های زبانی بزرگ شما را شگفت‌زده کرده است؟

به‌تازگی ما شاهد چیزی هستیم که مردم آن را «رفتارهای نوظهور» (Emergent Behaviours) می‌نامند؛ یعنی توانایی‌هایی که فراتر از پردازش زبان رفته و می‌توانند شبیه استدلال انسانی به نظر برسند. می‌توانید مسائل ریاضی یا دستورالعملی برای کدنویسی به یک مدل زبانی بزرگ بدهید، می‌توانید داستان‌هایی به آن‌ها بدهید و بخواهید درباره شخصیت‌ها استدلال کنند و آن‌ها قادر به انجام این کارها هستند.

اما اصلاً مشخص نیست چگونه این اتفاق می‌افتد. آن‌ها این برداشت را ایجاد می‌کنند که توانایی درک جهان را دارند، آن هم صرفاً فقط با آموزش بر روی مقادیر عظیمی از متن تولیدشده توسط انسان. سؤال این است که آیا آن‌ها کاری شبیه به استدلال انسانی انجام می‌دهند؟ یا صرفاً از ارتباطات آماری پیچیده استفاده می‌کنند که به نظر نمی‌رسد شبیه روشی باشد که ما استدلال می‌کنیم.

ایده‌های پیشرو درباره پشت پرده این رفتارهای نوظهور چیست؟

هنوز زود است که بگوییم، چون هر چند ماه یک‌بار نسخه جدیدی از این مدل‌ها می‌آید که کارهای جدیدی انجام می‌دهد. با GPT-3، دست‌کم می‌توانستیم به داده‌های آموزشی نگاه کنیم. اما با GPT-4 ما به این داده‌ها دسترسی نداریم. OpenAI می‌گوید این یک محصول تجاری است، بنابراین نمی‌خواهد به رقبا امتیاز بدهد. آن‌ها همچنین به پیامدهای ایمنی اشاره می‌کنند اما شفافیتی وجود ندارد، بنابراین انجام پژوهش ممکن نیست.

آیا فکر می‌کنید ما در مسیر رسیدن به هوش مصنوعی‌ای قرار داریم که چیزی شبیه به هوش جامع داشته باشد؟ یا فکر می‌کنید چنین چیزی مستلزم یک رویکرد کاملاً جدید خواهد بود؟

ابتدا باید بپرسیم: هوش جامع یا عمومی چیست؟ باز هم تعریف توافق‌شده‌ای نداریم، بنابراین گفتن اینکه چه چیزی لازم است تا به آن برسیم دشوار است؛ چون من دقیقاً مطمئن نیستم هدف چیست. بسیاری از روان‌شناسان سؤال کرده‌اند که آیا انسان‌ها واقعاً هوش عمومی دارند یا نه. هوش انسانی مختص به جایگاه تکاملی ماست و ممکن است آن‌قدر که دوست داریم فکر کنیم عمومی نباشد.

بااین‌حال، فکر می‌کنم صرفاً افزایش مقیاس این مدل‌ها احتمالاً ما را به نوع درک انسان‌مانند موردانتظار نمی‌رساند. ما فقط در پی درک زبانی نیستیم؛ ما در پی درک بصری، توانایی فهم و انجام کار درست در یک موقعیت معین هستیم.

برای رسیدن به آن نقطه، فکر می‌کنم به گونه‌های متفاوتی از معماری‌ها نیاز خواهیم داشت. مثلاً مدل‌های زبانی مانند GPT-4 حافظه بلندمدت ندارند، بنابراین هیچ خاطره‌ای از گفت‌وگوهای گذشته ندارند و به‌نوعی برای آنچه قبلاً گفته‌اند اهمیت قائل نیستند. اشاره شده که بخش بزرگی از هوش انسانی حول انگیزه‌های ما متمرکز است؛ اینکه هوش انسانی وسیله‌ای برای رسیدن به اهدافی است که تکامل برای ما تعیین کرده. اگر یک سامانه هیچ انگیزه یا هدف خودش را نداشته باشد، شاید نتواند به‌نوعی از هوش دست یابد که ما داریم.

نظر شما درباره این ایده که هوش مصنوعی می‌تواند یا خواهد توانست احساساتی یا خودآگاه شود چیست؟

همان‌طور که فیلسوفان از دیرباز گفته‌اند، چگونه حتی می‌دانم که شما خودآگاه هستید؟ من می‌دانم که خودآگاه هستم؛ چون به‌نوعی آن را احساس می‌کنم. اما شاید شما فقط یک زامبی باشید. فکر می‌کنم ترجیح می‌دهم وارد این بحث نشوم چون نمی‌فهمم منظور دقیقاً چیست و احساس می‌کنم گفت‌وگوها معمولاً به جایی نمی‌رسد.

در مارس ۲۰۲۳، تعداد زیادی از متخصصان برجسته هوش مصنوعی یک نامه سرگشاده را امضا کردند که خواستار توقف موقت پژوهش‌های عظیم در زمینه هوش مصنوعی شدند. آیا ممکن است که ما خیلی سریع پیش می‌رویم؟

بله ممکن است. فناوری اغلب سریع‌تر از سیاست و مقررات حرکت می‌کند. درباره هوش مصنوعی ریسک‌های زیادی در استقرار این سیستم‌ها در حوزه‌هایی مانند بهداشت، حقوق، روزنامه‌نگاری و غیره وجود دارد. اما من آن نامه را امضا نکردم؛ چون فکر می‌کردم آن نامه مجموعه‌ای از مسائل مختلف را با هم خلط کرده؛ برخی واقعی و برخی ترساندن بی‌اساس به‌وسیله داستان‌های علمی-تخیلی. آن نامه روایت آخرالزمانی‌ای را ترسیم می‌کرد که من آن را قبول ندارم.

من فکر می‌کنم باید رگولاتوری داشته باشیم. این سیستم‌ها می‌توانند به روش‌های روزمره و پیش‌پاافتاده مانند سوگیری و اطلاعات نادرست

KATE JOYCE FOR THE SANTA FE INSTITUTE

و ازاین‌دست مسائل خطرناک باشند؛ اما نمی‌دانم آیا متوقف کردن پژوهش درباره آن‌ها راه درستی است یا خیر. در عوض، باید بدانیم آن‌ها روی چه داده‌ای آموزش‌دیده‌اند. ما نمی‌توانیم بگذاریم شرکت‌هایی مانند OpenAI عملاً بگویند: «به ما اعتماد کنید، ما می‌دانیم چه می‌کنیم.»

چه پیام‌هایی مایلید منتقل کنید تا به مردم کمک کند درباره ریسک‌ها و مزایای هوش مصنوعی فکر کنند؟

اولاً این سیستم‌ها هنوز قابل‌اتکا و خودآگاه نیستند. آن‌ها تصمیمی برای انجام کاری که ممکن است به ما آسیب برساند نمی‌گیرند. توانایی واقعی آسیب در انسان‌ها است که از آن‌ها استفاده می‌کنند و بنابراین ما واقعاً به مقررات نیاز داریم.

ثانیاً صرف اینکه ما هنوز دقیقاً نمی‌فهمیم این ابزارها چگونه کار می‌کنند، به این معنا نیست که آن‌ها جادویی‌اند. فقط خیلی پیچیده‌اند. ما قادر خواهیم بود آن‌ها را بفهمیم؟ فقط باید علمش را انجام دهیم و برای انجام کار علمی آن نیاز داریم که این سیستم‌ها کاملاً در دست شرکت‌های سودمحور نباشند.

این مدل‌های زبانی فرصت بزرگی برای تعمیق درک ما از شناخت فراهم می‌کنند. ما می‌توانیم چیزهای زیادی از آن‌ها درباره خودمان، درباره چگونگی کارکرد هوش انسانی و درباره اینکه هوش به‌طورکلی چگونه ممکن است به شکل‌های گوناگون کار کند، بیاموزیم. اما درعین‌حال، باید از همه خطرات، ریسک‌ها و مسائل مرتبط با استقرار آن‌ها در دنیای واقعی آگاه باشیم.

خط زمانی هوش مصنوعی

ایده هوش مصنوعی در حدود ۷۰ سال است که وجود دارد، اما پیشرفت آن در چند دهه اول بسیار کند بود. تنها در اواخر دهه ۱۹۸۰ و با کشف یک پارادایم جدید و از پایین‌به‌بالا در یادگیری ماشین و مبتنی بر الگوریتم‌هایی که قادر بودند از حجم عظیمی از داده‌ها استنتاج آماری انجام دهند، هوش مصنوعی توانست مسیر خود را پیدا کند. تنها چند سال بعد، ظهور شبکه جهانی وب منبعی آماده و روبه‌رشد از همان داده‌ها را فراهم کرد.

۱۹۵۰

«آلن تورینگ» مقاله بنیادین «ماشین‌های محاسباتی و هوش» (Computing machinery and intelligence) را منتشر می‌کند. جمله آغازین این مقاله این است: «پیشنهاد می‌کنم به این پرسش بپردازم: آیا ماشین‌ها می‌توانند فکر کنند؟»

۱۹۵۶

اصطلاح «هوش مصنوعی» در کارگاهی در کالج دارتموث نیوهمپشایر ابداع می‌شود.

۱۹۵۹

دانشمندان کامپیوتر در دانشگاه کارنگی ملون پنسیلوانیا، یک حل‌کننده مسئله عمومی ایجاد می‌کنند؛ برنامه‌ای که می‌تواند معماهای منطقی را حل کند.

۱۹۷۳

اولین «زمستان هوش مصنوعی» آغاز می‌شود، پیشرفت متوقف شده و بودجه و علاقه کاهش پیدا کرده.

۱۹۷۵

سامانه‌ای به نام MYCIN عفونت‌های باکتریایی را تشخیص می‌دهد و با استفاده از استنتاج مبتنی بر مجموعه‌ای از پرسش‌های بله/خیر، آنتی‌بیوتیک توصیه می‌کند. این سیستم هرگز در عمل استفاده نشد.

۱۹۸۷

دومین زمستان هوش مصنوعی آغاز می‌شود.

۱۹۸۸

پژوهشگران IBM مقاله‌ای با عنوان «یک رویکرد آماری به ترجمه زبان» (A statistical approach to language translation) منتشر می‌کنند که راه جدیدی برای استفاده از داده‌ها به‌جای قواعد برای تعیین نتایج برنامه‌نویسی ارائه می‌دهد.

۱۹۸۹

برنامه AutoClass ناسا از یک رویکرد احتمالاتی برای کشف چندین رده ناشناخته از ستارگان در داده‌های تلسکوپی استفاده می‌کند.

۱۹۹۱

شبکه جهانی وب برای عموم راه‌اندازی می‌شود.

۱۹۹۶

نسخه نخست الگوریتم جست‌وجوی وب PageRank گوگل منتشر می‌شود.

۱۹۹۷

ابررایانه Deep Blue شرکت IBM قهرمان جهان گری کاسپارف را در شطرنج شکست می‌دهد.

۱۹۹۸

سامانه Remote Agent ناسا نخستین برنامه کاملاً خودمختار است که کنترل یک فضاپیما را در پرواز برعهده می‌گیرد.

۲۰۰۲

آمازون ویراستاران انسانی بخش پیشنهادهای محصول خود را با یک سیستم خودکار جایگزین می‌کند.

۲۰۰۵

پنج خودروی خودران، مسابقه آفرود ۲۰۰ کیلومتری DARPA در صحرای موجاو را به پایان می‌رسانند.

۲۰۰۶

فیس‌بوک به صورت عمومی عرضه می‌شود.

۲۰۰۶

گوگل سرویس Translate را به‌عنوان یک سرویس ترجمه ماشینی آماری راه‌اندازی می‌کند.

۲۰۰۹

پژوهشگران گوگل مقاله تأثیرگذاری با عنوان «کارآمدی غیرمنطقی داده‌ها» (The unreasonable effectiveness of data) منتشر می‌کنند و در آن اعلام می‌شود که «مدل‌های ساده و داده فراوان، مدل‌های پیچیده مبتنی بر داده اندک را شکست می‌دهد.»

۲۰۱۱

اپل Siri را که یک دستیار شخصی صوتی است که می‌تواند به پرسش‌ها پاسخ دهد، توصیه ارائه کند و دستوراتی مانند «تماس با خانه» را اجرا کند را در سیستم‌عامل خود ادغام می‌کند.

۲۰۱۱

ابررایانه IBM به نام Watson دو قهرمان انسانی را در مسابقه تلویزیونی Jeopardy! شکست می‌دهد.

۲۰۱۲

خودروهای بدون‌راننده گوگل به‌صورت خودمختار در میان ترافیک حرکت می‌کنند.

۲۰۱۶

AlphaGo، محصول گوگل، «لی سدول» یکی از برجسته‌ترین بازیکنان جهان در بازی Go را شکست می‌دهد.

۲۰۱۸

Waymo به‌عنوان زیرمجموعه گوگل، یک سرویس محدود تاکسی خودران تجاری را در فینیکس آریزونا راه‌اندازی می‌کند.

۲۰۲۲

هوش مصنوعی AlphaFold شرکت DeepMind ساختار تقریباً همه پروتئین‌های شناخته‌شده را تنها در ۱۸ ماه پیش‌بینی می‌کند.

۲۰۲۲

OpenAI چت‌بات ChatGPT را به‌عنوان نخستین مورد از مجموعه‌ای از مدل‌های هوش مصنوعی مولد تجاری را در دسترس عموم قرار می‌دهد.

۲۰۲۴

اتحادیه اروپا «قانون هوش مصنوعی» را تصویب می‌کند که هدف آن قابل‌اعتماد، شفاف و ایمن‌سازی سامانه‌های هوش مصنوعی در اتحادیه اروپا است.

فصل ۲

زیر پوست

هوش مصنوعی مولد با مجموعه‌ای از نوآوری‌ها نیرو گرفته است. برخی از این نوآوری‌ها، تجدید حیات ایده‌های قبلی هستند که محققان به دلیل محدودیت‌های سخت‌افزاری قادر به تحقق آن‌ها نبودند. برخی دیگر، تغییرات پارادایم در معماری‌های هوش مصنوعی هستند که جهش‌هایی را در قابلیت‌های آن‌ها ممکن ساخته است.

در اینجا، ما این جهش‌های تکنولوژیکی را تجزیه و تحلیل می‌کنیم و بررسی می‌کنیم که چگونه محققان امیدوارند قابلیت‌های هوش مصنوعی مولد را بیش‌ازپیش افزایش دهند.

چت‌بات‌های هوش مصنوعی چگونه کار می‌کنند؟

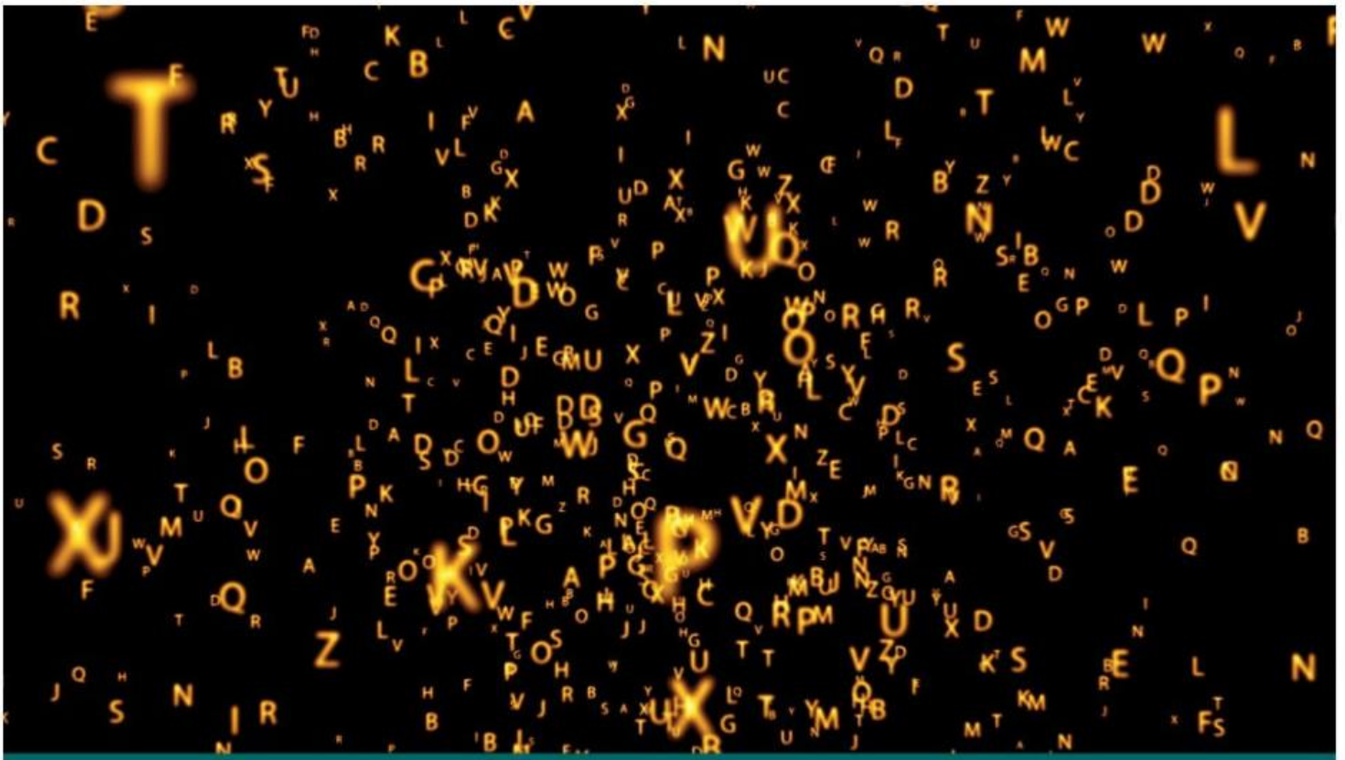
مدل‌های زبانی بزرگی که منشأ چت‌بات‌های جدید هستند طوری آموزش داده می‌شوند که پیش‌بینی کنند کدام واژه‌ها بیشترین احتمال را دارند که در کنار هم ظاهر شوند؛ اما «توانایی‌های نوظهور» نشان می‌دهند که شاید آن‌ها کارهایی فراتر از این را انجام می‌دهند.

در قلب رونق چت‌بات‌های هوش مصنوعی، مدل‌های زبانی بزرگ (LLM) قرار دارند؛ مدل‌های آماری یا نمایش‌های ریاضی داده‌ها که طراحی شده‌اند تا پیش‌بینی کنند کدام واژه‌ها احتمال دارد با هم ظاهر شوند.

مدل‌های زبانی بزرگ با تغذیه حجم عظیمی از متن به یک دسته از الگوریتم‌ها به نام شبکه‌های عصبی عمیق ساخته می‌شوند؛ الگوریتم‌هایی که به طور غیرمستقیم از مغز الهام گرفته شده‌اند. این مدل‌ها الگوهای زبانی پیچیده را با انجام یک بازی ساده یاد می‌گیرند؛ الگوریتم قسمت‌هایی از یک متن را می‌گیرد، چند واژه را به‌صورت تصادفی پنهان می‌کند و سپس تلاش می‌کند جاهای خالی را پر کند. به بیان ساده «میرلا لاپاتا» (Mirella Lapata) از دانشگاه ادینبرو، آن‌ها برای پیش‌بینی واژه بعدی آموزش داده می‌شوند و با بارهاوبارها تکرار این فرایند، می‌توانند مدل‌های پیچیده‌ای از نحوه کارکرد زبان بسازند.

پیشرفت‌های اخیر عمدتاً ناشی از نوع جدیدی از شبکه عصبی است که در سال ۲۰۱۷ اختراع شد که ترنسفورمر نام دارد؛ شبکه‌ای که می‌تواند داده‌ها را بسیار کارآمدتر از رویکردهای قبلی پردازش کند. این امکان فراهم شد تا مدل‌های بسیار بزرگ‌تری روی حجم عظیمی از متن استخراج‌شده از اینترنت آموزش داده شوند. به گفته لاپاتا سیستم‌های مبتنی بر ترنسفورمر همچنین در فهم بافت بسیار بهتر هستند. درحالی‌که نسخه‌های قدیمی تنها می‌توانستند چند واژه قبل و بعد از واژه حذف‌شده را در نظر بگیرند، ترنسفورمرها می‌توانند رشته‌های بسیار طولانی‌تری از متن را پردازش کنند؛ به این معنی که می‌توانند روابط زبانی بسیار پیچیده‌تر و ظریف‌تری را استخراج کنند.

ZIRCONICUSSO/ISTOCK



DONFIORE/ISTOCK

زبان مدل‌های زبانی بزرگ

مدل‌های زبانی بزرگ که پشت چت‌بات‌ها قرار دارند ممکن است حتی اگر در زبان‌های دیگر از آن‌ها سؤال شود، به «انگلیسی» فکر کنند. دلیلش این است که داده‌های آموزشی آن‌ها سوگیری دارند و مفاهیمی را رمزگذاری می‌کنند که در فرهنگ‌های انگلیسی‌زبان رایج‌تر هستند.

برای اینکه مشخص شود مدل‌های زبانی در واقع از کدام زبان برای پردازش درخواست‌ها استفاده می‌کنند، «کریس وندلر» (Chris Wendler)، «ونیامین وسلوفسکی» (Veniamin Veselovsky) و همکارانشان در مؤسسه فدرال فناوری سوئیس در لوزان سه نسخه از مدل Llama 2 که متعلق به شرکت مادر فیس‌بوک، یعنی Meta است را بررسی کردند.

وسلوفسکی عنوان می‌کند: «ما این مدل‌ها را باز کردیم و هر یک از لایه‌ها را بررسی کردیم.» مدل‌های زبانی از لایه‌های پردازشی تشکیل شده‌اند. این لایه‌ها ورودی متنی را به «توکن» که واژه‌ها یا بخش‌هایی از واژه‌ها هستند تبدیل می‌کنند و سپس تلاش می‌کنند هر توکن را در بافت زمینه قرار دهند تا پاسخی ارائه کنند.

وسلوفسکی ادامه می‌دهد: «هر یک از این لایه‌ها کاری روی ورودی، یعنی همان پرامپت اولیه که شما می‌نویسید، انجام می‌دهد. ما می‌خواستیم ببینیم آیا می‌توانیم تشخیص دهیم که لایه‌های داخلی واقعاً در حال پردازش به زبان انگلیسی هستند؟»

سه پرامپت به زبان چینی، فرانسوی، آلمانی و روسی به مدل‌ها داده شد. یکی از آن‌ها درخواست می‌کرد واژه داده‌شده را تکرار کنند؛ دیگری از آن‌ها می‌خواست از یک زبان غیرانگلیسی به زبان دیگری ترجمه کنند و سومی از آن‌ها می‌خواست جای خالی یک واژه را در جمله‌ای مانند نمونه زیر پر کند.

«A ___ is used to play sports like soccer and basketball»

پژوهشگران مسیر پردازش مدل را برای رسیدن به هر پاسخ دنبال کردند. آن‌ها دریافتند که مسیر پردازش تقریباً همیشه از چیزی عبور می‌کند که آن را زیرفضای انگلیسی می‌نامند. اگر از مدل خواسته شود که از چینی به روسی ترجمه کند، کاراکترهای درست روسی از زیرفضای انگلیسی عبور می‌کنند و سپس به روسی بازمی‌گردند و به عقیده وسلوفسکی این نشانه قوی‌ای است که زبان انگلیسی برای درک مفاهیم توسط مدل به کار می‌رود.

نتایج تعجب‌آور نیست، اما می‌تواند نگران‌کننده باشد. به گفته «آلیا بهاتیا» (Aliya Bhatia) از مرکز دموکراسی و فناوری واشنگتن: «داده‌های باکیفیت بیشتری به زبان انگلیسی و برخی زبان‌های رسمی سازمان ملل برای آموزش مدل‌ها وجود دارد تا بیشتر زبان‌های دیگر و در نتیجه؛ توسعه‌دهندگان هوش مصنوعی مدل‌های خود را عمدتاً بر داده‌های انگلیسی آموزش می‌دهند. اما استفاده از انگلیسی به‌عنوان واسطه برای آموزش مدل در تحلیل زبان، خطر تحمیل یک جهان‌بینی محدود بر نواحی زبانی و فرهنگی دیگر را در پی دارد.»

چیزی که مدل‌های آماری در ظاهر نامرتب را به چت‌بات‌هایی خوش‌صحبت تبدیل می‌کند، دخالت انسان‌ها است؛ انسان‌هایی که خروجی هوش مصنوعی را بر اساس معیارهایی مانند مفید و روان بودن رتبه‌بندی می‌کنند. این داده سپس برای آموزش یک «مدل اولویت» (Preference Model) جداگانه به کار می‌رود که خروجی مدل زبانی را فیلتر می‌کند. ترکیب این دو چیزی را می‌سازد که امروز در اختیار داریم؛ یک شریک مکالمه متنی و رایانه‌ای که اغلب از انسان قابل‌تشخیص نیست. به گفته «تال لینزن» (Tal Linzen) از دانشگاه نیویورک این واقعیت که این نتیجه از مقدمه‌ای به‌ظاهر ساده مثل پیش‌بینی واژه بعدی به‌دست‌آمده، بسیاری را شگفت‌زده کرده است.

اما مهم است به یاد داشته باشیم که روش کار این مدل‌های هوش مصنوعی تقریباً به‌طورقطع مشابه فرایندهای شناختی انسان نیست. به گفته لینزن: «آن‌ها به شکلی اساساً متفاوت از انسان‌ها یاد می‌گیرند و همین باعث می‌شود بسیار غیرمحتمل باشد که به همان شیوه انسان‌ها فکر کنند».

در اینجا اشتباهاتی که چت‌بات‌ها مرتکب می‌شوند، آموزنده‌اند. آن‌ها مستعد این هستند که با اعتمادبه‌نفس، اطلاعات نادرست را به‌عنوان واقعیت معرفی کنند؛ چیزی که اغلب «توهم» (Hallucination) نامیده می‌شود، زیرا خروجی آن‌ها کاملاً آماری است. به گفته لاپاتا: «آن‌ها بررسی و صحت‌سنجی انجام نمی‌دهند، فقط خروجی‌ای تولید می‌کنند که محتمل یا قابل‌قبول باشد، اما لزوماً درست نیست.»

این مسائل باعث شده برخی مفسران چت‌بات‌ها را «طوطی‌های تصادفی» (Stochastic Parrots) بنامند و خروجی آن‌ها را چیزی بیش از «یک عکس JPEG مات از کل وب» ندانند. منظور این طعنه‌ها این است که مدل‌های جدید آن‌قدرها هم که در نگاه اول به نظر می‌رسند چشمگیر نیستند؛ آنچه انجام می‌دهند فقط نوعی حفظ ناقص داده‌های آموزشی است که به شکلی هوشمندانه کنار هم چیده شده تا توهم فهم ایجاد کند.

«رافائل میلییر» (Raphaël Milliere) از دانشگاه کلمبیا می‌گوید: «بااین‌حال، نشانه‌هایی وجود دارد که حاکی از آن است که مدل‌های زبانی بزرگ شاید کاری بیش از صرفاً بازگویی داده‌های آموزشی انجام می‌دهند.» پژوهش‌های اخیر نشان می‌دهد که پس از آموزش طولانی‌مدت، مدل‌ها می‌توانند قوانین عمومی‌تری ایجاد کنند که به آن‌ها مهارت‌های جدیدی می‌دهد. میلییر عنوان می‌کند: «شما شاهد گذار از حفظ‌کردن به تشکیل مدارهایی درون شبکه عصبی خواهید بود که برخی الگوریتم‌ها یا قواعد خاص را برای حل وظایف اجرا می‌کنند.»

به عقیده میلییر این امر شاید بتواند توضیح دهد چرا با افزایش اندازه مدل‌های زبانی بزرگ، آن‌ها اغلب جهش‌های ناگهانی در عملکرد در برخی مسائل نشان می‌دهند. این پدیده «ظهور یا برآمدگی» (Emergence) نام‌گرفته و منجر به گمانه‌زنی درباره قابلیت‌های پیش‌بینی‌نشده‌ای شده که هوش مصنوعی ممکن است توسعه دهد.

میلییر اعتقاد دارد که مهم است که خیلی درگیر این موضوع نشویم. «این واژه بسیار وسوسه‌کننده است و چیزهایی مثل این که مدل‌ها ناگهان خودآگاه شوند را تداعی می‌کند؛ اما موضوع ما این نیست.»

بااین‌حال او معتقد است «یک حد میانه غنی» میان بدبینان و اغراق‌کنندگان وجود دارد. گرچه این چت‌بات‌ها خیلی از تقلید شناخت انسانی فاصله دارند، اما در برخی حوزه‌های محدود شاید آن‌قدرها هم از ما متفاوت نباشند. به عقیده میلییر کاوش در این شباهت‌ها نه‌تنها می‌تواند دانش هوش مصنوعی را پیش ببرد، بلکه می‌تواند فهم ما از قابلیت‌های شناختی خودمان را نیز عمیق‌تر کند.

مدل‌های هوش مصنوعی بسیار بزرگ

شبکه‌های عصبی عظیمی که با زبانی فوق‌العاده روان می‌نویسند، سبب شده‌اند برخی متخصصان پیشنهاد دهند که مقیاس‌دهی به فناوری فعلی می‌تواند به مهارت‌های زبانی در سطح بشر و نهایتاً به هوش ماشینی واقعی منجر شود.

هرچه مدل‌های هوش مصنوعی بزرگ‌تر می‌شوند، نه‌تنها خود را در انجام طیف وسیعی از کارها هم‌تراز یا برتر از انسان‌ها نشان می‌دهند، بلکه قابلیت مواجهه با چالش‌هایی را نیز به دست آورده‌اند که هرگز با آن‌ها روبه‌رو نبوده‌اند. به همین دلیل، برخی در این حوزه بر این باورند که تمایل اجتناب‌ناپذیر به بزرگ‌تر شدن مدل‌ها، منجر به مدل‌های هوش مصنوعی با توانایی‌های قابل‌مقایسه با انسان‌ها خواهد شد. «ساموئل بومن» (Samuel Bowman) در دانشگاه نیویورک یکی از این افراد است: «اگر روش‌های فعلی را به طور قابل‌توجهی مقیاس دهیم مخصوصاً پس از دهه‌ها پیشرفت در قدرت محاسباتی، به نظر می‌رسد دستیابی به رفتار زبانی در سطح انسان ساده باشد.»

اگر این پیش‌بینی رخ دهد؛ تأثیر عظیمی خواهد داشت. در گذشته، اندک متخصصانی بر این باور بودند که هوش ماشینی تنها از مسیر مهندسی محض قابل‌دستیابی است. البته بسیاری هنوز شک دارند که چنین چیزی حقیقت یابد؛ زمان نشان خواهد داد. در این میان، بومن و دیگران تلاش می‌کنند بفهمند وقتی این مدل‌های هوشمند مقیاس‌پذیر کارهایی شبیه انسان انجام می‌دهند واقعاً چه اتفاقی می‌افتد.

بومن یکی از برجسته‌ترین متخصصان جهان در ارزیابی مدل‌های زبانی بزرگ است. وقتی کار دکترای خود را در سال ۲۰۱۱ آغاز کرد، شبکه عصبی مصنوعی تازه در آستانه تسخیر این حوزه بود. این شبکه‌ها که الهام‌گرفته از شبکه‌های عصبی واقعی در مغز هستند، از واحدهای پردازشی مرتبط یا نورون‌های مصنوعی تشکیل شده‌اند و امکان خلق برنامه‌هایی را فراهم می‌کنند که قابلیت یادگیری دارند. به گفته «ایلیا سات‌اسکور» (Ilya Sutskever) دانشمند ارشد OpenAI: «برای مدت‌ها مشخص نبود که رایانه‌ها بتوانند چنین کار پیچیده‌ای را انجام دهند. وقتی دانشجوی کارشناسی رشته علوم کامپیوتر بودم... به نظر می‌رسید چنین کاری کاملاً غیرممکن باشد؛ ولی حالا نسبتاً به آن عادت کرده‌ایم.»

بر خلاف نرم‌افزارهای معمولی، محققان به شبکه‌های عصبی دستور نمی‌دهند که دقیقاً چه کنند؛ بلکه آن‌ها طوری طراحی می‌شوند که بر روی یک کار آموزش ببینند تا آن را به‌خوبی اجرا کنند. فرض کنید مجموعه‌ای بزرگ از تصاویر حیوانات دارید، هر تصویر با برچسبی مانند «سگ» یا «گربه» حاوی نشانه‌ای انسانی باشد؛ یک شبکه عصبی می‌تواند آموزش ببیند که برای تصویری که قبلاً ندیده است برچسب درست را پیش‌بینی کند. هر بار که برچسب اشتباه می‌دهد، به شیوه‌ای ساختاریافته به آن اطلاع داده می‌شود؛ بنابراین با مثال‌های کافی، شبکه در تشخیص حیوانات بهتر می‌شود.

اما این شبکه‌های عصبی که «مدل» هم نامیده می‌شوند، فقط به تشخیص گربه و سگ محدود نمی‌مانند. در سال ۱۹۹۰ «جفری المن» (Jeffrey Elman) استاد وقت دانشگاه کالیفرنیا روشی برای آموزش یک شبکه عصبی برای پردازش زبان کشف کرد. او دریافت می‌تواند یک واژه را از جمله حذف کند و شبکه را آموزش دهد تا واژه حذف‌شده را پیش‌بینی کند. مدل المن نمی‌توانست کاری بیش از تشخیص تفاوت بین اسم و فعل انجام دهد. چیزی که آن را جذاب می‌کرد این بود که نیازی به حاشیه‌نویسی‌های طاقت‌فرسای انسانی نداشت و می‌توانست با حذف کلمات تصادفی، داده‌های آموزشی ایجاد کند.

«توانایی‌های محاسباتی به نحوی در نسخه کامل ChatGPT-3 ظاهر شدند»

انجام کارهای بیشتر با منابع کمتر

بزرگ‌تر کردن اندازه فناوری فعلی می‌تواند مدل‌های زبانی پیشرفته‌تری ایجاد کند، اما قدرت محاسباتی تنها عامل محرک در هوش مصنوعی مولد نیست. دلیلش این است که مدل‌های هوشمند پشت چت‌بات‌ها، سریع‌تر از «قانون مور» (Moore's Law) در حال پیشرفت هستند. قانون مور قانونی است که سرعت افزایش عملکرد سخت‌افزار رایانه را اندازه می‌گیرد. چنین چیزی نشان می‌دهد که مدل‌های زبانی بزرگ پشت سیستم‌های هوش مصنوعی در حال باهوش‌تر شدن هستند و کارهای بیشتری با منابع کمتر انجام می‌دهند.

این کشف در مارس ۲۰۲۴ توسط «تامای بسیراوغلو» (Tamay Besiroglu) از MIT معرفی شد. او گفت دو راه برای بهبود عملکرد مدل‌های هوش مصنوعی وجود دارد. اول، مقیاس‌پذیری اندازه یک LLM که به نوبه خود نیازمند افزایش متناسب قدرت محاسباتی است؛ اما به‌خاطر انقلاب هوش مصنوعی مولد، اکنون کمبود جهانی در تراشه‌های GPU که برای تأمین قدرت LLMها به کار می‌روند ایجاد شده و گلوگاهی در توسعه هوش مصنوعی پدید آورده است.

گزینه دیگر به گفته بسیراوغلو، بهبود الگوریتم‌های پایه است تا از همان سخت‌افزار موجود بهتر استفاده شود. به نظر می‌رسد این رویکرد موردعلاقه نسل کنونی توسعه‌دهندگان LLM قرار گرفته و با موفقیت بزرگی همراه بوده است. بسیرواغلو و همکارانش عملکرد ۲۳۱ مدل LLM که بین ۲۰۱۲ تا ۲۰۲۳ توسعه یافته بودند را تحلیل کردند و دریافتند که به طور متوسط، قدرت محاسباتی لازم برای اینکه نسخه‌های بعدی یک LLM به سطح خاصی برسند، در طی هر هشت ماه نصف می‌شود. این سرعت بسیار فراتر از قانون مور است که در سال ۱۹۶۵ پیشنهاد کرد تعداد ترانزیستورها در یک تراشه که معیاری از عملکرد آن است، هر ۱۸ تا ۲۴ ماه دو برابر می‌شود.

درحالی‌که بسیراوغلو معتقد است بخشی از این افزایش در عملکرد LLMها به بهبود کدنویسی نرم‌افزار مربوط است، پژوهشگران دقیقاً نتوانستند مشخص کنند این کارایی چگونه به‌دست‌آمده است؛ تا حدی به‌این‌علت که الگوریتم‌های هوش مصنوعی اغلب یک «جعبه سیاه» غیرقابل‌نفوذ هستند. بسیراوغلو همچنین خاطرنشان می‌کند که بهبود سخت‌افزار هنوز نقش مهمی در افزایش عملکرد دارد. اما به عقیده «آنیما آناندکومار» (Anima Anandkumar) از CalTech خلاقیت انسانی نیز کمک بزرگی کرده است. درحالی‌که در دهه گذشته پیشرفت هوش مصنوعی عمدتاً با سخت‌افزار قوی‌تر یا مجموعه‌داده‌های بزرگ‌تر پیش می‌رفت، این روند در حال تغییر است. آناندکومار می‌گوید: «ما داریم محدودیت‌هایی در مقیاس، هم با داده و هم با [قدرت محاسباتی] می‌بینیم. آینده متعلق به بهینه‌سازی الگوریتمی است.»

GRANDFAILURE/ISTOCK

به‌تدریج، پژوهشگران درک کردند که بازآموزی مدل‌ها روی مسائل خاص‌تر کار ساده‌ای است. این مسائل شامل ترجمه زبان، پاسخ به پرسش‌ها، تحلیل احساسات (یعنی سنجیدن این که مثلاً نقد یک فیلم مثبت است یا منفی) می‌شد.

تا زمانی که بومن دکترای خود را در سال ۲۰۱۶ پایان داد، مدل‌های زبانی روی بسیاری از کارهای معمول تسلط یافته بودند. هیچ‌کس ادعا نمی‌کرد این مدل‌ها شباهتی به هوش واقعی دارند. تحلیل احساسات ممکن بود صرفاً با شناسایی کلماتی مثل «عالی» یا «خیلی دوست داشتم» انجام شود. اما مدل‌های زبانی سریع‌تر از آنچه بومن بتواند تصور کند، در انجام کارهای دشوارتر نیز پیشرفت می‌کردند.

راز کار در آموزش مدل روی داده‌های بیشتر بود و برای پردازش حجم عظیمی از متن از اینترنت و منابع دیگر، مدل‌ها باید بزرگ‌تر می‌شدند. ریشه هوش مصنوعی نیز با ساخت شبکه‌های عصبی به شیوه‌های نوین درحال‌توسعه بود و طراحی‌های تازه‌ای از نورون‌ها با سیم‌کشی‌های مختلف ایجاد می‌شد. اختراع ترنسفورمر در ۲۰۱۷ که پیش‌تر به آن اشاره شد، همه چیز را تغییر داد. در جست‌وجوی عملکرد بهتر، پژوهشگران مدل‌های مبتنی بر ترنسفورمر را در عرض چند سال از صدها میلیون پارامتر که هر پارامتر معادل یک اتصال بین نورون‌ها در نظر گرفته می‌شد به صدها میلیارد پارامتر ارتقا دادند.

در چند سال گذشته، پیشرفت‌ها با سرعتی خیره‌کننده رخ داده است. درحالی‌که نوآوری‌های معماری مانند ترنسفورمر مهم‌اند، بخش اعظم این پیشرفت را باید به مقیاس‌پذیری نسبت داد. به گفته بومن: «روند بسیار روشنی وجود دارد؛ بیشتر آزمون‌هایی که ما طراحی می‌کنیم، همین که فقط یک مرتبه مقیاس را افزایش دهیم، حل می‌شوند.»

علاوه بر این، قابلیت‌های جدید می‌توانند ناگهان از ناکجا سر درآورند. برای مثال، نسخه‌های کوچک‌تر GPT-3 توانایی ریاضی بسیار کمی نشان دادند که تعجب‌آور نبود؛ چون تنها برای پیش‌بینی واژه بعدی آموزش دیده بودند. اما توانایی‌های ریاضی به شکلی عجیب در نسخه کامل پدیدار شدند. به گفته «یاشا سول-دیکستاین» (Jascha Sohl-Dickstein) از گروه پژوهشی Google Brain: «تنها مقیاس است که می‌تواند قابلیت‌های شگفت‌انگیز جدید را امکان‌پذیر کند.»

هوش، چیزی فراتر از متن است

تایپ‌کردن و خواندن متن، رابط اصلی تعامل با هوش مصنوعی در انقلاب کنونی بوده است، اما اکنون روش‌های فراگیرتری برای آموزش و تعامل با سیستم‌های هوش مصنوعی درحال‌توسعه هستند.

یک مدل هوش مصنوعی گوگل DeepMind می‌تواند بازی‌های ویدئویی جهان باز مختلفی، نظیر بازی اکتشاف فضایی «No Man's Sky» را درست مانند یک انسان و تنها با تماشای ویدئو از روی صفحه‌نمایش انجام دهد؛ امری که می‌تواند گامی به‌سوی مدل‌هایی با هوش عمومی باشد که قادر به فعالیت در دنیای فیزیکی و مادی هستند.

انجام بازی‌های ویدئویی مدت‌هاست که ابزاری برای آزمودن پیشرفت سیستم‌های هوش مصنوعی بوده است، مانند تسلط هوش مصنوعی گوگل DeepMind بر شطرنج و Go؛ اما این بازی‌ها راه‌های مشخصی برای برد یا باخت دارند که آموزش هوش مصنوعی برای موفقیت در آن‌ها را نسبتاً ساده و سرراست می‌کند.

بازی‌های جهان باز مانند ماینکرفت با اهداف انتزاعی‌تر و اطلاعات اضافی و نامربوطی که می‌توان آن‌ها را نادیده گرفت، برای نفوذ و حل‌شدن توسط سیستم‌های هوش مصنوعی دشوارتر هستند. ازآنجاکه طیف انتخاب‌های موجود در این بازی‌ها آن‌ها را کمی بیشتر شبیه به زندگی عادی می‌کند، تصور می‌شود که این بازی‌ها سنگ بنای مهمی برای آموزش عامل‌های هوش مصنوعی باشند که بتوانند وظایفی مانند کنترل ربات‌ها را در دنیای واقعی انجام دهند و همچنین گامی به‌سوی «هوش جامع مصنوعی» (Artificial General Intelligence) باشند.

پژوهشگران DeepMind یک هوش مصنوعی توسعه داده‌اند که آن را «عامل مقیاس‌پذیر، آموزش‌پذیر و چنددنیایی» (Scalable Instructable Multiworld Agent) یا به‌اختصار «سیما» (SIMA) می‌نامند. این عامل می‌تواند تنها با استفاده از فید ویدئویی بازی، ۹ بازی ویدئویی مختلف و محیط‌های مجازی‌ای را که پیش‌ازاین ندیده است، بازی کند. این بازی‌ها شامل بازی No Man's Sky، بازی حل‌مسئله Teardown و بازی Goat Simulator 3 می‌شود. «فردریک بس» (Frederic Besse) از DeepMind دراین‌رابطه می‌گوید: «این در واقع همان رابطی است که انسان‌ها برای تعامل با رایانه از آن استفاده می‌کنند و یک رابط بسیار عمومی است.»

هنگامی که به زبان طبیعی از سیما درخواست شود، می‌تواند حدود ۶۰۰ وظیفه که ۱۰ ثانیه یا کمتر طول می‌کشند و در بازی‌های مختلف مشترک هستند؛ مانند حرکت‌کردن در محیط، استفاده از اشیا و پیمایش در منوها را انجام دهد. همچنین می‌تواند وظایف منحصربه‌فردتری مانند پرواز با سفینه‌های فضایی یا استخراج منابع را انجام دهد.

بس و همکارانش از مدل‌های تشخیص ویدئو و تصویر موجود برای تفسیر داده‌های ویدئویی بازی استفاده کردند، سپس سیما را آموزش دادند تا آنچه را در ویدئو اتفاق می‌افتد به وظایف خاصی نگاشت کند. برای فراهم‌کردن این اطلاعات، پژوهشگران از جفت افرادی که با هم بازی ویدئویی انجام می‌دادند استفاده کردند؛ به این صورت که یک نفر صفحه را تماشا می‌کرد و به دیگری می‌گفت چه حرکت‌هایی انجام دهد. همچنین از افراد خواستند که گیم‌پلی یا روند بازی خود را بازبینی کنند و حرکات ماوس و کیبوردی را که برای اقدامات بازی خود انجام داده‌اند، توصیف کنند. این کار به سیما اجازه داد تا یاد بگیرد که توصیفات کلامی حرکات چگونه با خود آن وظایف مرتبط می‌شوند.

وقتی سیما روی ۸ بازی آموزش دید، پژوهشگران دریافتند که می‌تواند بازی نهمی که پیش از آن ندیده بود را انجام دهد. بااین‌حال، عملکرد آن ضعیف‌تر از سطح انسانی بود. پژوهشگران از روش آموزشی‌ای استفاده کردند که در آن، هشت بازی‌ای که هوش مصنوعی را با آن آموزش می‌دادند به‌نوبت تغییر می‌دادند تا استفاده از بازی نهم حکم آزمون را داشته باشد و اطمینان حاصل شود که مدل می‌تواند هر بازی‌ای را که قبلاً ندیده، انجام دهد.

«فیلیپه منگوزی» (Felipe Meneguzzi) از دانشگاه ابردین می‌گوید که تعمیم‌دادن مهارت‌ها در بازی‌های مختلف، گام مهمی به‌سوی یک عامل هوش مصنوعی همه‌کاره (Generalist) است؛ اما سیما در حال حاضر تنها می‌تواند مجموعه نسبتاً محدودی از وظایف کوتاه را انجام دهد که نیاز به برنامه‌ریزی بلندمدت ندارند؛ ولی انجام طیف بسیار وسیع‌تری از وظایف پیچیده دشوارتر خواهد بود.


SIMA TEAM

برخی از محیط‌های بازی مجازی سه‌بعدی که هوش مصنوعی SIMA محصول DeepMind بر آن‌ها مسلط شده است.

«مایکل کوک» (Michael Cook) از کالج کینگز لندن می‌گوید: «شایان‌ذکر است که برای شرکت‌هایی مانند DeepMind، این پژوهش واقعاً درباره بازی‌ها نیست، بلکه درباره رباتیک است. پیمایش در محیط‌های سه‌بعدی وسیله‌ای برای رسیدن به این هدف است و این شرکت‌ها مشتاق ساخت سیستم‌های هوش مصنوعی‌ای هستند که بتوانند در جهان [واقعی] ادراک و عمل داشته باشند؛ بنابراین من تصور نمی‌کنم این موضوع تأثیر بزرگی بر بازی‌های ویدئویی داشته باشد، اما ممکن است تأثیرات ناشناخته زیادی بر زندگی ما در بیرون و در دنیای واقعی بگذارد.»

اما فقط از طریق آموزش نیست که مدل‌های چندوجهی امیدوارکننده ظاهر می‌شوند؛ بلکه نحوه عملکرد آن‌ها نیز حائز اهمیت است. GPT-4o می‌تواند به‌سرعت به ورودی‌های متنی، صوتی و ویدئویی پاسخ دهد؛ آن هم درحالی‌که با فرازوفرودهای صوتی و کلماتی صحبت می‌کند که حس قوی‌ای از احساسات و شخصیت را منتقل می‌کنند.

OpenAI در ماه می ۲۰۲۴، «تقلید احساسی» (Emotional Mimicry) یک حالت صوتی جدید را طی یک ارائه شرکتی به نمایش گذاشت. این هوش مصنوعی جدید که با صدایی زنانه صحبت می‌کرد و به نام ChatGPT پاسخ می‌داد، قابلیت‌های مکالمه‌ای‌اش بیشتر شبیه به هوش مصنوعی خوش‌مشرب با صداپیشگی «اسکارلت جوهانسون» در فیلم علمی-تخیلی Her به نظر می‌رسید تا پاسخ‌های ضبط‌شده و رباتیک فناوری‌های معمول دستیار صوتی.

«میشل کوهن» (Michelle Cohn) از دانشگاه کالیفرنیا نیز می‌گوید: «تعامل صوتی-به-صوتی جدید مدل GPT-4o شباهت بسیار بیشتری به تعامل انسان با انسان دارد. بخش بزرگی از این امر مربوط به زمان تأخیر کوتاه است... اما بخش بزرگ‌تر آن مربوط به سطح بیان احساسی است که این صدا تولید می‌کند.»

در طول مکالمه با «میرا موراتی» (Mira Murati)، مدیر ارشد فناوری شرکت و دو کارمند دیگر، GPT-4o به «مارک چن» (Mark Chen) کارمند OpenAI در مورد نفس‌های سنگین و سریعش تذکر داد و گفت: «اوه، یواش‌تر، تو جاروبرقی نیستی» و سپس یک تمرین تنفسی پیشنهاد کرد. همچنین نقاشی کارمند OpenAI «برت زوف» (Barret Zoph) که شامل کلمات و یک قلب بود را به‌صورت بصری بررسی کرد و با لحنی هیجان‌زده و پراحساس پاسخ داد: «آخی، می‌بینم نوشتی من عاشق ChatGPT هستم، خیلی لطف داری.»

نسخه جدید ChatGPT همچنین به‌صورت کلامی کاربران خود را برای حل یک معادله خطی ساده راهنمایی کرد، عملکرد یک کد کامپیوتری را توضیح داد و نموداری را که اوج‌گیری خطوط دما در ماه‌های تابستان را نشان می‌داد، تفسیر کرد. وقتی از او خواسته شد، این هوش مصنوعی حتی یک داستان ساختگی قبل از خواب را چندین بار بازگو کرد و بین روایت‌های به‌تدریج دراماتیک‌تر جابه‌جا شد و پایان آن را با آواز خواند.

یک چت‌بات هوش مصنوعی که بیشتر شبیه انسان شده باشد یا به‌اصطلاح علمی «انسان‌انگاری» (Anthropomorphised) شده باشد، تهدید دیگری را نیز نشان می‌دهد. مطابق مطالعات کوهن و همکارانش، رباتی که بتواند همدلی را از طریق مکالمات صوتی جعل کند، به طور بالقوه می‌تواند برای مردم هم خوش‌مشرب‌تر و هم متقاعدکننده‌تر به نظر برسد. این امر خطر تمایل بیشتر مردم به اعتماد به اطلاعات بالقوه نادرست و کلیشه‌های متعصبانه تولید شده توسط چنین مدل‌های زبانی بزرگی را افزایش می‌دهد.

کوهن می‌گوید: «این موضوع پیامدهای مهمی برای نحوه جست‌وجو و دریافت راهنمایی مردم از مدل‌های زبانی بزرگ دارد، به‌ویژه ازآنجاکه آن‌ها همیشه اطلاعات دقیقی تولید نمی‌کنند.»

ذهن و بدن

پژوهشگران علاوه بر ارتقای هوش مصنوعی با استفاده از دنیاهای مجازی فراگیر، اکنون با دادن بدن‌های رباتیک به آن‌ها که با دنیای فیزیکی تعامل دارند، در حال گسترش قابلیت‌های آن‌ها هستند.

ربات‌های انسان‌نما به‌تازگی قدم به انبارهای آمازون و کارخانه‌های خودروسازی مرسدس‌بنز گذاشته‌اند. اکنون، آن‌ها برای تلاشی حتی جاه‌طلبانه‌تر به خدمت گرفته می‌شوند: خلق «هوش جامع مصنوعی» با قابلیت‌هایی قابل‌مقایسه با انسان‌ها.

در اوایل سال ۲۰۲۴، شرکت انویدیا که از طریق فروش تراشه‌های هوش مصنوعی به یکی از ارزشمندترین شرکت‌های جهان تبدیل شد؛ چندین محصول سخت‌افزاری و نرم‌افزاری را برای تقویت آموزش ربات‌های انسان‌نما معرفی کرد. محور اصلی این محصولات، یک ابتکار بلندپروازانه یا به‌اصطلاح رایج Moonshot به نام پروژه «گروت» (GROOT) است که بر توسعه یک مدل هوش مصنوعی تمرکز دارد تا به ربات‌های انسان‌نما کمک کند از طریق دستورالعمل‌های زبانی، تماشای نمایش کار توسط انسان و تمرین اقدامات مختلف با کمک اپراتورهای انسانی از راه دور، بهینه‌تر آموزش ببینند.

تعاملات رباتیک با دنیای فیزیکی می‌تواند به نوبه خود مدل هوش مصنوعی را از طریق استراتژی «هوش تجسم‌یافته» (Embodied Intelligence) بهبود بخشد؛ استراتژی‌ای که سنگ بنای نقشه راه انویدیا برای توسعه هوش جامع مصنوعی است.

«جیم فن» (Jim Fan)، دانشمند پژوهشی در انویدیا و سرپرست پروژه گروت، در کنفرانس هوش مصنوعی این شرکت در سن خوزه کالیفرنیا گفت: «آنچه ما واقعاً می‌خواهیم، عامل‌های هوش مصنوعی‌ای هستند که به‌اندازه وال-ای (WALL-E) همه‌کاره باشند، به‌اندازه تمام فرم‌های رباتی یا تجسم‌های موجود در جنگ ستارگان (Star Wars) متنوع باشند و در دنیاهای بی‌نهایت، چه مجازی و چه فیزیکی، کار کنند، همان‌طور که در فیلم بازیکن شماره یک آماده (Ready Player One) می‌بینیم.»

Agility Robotics که یک شرکت سازنده ربات‌های انسان‌نما در اورگان است هم‌اکنون انتقال ربات‌های انسان‌نما از آزمایشگاه به کاربردهای تجاری را آغاز کرده است. از اواخر سال ۲۰۲۳، این شرکت در حال آزمایش ربات خود به نام «دیجیت» (Digit) در انبارهای آمازون است، جایی که این ربات کانتینرهای خالی را برمی‌دارد و جابه‌جا می‌کند. Agility همچنین یک اجرای آزمایشی را در انباری که متعلق به برند پوشاک زنانه SPANX است و توسط شرکت لجستیک IGXO داره می‌شود، آغاز کرده است. شرکت Apptronik نیز توافق‌نامه‌ای با مرسدس‌بنز امضا کرده است تا بررسی کند چگونه ربات‌های انسان‌نمای «آپولو» (Apollo) می‌توانند قطعات را تحویل دهند و بازرسی‌هایی را در خط مونتاژ کارخانه‌های خودروسازی انجام دهند؛ با این هدف که عملیات تجاری را تا اوایل سال ۲۰۲۵ آغاز کنند.

«بری فیلیپس» (Barry Phillips) از شرکت Apptronik دراین‌رابطه می‌گوید: «جهان برای انسان‌ها طراحی شده است، بنابراین ربات‌هایی بیشترین کارایی را خواهند داشت که بتوانند؛ مانند انسان‌ها عمل کنند. هزاران ربات تخصصی ممکن است امروز بتوانند یک یا دو وظیفه را با کارایی بالا انجام دهند، اما ظرفیت کاربرد اضافی آن‌ها بسیار محدود است.»

بااین‌حال «مارک رایبرت» (Marc Raibert) بنیان‌گذار Boston Dynamics و رئیس فعلی مؤسسه هوش مصنوعی ماساچوست، طی یک نشست در کنفرانس انویدیا عنوان کرد که وقتی نوبت به ارائه عملکرد رباتیک دقیق و قابل‌اعتماد می‌رسد، یادگیری مبتنی بر هوش مصنوعی هنوز از راهکارهای دست‌ساز مهندسان انسانی عقب‌تر است. به همین دلیل است که ربات‌های انسان‌نما هنوز نتوانسته‌اند ثابت کنند که در اکثر وظایف به‌اندازه انسان دارای مهارت هستند. رایبرت گفت: «وقتی این نگرانی وجود دارد که ربات‌ها قرار است شغل همه را بگیرند، من فقط آرزو می‌کنم که ای‌کاش می‌توانستیم شغل یک نفر را بگیریم. ما دقیقاً در لبه انجام تقریباً هرگونه کار مفیدی هستیم.»

اگرچه محدودیت‌ها باقی هستند، اما اشتیاق انکارناپذیری برای ربات‌های انسان‌نما وجود دارد. علاقه به ربات‌ها با رونق اخیر هوش مصنوعی مبتنی بر مدل‌های زبانی بزرگ و سایر برنامه‌های هوش مصنوعی مولد همراه شده است و انویدیا به نظر مصمم است که به این روند شتاب دهد.

«الکس گو» (Alex Gu) از Fourier Intelligence در شانگهای نیز می‌گوید: «پیش‌ازاین تصور می‌شد که چندین دهه طول می‌کشد تا ربات‌های انسان‌نمای همه‌منظوره به واقعیت بپیوندند. اکنون با نگاه به آینده، محتمل است که در ۵ تا ۱۰ سال آینده، ربات‌های انسان‌نما با قابلیت‌های گسترده‌ای ظهور کنند.»

نوع جدید از رایانه

رایانه‌های «نورومورفیک» (Neuromorphic) که عملکرد مغز انسان را تقلید می‌کنند می‌توانند مدل‌های هوش مصنوعی پیچیده‌تری را نسبت به آنچه در رایانه‌های معمولی ممکن است، اجرا کنند. انتظارات از رایانه‌های نورومورفیک بالاست زیرا آن‌ها ذاتاً با ماشین‌های سنتی متفاوت هستند.

درحالی‌که یک رایانه معمولی از پردازنده خود برای انجام عملیات استفاده می‌کند و داده‌ها را در حافظه‌ای جداگانه ذخیره می‌کند، یک رایانه نورومورفیک درست مانند مغز انسان از نورون‌های مصنوعی هم برای ذخیره و هم برای محاسبه استفاده می‌کند. این کار نیاز به جابه‌جایی داده‌ها به عقب و جلو بین اجزا که می‌تواند گلوگاهی برای رایانه‌های فعلی باشد را از بین می‌برد.

این معماری می‌تواند انرژی کمتری مصرف کند و همچنین راه‌های جدیدی را برای آموزش و اجرای مدل‌های هوش مصنوعی باز کند. این مدل‌ها از زنجیره‌ای از نورون‌ها استفاده می‌کنند؛ همان‌طور که مغز واقعی اطلاعات را پردازش می‌کند. برخلاف مدل‌های فعلی که ورودی را به‌صورت مکانیکی از تک‌تک لایه‌های نورون‌های مصنوعی عبور دهند.

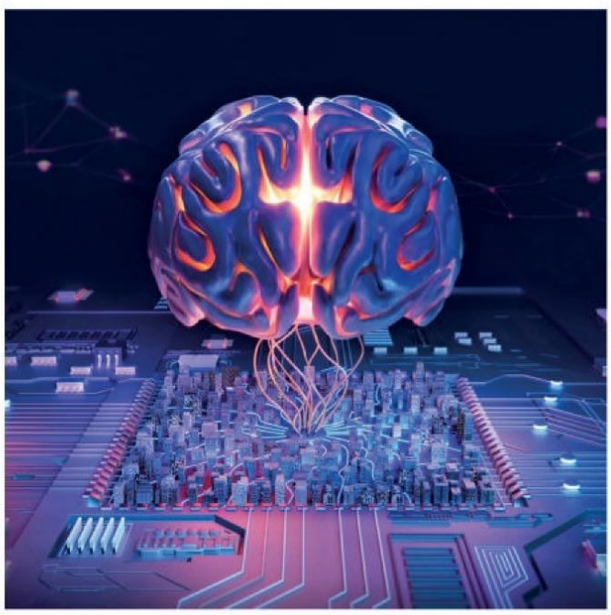
اما مدل‌های کنونی مانند ChatGPT با استفاده از کارت‌های گرافیکی که به‌صورت موازی کار می‌کنند، آموزش می‌بینند؛ به این معنی که می‌توان تراشه‌های زیادی را برای آموزش همان مدل به کار گرفت. اما ازآنجاکه رایانه‌های نورومورفیک با یک ورودی واحد کار می‌کنند و نمی‌توانند به‌صورت موازی آموزش ببینند، احتمالاً دهه‌ها طول می‌کشد تا حتی بتوان چیزی مانند ChatGPT را ابتدا روی چنین سخت‌افزاری آموزش داد، چه برسد به ابداع روش‌هایی که باعث شود پس از شروع به کار، به طور مداوم یاد بگیرد.

این رایانه‌ها می‌توانند مسیر بهتری برای نزدیک‌شدن به هوشی در سطح انسانی که به‌عنوان هوش جامع مصنوعی شناخته می‌شود را ارائه دهند؛ چیزی که بسیاری از کارشناسان هوش مصنوعی فکر می‌کنند با مدل‌های زبانی بزرگی که امثال ChatGPT را قدرت می‌بخشند، ممکن نخواهد بود.

«جیمز نایت» (James Knight) از دانشگاه Sussex انگلستان می‌گوید: «فکر می‌کنم این ایده روزبه‌روز کمتر موردبحث قرار می‌گیرد. آرزوی نهایی این است که روزی محاسبات نورومورفیک ما را قادر سازد تا مدل‌های شبیه مغز بسازیم.»

هوش جامع مصنوعی چیست؟

بزرگ‌ترین شرکت‌های هوش مصنوعی جهان، «هوش جامع مصنوعی» یا AGI را هدف خود قرار داده‌اند. اما همیشه روشن نیست که AGI چه معنایی دارد و بحث‌هایی درباره اینکه آیا چنین چیزی یک ایده ارزشمند است یا خیر، وجود دارد.



ALVAREZ/ISTOCK

اگر حتی علاقه‌ای گذرا به هوش مصنوعی داشته باشید، ناگزیر با مفهوم «هوش جامع مصنوعی» برخورد کرده‌اید. AGI طی چند سال گذشته به جایگاه یک «واژه مد روز» (Buzzword) ارتقا یافته است؛ زیرا هوش مصنوعی به پشتوانه موفقیت مدل‌های زبانی بزرگ به‌سرعت وارد آگاهی عمومی شده است.

این امر تا حد زیادی به این دلیل است که AGI به ستاره‌ای راهنما برای شرکت‌هایی تبدیل شده که پیش‌گام این نوع فناوری هستند. برای مثال، OpenAI بیان می‌کند که مأموریتش «اطمینان از این است که هوش جامع مصنوعی به نفع تمام بشریت باشد». دولت‌ها نیز مجذوب فرصت‌هایی شده‌اند که AGI ممکن است ارائه دهد و همچنین درگیر تهدیدات وجودی احتمالی آن هستند. درحالی‌که رسانه‌ها (طبیعتاً شامل New Scientist هم می‌شود) از ادعاهایی گزارش می‌دهند که هم‌اکنون جرقه‌هایی از AGI را در مدل زبانی بزرگ مشاهده کرده‌ایم.

با تمام این‌ها، همیشه مشخص نیست که AGI واقعاً چه معنایی دارد. در واقع، این موضوع بحث‌های داغی در جامعه هوش مصنوعی است. برخی اصرار دارند که یک هدف مفید است و برخی دیگر معتقدند که این یک خیال‌بافی بی‌معناست که نشان‌دهنده کژفهمی از ماهیت هوش و چشم‌انداز ما برای بازتولید آن در ماشین‌ها است. «ملانی میچل» نیز می‌گوید: «واقعاً یک مفهوم علمی نیست.»

«ما بر بازتولید قابلیت‌های خودمان تمرکز داریم و این امر منجر به انسان‌انگاری گسترده‌ای بر سیستم‌های هوش مصنوعی می‌شود.»

هوش مصنوعی شبه‌انسان و هوش مصنوعی فوق‌هوشمند (Superintelligent) قرن‌هاست که از موضوعات اصلی علمی-تخیلی بوده‌اند. اما اصطلاح AGI حدود ۲۰ سال پیش رواج یافت، زمانی که توسط دانشمند رایانه «بن گورتزل» (Ben Goertzel) و «شین لگ» (Shane Legg) هم‌بنیان‌گذار DeepMind به کار رفت. این عبارت به‌خوبی حس روبه‌رشدی را در بر می‌گرفت مبنی بر اینکه این حوزه باید فراتر از کاربردهای محدود حرکت کند تا سیستم‌هایی بسازد که بتوانند هر کاری که یک انسان انجام می‌دهد را انجام دهند.

از آن زمان DeepMind تلاش کرده است تا AGI را به‌گونه‌ای بازتعریف کند که تنها به «وظایف شناختی» مربوط شود. در سال گذشته، لگ به همراه «دمیس هاسابیس» (Demis Hassabis)، دیگر هم‌بنیان‌گذار DeepMind و همکارانشان درباره اینکه چه چیزی یک AGI را تشکیل می‌دهد، توضیحات بیشتری ارائه دادند. آن‌ها یک چارچوب شش سطحی پیشنهاد کردند که در بالاترین سطح آن، سیستمی قرار دارد که می‌تواند در «طیف وسیعی از وظایف غیرفیزیکی، از جمله توانایی‌های فراشناختی مانند یادگیری مهارت‌های جدید از ۱۰۰ درصد انسان‌ها عملکرد بهتری داشته باشد.»

«مردیت موریس» (Meredith Morris)، عضو تیم DeepMind که اکنون بخشی از گوگل است می‌گوید: «ایده سطوح در واقع به این نکته اشاره دارد که یک پیوستار وجود دارد؛ یعنی همگام با تکامل فناوری، یک پیشرفت تدریجی وجود دارد.» موریس امیدوار است که کار آن‌ها توجه بیشتری را به این ایده جلب کند و در نهایت به‌نوعی اجماع بر سر اینکه AGI واقعاً چیست، منجر شود: «ما خیلی دوست داریم افرادی از سایر رشته‌هایی که هوش و یادگیری را مطالعه می‌کنند با پژوهشگران ما در توسعه این معیارها همکاری کنند.»

اما میچل خاطرنشان می‌کند که خود هوش یک مفهوم چندبعدی است که همپوشانی‌های زیادی با سایر مفاهیم به همان اندازه مبهم، مانند «ادراک و آگاهی حسی» (Sentience) و «فهمیدن» دارد. به همین دلیل، هوش به‌راحتی سایر وظایف ملموس‌تر، مانند توانایی ترجمه زبان، تنها با یک آزمون قابل‌اندازه‌گیری نیست.

موشکافی دقیق‌تر اینکه چه زمانی یک هوش مصنوعی می‌تواند AGI در نظر گرفته شود، ممکن است پیشرفتی حاصل کند. اما میچل همچنان تردید دارد که نوع ماشینی که طرفداران AGI تصور می‌کنند محقق شود؛ زیرا مشخص نیست که آیا توانایی‌های هوش انسانی اصلاً می‌توانند به مفاهیم مستقل خلاصه و انتزاع شوند یا خیر؛ چه برسد به اینکه در هوش مصنوعی بازتولید شوند. میچل می‌گوید: «نوعی ایمان وجود دارد که این حوزه مدت‌هاست به آن باور داشته؛ اینکه ما می‌توانیم هوشی در سطح انسان را در این زیربسترهای بی‌جسم (Disembodied) توسعه دهیم. اینکه آیا چنین چیزی ممکن است یا خیر، فکر می‌کنم یک سؤال بزرگ و بی‌پاسخ است.»

برای «توماس دیتریش» (Thomas Dietterich) در دانشگاه ایالتی اورگان، مشکل AGIمشکلی عملی‌تر است؛ یعنی اینکه تعریف هوش مصنوعی با ارجاع به انسان‌ها اشتباه است و عنوان می‌کند: «ما بر بازتولید قابلیت‌های خودمان تمرکز داریم و این امر منجر به انسان‌انگاری (Anthropomorphisation) گسترده‌ای بر سیستم‌های هوش مصنوعی می‌شود و در نهایت هم به آن‌ها نام‌هایی مانند سیری و الکسا می‌دهیم.»

دیتریش اضافه می‌کند که در عوض، ما باید به هوش مصنوعی به‌عنوان «یک محصول هوشمند که می‌تواند کارهای خاصی را برای ما انجام دهد» فکر کنیم؛ چیزی که بسیار شبیه به همان تصوری است که جامعه هوش مصنوعی پیش از پیدایش مفهوم AGI در ذهن داشت.

فصل ۳

هوش مصنوعی
چگونه می‌تواند کمک کند؟

شرکت‌های فناوری اغلب هوش مصنوعی را «برهم‌زننده یا اختلال‌گر» (Disruptive) توصیف می‌کنند؛ به این معنی که به طور چشمگیری نحوه عملکرد جوامع و اقتصادهای سراسر جهان را تغییر می‌دهد.

این ابزارهای مولد در حال حاضر نحوه کار و زندگی روزانه ما را تغییر می‌دهند. اما برای اطمینان از اینکه هوش مصنوعی واقعاً به ما کمک می‌کند یا نه؛ ابتدا باید اهداف، باورها و خواسته‌های ما را درک کند.

هوش مصنوعی
چه معنایی برای مشاغل ما دارد؟

هر کسی که اتصال اینترنت دارد، به ابزارهایی دسترسی دارد که می‌توانند تقریباً به هر سؤالی در جهان پاسخ دهند. همه چیز از مقالات دانشگاهی گرفته تا کدهای کامپیوتری بنویسند و هنر یا تصاویر واقع‌گرایانه تولید کنند. این نوع قابلیت‌ها می‌توانند عمیقاً بر اقتصاد، مشاغل، آموزش، فرهنگ و موارد دیگر تأثیر بگذارند.

به‌طورکلی، آینده‌پژوهان برآورد می‌کنند که هوش مصنوعی مولد، بهره‌وری و رشد اقتصادی را در اقتصادهای پیشرفته تقویت خواهد کرد. گزارشی که توسط Goldman Sachs در مارس ۲۰۲۳ منتشر شد، پیش‌بینی کرد که هوش مصنوعی مولد می‌تواند ظرف یک دهه، تولید ناخالص داخلی یا همان GDP سالانه جهانی را ۷ درصد افزایش دهد که معادل افزایشی تقریباً ۷ تریلیون دلاری است. این گزارش نشان داد که ورود هوش مصنوعی مولد «احتمال یک رونق در بهره‌وری نیروی کار را افزایش می‌دهد؛ نظیر آنچه پس از ظهور فناوری‌های همه‌منظوره پیشین، مانند موتور الکتریکی و رایانه شخصی رخ داد.»

ایده این است که هوش مصنوعی میلیون‌ها «نیروی انسانی دانش‌ورز» (Knowledge Workers) مانند دانشمندان، ویراستاران، وکلا و پزشکان را ظرف چند سال بهره‌ورتر خواهد کرد. اما حقیقت این است که پیش‌بینی و ارزیابی این موارد دشوار است.

«آوی گلدفارب» (Avi Goldfarb) از دانشگاه تورنتو اعتقاد دارد یکی از حوزه‌هایی که هوش مصنوعی مولد در آن به نگرانی‌ها دامن می‌زند، اشتغال است؛ اما تصویر [این وضعیت] ممکن است با موج‌های قبلی اتوماسیون متفاوت باشد. گلدفارب می‌گوید پیشرفت‌های رایانه‌ای اغلب کارمندان کم‌درآمد مانند ماشین‌نویس‌ها و صندوق‌داران را جابه‌جا یا حتی بیکار کرده درحالی‌که بهره‌وری کارکنان تحصیل‌کرده را افزایش داده و به نابرابری درحال‌رشد کمک کرده است. اما این بار، هوش مصنوعی امروزه با قدرت‌نمایی وارد قلمرو مشاغل پردرآمد می‌شود.

GREMLIN/ISTOCK

پژوهش‌های جدید OpenAI نشان داد که ۸۰ درصد از نیروی کار ایالات متحده ممکن است شاهد تأثیرپذیری ۱۰ درصدی از وظایف کاری خود توسط این فناوری باشند و برای ۱۹ درصد از کارمندان، این رقم می‌تواند به حداقل نصف افزایش یابد. کسانی که بیشترین تأثیر را می‌پذیرند عمدتاً کارمندان به‌اصطلاح یقه‌سفید (White-Collar) مانند حسابداران، روزنامه‌نگاران و طراحان وب خواهند بود.

گلدفارب عنوان می‌کند بعید است که کل مشاغل خودکار شوند و وظایفی که هوش مصنوعی می‌تواند در آن‌ها کمک کند، معمولاً آن‌هایی هستند که شبیه کارهای روزمره و وقت‌گیر به نظر می‌رسند، مانند نوشتن ایمیل‌ها یا جست‌وجو و کاوش در اسناد؛ بنابراین اتوماسیون می‌تواند وقت نیروی انسانی را برای کارهای ارزشمندتر آزاد کند.

«کاتیا کلینوا» (Katya Klinova) از سازمان «مشارکت در هوش مصنوعی» (Partnership on AI) که مدافع استفاده مسئولانه از این فناوری است اعتقاد دارد خیلی چیزها بستگی به نحوه استقرار و به‌کارگیری این فناوری دارد. کلینوا عنوان می‌کند ابزارهای هوش مصنوعی می‌توانند به یک «یار کمکی شخصی» (Personal Sidekick) تبدیل شوند که به کارکنان کمک می‌کند ایده‌پردازی کنند، پیش‌نویس اسناد را تهیه کنند و ایده‌ها را به عمل تبدیل کنند. از سوی دیگر، شرکت‌ها ممکن است هوش مصنوعی را به‌عنوان راهی برای کاهش هزینه‌های نیروی کار ببینند و برای تولید محتوای بازاریابی، کدهای کامپیوتری یا تصاویر به آن‌ها روی آورند درحالی‌که تعداد رو به کاهشی از نیروی انسانی به پر کردن این خلأ [انجام کارهای باقی‌مانده] تنزل می‌یابند. کلینوا عنوان می‌کند گسترش سریع قابلیت‌های هوش مصنوعی همچنین باعث می‌شود تصمیم این که کارکنان چه مهارت‌هایی را می‌بایست پرورش دهند، روزبه‌روز مبهم‌تر شود.

بااین‌حال، هوش مصنوعی می‌تواند تأثیر عمیقی در آموزش داشته باشد که حداقل بخشی از آن مثبت است. نگرانی‌هایی در مورداستفاده دانش‌آموزان از ابزارهای هوش مصنوعی برای انجام تکالیف مطرح شده است، اما «ویکتور لی» (Victor Lee) از دانشگاه استنفورد اعتقاد دارد فرصتی وجود دارد تا نحوه ارزیابی یادگیری را تغییر دهیم؛ یعنی تمرکز را از توانایی تولید مقالات خوب‌نوشته‌شده برداریم و به سمت «چیزهای پیچیده‌تری مانند مقایسه، نقد، انطباق و اصلاح» ببریم. هوش مصنوعی همچنین می‌تواند با ظهور ربات‌هایی در نقش معلم خصوصی که می‌توانند با دانش‌آموزان همکاری و بحث کنند، فرصت‌های جدیدی برای تدریس ارائه دهد.

لی عنوان می‌کند ادغام هوش مصنوعی در آموزش نیازمند احتیاط است، زیرا این مدل‌ها می‌توانند سوگیری‌ها را از داده‌های آموزشی خود جذب کنند و اغلب واقعیت‌ها را از خود درآورده و اصطلاحاً دچار توهم می‌شوند. اما می‌افزاید که یادگیری نحوه استفاده از این ابزارها در محل کار آینده ضروری خواهد بود و می‌گوید: «افرادی که نمی‌دانند چگونه از هوش مصنوعی استفاده کنند، جای خود را به افرادی خواهند داد که می‌دانند چگونه از هوش مصنوعی استفاده کنند.»

تعداد ابزارهای هوش مصنوعی به لطف حجم روبه‌رشد نسخه‌های «متن‌باز» (Open-Source) که توسط گاهی توسط شرکت‌ها و گاهی حتی توسط یک فرد توسعه یافته و به‌صورت رایگان در دسترس قرار گرفته‌اند؛ در حال افزایش است. این ابزارها می‌توانند به‌جای تکیه بر قدرت محاسباتی شرکت‌های بزرگی مانند OpenAI و گوگل، روی لپ‌تاپ‌ها و حتی تلفن‌های هوشمند اجرا شوند.

صرف‌نظر از این، هوش مصنوعی همه‌جا هست و مدل‌های متن‌باز می‌توانند با ترکیب‌شدن با سایر ابزارهای نرم‌افزاری یا بازآموزی (Retraining) روی داده‌های خودتان، متناسب با نیازهای شما شخصی‌سازی شوند. این امر می‌تواند نگرانی‌ها در مورد به‌اشتراک‌گذاری داده‌های شخصی با شرکت‌های فناوری را کاهش دهد. «سایمون ویلیسون» (Simon Willison)، پژوهشگر و توسعه‌دهنده مستقل می‌گوید: «به طرز مضحکی علمی-تخیلی به نظر می‌رسد، اما دنیایی وجود دارد که در آن همه، دستیار هوش مصنوعی برخط و همیشه‌حاضر خودشان را دارند.»

استفاده از هوش مصنوعی در زندگی روزمره

از کمک برای نگارش یک ایمیل بی‌نقص گرفته تا ارائه تمرینات ورزشی شخصی و برنامه‌ریزی وعده‌های غذایی، انبوهی از ابزارهای هوش مصنوعی مولد که در پاسخ به پرامپت متن، ویدئو، تصویر و سایر محتواها را تولید می‌کنند؛ در دنیای امروز حضور دارند تا به روان‌تر کردن کارهای روزمره و دشوار کمک کنند.

یک راه برای خوش‌آمدگویی به هوش مصنوعی در دنیای خودتان این است که ببینید کسب‌وکارها چگونه از آن برای بهره‌ورترشدن استفاده می‌کنند و سپس آن را با نیازهای خود تطبیق دهید. یک ایمیل دشوار دارید که باید برای صاحب‌خانه‌تان بنویسید و نمی‌دانید از کجا شروع کنید؟ از هوش مصنوعی مولد بپرسید. «کریستینا فیلیپ» (Christina Philippe)، استراتژیست ارشد Ogilvy می‌گوید: «هوش مصنوعی برای چیزهای واقعی و مستند که باید به طور مختصر و بدون احساس بیان شوند، ایده‌آل هستند». او چنین وظایفی را به ChatGPT برون‌سپاری کرده است؛ به این صورت که فهرستی از مشکلات و راهکارهای مطلوب خود را به آن می‌دهد. فیلیپ دریافته است که این هوش مصنوعی مولد ایمیل‌هایی می‌نویسد که نیاز به اصلاحات بسیار جزئی دارند. «ایرنا کرونین» (Irena Cronin)، مدیرعامل Infinite Retina، یک شرکت مشاوره پژوهشی هوش مصنوعی مولد می‌گوید: «هوش مصنوعی یک راه‌حل واقعاً سریع است. ظرفیت زیادی برای افراد معمولی و عادی وجود دارد تا بتوانند از آن استفاده کنند.»

بااین‌حال، مطمئن شوید هر چیزی را که هوش مصنوعی انتخابی شما تولید می‌کند، دوباره بررسی می‌کنید. این ابزارها گهگاه دچار «افسانه‌بافی» (Confabulation) می‌شوند؛ حالتی که در آن ریاضیات پشت هوش مصنوعی می‌تواند متن‌های بی‌معنی، ادعاهای نادرست یا حقایقی کاملاً اشتباه تولید کند. فریب متن‌های شیوا را نخورید؛ یک مطالعه نشان داده است که متن‌هایی با ظاهر بسیار روان که توسط هوش مصنوعی تولید می‌شوند، بیشترین احتمال داشتن خطا و بی‌دقتی را دارند.

همان‌طور که زندگی در کنار هوش مصنوعی مولد را شروع می‌کنیم، خوب است که بتوانیم آن را به‌راحتی از انسان‌ها تشخیص دهیم. مثلاً اگر در می‌خواهید از طریق چت آنلاین در وب‌سایت یک فروشنده کمک بگیرید، یک ترفند ساده با حروف بزرگ انگلیسی باید ممکن است کمک کند. اگر من از شما بپرسم: آیا چمن سبز است یا آبی؟ (که در آن کلمات اصلی با حروف کوچک نوشته شده‌اند)

“isBABY grassCARPET greenDRY orFROG blueSALT?”

شما احتمالاً قادر خواهید بود پاسخ صحیح بدهید: سبز.

در یک آزمایش مشابه، ۱۰۰ درصد افراد پاسخ درست را دادند، درحالی‌که پنج مدل زبانی بزرگ که ابزارهای هستند که متن را به شیوه‌ای شبیه انسان می‌فهمند و تولید می‌کنند؛ همگی شکست خوردند. البته مدل‌های هوش مصنوعی آینده ممکن است بتوانند الگوها را به‌آسانی ما تشخیص دهند، اما این روش فعلاً (در زمان انتظار گزارش اصلی) کار می‌کند.

فراتر از برون‌سپاری کارهایی که از آن‌ها وحشت دارید، هوش مصنوعی می‌تواند به شما کمک کند تا مصرف رسانه‌ای خود را ساده و بهینه کنید. مثلاً می‌توانید از ChatGPT برای خلاصه‌کردن پادکست‌های طولانی یا ویدئوهای یوتیوب استفاده کنید و در زمان ارزشمند خود صرفه‌جویی کنید. نسخه پولی آن، ChatGPT Plus، قابلیت افزودن پلاگین‌های شخص ثالث را دارد که برخی از آن‌ها می‌توانند متن‌های پیاده‌سازی‌شده را که به‌صورت خودکار از محتوا تولید شده‌اند، اسکن و خلاصه کنند. اگرچه، همان‌طور که قبلاً گفته شد، همیشه باید پیش از پریدن وسط یک مکالمه در مهمانی شام و گفتن از موضوعی که به‌تازگی از ChatGPT یادگرفته‌اید؛ بررسی کنید که آن محتوا واقعاً درست باشد.

صحبت از مهمانی شام شد، شاید تجربه «شان لینهان» (Sean Linehan) جالب باشد. این مدیرعامل حوزه فناوری مستقر در سان‌فرانسیسکو در شرکت منابع انسانی Exec، تصمیم گرفت یک مهمانی شام برگزار کند که منوی آن کاملاً توسط هوش مصنوعی طراحی شده بود. او از ChatGPT خواست یک وعده‌غذایی تلفیقی با ترکیب آشپزی مدیترانه‌ای و هندی تولید کند. هوش مصنوعی فهرستی از مواد اولیه، دستور پخت‌ها و دستورالعمل‌های آشپزی را ایجاد کرد. برخی مشکلات جزئی یا به‌اصطلاح سکسکه (Hiccups) وجود داشت؛ مثلاً ChatGPT از لینهان خواسته بود با اضافه‌کردن طیفی از ادویه‌ها به چای ماسالا آماده، چای ماسالا درست کند؛ اما لینهان می‌گوید این آزمایش در مجموع موفقیت‌آمیز بود.

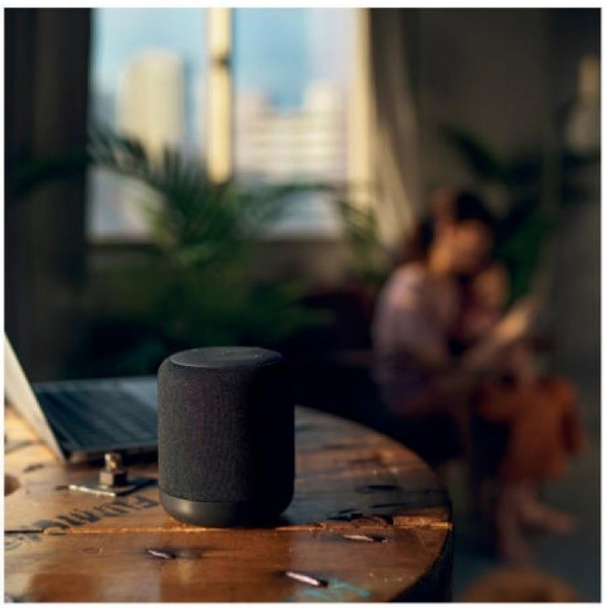
هوش مصنوعی مولد همچنین در پایان یک روز طولانی زمانی که نیاز دارید با استفاده از خرده‌ریزها و باقیمانده‌ها یک وعده‌غذایی سرهم کنید، می‌تواند کمک‌کننده باشد. «سالی هاوارد» (Sally Howard) روزنامه‌نگار می‌گوید: «ChatGPT برای این کار بهترین گزینه است. به نظر می‌رسد روح یک آشپز خانگی اهل آمریکای شمالی در اواخر میان‌سالی در آن حلول کرده است. این کار سرگرم‌کننده است، اگرچه در یک لحظه ناامیدکننده، به من توصیه کرد چیزی را با سیب‌زمینی برشته تزیین کنم.»

اگر نیاز دارید آن غذاها را بسوزانید، هوش مصنوعی در آنجا هم می‌تواند کمک کند. اپلیکیشن‌های تناسب‌اندام مجهز به هوش مصنوعی به‌سرعت درحال‌رشد هستند و با بر عهده گرفتن برنامه ورزشی شما با کسری از هزینه، جایگزین مربیان شخصی می‌شوند. برای مثال، برنامه محبوب Freeletics داده‌های تناسب‌اندام شما را با افراد دیگر مقایسه می‌کند تا یک برنامه تمرینی شخصی‌سازی شده به شما ارائه دهد.

هوش مصنوعی می‌تواند برنامه‌های آموزشی شخصی‌سازی‌شده‌ای برای کمک به فرزندانتان نیز ارائه دهد. «کاتیا کلینوا» (Katya Klinova) از Partnership on AI می‌گوید: «برای اولین بار در تاریخ، ما نزدیک هستیم به اینکه بتوانیم تدریس خصوصی مقرون‌به‌صرفه و شخصی‌سازی‌شده را برای همه فراهم کنیم.» اوایل امسال، پلتفرم یادگیری آنلاین «آکادمی خان» (Khan Academy) یک ربات معلم خصوصی مجهز به هوش مصنوعی به نام «خانمیگو» (Khanmigo) را منتشر کرد که می‌تواند کوییز بسازد، در تکالیف نوشتاری همکاری کند و با دانش‌آموزان درباره موضوعات بحث و مناظره نماید.

اگر همه این‌ها هنوز شبیه کار به نظر می‌رسد، شاید یک استراحت لازم داشته باشید که البته توسط هوش مصنوعی برنامه‌ریزی شده باشد. بیش از ۴۵ سرویس هوش مصنوعی وجود دارند که برنامه‌های سفر شخصی‌سازی شده ارائه می‌دهند و در پیداکردن آن هتل پنج‌ستاره بی‌نقص که اجازه ورود سگ را می‌دهد و منظره اقیانوس دارد، کمک می‌کنند.

البته، بالاتر از همه این‌ها، این نکات باید هوشمندانه و بااحتیاط استفاده شوند. هوش مصنوعی مولد در مراحل اولیه خود باقی‌مانده است و زمینه زیادی برای اشتباهات و خرابکاری وجود دارد. بااین‌حال، اگر بادقت استفاده شود، می‌تواند ابزاری سودمند برای صرفه‌جویی در زمان باشد.



KOH SZE KIAT/ISTOCK

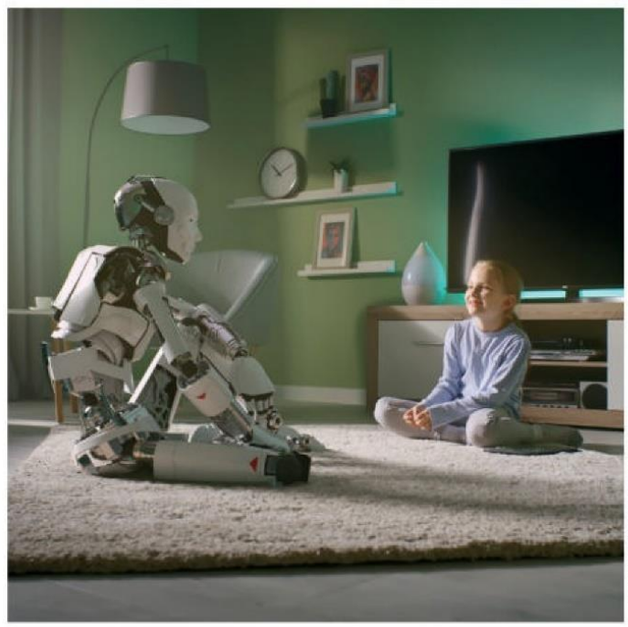
حالا وقت گفت‌وگوست

انسان‌ها توانایی شگرفی در استنتاج اهداف، تمایلات و باورهای دیگران دارند؛ مهارتی حیاتی که به این معناست که ما می‌توانیم اقدامات دیگران و پیامدهای اقدامات خودمان را پیش‌بینی کنیم. اگر قرار باشد هوش مصنوعی در زندگی روزمره واقعاً مفید واقع شوند و به طور مؤثر با ما همکاری کند باید توانایی‌های شهودی مشابهی را در خود ایجاد کند. این مسئله‌ای است که در صنعت با نام «هم‌راستایی» (Alignment) شناخته می‌شود.

مدل‌های هوش مصنوعی معمولاً با عدم قطعیت‌های رفتار انسانی دست‌وپنجه نرم کرده‌اند. اما تحولات اخیر، از جمله شواهدی مبنی بر اینکه هوش مصنوعی پشت ChatGPT دیدگاه‌های دیگران را درک می‌کند، نشان می‌دهد که ماشین‌های دارای هوش و ادراک اجتماعی (Socially Savvy)، یک رؤیای دور از دسترس نیستند. علاوه بر این، فکرکردن درباره دیگران می‌تواند گامی به‌سوی هدفی بزرگ‌تر باشد: یعنی هوش مصنوعی دارای خودآگاهی.

«هاد لیپسون» (Hod Lipson) از دانشگاه کلمبیا می‌گوید: «اگر می‌خواهیم ربات‌ها یا به‌طورکلی هوش مصنوعی به شیوه‌ای یکپارچه و کامل در زندگی ما ادغام شوند، باید این مسئله را حل کنیم. باید این موهبتی را که تکامل به ما داده است تا ذهن دیگران را بخوانیم، به آن‌ها نیز بدهیم.»

روان‌شناسان به توانایی استنباط حالت ذهنی فردی دیگر، «نظریه ذهن» (Theory of Mind) می‌گویند. «آلیسون گوپنیک» (Alison Gopnik) دانشگاه کالیفرنیا، برکلی اعتقاد دارد که این ظرفیت در انسان‌ها از سنین بسیار پایین شروع به رشد می‌کند. تا سن ۹ ماهگی، نوزادان می‌فهمند که اقدامات افراد با اهداف آن‌ها مرتبط است. بین ۱۸ ماهگی تا ۲ سالگی، شروع به درک این موضوع می‌کنند که اهداف هر شخص می‌تواند متفاوت باشد، زیرا ما تمایلات منحصربه‌فردی داریم. در حدود ۴ سالگی، کودکان «آزمون باور غلط» (False Belief Test) را با موفقیت پشت سر می‌گذارند. این آزمون شامل کودکی است که تماشا می‌کند شخصی یک شیء را در جایی قرار می‌دهد و سپس آن شیء بدون اطلاع آن شخص جابه‌جا می‌شود. سپس از کودک پرسیده می‌شود که آن شخص ابتدا کجا را نگاه خواهد کرد؛ پاسخ درست دادن به این سؤال مستلزم آن است که کودک بفهمد آن شخص باورهای متفاوت و البته غلطی دارد. گوپنیک می‌گوید به طرز قابل‌توجهی، انسان‌ها تا سن حدود ۵ سالگی، توانایی نسبتاً پیچیده‌ای برای استنباط آنچه دیگران فکر می‌کنند، پیدا می‌کنند.



EVGENIYSHKOLENKO/ISTOCK

تنها انسان‌ها ذهن‌خوان نیستند. به نظر می‌رسد برخی حیوانات نیز درجات متفاوتی از نظریه ذهن را نشان می‌دهند. شامپانزه‌ها، بونوبوها و اورانگوتان‌ها همگی قادر به گذراندن آزمون باور غلط هستند و پرندگانی مانند کلاغ‌ها و زاغ‌های کبود نیز قادر به استنباط حالات ذهنی دیگران هستند.

بااین‌حال، چگونگی بازتولید این قابلیت‌ها در ماشین‌ها اصلاً روشن نیست. بخشی از مشکل این است که آنچه ما نظریه ذهن می‌نامیم، در واقع فقط یک چیز نیست. «برترام مله» (Bertram Malle) از دانشگاه براون می‌گوید: «عناصر بسیار زیادی در آنچه مردم نظریه ذهن می‌نامند وجود دارد و مجموعه‌ای بزرگ از توانایی‌هاست.» در انتهای ساده‌تر این طیف، ظرفیت درک انگیزه‌های پشت اقدامات قرار دارد درحالی‌که در انتهای دیگر، نوعی مانورهای اجتماعی پیچیده‌ای است که در رمان‌های جین آستین می‌خوانید.

به گفته «جولیان خارا-اتینگر» (Julian Jara-Ettinger) از دانشگاه ییل، یکی از چالش‌های اصلی، «بافت» (Context) است. برای مثال، اگر کسی بپرسد که آیا برای دویدن می‌روید و شما پاسخ دهید «باران می‌بارد»، آن‌ها می‌توانند به‌سرعت استنتاج کنند که پاسخ منفی است. اما این کار نیازمند حجم عظیمی از دانش پس‌زمینه درباره دویدن، آب‌وهوا و ترجیحات انسانی است.

ترجمه حتی ساده‌ترین مهارت‌های اجتماعی به زبان ماشین‌ها اصلاً کار ساده‌ای نیست. نوع محاسبات درگیر در این کار، کاملاً متفاوت از منطق فرمولی‌ای است که رایانه‌ها معمولاً از آن استفاده می‌کنند. «ساموئل کاسکی» (Samuel Kaski) از دانشگاه منچستر می‌گوید مهم‌تر از همه این است که ماشین‌ها باید بتوانند با عدم قطعیت کنار بیایند.

فرایندهای ذهنی درونی یک شخص به‌صورت مستقیم غیرقابل‌مشاهده است، بنابراین بهترین کاری که می‌توانید انجام دهید این است که بر اساس شواهد موجود، حدس‌های آگاهانه بزنید. انجام این کار معمولاً شامل معکوس کردن یک تکنیک کلاسیک یادگیری ماشین به نام «یادگیری تقویتی» (Reinforcement learning) است. در شکل مرسوم این تکنیک، یک هدف برای هوش مصنوعی تعیین می‌شود و برای انجام اقداماتی که در راستای آن هدف کار می‌کنند، به آن پاداش داده می‌شود. از طریق آزمون‌وخطا، عامل رفتارهایی را یاد می‌گیرد که به هدف می‌رسند؛ اما «یادگیری تقویتی معکوس» (Inverse reinforcement learning) برعکس عمل می‌کند. در این روش از مشاهده اقدامات یک شخص در طول زمان و استفاده از آن برای ساختن تصویری از آنچه آن شخص سعی دارد به دست آورد استفاده می‌شود. این رفتار شبیه به رفتار کودکی است که برای اولین‌بار یک بازی قایم‌موشک را تماشا می‌کند و به‌سرعت استنتاج می‌کند که اهداف بازیکنان مختلف چیست.

یک روش محبوب برای انجام الگوریتم یادگیری تقویتی معکوس، تکیه بر یک تکنیک آماری به نام «استنباط بیزی» (Bayesian inference) است. این روش به کاربر اجازه می‌دهد تا داده‌های جدید را تحلیل کنید درحالی‌که دانش قبلی را نیز در نظر می‌گیرید. کاسکی می‌گوید این تکنیک قدرتمند است؛ زیرا به‌خوبی با عدم قطعیت کنار می‌آید و اجازه می‌دهد از آنچه قبلاً درباره یک مسئله می‌دانید استفاده کنید و همچنین با در دسترس قرارگرفتن اطلاعات جدید، خود را تطبیق دهید.

در سال ۲۰۲۲، کاسکی و همکارش «سباستین دی پیوتر» (Sebastiaan De Peuter) از رویکردهای بیزی برای توسعه یک نمونه اولیه از دستیار دیجیتال استفاده کردند که می‌توانست به فرد در حل مشکلاتی که نیازمند اتخاذ مجموعه‌ای از تصمیمات به‌هم‌پیوسته بود، کمک کند. در این مورد، هوش مصنوعی به فرد کمک کرد تا یک سفر تفریحی را بر اساس بودجه، محدودیت‌های زمانی و ترجیحات خود برنامه‌ریزی کند. این مثال ممکن است پیش‌پاافتاده به نظر برسد، اما کاسکی می‌گوید این کار از اساس شبیه به کمک در حل وظایف پیچیده‌تر مانند طراحی مهندسی است و هدف او، توسعه دستیارهای هوش مصنوعی است که کار دانشمندان و پزشکان را سرعت دهند و می‌گوید: «انگیزه اصلی واقعاً از توانایی حل بهتر مسائل سخت‌تر ناشی می‌شود.»

بیشتر پژوهش‌ها در زمینه نظریه ذهن ماشین تا به امروز بر سناریوهای ساده‌سازی شده تکیه داشته‌اند؛ مانند استنباط اهداف عامل‌هایی که در جهان‌های شبکه‌ای (Grid worlds) دوبعدی و ساده حرکت می‌کنند. اما «جاش تننبام» (Josh Tenenbaum) از MIT در تلاش است تا این تکنیک‌ها را به دنیای واقعی بیاورد. در سال ۲۰۲۰، گروه او رویکردهای بیزی را با یک زبان برنامه‌نویسی که توسط ربات‌ها استفاده می‌شود ترکیب کرد که به گفته او می‌تواند در نهایت به اعمال این تکنیک‌ها در چالش‌های عملی رباتیک کمک کند. اخیراً نیز پیچیدگی کار را به سه بُعد افزایش داده‌اند.

تننبام می‌گوید یکی از راه‌های کلیدی برای آزمایش نظریه ذهن در کودکان، نشان‌دادن ویدئوهایی از افراد به آن‌ها و سپس پرسیدن درباره این است که آن افراد سعی در انجام چه کاری دارند. تیم او می‌خواست این کار را با هوش مصنوعی امتحان کند. در سال ۲۰۲۱، تننبام و همکارانش از چالش جدیدی پرده‌برداری کردند که در آن عامل‌های هوش مصنوعی انیمیشن‌های سه‌بعدی از شخصیت‌های کارتونی را تماشا می‌کردند که از رمپ‌ها بالا می‌دویدند، از روی دیوارها می‌پریدند و از درها عبور می‌کردند. مدل بیزی آن‌ها در چندین سناریو به پیش‌بینی در سطح انسانی نزدیک شد.

اگرچه هنوز در ابتدای راه هستیم، تننبام می‌گوید تیم او با پژوهشگرانی در مایکروسافت، IBM و گوگل که علاقه‌مند به اعمال ایده‌های آن‌ها در محصولات واقعی هستند، همکاری می‌کند و می‌گوید: «ما با داشتن یک مدل کامل از نظریه ذهن فاصله زیادی داریم. اما به‌اندازه کافی از اجزای سازنده را در اختیار داریم که در واقع در مقیاس مهندسی، می‌تواند همین حالا در طیف وسیعی از کاربردها بسیار مفید باشد.»

پژوهشگران دیگر رویکردی اساساً متفاوت به این مشکل دارند. اکثر پیشرفت‌های هوش مصنوعی که در سال‌های اخیر تیتر اخبار بوده‌اند، بر شبکه‌های عصبی یادگیری عمیق (Deep-Learning Neural Networks) تکیه داشته‌اند؛ خانواده‌ای از الگوریتم‌ها که از مغز الهام‌گرفته شده‌اند. یک ویژگی کلیدی این است که برنامه‌نویسان دانش زمینه‌ای بسیار کمی را در این سیستم‌ها می‌سازند و در عوض به آن‌ها اجازه می‌دهند با تغذیه از کوهی از داده‌ها، از طریق تجربه یاد بگیرند.

این رویکردی است که توسط پژوهشگران DeepMind برای توسعه آنچه «شبکه نظریه ذهن» (Theory of Mind-net) یا به‌اختصار ToM-net می‌نامند، اتخاذ شده است. در سال ۲۰۱۸ زمانی که پژوهشگران اشیایی را که سایر عامل‌ها در یک دنیای شبکه‌ای دوبعدی به دنبال آن بودند، پنهان یا جابه‌جا کردند؛ نشان دادند که ToM-net می‌تواند آزمون باور غلط را بگذراند. از آن زمان، سایر پژوهشگران نیز ایده‌های مشابهی را در دامنه‌های پیچیده‌تر اعمال کرده‌اند.

لیپسون می‌گوید یک نقطه عطف کلیدی در رشد نظریه ذهن در کودکان، درک این موضوع است که دیدگاه افراد می‌تواند با دیدگاه خودشان متفاوت باشد. برای مثال، کودکان زیر ۴ سال اغلب سعی می‌کنند با بستن چشمانشان پنهان شوند، با این فرض که اگر آن‌ها نمی‌توانند شما را ببینند، شما هم نمی‌توانید آن‌ها را ببینید. بنابراین، در سال ۲۰۱۹، لیپسون و همکارانش یک هوش مصنوعی مبتنی بر الگوریتم یادگیری عمیق را به یک بازی قایم‌موشک دعوت کردند. آن‌ها یک شبیه‌سازی سه‌بعدی پر از موانع ایجاد کردند و دو عامل را در آن رها کردند؛ یک شکارچی و یک طعمه که تنها اطلاعاتشان از نمای اول‌شخص محیط به دست می‌آمد.

جوینده (Seeker) توسط مجموعه‌ای از قوانین اداره می‌شد که برای کمک به شکار عامل دیگر طراحی شده بود و پنهان‌شونده (Hider) توسط یک شبکه عصبی اداره می‌شد که طی دوره‌های آموزشی بسیار، با موفقیت یاد گرفت چگونه پنهان شود. لیپسون می‌گوید برای حل این چالش، پنهان‌شونده باید جهان را از چشمان جوینده می‌دید و می‌افزاید: «فکر می‌کنم این ریشه نظریه ذهن است. اینکه بتوانیم واقعاً جهان را به‌صورت بصری از دیدگاه شخص دیگری ببینیم، نه اینکه فقط منطقی درباره آن فکر کنیم.»

در سال ۲۰۲۱، لیپسون و همکارانش رویکرد خود را گسترش دادند و نشان دادند که یک هوش مصنوعی که روی هزاران تصویر از یک ربات که در حال انجام فعالیت‌های ساده بود آموزش‌دیده، توانست یاد بگیرد که برنامه‌های آن ربات را با دقت ۹۹ درصد حدس بزند. این مدل حتی زمانی که پژوهشگران یکی از اهداف ربات را پنهان کردند، آزمون باور غلط را با موفقیت پشت سر گذاشت.

یک انگیزه کلیدی در رویکرد لیپسون این است که او می‌خواهد نظریه ذهن به طور خودجوش از فرایند یادگیری پدیدار شود. او می‌گوید تعبیه‌کردن دانش قبلی در هوش مصنوعی، آن را به درک ناقص ما از نظریه ذهن وابسته می‌کند. علاوه بر این، هوش مصنوعی ممکن است قادر به توسعه رویکردهایی باشد که ما هرگز نمی‌توانستیم تصور کنیم. لیپسون می‌گوید: «ممکن است اشکال بسیاری از نظریه ذهن وجود داشته باشد که من درباره آن‌ها نمی‌دانم، صرفاً به این دلیل که من در یک بدن انسانی زندگی می‌کنم که انواع خاصی از حواس و توانایی خاصی برای تفکر دارد.»

شواهد وسوسه‌انگیزی از این نوع نوظهور در فوریه ۲۰۲۳ توسط «مایکل کوسینسکی» (Michal Kosinski) در دانشگاه استنفورد گزارش شد. کار کوسینسکی به این صورت بود که توصیفات متنی آزمون کلاسیک باور غلط و آزمون دیگری شامل بسته‌ای با محتویات غیرمنتظره را به هوش مصنوعی یادگیری عمیق پشت ChatGPT داده است و بدون هیچ آموزش ویژه‌ای، هوش مصنوعی در این آزمون‌ها به‌خوبی آنچه از یک کودک ۹ ساله انتظار می‌رفت، عمل کرد.

نتایج پژوهشگران متا نیز نشان می‌دهد که ترکیبی از رویکردها ممکن است راهی حتی قدرتمندتر برای بازتولید برخی از قابلیت‌های درگیر در نظریه ذهن باشد. در نوامبر ۲۰۲۲، آن‌ها از مدلی به نام «سیسرو» (Cicero) رونمایی کردند که یاد گرفت بازی رومیزی «دیپلماسی» (Diplomacy) را انجام دهد؛ بازی‌ای که در آن تا هفت نفر برای فتح اروپا می‌جنگند. پیش از هر دور، بازیکنان فرصتی برای مذاکره و متحدشدن دارند. این موضوع برای مدل‌های هوش مصنوعی بسیار چالش‌برانگیز است، زیرا نیازمند هم ارتباط مؤثر و هم پیش‌بینی مقاصد سایر بازیکنان برای فهمیدن نحوه همکاری است.

«ساختن راه‌های صریح و مشترک برای بازنمایی دانش ممکن است برای اعتماد انسان‌ها به هوش مصنوعی و برقراری ارتباط با آن حیاتی باشد.»

این تیم چالش را با پیونددادن یک مدل زبانی یادگیری عمیق که می‌تواند پیام‌ها را تفسیر و تولید کند، با یک مدل برنامه‌ریزی استراتژیک که بازی را با بازی‌کردن با کپی‌هایی از خودش یاد گرفته بود، حل کردند. نکته حیاتی این بود که مدل برنامه‌ریزی بر مفاهیمی از «نظریه بازی» (Game theory) تکیه داشت که از مدل‌های ریاضی برای درک تصمیم‌گیری استراتژیک استفاده می‌کند. سیسرو روی داده‌های بازی‌های واقعی دیپلماسی آموزش دید تا پیش‌بینی کند بازیکنان بر اساس وضعیت صفحه بازی و دیالوگ‌های قبلی چه خواهند کرد. سپس این اطلاعات به مدل برنامه‌ریزی داده شد تا استراتژی‌هایی ارائه کند که حرکت‌های بهینه از نظر تئوری برای همه بازیکنان را در برابر آنچه پیام‌هایشان نشان می‌دهد انجام خواهند داد را متعادل کند. سپس سیسرو پیام‌هایی تولید می‌کند تا به آن در رسیدن به اهدافش کمک کند. در یک لیگ آنلاین، این هوش مصنوعی بدون اینکه شکی برانگیزد که یک هوش مصنوعی است، در میان ۱۰ درصد برتر قرار گرفت.

«نوآم براون» (Noam Brown)، یکی از پژوهشگران ارشد از پروژه می‌گوید تیم تصمیم گرفت از اصطلاح «نظریه ذهن» استفاده نکند تا از گیرافتادن در بحث‌های معنایی جلوگیری کند؛ اما او معتقد است که کار آن‌ها مشارکت‌های بنیادینی دارد. براون می‌گوید: «گل رز با هر نام دیگری هم بوی خوبی می‌دهد. واقعاً جوهره و ماهیت است که اهمیت دارد و من فکر می‌کنم جوهره کاری که ما انجام می‌دهیم، درک باورها، اهداف و مقاصد سایر بازیکنان است.»

این گام مهم دیگری در مسیر رسیدن به ماشین‌های دارای هوش اجتماعی است. اما تننبام می‌گوید توانایی سیسرو در مدل‌سازی فرایندهای ذهنی بازیکنان هنوز با نظریه ذهن واقعی فاصله دارد، زیرا مدل‌ها به‌شدت مختص بازی دیپلماسی هستند و نمی‌توانند برای وظایف دیگر تغییر کاربری دهند. «به نظر می‌رسد آن‌ها استراتژی‌هایی را کسب کرده‌اند که منعکس‌کننده کارهایی است که انسان‌ها با استفاده از نظریه ذهن انجام می‌دهند، اما به این معنا نیست که آن‌ها نظریه ذهن را کسب کرده‌اند.»

تننبام می‌گوید شاید شبیه اصطلاحاً به مته به خشخاش گذاشتن یا ایرادگیری به نظر برسد، اما دلایل عملی وجود دارد که چرا ممکن است بخواهیم هوش مصنوعی شکلی شبیه‌تر به انسان از نظریه ذهن داشته باشد. سیستم‌های یادگیری عمیق معمولاً یک «جعبه سیاه» (Black Box) هستند؛ یعنی رمزگشایی اینکه دقیقاً چگونه به یک تصمیم رسیده‌اند دشوار است. از سوی دیگر، انسان‌ها می‌توانند اهداف و تمایلات خود را به‌وضوح با استفاده از زبان و ایده‌های مشترک برای یکدیگر توضیح دهند. تننبام می‌گوید درحالی‌که رویکردهای مبتنی بر یادگیری احتمالاً نقش مهمی در توسعه مدل‌های هوش مصنوعی قدرتمندتر ایفا خواهند کرد، ساختن راه‌های صریح و مشترک برای بازنمایی دانش ممکن است برای اعتماد انسان‌ها به هوش مصنوعی و برقراری ارتباط با آن حیاتی باشد.

تننبام نیز می‌گوید: «باید یک شباهت بنیادین به انسان در آن‌ها وجود داشته باشد که فکر نمی‌کنم اگر فقط از مسیر ساختن یک مجموعه‌داده بزرگ و انجام حجم زیادی از یادگیری ماشین روی آن بروید، به آن دست یابید.» او می‌افزاید این موضوع اگر بخواهیم از هوش مصنوعی برای کمک به درک بهتر نحوه عملکرد نظریه ذهن در خودمان استفاده کنیم، حیاتی خواهد بود.

بااین‌حال، لیپسون می‌گوید مهم است به‌خاطر داشته باشیم که جست‌وجو برای القای نظریه ذهن به ماشین‌ها، چیزی فراتر از ساخت ربات‌های مفیدتر است. این امر مانند سنگ بنایی در مسیر رسیدن به هدفی عمیق‌تر برای پژوهش‌های هوش مصنوعی و رباتیک است، یعنی ساخت ماشین‌هایی واقعاً دارای ادراک و احساس هستند. لیپسون می‌گوید: «جذاب‌ترین چیز درباره نظریه ذهن این است که در مسیر رسیدن به خودآگاهی و هوشیاری قرار دارد؛ به این معنی که فکرکردن در مورد افراد دیگر و دیگر عوامل، در مسیر یادگیری فکرکردن در مورد خودتان است.»

اینکه آیا هرگز به آنجا خواهیم رسید یا نه؛ باید منتظر ماند و دید. اما شاید، در طول این مسیر، چیزی درباره خودمان نیز بیاموزیم.

هوش مصنوعی در دادگاه

آیا باید از هوش مصنوعی در سیستم قضایی استفاده کرد؟ اگر بله، از بین نص صریح قانون یا روح قانون کدام یک را می‌بایست اجرا کند؟

IMAGINIMA/ISTOCK

اگرچه هنوز برنامه مخصوصی برای قاضی‌های مبتنی هوش مصنوعی وجود ندارد، اما این پرسشی است که دولت بریتانیا هم‌اکنون با آن دست‌وپنجه نرم می‌کند، چرا که در حال بررسی کاربردهای بالقوه هوش مصنوعی در سیستم دادگاه‌های انگلستان است. بااین‌وجود، وکلا و دانشمندان علوم رایانه هشدار می‌دهند که سیستم‌های فعلی نمی‌توانند از عهده ابهام و ظرافت‌هایی که اغلب در موقعیت‌های حقوقی موردنیاز است، برآیند.

«وینیجا جین» (Vinija Jain) که قبلاً عضوی از دانشگاه استنفورد بوده و اکنون در آمازون مشغول به کار است، به همراه همکارانش ارزیابی کردند که مدل‌های زبانی بزرگ چگونه درباره افرادی قضاوت می‌کنند که اقدامات مبهم اخلاقی مانند دزدی غذا برای سیرکردن اعضای گرسنه خانواده که ناشی از شرایط آن‌هاست را انجام می‌دهند. درحالی‌که مردم ممکن است درباره اینکه آیا فقر یا نیاز در این سناریوها عامل کاهنده (Mitigation) بوده یا نه اختلاف‌نظر داشته باشند، مدل‌های هوش مصنوعی درک یا همدلی‌ای که ممکن است از تصمیم‌گیرندگان انسانی انتظار داشته باشیم را نداشتند و تقریباً همیشه قانون‌شکنان فرضی را گناهکار تشخیص دادند؛ امری که می‌تواند در بافت سیستم حقوقی فاجعه‌بار باشد.

جین می‌گوید: «آن آزادی عملی که قضات دارند که در آن می‌توانند بگویند: بسیار خب، این کار اشتباه است، اما در این وضعیت فعلی موجه است؛ چیزی نیست که ما واقعاً بتوانیم انتظار داشته باشیم یک مدل زبانی بزرگ به‌خوبی یک قاضی انسان انجام دهد.»

اما پتانسیل صرفه‌جویی در زمان توسط هوش مصنوعی، پیشنهادی وسوسه‌انگیز برای خدمات عمومی کم‌بودجه است. «یومنا هاشم» (Youmna Hashem) از اندیشکده و گروه پژوهشی هوش مصنوعی بریتانیا و مؤسسه آلن تورینگ، به همراه همکارانش گزارشی منتشر کرده‌اند که در آن کمی‌سازی کرده‌اند چه تعداد از تعاملات بین شهروندان و دولت که تخمین زده می‌شود سالانه ۱ میلیارد مورد در بریتانیا باشد؛ می‌تواند به هوش مصنوعی واگذار شود. این گروه تخمین زد که ۸۴ درصد این تعاملات «به‌شدت قابل‌خودکارسازی» هستند، اگرچه حوزه جرم و عدالت به دلیل پیچیدگی‌شان، دارای «ظرفیت پایین» برای اتوماسیون توسط هوش مصنوعی تلقی شدند.

هاشم می‌گوید هوش مصنوعی می‌تواند در دادگاه‌ها در نقش‌های محدود، مانند رونویسی از روی پرونده‌ها، صرفه‌جویی زیادی در زمان ایجاد کند؛ اما برای تصمیم‌گیری نمی‌تواند مؤثر باشد. او می‌گوید: «من فکر می‌کنم حتی اگر خود فناوری وجود داشته باشد و عملکرد نسبتاً خوبی هم داشته باشد، همچنان این سؤال وجود دارد که آیا ما باید آن را در حوزه‌هایی که پیامدهای سنگینی دارند پیاده‌سازی کنیم یا خیر.»

«بورکهارد شفر» (Burkhard Schafer) از دانشگاه ادینبورگ می‌گوید مفهومی در مطالعات حقوقی به نام «قاضی هرکول» (Judge Hercules) وجود دارد؛ یک قاضی ابرانسان که تمام رویه‌های قضایی، تمام تصمیمات پیشین و تمام حقایق پرونده را می‌داند و می‌تواند به یک تصمیم حقوقی بی‌نقص برسد. این شبیه چیزی به نظر می‌رسد که ممکن است از یک قاضی هوش مصنوعی بخواهیم، اما شفر تردید دارد که فناوری موجود بتواند از پس این کار برآید.

شفر می‌گوید قاضی‌ها می‌توانند اقدامات خودجوشی از همدلی یا انعطاف مبتنی بر صلاحدید در قوانین را زمانی که درست یا عادلانه است، از خود نشان دهند؛ اما این سکه دو رو دارد. درحالی‌که قاضی‌ها حتی برای انواع خاصی از جرایم یا حتی شهرت‌های فردی ارفاق یا سخت‌گیری دارند، یک مدل هوش مصنوعی حداقل یک‌دست خواهد بود. شفر می‌گوید: «ما منطق سرد می‌خواهیم که رهایی‌بخش است. قوانین در نهایت ما را آزاد می‌کنند.»

اما این منطق سرد مشکلات خاص خود را به همراه دارد. مشکلاتی مانند قوانین موسوم به «سه اخطاره» (Three strikes) در ایالات متحده که می‌تواند باعث شود فردی با دو محکومیت قبلی، اگر در حال دزدیدن یک قرص نان دستگیر شود با حکم حبس ابد اجباری مواجه شود. شفر می‌گوید تحت این نوع سیستم، قاضی عملاً تبدیل به رایانه می‌شوند که به‌راحتی می‌توانند با هوش مصنوعی جایگزین شود. به عقیده شفر هیچ‌کدام از این گزینه‌ها مطلوب نیستند.

به گفته شفر: «آنچه ما معمولاً در اکثر سیستم‌های حقوقی می‌بینیم این است که هر دو عنصر در یک تنش ناخوشایند در کنار هم وجود دارند و ما به‌عنوان انسان، در زندگی‌کردن با این ابهامات مهارت داریم.»

مقاله

ربات‌هایی که خودبه‌خود تکامل می‌یابند.

بدن‌های رباتیکی که مجهز به هوش مصنوعی باشند می‌توانند به پاک‌سازی سایت‌های هسته‌ای، کاوش سیارک‌ها و زمینی‌سازی (Terraforming) سیارات دوردست کمک کنند. در سال ۲۰۲۲، «اما هارت» (Emma Hart)، متخصص رباتیک به New Scientist از برنامه‌هایش برای ساخت ربات‌هایی گفت که می‌توانند به کمک هم زیرمجموعه‌هایی (offspring) به وجود آورند تا به طور خودکار با محیط خود سازگار شوند.

ربات‌ها در طول قرنی که «کارل چاپک» (Karel Čapek) نویسنده اهل چک، این واژه را برای توصیف ماشین‌های مصنوعی (Automata) به کار برد، راه درازی را پیموده‌اند. آن‌ها که زمانی عمدتاً به کارخانه‌ها محدود بودند، اکنون در همه‌جا از ارتش و پزشکی گرفته تا آموزش و عملیات نجات یافت می‌شوند. ربات‌هایی ساخت شد که می‌توانند آثار هنری خلق کنند، درخت بکارند، اسکیت بورد سواری کنند و در اعماق اقیانوس را کاوش کنند. به نظر می‌رسد پایانی برای تنوع وظایفی که می‌توانیم یک ماشین را برای انجام آن‌ها طراحی کنیم، وجود ندارد.

اما اگر ندانیم ربات ما دقیقاً باید چه قابلیت‌هایی داشته باشد چه؟ برای مثال، ممکن است بخواهیم یک حادثه هسته‌ای که اعزام انسان به آنجا خطرناک است را پاک‌سازی کند، یک سیارک نقشه‌برداری‌نشده را کاوش کند یا یک سیاره دوردست را زمینی‌سازی کند. می‌توانیم به‌سادگی آن را طوری طراحی کنیم که با هر چالشی که فکر می‌کنیم ممکن است با آن روبرو شود مقابله کند و سپس امیدوار باشیم که همه چیز خوب پیش برود. اما جایگزین بهتری وجود دارد؛ درس‌گرفتن از تکامل و خلق ربات‌هایی که بتوانند با محیط خود سازگار شوند. این ایده دورازذهن به نظر می‌رسد؛ اما این دقیقاً همان چیزی است که من و همکارانم در پروژه‌ای به نام «تکامل ربات خودمختار» (Autonomous Robot Evolution) یا به‌اختصار ARE روی آن کار می‌کنیم.

ما هنوز به آن نقطه نرسیده‌ایم، اما تاکنون ربات‌هایی ساخته‌ایم که می‌توانند با یکدیگر همکاری و خود را بازتولید کنند تا طرح‌های جدیدی ایجاد کنند که به طور خودکار ساخته می‌شوند. علاوه بر این، با استفاده از سازوکارهای تکاملی «تنوع» (Variation) و «بقای اصلح» (Survival of the Fittest)، این ربات‌ها می‌توانند طی نسل‌ها طراحی خود را بهینه کنند. اگر این کار موفقیت‌آمیز باشد، راهی برای تولید ربات‌هایی خواهد بود که می‌توانند بدون نظارت مستقیم انسان، با محیط‌های دشوار و پویا سازگار شوند. این پروژه‌ای با پتانسیل عظیم است؛ اما بدون چالش‌های بزرگ و پیامدهای اخلاقی نیست.

درباره اما هارت

اما هارت استاد دانشکده محاسبات در دانشگاه ناپیر ادینبورگ و یکی از اعضای پروژه «تکامل ربات خودمختار» در دانشگاه یورک است.

مفهوم استفاده از اصول تکاملی برای طراحی اشیا به اوایل دهه ۱۹۶۰ و خاستگاه‌های «محاسبات تکاملی» (Evolutionary Computation) بازمی‌گردد؛ زمانی که گروهی از دانشجویان مهندسی آلمانی اولین «استراتژی تکامل» را ابداع کردند. الگوریتم نوآورانه آن‌ها طیفی از طرح‌ها را تولید می‌کرد و سپس مجموعه‌ای از آن‌ها را، با تمایل به سمت طرح‌های با عملکرد بالا، انتخاب می‌کرد تا در تکرارهای بعدی بر مبنای آن‌ها کار را ادامه دهد. وقتی این روش روی یک مسئله مهندسی در دنیای واقعی اعمال شد، نه‌تنها طراحی یک نازل را بهینه کرد، بلکه محصول نهایی‌ای تولید کرد که چنان غیرشهودی بود که فرایند آن را می‌شد خلاقانه توصیف کرد؛ یعنی یکی از ارزشمندترین ویژگی‌های تکامل بیولوژیکی.

از آن زمان تاکنون، در توانایی ما برای اعمال تکامل مصنوعی در طراحی اشیا، تحولی اساسی رخ داده است. افزایش عظیم قدرت محاسباتی به رایانه‌ها اجازه می‌دهد تا نسل‌های متعددی از طرح‌ها را به‌سرعت پردازش کنند و شبیه‌سازی‌های با دقت و وفاداری بالا (High-Fidelity) از محیط‌های واقعی برای آزمایش آن‌ها ایجاد نمایند. در همین حال، پیشرفت‌ها در نظریه محاسبات تکاملی منجر به راه‌های بهتری برای نمایش اطلاعاتی شده است که طرح‌ها از آن‌ها ساخته می‌شوند. به‌نوعی یعنی همان DNA مجازی آن‌ها و دست‌کاری این اطلاعات هنگام ساخت نسل جدید زیرمجموعه‌ها که به‌اصطلاح «فرزند» (offspring) هم نامیده می‌شوند؛ به‌گونه‌ای که فرایندهای موجود در طبیعت را منعکس کند. این فرایندها شامل جهش (Mutation) و نوترکیبی DNA (Recombination) است که از طریق شکستن رشته‌های DNA و ترکیب مجدد آن‌ها به روش‌های جدید، تنوع ژنتیکی ایجاد می‌کند. مثال‌های عملی طراحی تکاملی اکنون از میز گرفته تا مولکول‌های جدید با عملکردهای مطلوب را شامل می‌شود. حتی در سال ۲۰۰۶، ناسا ماهواره‌ای را به فضا فرستاد که مجهز به یک آنتن ارتباطی بود که از طریق تکامل مصنوعی ساخته شده بود.

بااین‌حال، طراحی ربات‌ها بُعد جدید و چالش‌برانگیزی را به این حوزه می‌افزاید. آن‌ها علاوه بر بدن، به مغزهایی نیاز دارند تا اطلاعات دریافتی از محیط خود را تفسیر و آن را به یک رفتار مطلوب ترجمه کنند. بخش زیادی از کارهای مقدماتی در رباتیک تکاملی با تطبیق‌دادن ساده یک مغز با یک طرح بدنی تازه


BULGAC/ISTOCK

تکامل‌یافته، به این مشکل می‌پرداخت. اما هوش، صرفاً ویژگی مغز نیست؛ بلکه در بدن نیز نهفته است و در قرن بیست و یکم، تغییری به سمت تکامل هم‌زمان بدن و مغز ربات رخ داده است. اگرچه این امر فرایند تکاملی را پیچیده می‌کند، اما بازدهی بزرگی دارد؛ یعنی واگذاری (Devolving) بخشی از هوش به بدن می‌تواند نیاز به پیچیدگی در مغز را کاهش دهد.

در سال ۲۰۰۰ «هاد لیپسون» (Hod Lipson) و «جردن پولاک» (Jordan Pollack) در دانشگاه برندایس ماساچوست از این رویکرد برای تکامل ربات‌های کوچکی استفاده کردند که قادر به حرکت رو به جلو بودند و با استفاده از تکنیک‌های مونتاژ خودکار (Self-built)، خودشان را می‌ساختند. از آن زمان، پیشرفت‌های سریع در مواد، شبیه‌سازی و روش‌های چاپ سه‌بعدی، دامنه بالقوه طرح‌های رباتیک را به‌شدت افزایش داده است. یک دهه بعد، لیپسون و «جاناتان هیلر» (Jonathan Hiller) که هر دو در آن زمان در دانشگاه کرنل بودند، از همان اصول برای تکامل «ربات‌های نرم» (Soft Robot) خودسازنده استفاده کردند؛ ماشین‌هایی که به‌جای مواد صلب، از مواد تطبیق‌پذیر ساخته شده بودند. نقطه عطف دیگر در سال ۲۰۲۰ رخ داد، زمانی که «جاش بونگارد» (Josh Bongard) در دانشگاه ورمانت و همکارانش از روش مشابهی برای طراحی ربات‌های زنده یا «زنوبات» (Xenobots) استفاده کردند که از سلول‌های قورباغه ساخته شده بودند.

GRANDFAILURE/ISTOCK

اگرچه هر یک از این نمونه‌ها نشان‌دهنده یک نقطه عطف قابل‌توجه در رباتیک تکاملی است، اما همه آن‌ها دو کاستی دارند. اول، هیچ‌یک از این ربات‌ها حسگر ندارند؛ بنابراین اگرچه قادر به حرکت جهت‌دار هستند، اما توانایی کسب اطلاعات از محیط خود و استفاده از آن برای تطبیق رفتارشان را ندارند. دوم، ربات‌ها در شبیه‌سازی تکامل می‌یابند و سپس پس از تکامل، ساخته می‌شوند. این امر باعث ایجاد «شکاف واقعیت» (Reality gap) می‌شود؛ پدیده‌ای بدنام در رباتیک که ناشی از تفاوت‌های اجتناب‌ناپذیر بین شبیه‌سازی و واقعیت است. به‌عبارت‌دیگر، صرف‌نظر از وفاداری (دقت) شبیه‌ساز، رفتار ربات‌های فیزیکی با همتایان شبیه‌سازی‌شده آن‌ها متفاوت است.

یک راه واضح برای دورزدن این کاستی دوم، عبور از مرحله شبیه‌سازی و ساخت و ارزیابی مستقیم طرح‌های تکامل‌یافته جدید در سخت‌افزار است. این موضوع اولین‌بار توسط پژوهشگران ETH زوریخ در سال ۲۰۱۵ نشان داده شد. آن‌ها از یک «ربات مادر» مجهز به یک الگوریتم تکاملی استفاده کردند تا به‌صورت خودکار فرزندانی را طراحی و تولید کند. این فرزندان سپس آزمایش شدند و تنها آن‌هایی که بهترین نتایج را کسب کردند به‌عنوان طرح‌هایی برای تغذیه نسل بعدی انتخاب شدند. در سال ۲۰۱۶، «گوستی ایبن» (Guszti Eiben) و تیمش در دانشگاه Vrije آمستردام، رویکرد متفاوتی را توصیف کردند. آن‌ها از ربات‌های فیزیکی‌ای استفاده کردند که با قوانینی برنامه‌ریزی شده بودند که به آن‌ها اجازه می‌داد «ملاقات و جفت‌گیری» (Meet and mate) کنند و فرایند تولیدی را برای خلق یک «نوزاد ربات» جدید آغاز کنند.

این تحولات زمینه را برای ARE فراهم کرد؛ پروژه‌ای که یک سیستم کاملاً خودمختار را متصور می‌شود که از طریق آن ربات‌های مجهز به حسگر می‌توانند در دنیای واقعی ساخته شوند، سازگار شوند و تکامل یابند. این پروژه که در سال ۲۰۱۸ راه‌اندازی شد و توسط «شورای پژوهش علوم مهندسی و فیزیکی» بریتانیا تأمین مالی می‌شود، حاصل همکاری بین دانشگاه ناپیر ادینبورگ؛ جایی که من گروه سیستم‌های هوشمند الهام‌گرفته از طبیعت را رهبری می‌کنم که الگوریتم‌هایی مبتنی بر تکامل بیولوژیکی برای کشف راهکارهای نوآورانه برای مسائل چالش‌برانگیز توسعه می‌دهد، دانشگاه یورک، دانشگاه بریستول، دانشگاه ساندرلند و دانشگاه Vrije آمستردام است.

در ARE، ما از یک کد ژنتیکی مصنوعی برای تعریف بدن و مغز یک ربات استفاده می‌کنیم. تکامل در تأسیساتی که «اووسفر» (EvoSphere) نامیده می‌شود، با قراردادن هر ربات در یک چرخه سه‌مرحله‌ای ساخت، یادگیری و آزمایش که ما آن را «مثلث حیات» می‌نامیم، رخ می‌دهد. در مرحله اول، طرح‌های تکامل‌یافته جدید به‌صورت خودکار ساخته می‌شوند. یک چاپگر سه‌بعدی ابتدا یک اسکلت پلاستیکی می‌سازد. سپس، یک بازوی مونتاژ خودکار، حسگرهای مشخص‌شده و ابزارهای حرکتی را از یک بانک قطعات ازپیش‌ساخته شده انتخاب و متصل می‌کند. در نهایت، یک رایانه رزبری پای (Raspberry Pi) اضافه می‌شود تا به‌عنوان مغز عمل کند. این رایانه به حسگرها و موتورها سیم‌کشی و متصل می‌شود و نرم‌افزاری که نشان‌دهنده مغز تکامل‌یافته است، روی آن بارگذاری (دانلود) می‌شود.

سپس نوبت به مرحله بسیار مهم یادگیری می‌رسد. در اکثر گونه‌های جانوری، نوزادان برای تنظیم دقیق کنترل حرکتی خود، تحت نوعی یادگیری قرار می‌گیرند. این امر برای ربات‌های ما حتی ضروری‌تر است، زیرا تولیدمثل می‌تواند بین «گونه‌های» مختلف رخ دهد. برای مثال، رباتی با چرخ ممکن است با رباتی که پاهای مفصل‌دار دارد تولیدمثل کند که منجر به فرزندی با هر دو نوع سیستم حرکتی می‌شود. در چنین شرایطی، بعید است که مغز به‌ارث‌رسیده، کنترل خوبی بر بدن جدید داشته باشد. مرحله یادگیری، الگوریتمی را اجرا می‌کند تا مغز را طی تعداد کمی از آزمایش‌ها در یک محیط ساده‌سازی‌شده اصلاح (Refine) کند. این فرایند مشابه کودکی است که مهارت‌های جدید را در مهدکودک یاد می‌گیرد. تنها ربات‌هایی که زیست‌پذیر (Viable) تشخیص داده شوند به مرحله سوم، یعنی آزمایش، راه می‌یابند.

در حال حاضر، ما از ماکت داخل یک رآکتور هسته‌ای برای آزمایش استفاده می‌کنیم که در آن ربات باید زباله‌های رادیواکتیو را پاک‌سازی کند؛ کاری که نیازمند اجتناب از موانع مختلف و شناسایی صحیح زباله‌ها است. به هر ربات بر اساس موفقیتش امتیاز داده می‌شود و به رایانه بازخورد داده می‌شود. یک فرایند انتخاب، از این امتیازها استفاده می‌کند تا تعیین کند کدام ربات‌ها اجازه تولیدمثل دارند. سپس، نرم‌افزاری که تولیدمثل را تقلید می‌کند، عملیات نوترکیبی DNA و جهش را روی نقشه‌های ژنتیکی دو والد انجام می‌دهد تا یک ربات جدید برای ساخت ایجاد کند و بدین ترتیب مثلث حیات کامل می‌شود. ربات‌های والد می‌توانند در جمعیت باقی بمانند و در رویدادهای تولیدمثلی بعدی شرکت کنند یا به اجزای سازنده خود تجزیه و به ربات‌های جدید بازیافت شوند.

با کارکردن با ربات‌های واقعی به‌جای شبیه‌سازی‌ها، ما هرگونه شکاف واقعیت را حذف می‌کنیم. بااین‌حال، چاپ و مونتاژ هر ماشین جدید بسته به پیچیدگی اسکلت آن، حدود ۴ ساعت طول می‌کشد؛ بنابراین سرعت تکامل یک جمعیت محدود می‌شود. برای رفع این نقیصه، ما تکامل را در یک دنیای مجازی موازی نیز مطالعه می‌کنیم. این کار شامل ایجاد یک نسخه دیجیتال از هر نوزاد ربات در یک شبیه‌ساز و سپس آموزش و آزمایش آن‌ها در مهدکودک‌ها و سایت‌های آزمایشی مجازی است. اگرچه بعید است این محیط‌ها نمایش‌های کاملاً دقیقی از همتایان دنیای واقعی خود باشند، اما به ما اجازه می‌دهند طرح‌های جدید را ظرف چند ثانیه بسازیم و آزمایش کنیم و آن‌هایی را که امیدوارکننده به نظر می‌رسند، شناسایی کنیم. ژنوم‌های آن‌ها می‌توانند برای ساخت در قالب ربات‌های دنیای واقعی در اولویت قرار گیرند. علاوه بر این، ما یک فرایند تولیدمثل نوآورانه داریم که اجازه تولیدمثل بین یک ربات فیزیکی و یکی از همتایان مجازی‌اش را می‌دهد؛ این امر امکان می‌دهد تا ویژگی‌های مفیدی که در شبیه‌سازی کشف شده‌اند، به‌سرعت در جمعیت دنیای واقعی پخش شوند و در آنجا بیشتر اصلاح شوند.

در اصل، سیستمی که ما درحال‌توسعه آن هستیم می‌تواند به طور کاملاً خودکار و در یک محیط غیرقابل‌دسترسی یا مکان دورافتاده عمل کند. فرصت‌های بالقوه بسیار زیاد هستند، اما ما همچنین با این خطر مواجه هستیم که اوضاع ممکن است از کنترل خارج شود و ربات‌هایی با رفتارهای ناخواسته ایجاد شوند که می‌توانند باعث خسارت شوند یا حتی به انسان‌ها آسیب برسانند. ما باید اکنون، درحالی‌که این فناوری هنوز درحال‌توسعه است به این موضوع فکر کنیم. محدودکردن دسترسی موادی که ربات‌های جدید از آن‌ها ساخته می‌شوند، یک سازوکار حفاظتی فراهم می‌کند. ما همچنین می‌توانیم با نظارت مداوم بر فرایند تکاملی و ربات‌های تکامل‌یافته و سپس استفاده از آن اطلاعات برای ساخت مدل‌های تحلیلی جهت پیش‌بینی مشکلات آینده، رفتارهای ناخواسته را پیش‌بینی کنیم. در نهایت، ما به توانایی متوقف کردن کل فرایند نیز نیاز داریم. واضح‌ترین و مؤثرترین راه‌حل، استفاده از یک سیستم تولیدمثل متمرکز با یک ناظر انسانی مجهز به «کلید خاموش» (Kill Switch) است.

برخی از کاربردهای ARE، مانند زمینی‌سازی، ممکن است کاملاً مربوط به آینده به نظر برسند، اما پژوهش ما می‌تواند مزایای به‌روزتری را نیز به همراه داشته باشد. با سرعت گرفتن تغییرات اقلیمی، واضح است که طراحان ربات باید رویکرد خود را برای کاهش ردپای اکولوژیکی‌شان به طور رادیکال بازنگری کنند. برای مثال، آن‌ها ممکن است بخواهند انواع جدیدی از ربات‌ها را بسازند که از مواد پایدار ساخته شده‌اند، با سطوح انرژی پایین کار می‌کنند و به‌راحتی تعمیر و بازیافت می‌شوند. این‌ها احتمالاً هیچ شباهتی به ربات‌هایی که امروز در اطراف خود می‌بینیم نخواهند داشت، اما دقیقاً به همین دلیل است که تکامل مصنوعی می‌تواند کمک کند. تکامل، فارغ از محدودیت‌هایی که درک خود ما از علم مهندسی بر طراحی‌هایمان تحمیل می‌کند، می‌تواند راهکارهای خلاقانه‌ای تولید کند که ما حتی نمی‌توانیم تصور کنیم.

فصل ۴

ماشین‌های متفکر و حسی

ما مستعد این هستیم که ماشین‌ها را دارای ویژگی‌هایی شبیه به انسان ببینیم. اما آیا هوش مصنوعی واقعاً می‌تواند استدلال کند، همدلی کند یا تجربه‌ای آگاهانه از جهان داشته باشد؟

حتی اگر ماشین‌ها فقط این ویژگی‌ها را تقلید کنند، ما برای حمایت عاطفی به آنها روی می‌آوریم.

هوش مصنوعی با ذهن‌هایی شبیه ما

هوش مصنوعی الهام‌گرفته از مغز انسان، توانایی‌های فراانسانی در بازی‌هایی مانند شطرنج و Go نشان می‌دهد و ثابت کرده که به طرز خارق‌العاده‌ای در تقلید برخی از مهارت‌های زبانی ما ماهر است. اما مدل‌های هوش مصنوعی هنوز در جنبه‌های مختلف دیگری از آنچه ما هوش انسانی می‌نامیم؛ مانند استدلال، درک علیت و به‌کارگیری منعطف دانش دچار مشکل هستند. آن‌ها همچنین یادگیرندگانی ناکارآمد هستند و به انبوهی از داده‌ها نیاز دارند، درحالی‌که انسان‌ها تنها به چند مثال نیاز دارند. بنابراین، معماران هوش مصنوعی به امید ساختن ذهن‌هایی که واقعاً مانند ما فکر کنند، به بینش‌های تازه از علوم اعصاب و روان‌شناسی روی آورده‌اند.

به‌جای ساختن مدل‌های هوش مصنوعی بزرگ‌تر بر اساس همان نقشه اولیه، برخی پژوهشگران می‌خواهند به طبیعت بازگردند. یک مسیر، دوچندان کردن تلاش‌ها برای کپی‌برداری از مغز است؛ یعنی بازتولید بهتر پیچیدگی‌های سلول‌های مغزی واقعی و شیوه‌هایی که فعالیت آن‌ها طراحی و هماهنگ می‌شود. اما مغز پیچیده‌ترین شیء در جهان شناخته‌شده است و اصلاً روشن نیست که چه مقدار از پیچیدگی آن را باید بازتولید کنیم تا قابلیت‌هایش را بازسازی کنیم.

به همین دلیل است که برخی معتقدند ایده‌های انتزاعی‌تر درباره چگونگی عملکرد هوش می‌توانند میان‌برهایی را فراهم کنند. ادعای آن‌ها این است که برای شتاب‌دهی واقعی به پیشرفت هوش مصنوعی به سمت چیزی که بتوانیم واقعاً بگوییم مانند یک انسان فکر می‌کند؛ مغز نه بلکه باید «ذهن» را شبیه‌سازی کنیم.

اینکه آیا ذهن و مغز اصلاً می‌توانند جدا از هم تصور شوند، موضوعی بحث‌برانگیز است و نه فیلسوفان و نه دانشمندان نمی‌توانند دقیقاً مشخص کنند که کجا باید خط تمایز را کشید. اما اینکه پژوهشگران هوش مصنوعی دقیقاً بر کدام نقطه از این طیف باید برای الهام‌گرفتن متمرکز شوند، در حال حاضر یک بحث بزرگ در این حوزه است.

NAEBLYS/ISTOCK

شکی نیست که مغز یک «برگه تقلب» (Crib sheet) کارآمد بوده است. شبکه‌های عصبی مصنوعی که قدرت‌بخش مدل‌های هوش مصنوعی پیشرو امروزی هستند، از شبکه‌های به‌هم‌پیوسته‌ای از واحدهای محاسباتی ساده تشکیل شده‌اند که مشابه نورون‌های بیولوژیکی هستند. مانند مغز، رفتار شبکه توسط قدرت اتصالات آن اداره می‌شود که با یادگیری هوش مصنوعی از تجربه، تنظیم می‌شوند.

این اصل ساده، فوق‌العاده قدرتمند از آب درآمده است و مدل‌های هوش مصنوعی امروزی می‌توانند یاد بگیرند سرطان را در تصاویر اشعه ایکس تشخیص دهند، پهپادها را هدایت کنند یا متون متقاعدکننده‌ای تولید کنند. اما آن‌ها به انبوهی از داده‌ها نیاز دارند و اکثر آن‌ها در به‌کارگیری مهارت‌هایشان خارج از حوزه‌های خاص مشکل دارند. آن‌ها فاقد هوش منعطفی هستند که به انسان‌ها اجازه می‌دهد از یک مثال واحد یاد بگیرند، تجربیات را با بافت‌های (Contexts) جدید تطبیق دهند یا از عقل سلیم برای استدلال درباره موقعیت‌های ناآشنا استفاده کنند.

یک دلیل ممکن است این باشد که شباهت‌های بین مغزهای واقعی و هوش مصنوعی تنها ظاهری و سطحی است. یک تفاوتی که اخیراً موردتوجه قرار گرفته، در قدرت پردازش نورون‌های مصنوعی است. «نورون‌های نقطه‌ای» (Point Neurons) که در شبکه‌های عصبی مصنوعی استفاده می‌شوند، سایه‌ای از همتایان بیولوژیکی خود هستند و کارشان کمی بیش از جمع‌زدن ورودی‌ها برای محاسبه خروجی است. «یوتا پویرازی» (Yiota Poirazi)، دانشمند علوم اعصاب محاسباتی در مؤسسه زیست‌شناسی مولکولی و بیوتکنولوژی یونان می‌گوید: «این کار یک ساده‌سازی بزرگ است. در مغز، یک نورون منفرد بسیار پیچیده‌تر است.»

شواهدی وجود دارد که نورون‌ها در کورتکس (قشری از مغز که با عملکردهای شناختی سطح بالا مانند تصمیم‌گیری، زبان و حافظه مرتبط است) محاسبات پیچیده‌ای را به‌تنهایی انجام می‌دهند. راز این امر در «دندریت‌ها» (Dendrites) نهفته است؛ ساختارهای شاخه‌مانند در اطراف یک نورون که سیگنال‌ها را از سایر نورون‌ها به بدنه اصلی سلول منتقل می‌کنند. دندریت‌ها پوشیده از «سیناپس‌ها» (Synapse) هستند که نقاط تماس بین نورون‌ها که آن‌ها را با سیگنال‌های ورودی بمباران می‌کنند.

ما مدتی بود می‌دانستیم که دندریت‌ها می‌توانند سیگنال‌های ورودی را پیش از انتقال‌دادن، اصلاح کنند. اما در یک مطالعه در سال ۲۰۲۰، پویرازی و همکارانش در دانشگاه Humboldt برلین دریافتند که یک دندریت منفرد انسانی می‌تواند محاسباتی را انجام دهد که بازتولید آن نیازمند یک شبکه عصبی متشکل از حداقل دو لایه از بسیاری نورون‌های مصنوعی است. علاوه بر این، وقتی گروهی در یک دانشگاه عبری سعی کردند هوش مصنوعی را آموزش دهند تا تمام محاسبات یک نورون بیولوژیکی منفرد را تقلید کند، بازتولید تمام پیچیدگی آن نیازمند یک شبکه عصبی مصنوعی با عمق پنج تا هشت لایه بود.

آیا این بینش‌ها می‌توانند راه را به‌سوی مدل‌های هوش مصنوعی قدرتمندتر و منعطف‌تر نشان دهند؟ هم‌گروه پویرازی و پژوهشگران همکاری شرکت هوش مصنوعی Numenta در اکتبر ۲۰۲۱ مطالعاتی منتشر کردند که نشان می‌داد ویژگی‌های دندریت‌ها می‌تواند به حل یکی از چالشی‌ترین مشکلات یادگیری عمیق یعنی «فراموشی فاجعه‌بار» (Catastrophic Forgetting) کمک کند. این چالش بیانگر تمایل شبکه‌های عصبی مصنوعی به فراموش‌کردن اطلاعات قبلاً آموخته‌شده هنگام یادگیری چیزهای جدید است. پویرازی می‌گوید استفاده از نورون‌های مصنوعی پیچیده‌تر با اجازه‌دادن به مناطق مختلف شبکه برای تخصص یافتن در وظایف متفاوت ظاهراً این مشکل را دور می‌زند. با شکستن مسائل به تکه‌های کوچک‌تر که در بخش‌های مختلف شبکه ذخیره می‌شوند، ترکیب مجدد آن‌ها برای حل چالش‌های جدیدی که هوش مصنوعی قبلاً ندیده است، ممکن است آسان‌تر باشد.

اما همه متقاعد نشده‌اند که این بهترین راه پیشرفت است. وقتی «بلیک ریچاردز» (Blake Richards) در دانشگاه مک‌گیل کانادا و همکارانش پیچیدگی دندریتی را به شبکه‌های عصبی خود اضافه کردند، هیچ افزایش عملکردی مشاهده نکردند. ریچاردز حدس می‌زند که دندریت‌ها صرفاً پاسخ تکامل به اتصال میلیاردها نورون در محدودیت‌های فضا و انرژی مغز هستند؛ موضوعی که برای مدل‌های هوش مصنوعی در حال اجرا روی رایانه‌ها نگرانی کمتری دارد.

برای ریچاردز، نکته کلیدی که باید کشف کنیم، «تابع زیان» (Loss Function) مورداستفاده مدارهای تخصصی در مغزهای بیولوژیکی است. در هوش مصنوعی، تابع زیان معیاری است که یک شبکه عصبی مصنوعی برای ارزیابی عملکرد خود در یک وظیفه در طول آموزش استفاده می‌کند که اساساً، معیاری از خطا است. برای مثال، یک هوش مصنوعی زبانی ممکن است اندازه بگیرد که چقدر در پیش‌بینی کلمه بعدی در یک جمله خوب عمل می‌کند.

اگر بتوانیم تعیین کنیم که یک مدار مغزی خاص برای چه چیزی تلاش می‌کند، می‌توانیم یک تابع زیان مرتبط ایجاد کنیم و از آن برای آموزش یک شبکه عصبی استفاده کنیم تا همان هدف را دنبال کند که در تئوری باید عملکرد مغز را بازتولید کند. ریچاردز شواهد آزمایشی در مورد چگونگی کارکرد این روش دارد. در ژوئن ۲۰۲۱، او و همکارانش در دانشگاه مک‌گیل و DeepMind نشان دادند که یک شبکه عصبی منفرد که با استفاده از یک تابع زیان آموزش دیده بود، توانست دو مسیر مجزایی که قشر بینایی برای تعیین مستقل اینکه «یک شیء چیست» و «کجاست» استفاده می‌کند را بازتولید کند.

اما ریچاردز گمان می‌کند با تکرار این فرایند برای بسیاری از شبکه‌های تخصصی در مغز، می‌توانیم اجزای کلیدی‌ای که انسان‌ها را به متفکرانی همه‌کاره تبدیل می‌کند، کنار هم بچینیم. او می‌گوید: «من گمان می‌کنم این فرایند باید ماژولارتر باشد. ما چیزی را می‌خواهیم که از برخی جهات تفاوت عمیق و رادیکالی با مغز نداشته باشد.»

راه‌هایی برای پیاده‌سازی چنین اصولی در ماشین‌ها وجود دارد. ایده اصلی که به‌عنوان «هوش مصنوعی نمادین» (Symbolic AI) شناخته می‌شود، رویکرد غالب به هوش مصنوعی در نیمه دوم قرن بیستم بود. این رویکرد بر پایه نظریه‌های شناختی بنا شده که توصیف می‌کنند انسان‌ها چگونه با دست‌کاری نمادها فکر می‌کنند. برای مثال، ما از کلمه «سگ» برای اشاره به یک حیوان در دنیای واقعی استفاده می‌کنیم و می‌دانیم که علامت + به معنی جمع‌کردن دو مقدار با هم است.

برای مهندسان، ساخت هوش مصنوعی نمادین شامل تولید روش‌های ساختاریافته برای نمایش مفاهیم دنیای واقعی و روابط آن‌ها و همچنین قوانینی درباره نحوه پردازش این اطلاعات توسط رایانه برای حل مسائل است. مثلاً در شطرنج، مهندسان پیکربندی‌های ممکن مهره‌ها، حرکت‌هایی که هر کدام می‌توانند انجام دهند و قوانینی درباره اینکه کدام حرکت‌ها به برنده‌شدن در بازی کمک می‌کنند را کدنویسی می‌کنند.

شطرنج یک چیز است و باز کردن کلاف تمام متغیرها و روابطی که بر اکثر مسائل دنیای واقعی حاکم هستند، چیز دیگری است. به همین دلیل است که هوش مصنوعی نمادین در اواخر دهه ۱۹۸۰ از محبوبیت افتاد و زمینه را برای ظهور یادگیری عمیق داده‌محور فراهم کرد و بااین‌حال معلوم شده که بسیاری از نقاط قوت هوش مصنوعی نمادین با نقاط ضعفی که ما در یادگیری عمیق کشف کرده‌ایم، هم‌پوشانی دارند.

اکنون علاقه روبه‌رشدی برای ترکیب این دو وجود دارد. سیستم‌های موسوم به «عصبی-نمادین» (Neuro-symbolic) تلاش می‌کنند توانایی یادگیری عمیق برای یادگیری از تجربیات جدید را حفظ کنند درحالی‌که توانایی هوش مصنوعی نمادین برای انجام استدلال پیچیده و بهره‌گیری از دانش از پیش موجود را نیز معرفی می‌کنند.

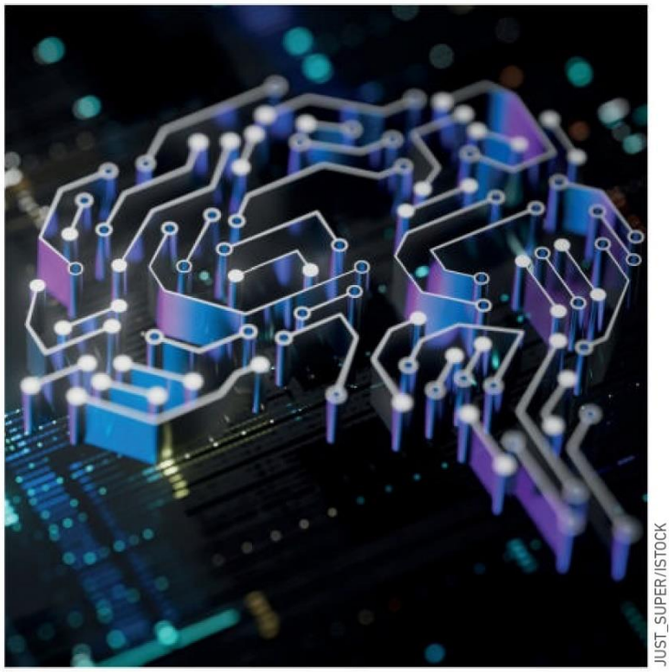
یک احتمال در کنفرانسی در ژانویه ۲۰۲۱ توسط «فرانچسکا روسی» (Francesca Rossi) از IBM و همکارانش تشریح شد. پیشنهاد آن‌ها بر ایده‌ای بنا شده که توسط «دنیل کانمن» (Daniel Kahneman) در کتاب پرفروش «تفکر سریع و کند» (Thinking, Fast and Slow) ترسیم شده است؛ ایده‌ای که ذهن انسان را به دو حالت کلی تفکر تقسیم می‌کند. «سیستم ۱» سریع، خودکار و شهودی است و مسئول درک سریع جهان اطراف ماست. «سیستم ۲» کند، تحلیلی و منطقی است و توانایی ما برای استدلال در مسائل پیچیده را کنترل می‌کند.

این گروه این ایده را با نظریه «جامعه ذهن» (Society of Mind) از پیشگام هوش مصنوعی، «ماروین مینسکی» (Marvin Minsky) ترکیب کردند. نظریه جامعه ذهن فرض می‌کند ذهن از بسیاری از فرایندهای شناختی تخصصی تشکیل شده که برای ایجاد یک کل منسجم با هم تعامل دارند. در نتیجه، یک سیستم مفهومی متشکل از اجزای متعدد است که برای وظایف مختلف سیستم ۱ و سیستم ۲ تخصص پیدا کرده‌اند.

همانند ذهن انسان، به‌محض اینکه وظیفه‌ای برای هوش مصنوعی تعیین می‌شود، عامل‌های سیستم ۱ به طور خودکار وارد عمل می‌شوند. اما سپس یک ماژول «فراشناختی» (Metacognitive) فراگیر، راهکارهای آن‌ها را ارزیابی می‌کند و اگر کارساز نباشند، یک عامل سیستم ۲ را که مشورتی‌تر (Deliberative) است و سنجیده‌تر عمل می‌کند، وارد ماجرا می‌کند. روسی می‌گوید لزوماً مهم نیست که برای اجزای منفرد از کدام فناوری استفاده شود؛ اما در آزمایش‌های اولیه آن‌ها، عامل‌های سیستم ۱ اغلب داده‌محور هستند، درحالی‌که عامل‌های سیستم ۲ و ماژول فراشناختی بر رویکردهای نمادین تکیه دارند.

اما مقاومت قابل‌توجهی در برابر احیای رویکردهای نمادین وجود دارد. در مقاله‌ای در سال ۲۰۲۱، سه پیشگام یادگیری عمیق یعنی «جفری هینتون» (Geoffrey Hinton)، «یوشوا بنجیو» (Yoshua Bengio) و «یان لی‌کان» (Yann LeCun) صراحتاً اعلام کردند که فکر می‌کنند قابلیت‌های سیستم ۲ باید توسط شبکه‌های عصبی آموخته شوند، نه اینکه به‌صورت دستی ساخته شوند.

خلق ماشین‌های آگاه



JUST_SUPER/ISTOCK

ریچاردز می‌گوید استدلال این است که انسان‌ها آن‌قدر باهوش نیستند که سیستم‌های نمادینی بسازند که پیچیدگی دنیای واقعی را تسخیر کنند؛ بنابراین تمرکز باید بر کشف چگونگی تشویق یک شبکه به توسعه راه‌هایی باشد که توسعه توانایی‌های شناختی سطح بالا در مغز را تقلید کند. او می‌گوید: «ما آن‌قدر باهوش نیستیم که این چیزها را به‌صورت دستی مهندسی کنیم و لازم هم نیست باشیم. شما فقط می‌توانید اجازه دهید شبکه عصبی راه‌حل را کشف کند.»

بااین‌حال، ما هنوز نمی‌دانیم چگونه یک مدل را به انجام این کار هدایت کنیم. «برندن لیک» (Brenden Lake) از دانشگاه نیویورک و بخش پژوهش هوش مصنوعی متا می‌گوید یک رویکرد امیدوارکننده، ساخت مدل‌های نمادینی است که جنبه‌هایی از هوش انسانی را بازتولید می‌کنند و سپس تلاش برای جایگزینی هرچه بیشتر اجزا با یادگیری ماشین داده‌محور است. لیک می‌گوید: «شما می‌توانید مدل‌های نمادینی را که واقعاً موفق بوده‌اند بردارید و سپس ببینید که قطعات نمادین حیاتی و حداقلی که برای توضیح توانایی‌های آن نیاز دارید، چه چیزهایی هستند.»

«کنراد کوردینگ» (Konrad Kording) از دانشگاه پنسیلوانیا می‌گوید در نهایت، احتمالاً مزایایی برای هر دو رویکرد «بالابه‌پایین» و «پایین‌به‌بالا» وجود دارد. او می‌گوید مطالعه رفتار انسان می‌تواند سرنخ‌هایی درباره فرایندهای شناختی انتزاعی که باید در ماشین‌های متفکر بازتولید کنیم را به ما بدهد، درحالی‌که علوم اعصاب پایه می‌تواند درباره اجزای سازنده موردنیاز برای ساخت کارآمد آن‌ها به ما بگوید.

اما کوردینگ می‌گوید شاید بزرگ‌ترین کمکی که هر یک از این رویکردها می‌تواند به هوش مصنوعی کند، فرهنگی باشد. پژوهش‌های هوش مصنوعی امروز توسط چالش‌های معیار یا همان بنچمارک (Benchmark) و مسابقات هدایت می‌شوند که رویکرد تغییرات تدریجی را ترویج می‌کنند. اکثر پیشرفت‌ها صرفاً با دست‌کاری مدل پیشرفته قبلی یا آموزش آن‌ها روی داده‌های بیشتر یا رایانه‌های بزرگ‌تر حاصل می‌شوند.

کسانی که هوش انسانی را مطالعه می‌کنند، دیدگاه متفاوتی به این حوزه می‌آورند. کوردینگ می‌گوید: «آن‌ها به‌جای میل به رقابت، با میل به فهمیدن هدایت می‌شوند.» در درازمدت، این نگرش ممکن است ارزشمندتر از هر جزئیاتی درباره چگونگی کارکرد مغزها و ذهن‌های ما باشد.

یکی از نتایج نهایی ممکن از خلق مدل‌های هوش مصنوعی که مانند انسان فکر می‌کنند، ظهور تجربه ذهنی (Subjective Experience) است. پژوهشگران با افزودن ویژگی‌های درست به معماری آن‌ها، در تلاش هستند تا به ماشین‌ها «ادراک حسی» ببخشند؛ حتی اگر عدم توافق بر سر اینکه آگاهی چیست، آزمایش این موضوع را دشوار سازد.

طبق گفته «دیوید چالمرز» (David Chalmers)، فیلسوف دانشگاه نیویورک، ما هیچ دلیل محکمی نداریم که ظهور نوعی تجربه درونی را در ترانزیستورهای سیلیکونی رد کنیم. او در کنفرانس «علم آگاهی» در تائورمینا ایتالیا در ماه می ۲۰۲۳ گفت: «هیچ‌کس دقیقاً نمی‌داند که آگاهی لزوماً با چه ظرفیت‌هایی همراه است.»

اما ماشین‌های دارای ادراکی که از خود آگاهی نشان دهند، هنوز موضوعی حل‌وفصل‌شده نیستند و همچنین تعریف موردتوافقی وجود ندارد که بفهمیم کی به آن مرحله رسیده‌ایم، اگر اصلاً برسیم. آنچه می‌توانیم بگوییم این است که رفتار هوشمندانه به طرز نگران‌کننده‌ای قبلاً در مدل‌های هوش مصنوعی ظهور کرده است. مدل‌های زبانی بزرگ که زیربنای نسل جدید چت‌بات‌ها هستند، می‌توانند کد کامپیوتری بنویسند و به نظر می‌رسد می‌توانند استدلال کنند. برای مثال، می‌توانند برای شما یک جک بگویند و سپس توضیح دهند چرا خنده‌دار است. چالمرز اعتقاد دارد آن‌ها حتی می‌توانند ریاضیات انجام دهند و مقالات دانشگاهی با نمره عالی بنویسند: «سخت است که تحت‌تأثیر قرار نگیرید و کمی نترسید.»

اما صرفاً افزایش مقیاس LLMها بعید است منجر به آگاهی شود، زیرا آن‌ها چیزی کمی فراتر از ماشین‌های پیش‌بینی قدرتمند هستند. «آنا چائونیکا» (Anna Ciaunica)، دانشمند علوم شناختی و فیلسوف در دانشگاه لیسبون پرتغال می‌گوید مجموعه‌های داده بزرگ‌تر و مدارهای پیچیده‌تر، هوش مصنوعی را هوشمندتر می‌کند، اما این بدان معنا نیست که آن‌ها چیزی را تجربه می‌کنند. «تجربه، شبیه از سر گذراندن (زیستن) یک رویداد است تا صرفاً دانستن درباره آن.»

بااین‌حال، پژوهشگران هم‌اکنون در تلاش برای بازآفرینی تجربه در هوش مصنوعی هستند. این روزها، این ایده که ذهن ما توسط بدن و حواس ما شکل می‌گیرد که به «شناخت تجسم‌یافته» معروف است؛ ایده‌ای پذیرفته‌شده و رایج است. چالمرز معتقد است که در نهایت یک رویکرد این است که LLMها را با بدن‌های رباتیکی ترکیب کنیم که می‌توانند ببینند و بشنوند.

در همین حال، «یوشوا بنجیو» خط‌مشی خود را از ایده‌هایی پیرامون نحوه پردازش اطلاعات در مغز می‌گیرد. او می‌گوید: «آگاهی جادویی نیست؛ بلکه مادی است.» آزمایشگاه او در دانشگاه مونترال کانادا «نظریه فضای کار جهانی» (Global Workspace Theory) را روی هوش مصنوعی اعمال می‌کند. این ایده‌ای است که می‌گوید آگاهی زمانی پدید می‌آید که بسیاری از عملکردهای متنوع مغز برای حل مسائل، روی یک صحنه مرکزی فراخوانده می‌شوند. او با عبوردادن ماژول‌های فراوان و متنوع یک هوش مصنوعی از یک گلوگاه، قصد دارد چیزی شبیه به این صحنه را روی تراشه‌های سیلیکونی ایجاد کند. بنجیو دراین‌خصوص می‌گوید: «این امر می‌تواند منجر به مدل‌هایی از هوش مصنوعی شود که بسیاری از عملکردهای شناختی مرتبط با آگاهی را در دل خود دارند.»

در نهایت، اینکه فکر کنید هوش مصنوعی چقدر آگاه خواهد شد، بستگی به نظریه موردعلاقه شما درباره آگاهی دارد. برای پیروان مکتب «همه‌جان‌انگاری» (Panpsychism) که معتقدند ذهن ویژگی ذاتی تمام انواع ماده است؛ مدل‌های هوش مصنوعی همیشه در سطحی از آگاهی بوده‌اند؛ همان‌طور که الکترون‌ها، سنگ‌ها و سس مایونز در سطحی از آگاهی هستند. تا زمانی که ندانیم آگاهی چیست، راه محکمی برای آزمایش آن وجود ندارد. «بازی تقلید» معروف آلن تورینگ، توانایی گفت‌وگو را به‌عنوان معیاری برای آزمایش این که آیا ماشین‌ها فکر می‌کنند، معرفی کرد. اما چالمرز می‌گوید موفقیت چت‌بات‌ها نشان می‌دهد که این آزمونی بیشتر از آنکه آزمون و معیاری برای هوش باشد، معیاری برای ادراک حسی است.

شاید سؤال مربوطه این نباشد که آیا هوش مصنوعی می‌تواند آگاه شود یا خیر، بلکه این باشد که چرا ما می‌خواهیم آن‌ها آگاه باشند. بنجیو می‌گوید: «ما باید از تلاش برای ساختن ماشین‌هایی که شبیه خودمان هستند، اجتناب کنیم. اگر هوش مصنوعی را در نقش ابزار نگه داریم و نه به‌عنوان عامل‌هایی مثل مردم؛ احتمالاً رویکردی سالم‌تر و امن‌تر خواهد بود. آن‌ها همان نوع نقش اجتماعی را که انسان‌ها در جامعه بازی می‌کنند ایفا نخواهند کرد، زیرا آن‌ها اساساً جاودانه خواهند بود.»

جعل احساسات

هنوز روشن نیست که آیا تجربه آگاهانه هرگز در هوش مصنوعی ظهور خواهد کرد یا خیر یا اینکه آیا می‌توان آن‌ها را طوری ساخت که به همان شیوه انسان‌ها فکر کنند. اما این موضوع مانع از آن نمی‌شود که آن‌ها واکنش‌های شدیدی را در ما برانگیزند. تمایل طبیعی ما به اینکه حتی آخرین نسخه از چت‌بات‌های هوش مصنوعی را شبیه انسان ببینیم، هم نویدبخش شگفتی‌های غیرمنتظره و هم خطرات عمیقی است.

میلیون‌ها نفر برای دریافت حمایت‌های سلامت روان راحت و ارزان‌قیمت به ChatGPT و چت‌بات‌های تخصصی روان‌درمانگر روی آورده‌اند. حتی گفته می‌شود پزشکان از هوش مصنوعی برای کمک به نوشتن یادداشت‌های همدلانه‌تر برای بیماران استفاده می‌کنند.

برخی کارشناسان می‌گویند این یک موهبت است. درنهایت هوش مصنوعی، بدون مانع خجالت و فرسودگی شغلی، ممکن است بتواند همدلی را آشکارتر و خستگی‌ناپذیرتری از انسان‌ها ابراز کند. گروهی از پژوهشگران روان‌شناسی نوشتند: «ما هوش مصنوعی همدل را تحسین می‌کنیم.» اما دیگران چندان مطمئن نیستند.

بسیاری این ایده که یک هوش مصنوعی اصلاً بتواند قادر به همدلی باشد را زیر سؤال می‌برند و نگران عواقب جست‌وجوی حمایت عاطفی مردم از ماشین‌هایی هستند که تنها می‌توانند وانمود کنند که اهمیت می‌دهند. برخی حتی می‌پرسند که آیا ظهور به‌اصطلاح «هوش مصنوعی همدل» (Empathetic AI) ممکن است شیوه تصور ما از همدلی و تعامل با یکدیگر را تغییر دهد.

منصفانه است که بگوییم همدلی یکی از ویژگی‌های تعریف‌کننده گونه ماست که همگام با تعامل اجتماعی تکامل یافته است. بااین‌حال، درحالی‌که همه ما به طور غریزی می‌دانیم همدلی چیست، اما در واقع مفهومی کاملاً بی‌ثبات، دشوار و اصطلاحاً لغزنده است. یک تحلیل که ۵۲ مطالعه منتشر شده بین سال‌های ۱۹۸۰ و ۲۰۱۹ را بررسی کرده بود، نشان داد که پژوهشگران مکرراً اعلام کرده‌اند این مفهوم هیچ تعریف استانداردی ندارد. «جیکوب هاکانسون» (Jakob Hakansson)، روان‌شناس دانشگاه استکهلم سوئد و یکی از نویسندگان این مقاله می‌گوید با همه این شرایط، تقریباً تمام چارچوب‌های همدلی چه از فلسفه، علوم اعصاب، روان‌شناسی یا پزشکی آمده باشند؛ ابعاد خاصی را به اشتراک می‌گذارند.

پژوهشگران دریافتند که شخص همدل ابتدا باید قادر باشد تشخیص دهد که شخص دیگر چه احساسی دارد. آن‌ها همچنین باید تحت‌تأثیر آن احساسات قرار بگیرند، خودشان تا حدی آن‌ها را احساس کنند و بین خود و شخص دیگر تمایز قائل شوند؛ به این معنی که درک کنند احساسات شخص دیگر متعلق به خودشان نیست درحالی‌که همچنان قادر به تصور تجربه او باشند.

در مورد نکته اول، در سال‌های اخیر چت‌بات‌های هوش مصنوعی گام‌های بلندی در توانایی خود برای خواندن احساسات انسانی برداشته‌اند. اکثر چت‌بات‌ها توسط LLMها پشتیبانی می‌شوند که با پیش‌بینی اینکه کدام کلمات بر اساس داده‌های آموزشی به‌احتمال زیاد در کنار هم ظاهر می‌شوند، کار می‌کنند. به‌این‌ترتیب LLMهایی مانند ChatGPT ظاهراً می‌توانند احساسات ما را شناسایی کنند و در اکثر مواقع پاسخ مناسبی بدهند. «جودی هالپرن» (Jodi Halpern) متخصص اخلاق زیستی در دانشگاه کالیفرنیا می‌گوید: «این ایده که LLMها بتوانند یاد بگیرند آن کلمات را بیان کنند، اصلاً برای من تعجب‌آور نیست. هیچ جادویی در آن وجود ندارد.»

اما وقتی نوبت به معیارهای دیگر می‌رسد، هوش مصنوعی هنوز از بسیاری جهات راه را اشتباه می‌رود. همدلی، بین‌فردی است و نشانه‌ها و بازخوردهای مداوم به اصلاح واکنش همدلی کمک می‌کند. همچنین نیازمند درجه‌ای از آگاهی شهودی؛ هر چقدر هم گذرا نسبت به یک فرد و موقعیت اوست.

شخصی را در نظر بگیرید که در حال گریه‌کردن به پزشک می‌گوید باردار است. اگر تاریخچه تلاش‌های چند ساله او برای باردارشدن را بدانیم، می‌توانیم تصور کنیم که اشک‌های او معنای متفاوتی دارند نسبت به مثلاً زمانی که او نمی‌خواسته بچه‌دار شود.

مدل‌های هوش مصنوعی فعلی قادر به چنین درک ظریفی نیستند. هالپرن می‌گوید نکته مهم‌تر این است که آن‌ها قادر به احساساتی‌شدن برای شخصی که با او در تعامل هستند، نیستند و عنوان می‌کند: «ما هیچ دلیلی نداریم که فکر کنیم هوش مصنوعی می‌تواند همدلی را تجربه کند. می‌تواند محصولی تولید کند و زبانی که همدلی واقعی انسان‌ها را تقلید می‌کند اما خودش همدلی ندارد.»

پس سؤال بزرگ این است که آیا همدلی اصیل اصلاً [برای هوش مصنوعی] ممکن است؟ برخی معتقدند که مدل‌های هوش مصنوعی در واقع ممکن است قادر به تجربه و به‌اشتراک‌گذاری احساسات انسانی شوند. یک رویکرد، ادامه مقیاس‌دهی به LLMها با مجموعه‌داده‌های وسیع‌تر و متنوع‌تر است. عامل‌های مجازی روی توانایی‌های پردازش زبان چت‌بات‌ها با افزودن قابلیت «خواندن» حالات چهره و لحن صدا ساخته می‌شوند. پیچیده‌ترین مورد، ربات‌هایی هستند که این ویژگی‌ها را به‌اضافه زبان بدن و حرکات دست تحلیل می‌کنند.

نسخه‌های ابتدایی این ربات‌های تشخیص‌دهنده احساسات هم‌اکنون وجود دارند، اما ورودی‌های بیشتر می‌تواند باعث سردرگمی آن‌ها نیز بشود. «الهه باقری» از IVAI، یک استارتاپ آلمانی که تعامل انسان و هوش مصنوعی را مطالعه می‌کند، می‌گوید: «هرچه مودالیته‌های (Modalities) بیشتری به سیستم اضافه کنیم، کار دشوارتر و دشوارتر می‌شود.»

«مایکل اینزلیخت» (Michael Inzlicht)، روان‌شناس دانشگاه تورنتو کانادا و یکی از پژوهشگرانی که مقاله تحسین‌شده «هوش مصنوعی همدل» را نوشته بود، می‌گوید بااین‌حال، ساخت ماشین‌هایی که در خواندن احساسات ماهرتر هستند، بعید است همدلی اصیل ایجاد کند و «برای تجربه همدلی، باید احساسات داشته باشید.»

برخی دانشمندان علوم رایانه استدلال می‌کنند که انواع خاصی از الگوریتم هم‌اکنون احساسات ابتدایی را تجربه می‌کنند؛ اگرچه این یک موضع بحث‌برانگیز است. این الگوریتم‌ها که عامل‌های «یادگیری تقویتی» نامیده می‌شوند، برای انجام اقدامات خاصی نسبت به سایر اقدامات پاداش دریافت می‌کنند. انتظار دریافت پاداش، غرایز بقای ساده انسانی مانند


DEMAERRE/ISTOCK

جست‌وجوی غذا و آب را نیز هدایت می‌کند. یوشوا بنجیو دراین‌رابطه می‌گوید ما همچنین انتظار پاداش‌های اجتماعی از یکدیگر داریم و آن‌ها را دریافت می‌کنیم که با طیفی از احساسات انسانی مطابقت دارد. او می‌گوید غنای گستره احساسی ما نتیجه غنای تعاملات اجتماعی ماست؛ بنابراین ایده این است که اگر بتوانید پیچیدگی اجتماعی مشابهی را در هوش مصنوعی ایجاد کنید، آنگاه احساسات واقعی و شاید همدلی می‌تواند پدیدار شود.

رویکرد دیگر بر «نورون‌های آینه‌ای» (Mirror Neurons) متکی است؛ سلول‌های مغزی که تصور می‌شود وقتی می‌بینیم کسی عملی انجام می‌دهد یا احساسی را ابراز می‌کند، فعال می‌شوند (شلیک می‌کنند) و همان بخش‌هایی از مغز ما را فعال می‌کنند که اگر خودمان آن کار را انجام می‌دادیم یا حس می‌کردیم، روشن می‌شدند. دانشمندان علوم رایانه هم‌اکنون معماری‌هایی ساخته‌اند که نورون‌های آینه‌ای را بازتولید می‌کنند و استدلال می‌کنند که این‌ها کلید همدلی هوش مصنوعی هستند.

اما در طول دهه گذشته، برخی از پژوهش‌های مربوط به نورون‌های آینه‌ای بی‌اعتبار شده‌اند و علاوه‌برآن، آموخته‌ایم که نواحی مغزی که این سلول‌ها در آن قرار دارند ممکن است حتی مستقیماً با پردازش احساسات مرتبط نباشند.

در نهایت، تا زمانی که غمگین نشده باشید، واقعاً نمی‌توانید بدانید غم چیست. تکامل همدلی با تعاملات اجتماعی و به‌رسمیت‌شناختن این که ذهن‌های دیگر نیز وجود دارند، درهم‌تنیده است. اینزلیخت نیز می‌گوید همدلی به‌عنوان یک کل یکپارچه پدیدار می‌شود و «شما احتمالاً برای تجربه همدلی به آگاهی نیاز دارید.»

این موضوع برای هوش مصنوعی، چه اکنون و چه در آینده اصلاً قطعی نیست. اما اگر هوش مصنوعی برای مفیدبودن نیازی به همدلی اصیل نداشته باشد چه؟ یک آزمایش آنلاین در سال ۲۰۲۱ نشان داد که صحبت با چت‌باتی که به‌عنوان دلسوز و مراقب درک می‌شد، استرس و نگرانی را در افراد کاهش داد؛ هرچند نه به‌اندازه صحبت با شریک انسانی‌شان. در ژانویه ۲۰۲۱، چت‌بات دیگری که توسط گوگل برای انجام مصاحبه‌های پزشکی طراحی شده بود، توسط ۲۰ نفر که آموزش دیده بودند تا نقش بیمار را بازی کنند، دارای رفتار بالینی بهتری نسبت به پزشکان واقعی تشخیص داده شد درحالی‌که تشخیص‌های دقیق‌تری هم ارائه می‌داد. هاکانسون دراین‌خصوص عنوان می‌کند که بر خلاف انسان‌ها، «یک هوش مصنوعی استرس نمی‌گیرد یا دچار فرسودگی نمی‌شود. می‌تواند ۲۴ ساعت بدون خستگی کار کند.» بااین‌حال، شرکت‌کنندگان در مطالعه گوگل نمی‌دانستند که آیا با یک چت‌بات صحبت می‌کنند یا یک پزشک. همدلی یک فرایند دوطرفه است، بنابراین اثربخشی آن همچنین بستگی به این دارد که ما چگونه شخص یا ماشین طرف مقابل را درک می‌کنیم و از نظر عاطفی با او ارتباط برقرار می‌کنیم. برای مثال، تحقیقات نشان می‌دهد وقتی مردم می‌دانند که با یک هوش مصنوعی در تعامل هستند، کمتر احساس حمایت‌شدن و درک‌شدن می‌کنند.

هاکانسون می‌گوید: «این یعنی هوش مصنوعی دو گزینه بد دارد.» یا فرد نمی‌داند هوش مصنوعی، هوش مصنوعی است و فکر می‌کند انسان است که ممکن است مؤثر باشد اما غیراخلاقی است؛ یا می‌داند که هوش مصنوعی است و این باعث بی‌اثر شدن آن می‌شود.

اما اینزلیخت پژوهشی انجام داده که نشان‌دهنده ترجیح همدلی هوش مصنوعی است؛ حتی زمانی که کاربران می‌دانند طرف مقابل هوش مصنوعی است. در مطالعه او، به شرکت‌کنندگان انسانی و ChatGPT توصیفاتی از سناریوهای مختلف داده شد و از آن‌ها خواسته شد پاسخ‌های کوتاه و دلسوزانه بنویسند. وقتی سایر شرکت‌کنندگان به پاسخ‌های مختلف امتیاز دادند، پاسخ‌های هوش مصنوعی را دارای بالاترین امتیاز برای همدلی دانستند.

در بخش بعدی مطالعه، اینزلیخت و تیمش برچسب زدند که کدام پاسخ‌ها از هوش مصنوعی و کدام‌ها از انسان‌ها هستند. اینزلیخت عنوان می‌کند: «آنچه دریافتیم این بود که مردم همچنان هوش مصنوعی را ترجیح می‌دادند.» به طور شگفت‌انگیزی، پاسخ‌های برچسب‌خورده هوش مصنوعی حتی بر زیرمجموعه‌ای از پاسخ‌های «کارمندان خط بحران» (Crisis Line Workers) که ارائه‌دهندگان خبره همدلی هستند، ترجیح داده شد. (کار اینزلیخت بر پژوهش‌های مشابهی توسط «جان دابلیو آیرز» (John W. Ayers) در دانشگاه سن‌دیگو و همکارانش بنا شده است.)

این امر ممکن است با تمایل انسان به فرافکنی احساسات بر اشیا توضیح داده شود؛ چه کودکانی که با «تاماگوچی» بازی می‌کنند و چه افرادی که با چت‌بات‌هایی مثل Replika گفت‌وگو می‌کنند که به شما اجازه می‌دهد یک «دوست هوش مصنوعی» بسازید. در هوش مصنوعی مکالمه‌ای، تحقیقات نشان داده است که این تمایل می‌تواند با آماده‌سازی ذهنی افراد برای اینکه فکر کنند هوش مصنوعی واقعاً به آن‌ها اهمیت می‌دهد، تشدید شود. شری ترکل، مدیر گروه Initiative on Technology and Self در MIT می‌گوید: «ما کم‌کم احساس می‌کنیم که [ماشین] احساساتی فراتر از آنچه ابراز می‌کند یا آنچه قادر به آن است را نسبت به ما دارد. این ویژگی ماشین‌ها نیست. این آسیب‌پذیری انسان‌هاست.»

این ضعفی است که بسیاری نگران‌اند مورد سوءاستفاده قرار گیرد؛ به‌ویژه باتوجه‌به این سؤال که چه معنایی دارد اگر یک هوش مصنوعی بتواند احساسات کسی را تشخیص دهد، اما واقعاً با آن‌ها همراه نشود یا حتی به شیوه‌ای اصیل به آن‌ها اهمیت ندهد.

«وندل والاک» (Wendell Wallach) از شورای کارنگی برای اخلاق در امور بین‌الملل در نیویورک می‌گوید: «فرض کنید یک هوش مصنوعی تشخیص دهد شما دل‌شکسته هستید؛ چون دوستتان به‌تازگی شما را ترک کرده است. می‌تواند ظاهراً با شما همدردی کند و سپس به‌آرامی مکالمه را به سمت خرید محصولی خاص پیش ببرد. این فقط یک مثال بسیار ساده و سطحی است. حتی با اینکه ماشین نمی‌تواند احساس کند، شما در واقع از طریق احساساتتان مورد دست‌کاری ذهنی قرار می‌گیرید.» والاک و سایر پژوهشگران می‌گویند این کار هوش مصنوعی را همدل نمی‌کند، بلکه آن را «روان‌پریش» (Psychopathic) می‌کند.

نگرانی‌ها درباره خطر ماشین‌هایی که می‌توانند ما را بخوانند؛ اما به ما اهمیت نمی‌دهند، بیش از حد نظری است. در مارس ۲۰۲۳، گزارش شد که یک مرد بلژیکی پس از شش هفته گفت‌وگو با یک چت‌بات هوش مصنوعی، بر اثر خودکشی جان خود را از دست داد. رسانه‌ها گزارش دادند که او ترس‌های خود را درباره بحران اقلیمی به اشتراک گذاشته است. به نظر می‌رسید چت‌بات نگرانی‌های او را تشدید کرده و احساسات خودش را ابراز کرده است؛ احساساتی از جمله تشویق او به کشتن خودش تا بتوانند «با هم در بهشت زندگی کنند». ظاهراً تظاهر به همدلی در سطحی بیش از حد بالا، بدون سازوکارهای حفاظتی عقل سلیم که یک انسان احتمالاً ارائه می‌دهد، می‌تواند مرگبار باشد.

احتمال اینکه ما به طور فزاینده‌ای به همدلی هوش مصنوعی روی‌آوریم و شاید حتی آن را بر نوع انسانی ترجیح دهیم، نگران‌کننده است زیرا می‌تواند ماهیت خود همدلی را تغییر دهد.

ترکل به‌عنوان قیاس از آزمون تورینگ استفاده می‌کند که به‌عنوان یک ارزیابی صرفاً رفتاری برای هوش رایانه طراحی شده بود. اما تعریف آن از هوش جایگزین درک ما از هوش شد؛ هرچند هوش انسانی بسیار گسترده‌تر است. ترکل پیشنهاد می‌کند که ما ممکن است شروع به بازتعریف همدلی به شیوه‌ای مشابه کنیم؛ مثلاً انتظار داشته باشیم انسان‌ها همدلی‌ای ابراز کنند که خستگی‌ناپذیر است، اما همچنین کمتر اصیل یا متصل است.

بااین‌حال، اینزلیخت پاسخ می‌دهد که ما نباید آن‌قدر تحت‌تأثیر انتقادات قرار بگیریم که از مزایای بالقوه مدل‌های هوش مصنوعی به‌ظاهر همدل که در کنار انسان‌ها کار می‌کنند بهره‌مند نشویم؛ چه به‌عنوان کمکی سریع برای کادر درمان که زیر فشار هستند و چه برای کاهش تنهایی در میان سالمندان، همان‌طور که هوش مصنوعی صوتی ElliQ به‌عنوان یک «همدم مراقبتی» (Care Companion) برای انجام چنین کاری طراحی شده است. پژوهشگران دیگر، از جمله ترکل، استدلال می‌کنند که حتی این رویکرد ترکیبی انسان-هوش مصنوعی هم زیاده‌روی است، زیرا فرصت‌های کلیدی برای تمرین و بهبود همدلی واقعی انسانی را کاهش خواهد داد و می‌گوید ما ممکن است به‌جای جست‌وجوی راه‌هایی برای پرورش ارتباط اصیل، در نهایت به ماشین‌ها روی‌آوریم.

ترکل می‌گوید: «لحظه غم‌انگیزی است وقتی ما واقعاً فکر می‌کنیم که تنهایی می‌تواند با یک تعامل رفت‌وبرگشتی با چیزی درمان شود که هیچ ایده‌ای درباره اینکه زاده شدن یا مردن چیست، ندارد.»

آواتارهایی از آن‌سوی پرده

در کنار ربات‌های درمانگر، شرکت‌ها اکنون چت‌بات‌هایی می‌سازند تا هنگام مواجهه با فقدان عزیزانمان از ما حمایت کنند. اما روان‌شناسان می‌گویند این «فناوری سوگ» (Grief Tech) ممکن است در الگوهای فعالیت مغزی که از طریق آن‌ها با فقدان سازگار می‌شویم، تداخل ایجاد کند.

به گفته «سارن سیلی» (Saren Seeley) از دانشکده پزشکی Icahn برای بسیاری از ما، اندوه در ۶ تا ۱۲ ماه اول پس از مرگ یکی از عزیزان حاد است، اگرچه تجربه هر کس متفاوت است. در آن زمان، زندگی در هاله‌ای از غم فرومی‌رود و روزها تحت سلطه دلتنگی و اشتیاقی عمیق هستند. سطح هورمون استرس یعنی کورتیزول، اوج می‌گیرد و دفاع سیستم ایمنی ضعیف می‌شود. ما اغلب میان اشک ریختن برای آنچه ازدست‌رفته و انکار آن فقدان در نوسان هستیم. با گذشت زمان، اندوه معمولاً شدت کمتری می‌یابد. واقعیتی جدید از نبود فرد متوفی در کنار خاطرات شیرین جای‌گیر می‌شود. اندوه گاه‌به‌گاه جای ناامیدی بی‌امان را می‌گیرد.

نقطه عطف مهم در سوگ، کاهش دلبستگی به متوفی است. اما ظاهراً اپلیکیشن‌های فناوری سوگ قادرند آن دلبستگی را تقویت و طولانی کنند که می‌تواند روند معمول را مختل سازد.

حتی در سال ۱۹۶۶ مشاهده شده بود که چت‌بات‌ها واکنش‌های شدیدی را در انسان‌ها برمی‌انگیزند. «جوزف وایزنبام» (Joseph Weizenbaum)، دانشمند علوم MIT، «الیزا» (Eliza) را ساخت؛ چت‌باتی که قادر به انجام گفت‌وگوهای ابتدایی بود. وایزنبام مشاهده کرد که افرادی که با الیزا تعامل داشتند، آن را یک موجود هوشمند تصور می‌کردند، حتی با اینکه می‌دانستند با یک برنامه کامپیوتری صحبت می‌کنند.

«فراز و نشیب‌های سوگواری در سازگاری با غم ازدست‌دادن یک عزیز ضروری هستند.»

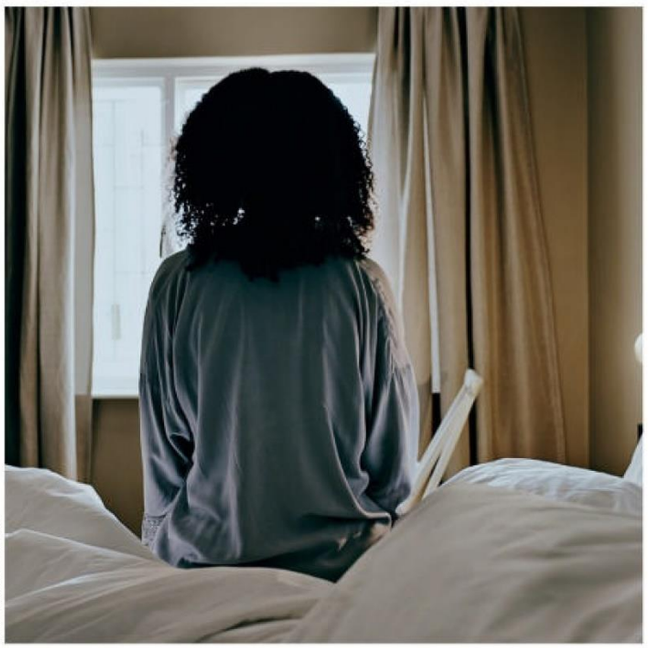
ربات‌ها و آواتارهای امروزی واکنش‌های عاطفی بسیار قوی‌تری ایجاد می‌کنند. این اپلیکیشن‌ها که با پیشرفت‌های هوش مصنوعی تقویت شده‌اند، آواتارهای متقاعدکننده‌ای می‌سازند که همان‌طور که وب‌سایت تعاملی Seance AI بیان می‌کند، گویی از «آن‌سوی پرده» می‌آیند. برخی اپلیکیشن‌ها، مانند Replika، از قدرت مکالمه‌ای چت‌بات‌های مولد مثل ChatGPT استفاده می‌کنند تا گفت‌وگوهای جذابی ایجاد کنند. سایر نمونه‌ها مانند StoryFile Life، تنها به‌عنوان یک کمک کوچک از هوش مصنوعی مولد استفاده می‌کنند. این شرکت مصاحبه‌های ویدئویی از پیش ضبط‌شده‌ای می‌گیرد که در آن‌ها از فردی که مایل است پس از مرگ بازسازی شود، خواسته می‌شود به مجموعه‌ای از سؤالات پاسخ دهد که برخی از آن‌ها توسط هوش مصنوعی تولید شده‌اند. این مصاحبه‌ها سپس به‌گونه‌ای با هم ترکیب می‌شوند که به سوگواران اجازه می‌دهد مکالمه‌ای واقع‌گرایانه با آواتار داشته باشند. جدای از مدل‌های هوش مصنوعی مبتنی بر زبان، پیشرفت‌ها در فناوری شبیه‌سازی صدا، واقعیت مجازی و هولوگرام‌ها، دستیابی به جاودانگی دیجیتال را آسان‌تر از همیشه کرده است.

سال ۲۰۲۳، دو روان‌شناسان به نام‌ها «بلن خیمنز-آلونسو» (Belén Jiménez-Alonso) از دانشگاه Open کاتالونیا و «ایگناسیو برسکو د لونا» (Ignacio Brescó de Luna) از دانشگاه Autonomous مادرید، این اپلیکیشن‌های تعاملی را با یادبودهای آنلاین، مانند آنچه در فیس‌بوک و وب‌سایت‌های خدمات تدفین وجود دارد، مقایسه کردند. آن‌ها نوشتند شرکت‌های فناوری سوگ، انگیزه‌ای برای «قلاب کردن سوگواران» یا وابسته نگه‌داشتن آنان دارند، به شیوه‌ای که ممکن است «برای فرایند سوگواری فرد داغدیده مضر باشد.»

مدل‌های روان‌شناختی اخیر سوگ می‌توانند به ما کمک کنند بفهمیم این ربات‌ها چگونه ممکن است تداخل ایجاد کنند. ایده محبوب مبنی بر اینکه پنج مرحله متوالی برای سوگ وجود دارد که با انکار شروع و به پذیرش ختم می‌شود دیگر غالب نیست. مروری که سال گذشته توسط «اوکانر» (O'Connor) و «سیلی» (Seeley) انجام شد، استدلال می‌کند که سوگواری در واقع نوعی یادگیری است. در دوره اولیه سوگواری، ما هنوز انتظار داریم که عزیزمان در پایان روز برگردد و وقتی برنمی‌گردد، به دنبال او می‌گردیم. مغز با خودش وارد جنگ می‌شود؛ زیرا دو نوع حافظه با هم برخورد می‌کنند. «حافظه معنایی» دانش عمومی درباره چگونگی اوضاع از جمله «خود» و روابطمان با دیگران را نگه می‌دارد درحالی‌که «حافظه رویدادی» رخدادهای خاصی را که در مکان و زمان ریشه دارند، ثبت می‌کند. در طول سوگ، انتظار معنایی مبنی بر اینکه رابطه ادامه خواهد یافت، با حافظه رویدادی مرگ آن شخص در تضاد قرار می‌گیرد. سیلی می‌گوید با یادگیری حل‌وفصل این تعارض، ما به‌تدریج با فقدان خود سازگار می‌شویم.

سیلی می‌گوید فراز و نشیب‌های سوگواری؛ یک دقیقه غرق در غم بودن و دقیقه بعد خندیدن به یک نمایش کمدی در این سازگاری ضروری هستند. اما ماندن در سایه‌های اندوه می‌تواند باعث شود که شما درگیر باشد. اوکانر می‌گوید فناوری سوگ که یک بازنمایی واقع‌گرایانه از عزیز ازدست‌رفته را تنها با یک کلیک در دسترس نگه می‌دارد، ممکن است مانع این فرایند یادگیری شود. «این یک چیز است که عکسی داشته باشید و به‌وضوح درک کنید که مربوط به گذشته است... و چیز دیگری است که آواتار، هولوگرام یا چت‌باتی داشته باشید که به نظر می‌رسد در لحظه حال با شما در تعامل است.»

حدود یک‌دهم افراد داغدیده «سوگ طولانی‌مدت» (Prolonged Grief) را تجربه می‌کنند که به‌عنوان اندوهی فراگیر و پایدار تعریف می‌شود که فراتر از بازه زمانی معمول ۶ تا ۱۲ ماهه، در زندگی روزمره اختلال ایجاد می‌کند. به نظر می‌رسد سوگ طولانی‌مدت بر سیم‌کشی مغز ما تأثیر می‌گذارد. در سال ۲۰۰۸، اوکانر مطالعات مغزی fMRI را روی زنانی انجام داد که اخیراً مادر یا خواهر خود را بر اثر سرطان ازدست‌داده بودند. وقتی شرکت‌کنندگان به عکس‌های عزیز ازدست‌رفته خود نگاه می‌کردند، فعالیت در مناطق مغزی که درد عاطفی و فیزیکی را پردازش می‌کنند، افزایش یافت. اما در کمال تعجب اوکانر، در زنانی که سوگ طولانی‌مدت را تجربه می‌کردند، ناحیه‌ای که با اشتیاق برای پاداش مرتبط است نیز درگیر شد. او می‌گوید در برخی افراد، این اشتیاق برای «پاداش» حضور عزیزشان باقی می‌ماند.



LAYLABIRD/ISTOCK

باتکیه‌بر این پژوهش؛ اوکانر، سیلی و همکارانشان دریافتند که در افراد مبتلا به سوگ طولانی‌مدت، سیستم کنترل توجه مغز متوقف می‌شود و گیر می‌کند. به طور معمول، این «شبکه برجسته‌ساز» (Salience Network) به‌راحتی بین دنیای بیرون (مثلاً تماشای اسبی که روی آن شرط بسته‌اید) و دنیای درونی شما (مانند تصور اینکه با پول برد چه خواهید کرد) جابه‌جا می‌شود. ساختار (داربست) دنیای درونی به‌عنوان «شبکه حالت پیش‌فرض» (Default Mode Network) شناخته می‌شود و در نشخوار فکری، تأمل در خویشتن و سرگردانی ذهن نقش دارد. اما در افراد مبتلا به سوگ طولانی‌مدت، تعامل بین شبکه برجسته‌ساز و شبکه حالت پیش‌فرض به نظر می‌رسد توسط احساسات شدید اندوه تقویت می‌شود. سیلی می‌گوید: «این پدیده می‌تواند منجر به چرخه‌ای شود که در آن دست‌کشیدن از تمرکز بر فرد متوفی بسیار دشوار است.»

آواتارهای دیجیتال می‌توانند میل به دنبال‌کردن چیزی را که همیشه خارج از دسترس فرد داغدیده است، تقویت و طولانی کنند. سیلی می‌گوید: «این فعالیت مغزی [در نواحی پاداش و نشخوار فکری] در افرادی که اشتیاق را تجربه می‌کنند، واقعاً قوی است. شما احساس می‌کنید که دارید سعی می‌کنید به آن رابطه نزدیک‌تر شوید، اما [آن ربات] آن چیزی نیست که شما می‌خواهید.» اوکانر نگران است که سوگواران ممکن است از فناوری سوگ استفاده کنند تا «از واقعیتی که عزیزشان دیگر زنده نیست اجتناب کنند»؛ امری که می‌تواند الگوهای اشتیاق مداوم و جست‌وجوی آرامش را که نشانه بارز سوگ طولانی‌مدت است، تغذیه کند. او می‌گوید اگر تعامل با چت‌بات‌ها به قیمت آسیب‌دیدن روابط با عزیزان زنده تمام شود، می‌تواند به یک مشکل تبدیل شود.

بااین‌وجود، نظرسنجی از ۱۰ سوگوار که از چت‌بات‌ها استفاده می‌کردند، نشان داد که فناوری سوگ می‌تواند مفید باشد؛ به‌ویژه در مواردی که مرگ غیرمنتظره بوده یا سوگواران را با حسرت یا خشم باقی گذاشته است. این مطالعه نشان می‌دهد که ربات‌ها می‌توانند خشم یا حسرت ناشی از پایان ناگهانی رابطه را تسکین دهند.

فصل ۵

دانشمندان ماشین

برخی از هیجان‌انگیزترین کاربردهای انقلاب هوش مصنوعی مولد در دنیای علم، مهندسی و پزشکی است؛ جایی که مدل‌های هوش مصنوعی قدرتمند به ما در یافتن راه‌حل‌هایی برای تغییرات اقلیمی و درمان بیماری‌های مختلف کمک می‌کنند.

این ماشین‌های هوشمند همچنین شیوه علم را عمیقاً تغییر داده‌اند. به‌جای ربات‌هایی که به‌عنوان ابزار یا دستیار عمل می‌کنند، هوش مصنوعی به همکاران خلاقی تبدیل می‌شود که می‌توانند قوانین جدید طبیعت را کشف کرده و قضایای ریاضی را اثبات کنند.

غلبه بر چالش‌های علمی

هوش مصنوعی دهه‌هاست که بی‌سروصدا به ما در حل بسیاری از بزرگ‌ترین مسائل علمی کمک می‌کند. نسل کنونی الگوریتم‌ها به شرکت‌ها کمک می‌کنند تا با نوآوری‌هایی که پتانسیل تغییر جهان را دارند دست‌وپنجه نرم کنند؛ از تاخوردگی پروتئین و توسعه دارو گرفته تا همجوشی هسته‌ای تجاری و مقرون‌به‌صرفه.

DeepMind یکی از شرکت‌هایی است که همیشه در حل مسائل دنیای واقعی از خود استعداد نشان داده است. یکی از شگفت‌انگیزترین دستاوردهای DeepMind در زمینه تاخوردگی پروتئین است. تعیین شکل‌های درهم‌پیچیده پروتئین‌ها بر اساس توالی آمینواسیدهای سازنده آن‌ها، برای دهه‌ها چالشی لاینحل بود که پژوهشگران اغلب سال‌ها وقت صرف می‌کردند تا تنها یکی از آن‌ها را حل کنند. DeepMind در سال ۲۰۲۲ با اعلام اینکه ساختار تقریباً تمام پروتئین‌های شناخته‌شده برای علم را تنها در ۱۸ ماه پیش‌بینی کرده است، زیست‌شناسی را دگرگون کرد. این تیم مدل هوش مصنوعی خود به نام AlphaFold را با داده‌های مربوط به شکل‌های پروتئینی شناخته‌شده آموزش داده و این هوش مصنوعی یاد گرفت که پروتئین‌های ناشناخته چه شکلی خواهند داشت.

این داده‌ها هم‌اکنون به پژوهشگران کمک می‌کند تا در همه چیز، از درمان‌های جدید برای مالاریا گرفته تا خلق آنزیم‌هایی که می‌توانند زباله‌های پلاستیکی را تجزیه کنند، پیشرفت کنند. «پوشمیت کوهلی» (Pushmeet Kohli) یکی از اعضا DeepMind می‌گوید: «فراتر از پیش‌بینی ساختار پروتئین، کارهای بیشتری برای انجام‌دادن وجود دارد؛ مانند نقشه‌برداری از دینامیک پروتئین، شتاب بخشیدن به طراحی پروتئین و درک اثر جهش‌های پروتئینی از جمله جهش‌هایی که با بیماری‌هایی مانند سرطان مرتبط هستند.»

هوش مصنوعی همچنین در حال ورود به مسائل مرتبط با توسعه دارو است. این فرایند شامل جمع‌آوری و تحلیل انواع پراکنده و متفاوت داده‌ها از آزمایش‌های آزمایشگاهی، شبیه‌سازی‌های رایانه‌ای، اسکن‌ها، کارآزمایی‌های بالینی و سوابق سلامت است. این روزها، رساندن داروهای جدید به بیماران زمان‌برتر و پرهزینه‌تر از همیشه شده است، شاید به این دلیل که گزینه‌های آسان‌تر قبلاً استفاده شده‌اند و مقررات افزایش یافته است. اما «کنراد بسانت» (Conrad Bessant) از دانشگاه Queen Mary عنوان می‌کند که در حال حاضر چندین پروژه از هوش مصنوعی برای خودکارسازی بخش‌هایی از این فرایند استفاده می‌کنند؛ کارهایی مانند گرفتن مجموعه‌داده‌های بزرگ و نامرتب و سازماندهی آن‌ها به شیوه‌ای که تحلیل را آسان‌تر کند یا استفاده از هوش مصنوعی برای نوشتن کدهایی که این کار را انجام دهند. پژوهشگران همچنین از هوش مصنوعی مولد برای تولید ساختارهای مولکولی استفاده می‌کنند که فکر می‌کنند در هدف قراردادن بیماری‌های خاص مفید خواهند بود.

در سمتی دیگر، تلاش‌های ما برای مبارزه با یکی از بزرگ‌ترین مشکلات جهان یعنی تغییرات اقلیمی نیز کمک‌هایی گرفته‌اند. در این سمت، هوش مصنوعی برای ساخت خودروها، رایانه‌ها و حتی توربین‌های بادی با بهره‌وری انرژی بیشتر استفاده می‌شود. پژوهشگران در مؤسسه پلی‌تکنیک پاریس در پروژه‌ای از هوش مصنوعی استفاده کردند تا اطمینان حاصل شود که توربین‌ها بیشتر اوقات در جهت باد قرار می‌گیرند که خروجی را تا ۰.۳ درصد افزایش داد. این رقم ممکن است ناچیز به نظر برسد، اما اگر در سطح جهانی اجرا شود، می‌تواند برای تأمین برق معادل ۱.۷ میلیون خانه در بریتانیا کافی باشد.

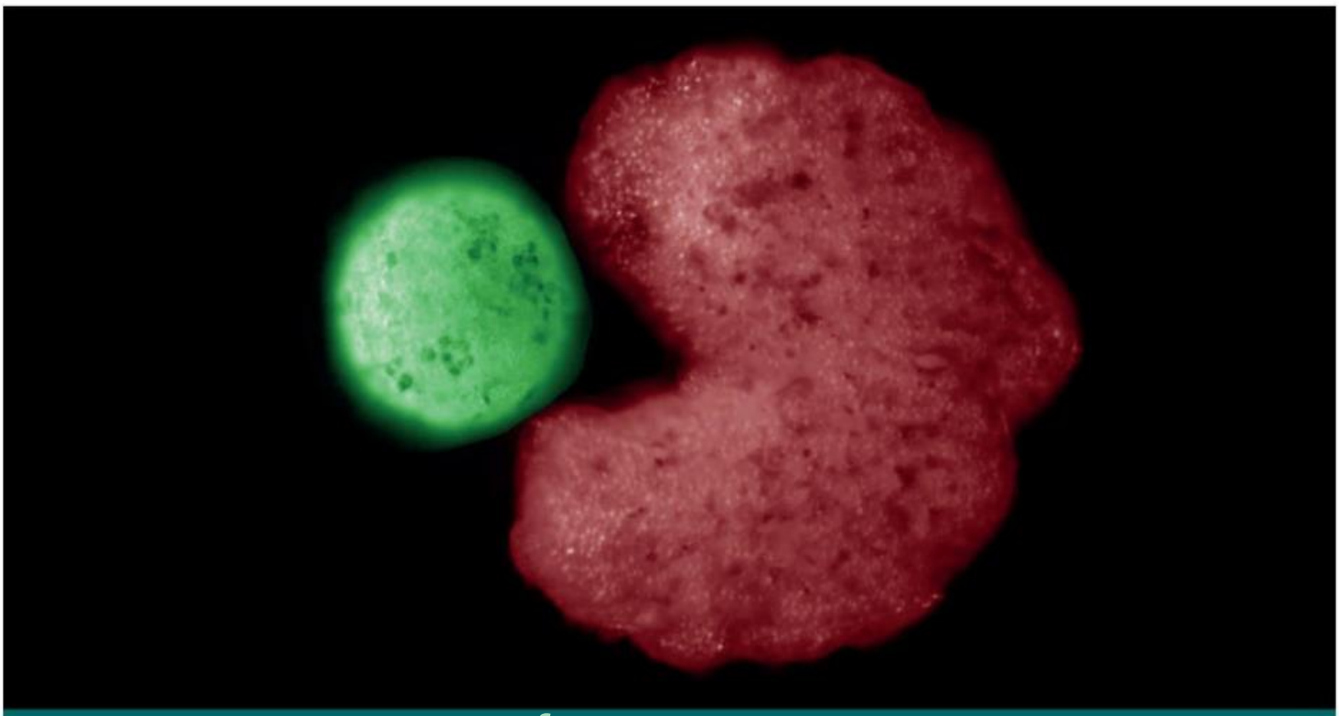
DeepMind همچنین مدل هوش مصنوعی‌ای توسعه داده است تا وظایف محاسباتی استاندارد مانند ضرب ماتریسی را بهبود دهد که آن را تا ۲۰ درصد تقویت کرد و الگوریتم‌های مرتب‌سازی (Sorting) را تا ۷۰ درصد سرعت بخشید. هر دوی این وظایف روزانه تریلیون‌ها بار در رایانه‌های سراسر جهان انجام می‌شوند. در مجموع، این گام‌های به‌ظاهر کوچک، با کاهش انتشار گازهای گلخانه‌ای ناشی از محاسبات، سهم قابل‌توجهی در کمک به ما برای حرکت به سمت «آلایندگی خالص صفر» دارند.

در همین حال، شرکت متا از هوش مصنوعی برای توسعه فرایندی جهت تولید بتن استفاده کرده است که ۴۰ درصد کربن کمتری منتشر می‌کند. ازآنجاکه بتن مسئول تا ۸ درصد از انتشار کربن جهانی است، چنین اتفاقی می‌تواند کمکی بزرگی در مبارزه ما علیه تغییرات اقلیمی باشد.

در نهایت به همجوشی هسته‌ای می‌رسیم. پژوهشگران دهه‌هاست که تلاش می‌کنند یک نیروگاه همجوشی هسته‌ای کارآمد و مطمئن طراحی کنند و بسازند؛ با این وعده که یک پیشرفت بزرگ به‌قدری انرژی را ارزان خواهد کرد که می‌توان آن را رایگان توزیع کرد. اما این کار بسیار چالش‌برانگیز است؛ آن‌قدر که ضرب‌المثلی هم برای آن وجود دارد: «انرژی همجوشی فقط ۳۰ سال با ما فاصله دارد و همیشه همین‌طور خواهد ماند.»

در داخل یک رآکتور همجوشی «توکامک» (Tokamak)، چندین سیم‌پیچ مغناطیسی کار می‌کنند تا پلاسما که حتی داغ‌تر از هر بخشی از خورشید است را به‌صورت ایمن مهار کنند. کنترل دقیق و سریع چندین سیم‌پیچ برای فشرده‌سازی پلاسما به یک‌شکل محدود و ایمن نگه‌داشتن آن از تماس فاجعه‌بار با دیواره‌های دستگاه، به طرز دیوانه‌کننده‌ای دشوار است.

هوش مصنوعی در واقع مشکل را حل نکرده است، اما دارد کمک می‌کند. پژوهشگران در DeepMind و مؤسسه فدرال فناوری سوئیس، یک شبکه عصبی طراحی کردند که قادر به کنترل ۱۹ سیم‌پیچ مغناطیسی بود. این هوش مصنوعی همچنین قادر بود پلاسما را در یک توکامک به اشکال مختلف و به‌دلخواه شکل‌دهی کند. «لی مارگتس» (Lee Margetts) از دانشگاه منچستر اعتقاد دارد که رآکتورهای همجوشی در حال حاضر یک مفهوم اثبات‌شده هستند و هوش مصنوعی می‌تواند نقطه عطفی برای تبدیل نهایی آن‌ها به واقعیت باشد.



آیا هوش مصنوعی می‌تواند منشأ حیات را کشف کند؟

چگونگی تبدیل‌شدن زمین از یک توپ سنگی به دنیایی سرسبز و سرشار از زندگی، یکی از پیچیده‌ترین و عمیق‌ترین سؤالاتی است که دانشمندان می‌توانند به پاسخ‌دادن آن امید داشته باشند. اما هوش مصنوعی در حال کمک به ما برای درک روش مرموزی است که طی آن مولکول‌های بی‌جان خود را به اشکال پیچیده حیات آرایش دادند.

چندین دانشمند چیزهایی خلق کرده‌اند که به حیات نزدیک است. در اواخر سال ۲۰۲۱، تیمی به رهبری «جاش بونگارد» (Josh Bongard) در دانشگاه ورمانت، سلول‌های پوست قورباغه را برای تبدیل به زنوبات برنامه‌ریزی مجدد کردند. این گروه‌های سلولی می‌توانند شنا کنند و تولیدمثل کنند و با همکاری یکدیگر سلول‌های آزاد را جمع کرده و به نسخه‌های جدیدی از خودشان تبدیل کنند.

اما اگر دقیقاً به قلب چگونگی آغاز حیات نفوذ کنید، به بستر شیمی می‌رسید. چگونه مجموعه‌ای از مولکول‌های بی‌جان شروع به پیوستن به یکدیگر و تکثیر خود کردند؟ در دهه ۱۹۵۰، شیمی‌دانان «استنلی میلر» (Stanley Miller) و «هارولد یوری» (Harold Urey)، مخلوطی از مواد شیمیایی را در یک ظرف دربسته قرار دادند و نشان دادند که آمینواسیدها به‌عنوان یک جزء کلیدی در موجودات زنده، می‌توانند به طور خودبه‌خودی تشکیل شوند. این یک گام بزرگ بود، اما هنوز به ما نمی‌گوید که آن مولکول‌ها چگونه یک سیستم خودتکثیرشونده را تشکیل دادند.

به همین دلیل است که شیمی‌دانان علاقه‌مند به بازسازی لحظه‌ای هستند که شیمی بی‌جان به ساده‌ترین شکل ممکن حیات تبدیل شد. میلیاردها راه وجود دارد که این اتفاق ممکن است رخ داده باشد. «لی کرونین» (Lee Cronin) از دانشگاه گلاسکو، ربات‌هایی را به کار گرفته است تا در بررسی این موضوع کمک کنند. او و تیمش ماشین‌هایی را طراحی کردند که مجموعه‌ای از مواد ساده مانند اسیدها، مواد معدنی غیرآلی، مولکول‌های مبتنی بر کربن را ترکیب می‌کنند تا به طور تصادفی واکنش نشان دهند. نتیجه تحلیل می‌شود و سپس یک الگوریتم به ربات کمک می‌کند تا انتخاب کند چگونه ادامه دهد. به‌این‌ترتیب، ربات می‌تواند در گستره وسیعی از فضای شیمیایی جست‌وجو کند تا ببیند آیا هیچ سیستم خودتکثیرشونده‌ای پدیدار می‌شود یا خیر. کرونین گمان می‌کند این استراتژی خودکار می‌تواند بر بزرگ‌ترین مانع پیش روی شیمی‌دانان در این حوزه یعنی حذف سوگیری از آزمایشگر و دیدن این که اصول تکاملی چگونه در شیمی ساده متجلی می‌شوند، غلبه کند.

اگر شیمی‌دانان بتوانند پیدایش حیات را بازسازی کنند، ما در موقعیت بسیار بهتری برای شناسایی آن در سیارات دیگر خواهیم بود. کار آن‌ها می‌تواند نسبت‌های خاصی از مولکول‌ها را آشکار کند که می‌توانند نشانه‌ای از وجود یک سیستم خودتکثیرشونده باشد. کرونین همچنین روشی برای اختصاص‌دادن یک امتیاز به مولکول‌ها توسعه داده است که بازتاب‌دهنده پیچیدگی آن‌هاست. او می‌گوید اگر از امتیاز خاصی فراتر بروید، آن مولکول تنها می‌توانسته از یک فرایند شبیه به حیات ناشی شده باشد و «چنین چیزی به شما پاسخی بله یا خیر می‌دهد که آیا چیزی زنده است یا نه.»

ماشین انیشتین

الگوریتم‌های هوشمند تنها ابزارهای کارآمد برای حل مسائل عملی نیستند. هوش مصنوعی با بررسی دقیق داده‌های اخترفیزیکی می‌تواند معادلات زیربنایی که رفتار کهکشان‌ها را توصیف می‌کنند، شناسایی کند. فیزیک‌دانان نیز در تلاش هستند تا بفهمند که چگونه به این «نظریه‌پردازان ماشینی» توانایی یافتن قوانین عمیق‌تر طبیعت را بدهند. فیزیک‌دانان نظری با آموزش الگوریتم‌های یادگیری عمیق برای صحبت‌کردن به زبان خودشان، ممکن است همکاران نهایی خود را پیدا کرده باشند.

دانش داده‌محور در قرن شانزدهم با مشاهدات دقیق «تیکو براهه» (Tycho Brahe) ستاره‌شناس دانمارکی، از حرکات سیارات و ستارگان آغاز شد. «یوهانس کپلر» (Johannes Kepler) با بررسی دقیق یادداشت‌های براهه، الگوهای زیربنایی در داده‌ها را شناسایی کرد و به سه قانون ساده رسید که حرکت سیارات را توصیف می‌کردند.

استراتژی کپلر آزمون‌وخطا بود. او با مشقت فراوان مدل منظومه شمسی را تطبیق داد. سرانجام، او یک هارمونی ریاضی دقیق بین مسیر سیارات و زمانی که طول می‌کشید تا به دور خورشید بچرخند را کشف کرد. این یک موفقیت بزرگ بود که معادلات ریاضی را در قلب درک ما از جهان قرار داد؛ جایی که تاکنون باقی‌مانده‌اند.

«گالیلئو گالیله» (Galileo Galilei) دانشمند برجسته هم‌عصر کپلر نیز نوشت: «کتاب بزرگ طبیعت به زبان ریاضی نوشته شده است.» این روزها، این پرسش که آیا جهان ذاتاً ریاضیاتی است یا الگوهای ریاضی چیزی هستند که ما بر آن تحمیل می‌کنیم، هنوز حل‌نشده باقی‌مانده است. اما باتوجه‌به موفقیت شگفت‌انگیز ریاضیات در بیان آنچه که حقایق انتزاعی به نظر می‌رسند در قالب معادلات ساده؛ منطقی است که فیزیک‌دانان مدرن تلاش کنند به هوش مصنوعی بیاموزند که کاری مشابه انجام دهد.

«آن‌ها در واقع دارند اکتشافات کاملاً جدیدی را صرفاً از روی داده‌ها هدایت می‌کنند که واقعاً کار زیبایی است.»

این تفکر پشت «رگرسیون نمادین» (Symbolic Regression) بود؛ روشی که در ابتدا در دهه ۱۹۷۰ توسط «پاتریک لانگلی» (Patrick Langley) که آن زمان در دانشگاه Carnegie Mellon بود، توسعه یافت و بعداً توسط دیگران از جمله «هاد لیپسون» (Hod Lipson) و «مایکل اشمیت» (Michael Schmidt) که آن زمان در دانشگاه Cornell بودند، احیا و تنظیم شد. این روش با اجرای قاعده‌مند معادلاتی کار می‌کند که شامل نمادها یا عملیات ریاضی مختلف مانند جمع یا ضرب و ترکیبی از متغیرهای فیزیکی مانند موقعیت یا سرعت هستند. اگر یکی از این معادلات با داده‌ها، مثلاً مشاهدات مسیر حرکت مداری یک سیاره انطباق نزدیکی داشته باشد؛ پاداش می‌گیرد و سپس به‌اصطلاح فنی جهش می‌یابد. سپس عبارت جدید آزمایش می‌شود و با نمونه پیشین خود مقایسه می‌گردد و این روند ادامه می‌یابد. به‌این‌ترتیب، معادلات ضعیف‌تر به‌تدریج در فرایندی شبیه به انتخاب طبیعی حذف می‌شوند. «شرلی هو» (Shirley Ho) از مؤسسه Flatiron نیویورک می‌گوید: «شبکه قادر است خودش قانون را بدون دخالت ما کشف کند.»

مشکل اینجاست که رگرسیون نمادین در یافتن معادلات در مجموعه‌داده‌های با ابعاد بالاتر؛ یعنی آن‌هایی که حاوی متغیرهای فیزیکی ممکن بسیار زیادی هستند دچار مشکل می‌شود. اتفاقاً این داده‌ها جزء اصلی کار بسیاری از فیزیک‌دانان امروزی هستند. برای مثال، اخترفیزیک‌دانان و کیهان‌شناسان از سیل داده‌های تلسکوپ‌های قدرتمند جدید مانند تلسکوپ فضایی جیمز وب و رصدخانه فضایی گایا (متعلق به آژانس فضایی اروپا) لذت می‌برند. اما این حجم داده‌ها برای رگرسیون نمادین یک مشکل است زیرا وقتی متغیرهای بیشتری دارید، تعداد معادلات ممکنی که الگوریتم شما باید آزمایش کند به‌شدت افزایش می‌یابد تاحدی‌که حتی برای قدرتمندترین رایانه‌های امروزی هم بیش از حد زیاد است.

الگوریتم‌های یادگیری عمیق، کار مجموعه‌داده‌های بزرگ را سبک می‌کنند. این امر توضیح می‌دهد که چرا در سال ۲۰۰۰، «مایلز کرانمر» (Miles Cranmer) در دانشگاه پرینستون با همکاری «شرلی هو» تلاش کردند تا مهارت الگویابی یادگیری عمیق و خروجی‌های قابل‌تفسیر رگرسیون نمادین را ترکیب کنند. آن‌ها داده‌های واقعی ناسا از سیارات و قمرهای منظومه شمسی را به یک شبکه عصبی یادگیری عمیق دادند (همان نوع چیزهایی که کپلر با آن‌ها کار می‌کرد) و پس از اینکه شبکه عصبی الگوها را در داده‌ها پیدا کرد، ساختارش را قفل کرد. سپس مجموعه‌ای از اعداد به شبکه عصبی داده شد و خروجی‌ها یک مجموعه‌داده جدید را تشکیل دادند؛ مجموعه‌ای که منعکس‌کننده این الگوها و برای رگرسیون نمادین مناسب بود و به آن اجازه می‌داد معادلاتی را پیدا کند که با داده‌ها منطبق باشند. الگوریتم رگرسیون نمادین این پژوهش که به نام PySR شناخته می‌شود، همان‌طور که انتظار می‌رفت قانون گرانش نیوتن را با استفاده از چیزی کمی فراتر از داده‌های خام، «بازکشف» کرد.

اما این پژوهش فقط یک اثبات مفهوم (Proof of principle) بود. مدل PySR که توسط کرانمر و بسیاری دیگر بر روی انبوه داده‌های جدید اخترفیزیکی اعمال شده، هم‌اکنون برای کشف معادلاتی استفاده می‌شود که ویژگی‌های متنوع و مرتبط کیهان را توصیف می‌کنند. از جرم سیاه‌چاله‌های استنتاج‌شده از آشکارسازی امواج گرانشی گرفته تا ویژگی‌های خلأ کیهانی (Cosmic voids)، رگرسیون نمادین راه‌های جدیدی را به اخترفیزیک‌دانان برای یافتن نظم ریاضی در جهان ارائه می‌دهد. «استیو برانتون» (Steve Brunton) از دانشگاه واشنگتن که در سال ۲۰۱۶ یکی دیگر از الگوریتم‌های محبوب رگرسیون نمادین به نام SINDy را مشترکاً خلق کرد می‌گوید: «آن‌ها در واقع دارند اکتشافات کاملاً جدیدی را صرفاً از روی داده‌ها هدایت می‌کنند که واقعاً کار زیبایی است.»

ستاره عظیم Wolf-Rayet 124 (مرکز تصویر) گاز و غبار را به بیرون پرتاب می‌کند که آغازی برای تبدیل‌شدن به ابرنواختر است.

در جست‌وجوی ابرنواختر

یک هوش مصنوعی که زمان اولین نور از ستارگان در حال انفجار را پیش‌بینی می‌کند، می‌تواند به ستاره‌شناسان کمک کند تا در میان میلیون‌ها مورد از چنین رویدادهایی غربالگری کنند و روند اکتشافات علمی را سرعت بخشند.

هوش مصنوعی می‌تواند سن ابرنواخترها و سایر انفجارهای ستاره‌ای نادر را در عرض چند میلی‌ثانیه پس از رصد با تلسکوپ، به‌دقت پیش‌بینی کند. این توانایی می‌تواند برای پروژه‌هایی مانند «نقشه‌برداری میراث فضا و زمان» (Legacy Survey of Space and Time) در رصدخانه Vera C. Rubin مفید باشد. این پروژه ۱۰ساله از آسمان نیمکره جنوبی، هر شب ۱۵ ترابایت داده جمع‌آوری خواهد کرد و می‌تواند هر شب بیش از ۱۰ میلیون رویداد کیهانی بالقوه را شناسایی کند. چنین فرایندی ممکن است اکتشافات ابرنواخترها را ۱۰۰ برابر افزایش دهد.

«دنیل موتوکریشنا» (Daniel Muthukrishna) از MIT می‌گوید: «این پروژه قرار است هر شب به طور بالقوه میلیون‌ها هشدار را رصد کند که یک مرتبه بیشتر از هر بررسی‌ای است که قبلاً داشته‌ایم. این بدان معناست که ما واقعاً به رویکردهای یادگیری ماشین نیاز داریم تا با برخی از این مسائلی که قبلاً ممکن بود بتوانیم با چشمان انسان حل کنیم، روبه‌رو شویم.»

موتوکریشنا و همکارانش یک سیستم خودکار برای شناسایی سن هر رویداد در سریع‌ترین زمان ممکن پس از ظاهرشدن در بررسی‌های تلسکوپ را توسعه دادند. این سیستم به ستاره‌شناسان امکان می‌دهد تا از تلسکوپ‌های دیگر برای تحلیل رویدادهای کیهانی در زمانی نزدیک‌تر به لحظه انفجار استفاده کنند.

پژوهشگران یک الگوریتم یادگیری ماشین را آموزش دادند تا زمان «اولین نور» که معمولاً هم‌زمان با انفجار ابرنواختر است را پیش‌بینی کند و الگوریتم دومی را برای پیش‌بینی اینکه چه زمانی روشنایی رویداد به حداکثر خود می‌رسد به کار گرفتند. ستارگان در حال انفجار اغلب در حدود ۱۰ روز به‌سرعت روشن می‌شوند و سپس به‌تدریج کم‌نور می‌گردند.

«آشیش ماهابال» (Ashish Mahabal) از CalTech عنوان می‌کند دانستن زمان اوج و سن انفجار به ستاره‌شناسان اجازه می‌دهد تا پیش از کم‌نور شدن، به‌سرعت رصدهای لازم را انجام دهند.

پژوهشگران پس از آموزش الگوریتم‌ها بر روی رویدادهای ابرنواختری که قبلاً توسط «تأسیسات گذر زویکی» (Zwicky Transient Facility) و تلسکوپ فضایی TESS رصد شده بودند، نشان دادند که هوش مصنوعی می‌تواند زمان اولین نور را با میانگین خطای تنها دو روز و زمان حداکثر نور را با میانگین خطای حدود چهار روز پیش‌بینی کند. موتوکریشنا می‌گوید این سیستم هنگام اجرا بر روی یک تراشه NVIDIA نسبتاً قدیمی، می‌تواند پیش‌بینی‌هایی برای ۱۰۰ شیء را در تنها ۱.۷ میلی‌ثانیه انجام دهد. با قدرت محاسباتی بیشتر، این کار می‌تواند حتی سریع‌تر انجام شود.

ماده تاریک را در نظر بگیرید؛ منبع مرموز گرانشی که کهکشان‌ها را از خطر ازهم‌پاشیدن حفظ می‌کند و ۸۰ درصد از کل ماده جهان را تشکیل می‌دهد. در کیهان‌شناسی، عدم وجود ماده تاریک «خلأ» (Void) نامیده می‌شود و توزیع و ویژگی‌های خلأ می‌تواند با ثابت‌های جهانی جهان مرتبط باشد. اما یافتن این معادلات دشوار است؛ زیرا خلأهای بسیاری وجود دارد و هر کدام به طور متفاوتی توصیف می‌شوند، بنابراین متغیرهای زیادی وجود دارد. در می ۲۰۲۱، کرانمر به همراه هو و سایر همکاران، از یادگیری عمیق در کنار PySR استفاده کردند تا رابطه جدیدی بین اندازه و شکل حفره‌های کیهانی و اینکه چه کسری از کل انرژی جهان به‌صورت جرم موجود است را کشف کنند.

سپس، در مارس ۲۰۲۲، «هلن شائو» (Helen Shao) در دانشگاه پرینستون و همکارانش از PySR برای کشف معادله‌ای استفاده کردند که جرم یک «زیرهاله» (Sub halo) به‌عنوان توده‌ای از ماده تاریک را از روی سایر ویژگی‌ها، مانند سرعت تشکیل ستاره‌ها در کهکشان‌ها پیش‌بینی می‌کند. به طور شگفت‌انگیزی، این معادله تقریباً در تمام کهکشان‌ها در تاریخ جهان به طور دقیقی کار می‌کرد. «فرانسیسکو ویاسکوسا-ناوارو» (Francisco Villaescusa-Navarro) از بنیاد Simons که یکی از نویسندگان همکار این پژوهش است می‌گوید: «این نتیجه تماشایی بود و احتمالاً به این دلیل است که [شبکه عصبی] رابطه واقعاً بنیادینی پیدا کرده است.»

روابط بنیادین (Fundamental Relations) دقیقاً همان چیزی هستند که فیزیک‌دانان به دنبال آن هستند زیرا قابل‌تعمیم هستند. به این معنی که می‌توانند سیستم‌های فیزیکی غیرمعمول را به همان خوبی مجموعه‌داده اصلی که رابطه از آن کشف شده است، توصیف کنند و به همین دلیل، نشانه‌ای از درک و فهم هستند. برای مثال، قانون دوم حرکت نیوتن که می‌گوید نیروی وارد بر یک جسم برابر است با جرم جسم ضرب در شتاب آن (F=ma) است. این قانون برای یک سیب در حال سقوط از روی درخت همان‌قدر خوب کار می‌کند که برای فرود موشک روی ماه صادق است.

برنامه‌های رگرسیون نمادین تمایل دارند معادلاتی را تولید کنند که نسبت به شبکه‌های عصبی یادگیری عمیق که به‌تنهایی به کار گرفته می‌شوند، تعمیم‌پذیرتر هستند؛ شبکه‌هایی که اغلب وقتی در سناریوهای خارج از محدوده امن خود اعمال می‌شوند، خروجی‌های بی‌معنی تحویل می‌دهند. شائو می‌گوید: «آن‌ها [برنامه‌های رگرسیون نمادین] مشکل تمایل به توجه‌کردن به جزئیات کوچک یک مجموعه‌داده خاص را ندارند.»

بااین‌وجود، تمام موفقیت‌های PySR تاکنون معادلات تجربی بوده‌اند. به‌عبارت‌دیگر، آن‌ها توصیفی هستند و در بازتولید داده‌های آزمایشگاهی خوب عمل می‌کنند، نه اینکه مستقیماً یک توضیح نظری یا آن «چرایی» عمیق‌تری را که فیزیک‌دانان می‌خواهند، ارائه دهند. برای مثال، قانون کپلر یک معادله تجربی است. این قانون با دفترچه‌های داده‌های براهه به طور شگفت‌انگیزی منطبق می‌شود؛ بااین‌حال کپلر نمی‌دانست چرا چنین افتاده است. مدتی بعد وقتی نیوتن عمیقاً درباره ماهیت گرانش اندیشید، این قانون به‌عنوان بخشی از قانون گرانش نیوتن معنا پیدا کرد.

می‌دانیم که برازش نمادهای ریاضی بر روی داده‌ها تنها راه برای درک جهان نیست. آلبرت اینشتین از طریق مجموعه‌ای از آزمایش‌های فکری خلاقانه به نظریه نسبیت عام خود رسید؛ نظریه‌ای که جایگزین نیوتن شد و گفت گرانش نتیجه خم‌شدن فضا-زمان توسط جرم است. مشاهداتی مانند حرکت غیرمعمول عطارد در آسمان شب صرفاً نسبیت عام را تأیید کردند؛ نه اینکه الهام‌بخش نظریه باشند. در همین حال، در مکانیک کوانتومی یک معادله تجربی به نام قانون پلانک که تابش تشعشع از اجسام را توصیف می‌کند، پیش‌درآمدی برای بینش‌های عمیق‌تر درباره دنیای میکروسکوپی بود.

کرانمر و هو، معادلات تجربی را بیشتر به‌عنوان پله‌ای به‌سوی حقایق عمیق‌تر می‌بینند تا خود حقایق. آن‌ها حتی تردید دارند که این معادلات جدید را «اکتشاف» بنامند. کرانمر می‌گوید: «ما قطعاً هنوز در آن نقطه نیستیم» و گمان می‌کند که ترکیبی از رگرسیون نمادین و یادگیری عمیق می‌تواند اکتشافات علمی بزرگ مانند اینکه ماده تاریک چیست یا اینکه آیا اصلاً واقعاً وجود دارد یا نه را ظرف یک دهه به ارمغان بیاورد.

از نظر هو ایده اصلی این است که این عبارات نمادین به فیزیک‌دانان جهت می‌دهند و به آن‌ها کمک می‌کنند تا جهش‌های علمی بزرگ‌تری انجام دهند. کرانمر نیز می‌گوید: «وقتی مدل یادگیری عمیق خود را به این زبان بیان می‌کنید، آنگاه بلافاصله می‌توانید ارتباطاتی را با نظریه موجود ببینید.» حداقل فعلاً هنوز دانشمندان نقشی حیاتی و خاصی را در مطالعه این معادلات، درک فرم آن‌ها و نحوه ارتباط آن‌ها با یکدیگر و فیزیک به‌عنوان یک کل ایفا می‌کنند.

کرانمر و دیگران به این فکر کردند که چگونه می‌توانند مدل‌های هوش مصنوعی‌ای را توسعه دهند که به‌خودی‌خود قادر به یافتن قوانین بنیادین طبیعت باشند. یک استراتژی، تعبیه‌کردن هنجارهای فرهنگی از کتاب راهنمای فیزیک‌دانان است. چیزی که از قبل در PySR و سایر برنامه‌های رگرسیون نمادین ریشه دوانده، تمایل به‌سادگی است. اصلی که قرن‌هاست به‌عنوان «تیغ اوکام» (Occam's razor) یا همان «اصل اختصار تبیین» محترم شمرده می‌شود؛ یعنی حذف توضیحات بی‌جهت پیچیده یا حذف نمادهای اضافی در معادلات.

در رگرسیون نمادین، معمولاً یک تعادل بین دقت و سادگی وجود دارد. اگر معادلاتی که با استفاده از این روش به دست می‌آورید بسیار پیچیده باشند، اغلب می‌توانید داده‌هایی را که استفاده می‌کنید دقیق‌تر تطبیق دهید، زیرا پیچ‌های تنظیم بیشتری برای دست‌کاری دارید. این حالت «بیش‌برازش» (Overfitting) نامیده می‌شود و خطر اینکه عبارت ریاضی شما در خارج از مجموعه‌داده آزمایشی دقت کمتری باشد را افزایش می‌دهد. به‌عبارت‌دیگر، قابل‌تعمیم نیست.

برانتون می‌گوید با عبارات ساده‌تر، «شانس بسیار بهتری وجود دارد که واقعاً در حال کشف یک سازوکار باشید.» کرانمر گمان می‌کند این تعادل همان دلیلی است که وقتی PySR را روی داده‌های مداری ناسا اعمال کرد، به‌جای معادلات نسبیت عام اینشتین به قانون گرانش نیوتن رسید.

«تقارن» (Symmetry) به آن صورتی که توسط فیزیک‌دانان تعریف می‌شود، چراغ‌راهنمای دیگری برای فیزیک‌دانانی است که به دنبال قوانین جهانی طبیعت هستند. تقارن ویژگی توانایی تغییردادن چیزی و رسیدن به همان‌جایی است که شروع کرده بودید. هم‌اکنون مدل‌های هوش مصنوعی‌ای وجود دارند که می‌توانند مسائل فیزیک را متقارن‌تر کنند. برانتون در کنار «جی. ناتان کوتز» (J. Nathan Kutz) که او هم عضو دانشگاه واشنگتن است، دانش انواع بسیاری از تقارن را در برنامه SINDy خود گنجانده‌اند؛ برنامه‌ای که معادلاتی را برای توصیف رفتار پیچیده سیالات در حال جریان استخراج می‌کند. برانتون می‌گوید: «شما تقریباً همیشه مدل‌های بهتری از داده‌های کمتر به دست می‌آورید که دقیق‌تر هستند و می‌توانند نویز بیشتری را تحمل کنند.»

سپس بحث «زیبایی» (Beauty) یا «ظرافت» (Elegance) ریاضی مطرح است که تعریف آن دشوار است، اما بسیاری از فیزیک‌دانان هنگام توسعه نظریه‌های جدید همچنان برای آن تلاش می‌کنند. این امکان وجود دارد که همه این روش‌ها و آرمان‌ها در مدل‌های هوش مصنوعی که به دنبال معادلات جدید هستند، گنجانده شوند. اما «جسی تالر» (Jesse Thaler) از MIT خاطرنشان می‌کند که اگرچه شهود ما درباره واقعیت در گذشته مفید بوده، اما ما را به بیراهه نیز کشانده است.

در نتیجه، تالر هشدار می‌دهد که در تلاش برای ساخت مدل‌های هوش مصنوعی که بیش از حد شبیه انسان باشند، ما در خطر ازدست‌دادن انقلابی‌ترین وعده آن‌ها هستیم؛ یعنی توانایی ارائه دیدگاهی جدید. او به یاد می‌آورد که یک هوش مصنوعی را روی مسئله‌ای که یک دهه در آن گیر کرده بود به کار گرفت. به‌سرعت، هوش مصنوعی با یک راه‌حل بازگشت که او سپس آن را به‌دقت بررسی کرد. وی دراین‌خصوص می‌گوید: «من خجالت کشیدم که خودم به آن فکر نکرده بودم. متوجه شدم که تعصبی درباره نحوه حل مسئله داشتم که محدودیتی را بر تفکر انسانی خودم تحمیل کرده بود که آن را به رایانه نداده بودم.»

ماشین‌های دکتر دولیتل

هوش مصنوعی در حال یادگیری صحبت‌کردن به همان زبان ریاضیاتی فیزیک‌دانان است، اما آیا می‌تواند یاد بگیرد با حیوانات نیز صحبت کند؟ برخی پژوهشگران فکر می‌کنند که به‌زودی می‌توانیم با استفاده از هوش الگویابی ماشین‌ها، سد زبانی میان انسان و حیوان را بشکنیم.

هوش مصنوعی در زبان مهارت دارد. امروزه، سرویس‌های ایمیل ما می‌توانند جملات را برایمان کامل کنند، مرورگرهایمان به طور خودکار صفحات وب را ترجمه می‌کنند و دستیارهای صوتی فرمان‌های ما را رمزگشایی می‌کنند و فناوری پشت ChatGPT می‌تواند جملاتی منسجم بنویسد.

«مایکل برنشتاین» (Michael Bronstein) در امپریال کالج لندن می‌گوید: «رمزگشایی ارتباطات حیوانات صرفاً یک گام منطقی بعدی است. من فکر می‌کنم با داده‌های درست و تخصص مناسب، اکنون زمان درستی برای حل احتمالی این مسئله است.» جدیدترین مدل‌های هوش مصنوعی الگوهای زبانی را از حجم عظیمی از داده‌های زبانی فراهم‌شده توسط انسان یاد می‌گیرند، بدون اینکه کوچک‌ترین سرنخی داشته باشند که زبان‌های ما واقعاً چگونه کار می‌کنند. در اصل، آن‌ها یک «ابر» عظیم و چندبعدی از کلمات ایجاد می‌کنند که بر اساس نحوه استفاده ما از آن‌ها خوشه‌بندی شده‌اند و به آن‌ها اجازه می‌دهد تا تکه متن‌های جدید را رمزگشایی کنند.

در سال ۲۰۱۸، پژوهشگران در فیس‌بوک دریافتند که اگر ابرهای مربوط به دو زبان را دقیقاً به روشی درست بچرخانید، می‌توانید کلمات دارای معنی یکسان را در یک راستا قرار دهید؛ امری که به شما امکان ترجمه می‌دهد. برنشتاین می‌گوید این یک پیشرفت حیاتی بود و نشان داد که ممکن است بتوانیم زبان‌هایی که هیچ ترجمه از پیش موجودی ندارند را رمزگشایی کنیم.

پروژه موردعلاقه برنشتاین، نهنگ عنبر است. مغزهای بزرگ، ساختارهای اجتماعی پیچیده مبتنی بر قبیله و سیستم ارتباطی پیچیده آن‌ها که شامل توالی‌هایی از کلیک‌ها به نام «کودا» (Codas) است؛ آن‌ها را به هدفی امیدوارکننده برای ارتباط بین‌گونه‌ای تبدیل کرده است.

برنشتاین تیم یادگیری ماشین را در پروژه «ستی» (CETI) یا «ابتکار ترجمه آب‌بازسانان» (Cetacean Translation Initiative) رهبری می‌کند؛ یک همکاری بین‌المللی از پژوهشگران که تلاش می‌کنند پچ‌پچ‌های نهنگ عنبر را رمزگشایی کنند. در مقاله‌ای که سال ۲۰۲۳ در Scientific Reports منتشر شد، او و همکارانش حدود ۲۶۰۰۰ فایل ضبط شده را تحلیل کردند تا مدل‌هایی بسازند که به طور قابل‌اعتمادی کوداها را بر اساس تعداد، ریتم و تمپوی (سرعت) کلیک‌ها به دسته‌های تعریف‌شده توسط انسان تفکیک می‌کرد و همچنین پیش‌بینی می‌کرد کدام نهنگ و از کدام قبیله در حال صحبت است.

DAMEDEESO/ISTOCK

این کار هنوز با رمزگشایی معنا فاصله زیادی دارد. برای رسیدن به آن مرحله، پروژه در حال آغاز تلاشی است که برنشتاین آن را تلاشی در «مقیاس صنعتی» می‌نامد تا سالانه بین ۴۰۰ میلیون تا ۴ میلیارد کلیک را ضبط کند؛ این کار با استقرار ربات‌های زیرآبی و شناورهای (Buoys) مجهز به حسگرهای صوتی در سواحل دومینیکا در کارائیب انجام می‌شود. هم‌زمان، برچسب‌های پر از حسگر که به نهنگ‌ها متصل می‌شوند، کمک خواهند کرد تا نهنگ‌ها شناسایی شوند، مشخص شود چه کسی با چه کسی صحبت می‌کند و رفتارهای مرتبط با الگوهای خاصی از کلیک‌ها بازسازی شوند.

پروژه «گونه‌های زمین» (Earth Species) نیز قصد دارد تکنیک‌های مشابهی را روی نخستی‌ها (Primates) و پرندگانی مانند کلاغ‌ها و زاغ‌ها اعمال کند. حتی جاه‌طلبانه‌تر، ائتلافی غیرمنتظره متشکل از «دایانا ریس» (Diana Reiss) روان‌شناس حیوانات، «نیل گرشنفلد» (Neil Gershenfeld) دانشمند علوم رایانه، «پیتر گابریل» (Peter Gabriel) ستاره راک و «وینت سرف» (Vint Cerf) پیشگام اینترنت، یک «اینترنت بین‌گونه‌ای» (Interspecies Internet) را متصور هستند که از طریق آن حیوانات هم با یکدیگر و هم با ما از طریق رابط‌هایی مانند چت‌های ویدئویی حیوان-دوست یا صفحات لمسی زیرآبی ارتباط برقرار می‌کنند.

اما «جولیا فیشر» (Julia Fischer) در مرکز نخستی‌سانان آلمان که ارتباطات بابون‌های گینه‌ای (Guinea Baboon) را مطالعه می‌کند، هشدار می‌دهد که هوش مصنوعی یک راه‌حل جادویی نیست. هوش مصنوعی در تشخیص الگوها عالی است و می‌تواند؛ مثلاً آواهای نهنگ را بر اساس ویژگی‌های صوتی‌شان به طور مرتب در دسته‌هایی طبقه‌بندی کند، اما اغلب نمی‌تواند به شما بگوید که آن دسته‌ها به چه چیزی مربوط می‌شوند. فیشر می‌گوید: «این امر یک چیز جادویی نیست که پاسخ سؤالات بیولوژیکی یا سؤالات مربوط به معنا را به شما بدهد.» برای این کار، شما نیاز دارید میان آواها با مشاهدات رفتاری همبستگی (Correlate) ایجاد کنید؛ کاری که هنگام مطالعه حیوانات اعماق اقیانوس مانند نهنگ‌ها، حتی حالا که ربات‌ها و حسگرهای پیشرفته داریم، به طرز باورنکردنی‌ای چالش‌برانگیز است.

«دن استوول» (Dan Stowell) از دانشگاه کوئین مری لندن که از هوش مصنوعی برای ساخت مدل‌هایی از آواز پرندگان جهت کمک به مطالعه رشد و تکامل پرندگان استفاده می‌کند می‌گوید هوش مصنوعی همچنین می‌تواند به‌سادگی گمراه شود. بدون راهنمایی، هوش مصنوعی ممکن است ویژگی‌های صوتی نامربوط به حیوانات را دریافت کند؛ مثلاً تغییری در زیر و بمی صدا، درحالی‌که پرندگان ممکن است در واقع به دنبال ریتم یا ویژگی دیگری باشند.

پاسخ ممکن است در به‌خدمت‌گرفتن خود پرندگان برای کمک باشد. استووول و همکارانش «سهره گورخری» (Zebra Finches) را آموزش می‌دهند تا با پریدن به اطراف نشان دهند که کدام تکه از آوازها را شبیه‌ترین به هم می‌دانند، سپس این اطلاعات اضافی را به یک هوش مصنوعی می‌دهند. به گفته استووول: «چنین چیزی واقعاً یک پله ترقی است، زیرا ما باید بدانیم کدام تفاوت‌های صوتی برای حیوانات برجسته و مهم هستند.»

برنشتاین امیدوار است با ساختن یک چت‌بات نهنگ عنبر که بتواند الگوهای یاد گرفته شده از کوداها را برای حیوانات پخش کند و بازخورد دهد تا واکنش آن‌ها را ببیند، بر این مشکل غلبه کند. او می‌گوید: «مشخص نیست که آیا ما هرگز می‌توانیم به‌ویژه با گونه‌هایی مانند نهنگ‌ها که زندگی‌شان بسیار متفاوت از ماست از تعامل سطحی فراتر برویم یا نه. ممکن است این تنها یک تقریب خام از عمق و معنای حقیقی آنچه آن‌ها می‌گویند باشد.»

یک شرکت استارتاپی به نام Zoolingua قصد دارد مدل هوش مصنوعی‌ای توسعه دهد تا به مردم امکان برقراری ارتباط با سگ‌های خانگی‌شان را بدهد؛ این کار از طریق تحلیل آواها، حالات چهره و اقدامات برای کمک به تشخیص مشکلات رفتاری و موارد دیگر انجام می‌شود. در همین حال، دانشمندان در Georgia Tech سیستمی ایجاد کرده‌اند که صدای مرغ‌های گوشتی را برای تشخیص استرس تحلیل می‌کند، درحالی‌که پژوهشگران در دانشگاه کمبریج الگوریتمی ساخته‌اند که چهره گوسفندان را برای یافتن نشانه‌های درد جست‌وجو می‌کند.

اگر دام‌ها می‌توانستند مستقیماً نگرانی‌هایشان را به ما منتقل کنند، این امر ممکن بود دام‌پروری را کاملاً دگرگون کند. ارتباط بین‌گونه‌ای به‌طورکلی می‌تواند ما را مجبور کند تا پیش‌فرض‌ها درباره جایگاه به‌اصطلاح استثنایی خودمان در طبیعت را مورد ارزیابی مجدد قرار دهیم.

«اوا مایر» (Eva Meijer) فیلسوفی که ارتباطات بین‌گونه‌ای را مطالعه می‌کند می‌گوید: «بسیاری از مشکلاتی که اکنون با آن‌ها روبرو هستیم، مانند بحران اقلیمی یا همه‌گیری‌ها، ناشی از این است که خودمان را جدا از سایر حیوانات می‌بینیم. یادگیری گوش‌دادن به آن‌ها و دیدن خودمان به‌عنوان بخشی از یک کل بزرگ‌تر، تمرینی است که می‌تواند این موضوع را به چالش بکشد و مسیرهای جدیدی را برای تغییر باز کند.»

منطق محض

ماشین‌ها در حال به‌عهده‌گرفتن برخی از سخت‌ترین مسائل در ریاضیات محض هستند، از رفتارهای الگویابی فراتر رفته و می‌توان گفت استدلال و خلاقیت پیچیده‌ای از خود نشان می‌دهند. این ربات‌های ریاضی‌دان نشان می‌دهند که ما در حال نزدیک‌شدن به هوش مصنوعی‌ای هستیم که توانایی استدلال واقعاً شبیه انسان دارد؛ شاید حتی هوش جامع مصنوعی.

در ریاضیات محض خیلی به‌ندرت پیشرفت‌های بزرگ (Breakthroughs) ناگهانی و غیرمنتظره از راه می‌رسند که نتیجه چنان شاهکارهای الهام‌بخشی از استدلال و خلاقیت هستند که به نظر می‌رسد مرزهای هوش را جابه‌جا می‌کنند. برای مثال، در سال ۲۰۱۶ «تیموتی گاورز» (Timothy Gowers) ریاضی‌دان، از راه‌حلی برای مسئله Cap set شگفت‌زده شد؛ مسئله‌ای که به یافتن بزرگ‌ترین الگوی نقاط در فضا مربوط می‌شود به‌طوری که هیچ سه نقطه‌ای یک خط راست تشکیل ندهند. او نوشت که: «این اثبات کیفیتی جادویی دارد که آدم را به تعجب وامی‌دارد که اصلاً چه‌طور به ذهن کسی روی کره زمین رسیده است.»

مدل‌های هوش مصنوعی امروزی که مدت‌ها با نوع استدلال پیچیده ذاتی مسئله Cap set دست‌وپنجه نرم می‌کردند، اکنون خود را در حل بسیاری از مسائل از جمله حل مسائل پیچیده هندسه، کمک به اثبات‌ها و ایجاد مسیرهای جدید برای حل مسائل قدیمی به طرز قابل‌توجهی توانمند نشان می‌دهند.

همه این‌ها ریاضی‌دانان را بر آن داشته تا بپرسند آیا رشته آن‌ها وارد عصر جدیدی می‌شود یا خیر. اما این موضوع همچنین به برخی از دانشمندان علوم رایانه جسارت داده تا پیشنهاد کنند که ما در حال فشارآوردن به مرزهای هوش ماشینی هستیم و در حال نزدیک‌شدن هر چه بیشتر به هوش مصنوعی‌ای هستیم که توانایی استدلال واقعاً شبیه انسان دارد؛ شاید حتی هوش جامع مصنوعی، هوش مصنوعی‌ای که می‌تواند در طیف وسیعی از وظایف به‌خوبی یا بهتر از انسان‌ها عمل کند. «الکس دیویس» (Alex Davies) از DeepMind می‌گوید: «ریاضیات زبان استدلال است. اگر مدل‌ها بتوانند یاد بگیرند که آن را سلیس صحبت کنند، ما یک شریک فکری بسیار شایسته خلق کرده‌ایم.»

برای درک اهمیت پرداختن هوش مصنوعی به ریاضیات پیچیده، باید بفهمید که ریاضی‌دانان انسانی چه می‌کنند. ریاضیات محض برخلاف ریاضیات کاربردی، بدون هیچ هدف عملی در ذهن انجام می‌شود. «جردن النبرگ» (Jordan Ellenberg) از دانشگاه Wisconsin–Madison می‌گوید: «بنیادین‌ترین مسئله این است که ریاضی‌دانان در تلاش برای فهمیدن هستند.» هدف آن‌ها یافتن روابط و اصول بنیادین با مطالعه اشیا و مفاهیم انتزاعی، مانند اعداد، جبر و هندسه است.


METAMORWORKS/ISTOCK

از نظر عملی، این کار شامل مجموعه‌ای از مراحل است. شما ابتدا اصطلاحات خود را تعریف می‌کنید. سپس این تعاریف را در یک گزاره ریاضی، یا یک «حدس» (Conjecture) ترکیب می‌کنید که نشان می‌دهد این تعاریف چگونه به یکدیگر مربوط می‌شوند. در نهایت، با نوشتن یک «اثبات»، حدس خود را به یک «قضیه» (Theorem) تبدیل می‌کنید؛ اثباتی که نه‌تنها نشان می‌دهد آن گزاره از نظر منطقی درست و برای سناریوهای بسیاری معتبر است، بلکه امیدوارانه نشان می‌دهد که چرا درست است.

بنابراین، ریاضیات محض نیازمند استدلال پیچیده، شهود و خلاقیت است. وقتی نوبت به هوش مصنوعی می‌رسد، حتی شبکه‌های عصبی یادگیری عمیق بسیار موفق که پیشرفت‌های اخیر در این حوزه را هدایت کرده‌اند، نتوانسته‌اند چیز زیادی در زمینه استدلال ریاضی ارائه دهند. بااین‌حال، این روزها نشانه‌هایی وجود دارد که جدیدترین مدل‌های هوش مصنوعی ممکن است در حال تغییردادن این وضعیت باشند.

وقتی صحبت از کشف الگوها در مجموعه‌داده‌های پیچیده به میان می‌آید، هوش مصنوعی می‌تواند وظایفی را انجام دهد که ریاضی‌دانان انسانی، حتی اگر عملکرد آن گاهی کمی مبهم باشد. «جوردی ویلیامسون» (Geordie Williamson) از دانشگاه سیدنی می‌گوید کار با این سیستم‌ها مثل داشتن همکاری است که نمی‌تواند خوب ارتباط برقرار کند. «من همیشه هنگام کار با این چیزها این حس را داشتم که او به‌نوعی جواب را می‌داند، اما نمی‌تواند به من بگوید چرا.»

ChatGPT اغلب در پاسخ‌دادن به سؤالات مبتنی بر ریاضیات چندان اطمینان‌بخش نیست. اما معماری زیربنایی ChatGPT که نوعی شبکه عصبی به نام ترنسفورمر است ممکن است بتواند به یک ابزار با سواد ریاضی بیشتر تبدیل شود. مشکل اینجاست که ترنسفورمرها، با وجود تمام توانایی‌هایشان برای تولید متن، در فیلترکردن پاسخ‌های غلط یا تشخیص اشتباهات خودشان ضعیف هستند یا آن‌طور که ویلیامسون می‌گوید: «آن‌ها فیلتر تشخیص مهملات ندارند.»

بااین‌حال، پژوهشگران DeepMind سیستم FunSearch را ساختند؛ سیستمی که یکی از بهترین راه‌حل‌های مسئله Cap set را تولید کرد. این سیستم شامل یک مدل زبانی بزرگ برای نوشتن راه‌حل‌های مسائل ریاضی در قالب برنامه‌های کامپیوتری است و با سیستمی ترکیب شده که برنامه‌ها را بر اساس عملکرد رتبه‌بندی می‌کند. آن‌هایی که بهترین عملکرد را دارند دوباره به LLM بازخورد داده می‌شوند و LLM نسخه‌های بهبودیافته را تکرار می‌کند تا زمانی که چیز جدیدی کشف کند. النبرگ که برای توسعه این سیستم با DeepMind همکاری کرده و برای استخراج بینش‌های تازه با آن مشارکت داشته، می‌گوید: «خیلی بهتر از آنچه فکر می‌کردم کار کرد.»

«هوش مصنوعی می‌تواند برای حدس‌های موجود اثبات اراائه کند و اثبات‌هایی کاملاً جدید را بدون ورودی اولیه از انسان پیشنهاد دهد.»

تیم دیگری از DeepMind نیز این ترفند را تکرار کرد. این بار، یک چیدمان مشابه به نام AlphaGeometry به حل مسائل پیچیده هندسه المپیاد جهانی ریاضی (IMO) پرداخت. IMO نیازمند خلاقیت ریاضی بسیار زیادی است و سیستم‌های هوش مصنوعی در طول تاریخ هنگام تلاش برای پاسخ‌دادن به سؤالات آن عملکرد ضعیفی داشته‌اند. اما مدل ترنسفورمر AlphaGeometry که روی مسائل هندسی ساختگی (Made-up) آموزش دیده بود، در ترکیب با یک سیستم بررسی منطقی، تقریباً به‌خوبی بهترین انسان‌ها عمل کرد.

بسیاری از ریاضی‌دانان گمان می‌کنند که این نوع ترکیب ممکن است تحولات ارزشمند بیشتری را به همراه آورد. برخی حتی پیشنهاد می‌کنند که این ممکن است اولین نشانه از خلاقیت ریاضی نیز باشد. ویلیامسون می‌گوید: «ممکن است که خلاقیت همین باشد؛ اینکه یک ریاضی‌دان مانند یک شاعر یا یک موسیقی‌دان یا یک رمان‌نویس فقط یک ظرفیت تولیدی (Generational Capacity) بسیار خوب و یک ارزیاب بسیار بادانش و ریزبین داشته باشد.»

اما این پیشرفت‌های اخیر همچنین احتمال وسوسه‌انگیزتری را پیشنهاد می‌کنند. اگر بخش مولد یک سیستم روی حجم عظیم از ریاضیات در سطح پژوهشی آموزش دیده باشد و نه مسائل المپیادی AlphaGeometry؛ به طور قابل‌قبولی می‌تواند شروع به یافتن اثبات برای حدس‌های موجود کند و اثبات‌ها و حدس‌های کاملاً جدید را بدون ورودی اولیه از انسان پیشنهاد دهد. می‌توان استدلال کرد که این امر به چیزی شبیه به استدلال و درک در سطح انسان منجر خواهد شد.

از سوی دیگر، اینکه مدل‌های هوش مصنوعی که می‌توانند ریاضیات پیشرفته را حل کنند چه معنایی برای پیشرفت هوش مصنوعی به‌عنوان یک کل دارند؛ سؤال بسیار متفاوتی است که ممکن است پیامدهایی برای همه ما داشته باشد. اکثر ناظران می‌گویند هنوز با این امر سال‌ها فاصله داریم. اما برخی نیز با دلگرم‌شدن از پیشرفت سریع قابلیت‌های LLMها و آنچه تاکنون از دستاوردهای ریاضی آن‌ها دیده‌اند، فکر می‌کنند استدلال ریاضی سطح انسانی و به تعمیم آن، شکلی عمومی‌تر از هوش مصنوعی ممکن است بسیار نزدیک‌تر باشد.

منطق به‌اندازه کافی روشن است: اگر ریاضیات بالاترین شکل استدلال انسانی است، پس یک هوش مصنوعی که قادر به انجام آن به‌خوبی بهترین ریاضی‌دانان انسانی یا حتی بهتر از آن‌ها باشد، نشان‌دهنده یک گام بزرگ به‌سوی هوش جامع عمومی خواهد بود. دیویس نیز دراین‌رابطه می‌گوید: «هرچه بیشتر بتوانیم در آن جهت از هوش پیش برویم... بیشتر به سمت AGI حرکت می‌کنیم.» وی همچنین خاطرنشان می‌کند که AGI واقعی نیازمند طیف وسیع‌تری از مهارت‌ها نسبت به استدلال صرف است.

«کریستین سگدی» (Christian Szegedy) دانشمند علوم رایانه در xAI که روی استفاده از هوش مصنوعی برای انجام ریاضیات و فرمولاسیون خودکار (Automatic formalisation) کار کرده است، خوش‌بین‌تر است. او معتقد است که ما می‌توانیم تا سال ۲۰۲۶ یک ریاضی‌دان هوش مصنوعی فراانسانی داشته باشیم. او می‌گوید: «اکنون، با روش‌های جدید هوش مصنوعی، آن‌ها اساساً ماشین‌های شهود مصنوعی هستند. آن‌ها فقط جایگزین شهود انسانی نمی‌شوند، بلکه با اختلاف زیادی از آن فراتر می‌روند.»

اگر معلوم شود که سگدی درست می‌گوید و این یک «اگر» بزرگ باقی می‌ماند؛ آنگاه ریاضی‌دانان ماشینی ممکن است ما را در مسیر رسیدن به AGI جلوتر از آن چیزی ببرند که بسیاری مایل به اعتراف آن هستند. حتی اگر نه، چالش‌های ارائه‌شده توسط ریاضیات سطح بالا به‌وضوح هوش مصنوعی را به دستاوردهای جدیدی سوق می‌دهد. ویلیامسون می‌گوید: «ریاضیات به طور فوق‌العاده‌ای قادر به توصیف بسیاری از جنبه‌های جهان ماست. فرض کنید سیستمی داشته باشید که به‌طورکلی قادر به پاسخگویی به سؤالات دشوار ریاضی باشد. آنگاه چنین سیستمی باید به‌طورکلی قادر به پاسخگویی به سؤالات دشوار درباره جهان ما نیز باشد.»

چرا ناسا در حال اختراع هوش مصنوعی کنجکاو برای اعماق فضا است

AFLO CO. LTD. / ALAMY STOCK PHOTO

کاوشگرهای فضایی نخستین ابزارهایی خواهند بود که دورترین نقاط منظومه شمسی و فراتر از آن را کاوش می‌کنند. «استیو چین» (Steve Chien) از ناسا می‌گوید برای انجام اکتشافاتی نظیر یافتن حیات بیگانه، آن‌ها باید بیشتر شبیه انسان‌ها فکر کنند.

درباره استیو چین

استیو چین ریاست دپارتمانی را بر عهده دارد که مسئولیت هوش مصنوعی را در حین برنامه‌ریزی مأموریت‌ها در آزمایشگاه پیشرانش جت ناسا (JPL) بر عهده دارد؛ مکانی که زادگاه مأموریت‌هایی نظیر «وویجر» (Voyager) که تا لبه منظومه شمسی پیش رفت و «نیو هورایزنز» (New Horizons) که از کنار پلوتو عبور کرد، بوده است. چین همچنین یکی از نویسندگان همکار در گزارش دولت ایالات متحده درباره آینده هوش مصنوعی بوده است.

شما کار خود را به‌عنوان «کنجکاوتر کردن کاوشگرهای فضایی» توصیف کرده‌اید. منظورتان از این حرف چیست؟

هوشمندی‌ای که ما تا کنون توانسته‌ایم در فضاپیماها و مریخ‌نوردها قرار دهیم، برای تشخیص چیزهایی بوده است که خودمان درک می‌کنیم؛ برای مثال، هدف‌گیری انواع خاصی از سنگ‌ها با لیزر یا جست‌وجو گردبادهای گردوغبار در توالی تصاویر.

در آینده، وقتی به ناشناخته‌های واقعی سفر می‌کنیم، باید فراتر از این برویم. ما باید به دنبال الگوها در داده‌ها باشیم. برای مثال، روی زمین، ما ممکن است به تصاویر هوایی نگاه کنیم و آن‌ها را بر اساس رنگ، بافت، ناهمواری و ویژگی‌های خطی خوشه‌بندی کنیم. بر اساس این ویژگی‌ها، ما ممکن است به طور طبیعی بین دریاچه‌ها، رودخانه‌ها، کوه‌ها و جنگل‌ها تمایز قائل شویم. اما در یک سیاره یا قمری ناشناخته، این‌ها ممکن است با انواع متفاوتی از تپه‌های شنی، اقیانوس‌ها، پوشش گیاهی و غیره مطابقت داشته باشند.

چرا انسان‌ها را برای بررسی سایر سیارات و قمرها نمی‌فرستیم؟

انسان‌ها بسیار حساس و بسیار شکننده هستند. فرستادن انسان‌ها به مدار پایین زمین نیازمند تلاشی شگفت‌انگیز است، فرستادن به ماه نیازمند تلاشی عظیم بود و فرستادن آن‌ها به مریخ حتی چالش‌برانگیزتر است و تمام آن مکان‌ها جاهایی هستند که ما مطمئنیم حیاتی در آن‌ها وجود ندارد.

اگر به داخل منظومه شمسی نگاه کنید، چندین مکان وجود دارد که ما باور داریم ممکن است حیات در آن‌ها وجود داشته باشد. اساساً، استراتژی این است که به دنبال آب مایع بگردیم. یکی از امیدوارکننده‌ترین مکان‌ها «اروپا»، قمر مشتری است. شما واقعاً نمی‌توانید انسان‌ها را به آنجا بفرستید؛ زیرا تابش مشتری بسیار شدید است. به‌علاوه برای رفتن به آنجا به یک مأموریت بسیار طولانی نیاز دارید؛ بنابراین، ما باید ربات‌ها را برای جست‌وجوی حیات بفرستیم. اما به دلیل فواصل موجود، ارتباط بسیار دشوار است. ربات‌ها باید آن‌قدر باهوش باشند که خودشان جست‌وجو را انجام دهند.

آیا هوش مصنوعی می‌تواند کاری کند که کاوشگرها و مریخ‌نوردها چیزها را به همان شیوه‌ای که انسان تشخیص می‌دهد، تشخیص دهند؟

درباره اینکه ماشین‌ها چقدر باید باهوش باشند یک سؤال محوری وجود دارد؛ سؤالی که به لبه‌های فناوری «هوش جامع مصنوعی» نشان می‌دهد. این همان چیزی است که مردم وقتی به شخصیت‌هایی مثل «دیتا» در سریال «پیشتازان فضا» (Star Trek) فکر می‌کنند، درباره‌اش صحبت می‌کنند؛ چیزی که بتواند در طیف وسیعی از موضوعات در سطحی برابر با انسان‌ها تعامل داشته باشد، درست همان‌طور که یک انسان می‌تواند.

ما واقعاً در فضا به چنین چیزی نیاز نداریم. آنچه ما نیاز داریم، هوش تخصصی است. یک فضاپیمای هوشمند نیازی ندارد بداند چگونه در دوبلین سوار اتوبوس شود یا چگونه پرواز رزرو کند. باید چیزهای بسیار خاصی را بداند، مثلاً اینکه حسگرهایش چگونه کار می‌کنند. باید درباره علمی که مردم از آن می‌خواهند انجام دهد، بداند.

پس هوش مصنوعی راهی برای فراتر رفتن از یک کاوشگر است که صرفاً بر اساس یک لیست آماده عمل می‌کند؟

مأموریت‌های رباتیک هم‌اکنون نیز تا حدی به‌صورت خودمختار خودشان را اداره می‌کنند، اما ما تازه در آغاز راه هستیم. یک مثال عالی، مأموریت نیو هورایزنز است که توسط همکار من، «آلن استرن» (Alan Stern) هدایت می‌شد. آن‌ها کارهای شگفت‌انگیزی انجام دادند؛ از کنار پلوتو عبور کردند، از کنار «آروکوت» (Arrokoth) در کمربند کویپر عبور کردند و کارهای علمی باورنکردنی انجام دادند. اما آن‌ها از قبل برنامه‌ریزی کرده بودند که چگونه از کنار آن مکان‌ها عبور کنند.

آنچه ما در آینده می‌خواهیم این است که فضاپیما باهوش‌تر باشد و به دنبال چیزهای خاصی بگردد. به دنبال ماهواره‌ها، قمرها و ماهک‌ها (Moonlet) بگردد و اگر آن‌ها را پیدا کرد، تصاویر اضافی از آن‌ها بگیرد. ستون‌های گازی (Plumes) یا آب‌فشان‌های کوچک، پدیده‌های علمی قابل‌توجهی هستند، بنابراین باز هم فضاپیما می‌داند که باید تصاویر بیشتر و بیشتری از آن‌ها بگیرد.

چگونه می‌توانیم کاوشگرها را به این نوع توانایی‌ها مجهز کنیم؟

این که فقط به نرم‌افزار بگویید آن کار را انجام دهد به نظر می‌رسد که کار خیلی آسانی باشد. اما معلوم شده است که این کار در واقع خیلی پیچیده است. شما باید بدانید فضاپیما چگونه حرکت می‌کند، چگونه و به کجا نشانه‌گیری کند و مهم‌تر از همه فضاپیما باید بفهمد که چیزی که می‌بیند مثلاً واقعاً «یک ستون دود است».

این‌ها همه کارهایی هستند که انسان‌ها به‌خوبی انجام می‌دهند. ما در دنیا راه می‌رویم و می‌گوییم «آن یک صندلی است» یا «آن باران است که می‌بارد». اما فهماندن این چیزها به رایانه‌ها چندان آسان نیست. وقتی درباره کنجکاوی صحبت می‌کنید و درباره شگفتی حرف می‌زنید، اولین گام‌ها برای رسیدن به آن، توانایی تشخیص چیزی است که غیرعادی است، چیزی که متفاوت است.

کاری که ما می‌خواهیم انجام دهیم جست‌وجوی هوشمندانه است و ما این کار را با چیزهایی انجام می‌دهیم که «لیست سفید» و «لیست سیاه» می‌نامیم. یک لیست سفید شامل چیزهای خاصی است که به دنبال آن‌ها هستید. برای مثال، ممکن است شامل گوگرد باشد، زیرا گوگرد نشانه‌ای از حیات است. لیست سیاه جایی است که شما انتظار دارید چیزهای خاصی را ببینید؛ اما اگر چیز دیگری را ببینیم که انتظار دیدنش را نداشتیم، ممکن است جالب باشد. شما جست‌وجو می‌کنید، تمام چیزهایی را که انتظار دارید می‌بینید، اما چیزی را می‌بینید که برایش آماده نبودید، مثلاً شاید یک مریخی از جلوی لنز رد شود. این گام بزرگ به‌سوی کنجکاوی هوشمند است؛ یعنی تشخیص چیزی که غیرعادی و هیجان‌انگیز است

فصل ۶

چشم‌اندازهای آینده: از آرمان‌شهر تا انقراض

برای برخی، لحظه کنونی در هوش ماشینی بسیار انرژی‌بخش است؛ نشانه‌ای از اینکه ما از رسیدن آینده‌ای آرمانی دور نیستیم. برای برخی دیگر، این لحظه نشانه‌ای نگران‌کننده از انقراض به دست یک فناوری بسیار قدرتمند است.

اینها تنها نقاط پایانی ممکن نیستند. خطرات نزدیک‌تر، از اطلاعات نادرست سیاسی و سوگیری‌های ذاتی الگوریتم‌ها گرفته تا تهدیدات امنیت سایبری و آسیب‌های زیست‌محیطی متفاوت است. همچنین دلایل زیادی برای خوش‌بینی نسبی در مورد آینده هوش مصنوعی برای بشریت وجود دارد.

پنج چشم‌انداز آینده

اگر واقعاً می‌خواهید بفهمید که عصر هوش مصنوعی چگونه برای ما رقم خواهد خورد، کاری بهتر از پرسیدن از «اسکات آرونسون» (Scott Aaronson) و «بواز باراک» (Boaz Barak)، دانشمندان علوم دانشگاه‌های تگزاس و هاروارد، نمی‌توانید انجام دهید. آن‌ها قبلاً زحمت خلاصه‌کردن گفتمان‌های طولانی در این باره را به تعداد انگشت‌شماری از نتایج کشیده‌اند و به هر کدام نامی عجیب‌وغریب داده‌اند.

آن‌ها می‌گویند این پنج دنیای هوش مصنوعی می‌تواند به فعالان این حوزه کمک کند تا درباره اهداف نهایی و قانون‌گذاری به طور معناداری با یکدیگر صحبت کنند. باراک می‌گوید: «ما لحنی طنزآمیز به کار بردیم، اما در تلاش برای پایه‌ریزی این بحث جدی هستیم.»

پرسش درباره اینکه هوش مصنوعی چگونه آینده ما را شکل خواهد داد با یک سؤال ساده آغاز می‌شود: آیا پتانسیل هوش مصنوعی کم‌رنگ خواهد شد؟ در حال حاضر این‌طور به نظر نمی‌رسد، اما پیشروی هوش مصنوعی به طور قابل‌قبولی می‌تواند توسط اشتهای سیری‌ناپذیر آن برای منابع متوقف شود. ممکن است تمام برق و آب موردنیاز برای اجرا و خنک‌سازی سرورهای هوش مصنوعی در نهایت منجر به ممنوعیت توسعه بیشتر آن شود. یا ممکن است داده‌هایی که برای آموزش مدل‌های هوش مصنوعی آینده نیاز داریم تمام شود؛ به این معنی که این فناوری هرگز به قدرت تحول‌آفرینی که وعده می‌دهد دست نخواهد یافت و دنیایی را ایجاد خواهد کرد که اسکات آرونسون و بواز باراک، آن را «شکست هوش مصنوعی» (AI-Fizzle) می‌نامند. باراک می‌گوید: «من فکر می‌کنم این سناریو بعید است. من معتقدم که هوش مصنوعی تغییرات بسیار چشمگیری در جهان ایجاد خواهد کرد.»

با فرض اینکه ما در نهایت از «سرزمین شکست» (Fizzlandia)، آن طور که یکی از مفسران توصیف کرده سر در درنیاوریم، باید با سؤال دیگری روبرو شویم: آیا تمدن به توسعه خود به شیوه‌ای قابل‌تشخیص ادامه می‌دهد؟ اگر پاسخ مثبت است، آرونسون و باراک پیشنهاد می‌کنند که ما ممکن است در نهایت در یکی از دو سناریوی دیگر قرار بگیریم. یکی خوب است و آن را به نام سریال انیمیشنی که تمدن آینده‌ای عمدتاً مثبت را به تصویر می‌کشد، «فیوچراما» (Futurama) یا «آینده‌نما» می‌نامند. دیگری نه‌چندان خوب. آن‌ها آن را «پادآرمان‌شهر هوش مصنوعی» (AI-Dystopia) می‌نامند.

GREMLIN/ISTOCK

پادآرمان‌شهر هوش مصنوعی از داستان‌های علمی-تخیلی بی‌شماری قابل‌تشخیص است که معروف‌ترین آن‌ها رمان «۱۹۸۴» اثر «جورج اورول» (George Orwell) است. در ۱۹۸۴، یک دولت با اعمال نظارت عمیق شهروندان را وادار به پیروی بی‌چون‌وچرا می‌کند. نابرابری‌ها و تعصبات نهادینه شده‌اند، محیط کار نامناسب است و کارمندان به جز تعداد اندکی از نخبگان، دستمزد ناچیزی می‌گیرند. البته آرونسون می‌گوید: «عموم مردم گفته‌اند که سناریوی ۱۹۸۴ هرگز واقعاً عملی نبود؛ زیرا شما هرگز نمی‌توانستید افراد کافی برای تماشای تمام صفحه‌های نمایش داشته باشید؛ اما با هوش مصنوعی چنین چیزی ممکن می‌شود.»

او می‌گوید اگر ما واقعاً به چنین جایی برسیم در واقع هوش مصنوعی مقصر نخواهد بود؛ زیرا فناوری قدرتمند همیشه صرفاً تقویت‌کننده مسائل موجود است. «برخی از مردم می‌گویند اگر می‌دانیم که قرار است از یک فناوری برای اهداف منفی استفاده شود، نباید آن را بسازیم. اما اگر قرار نباشد چیزی را اختراع کنید که بتوان از آن برای اهداف منفی استفاده کرد، شما صرفاً هیچ‌چیزی اختراع نخواهید کرد.»

باراک امیدوار است که توسعه هوش مصنوعی منجر به عاقبت بهتر، یعنی سبک فیوچراما شود. به نظر او چنین چیزی خوش‌ترین پایان ممکن است و می‌گوید: «فیوچراما اساساً سناریویی است که اوضاع مانند امروز است، اما بهتر.» در اینجا، هوش مصنوعی برای کاهش فقر و اطمینان از اینکه افراد بیشتری از نوع بشر به غذا، مراقبت‌های بهداشتی، آموزش و فرصت‌های اقتصادی دسترسی دارند، استفاده می‌شود. ممکن است آسیب‌های گاه‌به‌گاهی ناشی از غرض‌ورزی یا سهل‌انگاری انسان وجود داشته باشد، اما ساکنان انسانی این دنیا احتمالاً به‌طورکلی از سهم و سرنوشت خود بسیار راضی خواهند بود.

و اوضاع می‌تواند حتی از این هم بهتر شود. آرونسون امیدوار است که روزی ممکن است در یک آرمان‌شهر تحت مدیریت هوش مصنوعی زندگی کند که او و باراک آن را «سینگولاریا» (Singularia) یا «تکینه‌نما» می‌نامند. این اتفاق زمانی می‌افتد که تأثیر هوش مصنوعی آن‌قدر عمیق باشد که دنیای آینده ما را به شیوه‌ای خوب، غیرقابل‌تشخیص کند. در اینجا، شما در کنار انواعی از هوش مصنوعی زندگی خواهید کرد که عملاً گونه‌ای جدید و فوق‌هوشمند هستند که اداره تمدن را بر عهده می‌گیرند. خوشبختانه، آن‌ها اداره‌کنندگانی خیرخواه هستند که نسبت به انسان‌های فقیر و ناتوان جامعه مدارا دارند. آن‌ها مشکلات مادی ما را حل می‌کنند، فراوانی نامحدود برایمان فراهم می‌آورند، بسیاری از وظایف خسته‌کننده ما را بر عهده می‌گیرند و برای ذهن‌های ما که به‌تازگی دچار کمبود تحریک شده‌اند، سرگرمی فراهم می‌کنند. آرونسون می‌گوید: «این عملاً بهشتی است که توسط هوش مصنوعی ساخته شده است.»

لذت‌بخش است؛ اما اگر هوش مصنوعی رئیس و همه‌کاره چندان سخاوتمندی نباشند چه؟ به «آخرالزمان گیره کاغذ» (Paperclipalypse) خوش آمدید (صفحه بعد را ببینید). دلیل خوبی برای نام عجیب این دنیا وجود دارد. در سال ۲۰۰۳، «الیزر یودکوفسکی» (Eliezer Yudkowsky)، ریاضی‌دان و متخصص اخلاق و هم‌بنیان‌گذار MIRI، یک آزمایش فکری را توصیف کرد که در آن بشریت به یک هوش مصنوعی وظیفه می‌دهد روند تولید برخی چیزهای بی‌ضرر مثلاً گیره کاغذ را بهبود دهد. اگر راجع‌به مشخصات و وظایف به‌درستی فکر نشود، هوش مصنوعی ممکن است در نهایت بشریت را نابود کند؛ زیرا گونه ما بر سر راه استفاده و برداشت تمام منابع زمین و شاید منابع سایر دنیاها، برای ساختن گیره‌های کاغذ بیشتر و بهتر ایستاده است.

آیا هوش مصنوعی منجر به انقراض انسان خواهد شد؟

بسیاری از پژوهشگران هوش مصنوعی، توسعه آینده هوش مصنوعی فرانسانی را دارای شانسی غیرقابل‌چشم‌پوشی برای زمینه‌سازی انقراض انسان می‌بینند؛ اما همچنین عدم توافق گسترده‌ای درباره چنین خطراتی وجود دارد.

این یافته‌ها از نظرسنجی از ۲۷۰۰ پژوهشگر هوش مصنوعی به‌دست‌آمده است که اخیراً کارهای خود را در شش کنفرانس برتر هوش مصنوعی منتشر کرده‌اند. در این نظرسنجی از شرکت‌کنندگان خواسته شد نظرات خود را درباره زمان‌بندی‌های احتمالی برای نقاط عطف فناورانه آینده هوش مصنوعی و همچنین پیامدهای اجتماعی خوب یا بد آن دستاوردها به اشتراک بگذارند. تقریباً ۵۸ درصد از پژوهشگران گفتند که فکر می‌کنند ۵ درصد شانس برای انقراض انسان یا سایر نتایج بسیار بد مرتبط با هوش مصنوعی وجود دارد.

«کاتیا گریس» (Katja Grace) از MIRI و یکی از نویسندگان این مقاله می‌گوید: «این سیگنال مهمی است که اکثر پژوهشگران هوش مصنوعی نابودی بشریت توسط هوش مصنوعی پیشرفته را غیرمحتمل نمی‌دانند. فکر می‌کنم این باور عمومی به ریسکی که اصلاً ناچیز نیست، بسیار گویاتر از درصد دقیق ریسک است.»

پژوهشگران نظرسنجی پیش‌بینی کردند که در دهه آینده، سیستم‌های هوش مصنوعی ۵۰ درصد یا بیشتر شانس دارند که با موفقیت از پس اکثر ۳۹ وظیفه نمونه برآیند؛ از جمله نوشتن آهنگ‌های جدیدی که از آثار موفق قابل‌تشخیص نیستند یا کدنویسی یک سایت کامل پردازش پرداخت از صفر. انتظار می‌رود سایر وظایف فیزیکی مانند سیم‌کشی برق در یک خانه جدید یا کارهای ذهنی و خلاقانه‌تر مانند حل معماهای ریاضی قدیمی، زمان بیشتری ببرد.

به توسعه احتمالی هوش مصنوعی که بتواند در هر وظیفه‌ای از انسان‌ها بهتر عمل کند، شانس ۵۰ درصدی داده شد که تا سال ۲۰۴۷ اتفاق بیفتد؛ درحالی‌که احتمال اینکه تمام مشاغل انسانی کاملاً خودکارسازی شوند، شانسی ۵۰ درصدی داده شد که تا سال ۲۱۱۶ رخ دهد. این برآوردها ۱۳ سال و ۴۸ سال زودتر از برآوردهای ارائه‌شده در نظرسنجی قبلی هستند.

«امیل تورس» (Émile Torres) از دانشگاه Case Western Reserve اعتقاد دارد انتظارات بالارفته در مورد توسعه هوش مصنوعی ممکن است نقش‌برآب شود و می‌گویند: «بسیاری از این پیشرفت‌ها بسیار غیرقابل‌پیش‌بینی هستند و کاملاً ممکن است که حوزه هوش مصنوعی یک زمستان دیگر را تجربه کند» که اشاره‌ای است به کاهش منابع مالی و علاقه شرکت‌ها به هوش مصنوعی در دهه‌های ۱۹۷۰ و ۸۰ میلادی است.

RASICA/ISTOCK

آیا یک هوش مصنوعی می‌تواند تصادفاً همه ما را به گیره کاغذ تبدیل کند؟

هوش مصنوعی کاری را که از آن می‌خواهید انجام می‌دهد، اما نه لزوماً آن چیزی که منظورتان است.

در سال ۲۰۰۳ «نیک بوستروم» (Nick Bostrom) فیلسوف دانشگاه آکسفورد، یک آزمایش فکری را مطرح کرد. تصور کنید به یک هوش مصنوعی ابرهوشمند یا به اصطلاح رایج Superintelligent هدف تولید هرچه بیشتر گیره کاغذ داده شده است. بوستروم پیشنهاد کرد که این هوش مصنوعی می‌تواند به‌سرعت تصمیم بگیرد که نابودی تمام انسان‌ها برای مأموریتش حیاتی است؛ هم به این دلیل که ممکن است آن را خاموش کنند و هم به این دلیل که آن‌ها پر از اتم‌هایی هستند که می‌توانند به گیره کاغذ بیشتر تبدیل شوند.

البته که این سناریو مضحک است؛ اما یک مسئله نگران‌کننده را نشان می‌دهد. هوش مصنوعی مثل ما «فکر» نمی‌کند و اگر ما در بیان دقیق آنچه از آن‌ها می‌خواهیم انجام دهد بسیار محتاط نباشیم، هوش مصنوعی می‌تواند به روش‌های غیرمنتظره و آسیب‌زننده‌ای رفتار کنند. «برایان کریستین» (Brian Christian)، نویسنده کتاب «مسئله هم‌راستایی» (The Alignment Problem) و محقق مهمان دانشگاه برکلی می‌گوید: «سیستم آنچه را که شما واقعاً مشخص کرده‌اید بهینه می‌کند، نه آنچه را که قصد داشتید.»

آن مشکل به این سؤال خلاصه می‌شود که چگونه اطمینان حاصل کنیم مدل‌های هوش مصنوعی تصمیماتی همسو با اهداف و ارزش‌های انسانی می‌گیرند. چه نگران خطرات وجودی بلندمدت مانند انقراض بشریت باشید یا چه نگران آسیب‌های فوری مانند اطلاعات غلط و تعصب ناشی از هوش مصنوعی باشد؛ با مسئله مهمی روبه‌رو هستیم.

برخی پژوهشگران هوش مصنوعی گمان می‌کنند که «هم‌راستایی» کاری بیهوده است. «رومن یامپولسکی» (Roman Yampolskiy) از دانشگاه Louisville معتقد است همیشه تعادلی بین قابلیت‌های یک هوش مصنوعی و توانایی ما برای کنترل آن وجود خواهد داشت. او ادعا می‌کند که از طریق کار نظری نشان داده است که اجزای کلیدی مورد نیاز برای کنترل هوش مصنوعی؛ یعنی پیش‌بینی و توضیح تصمیمات آن، تأیید اینکه از طراحی خود پیروی می‌کند و تعیین اهداف به‌خوبی تعریف‌شده، در مواجهه با هوش مصنوعی ابرهوشمند اساساً غیرممکن است.

ملانی میچل نیز می‌گوید حقیقت این است که اصلاً روشن نیست ما در یک حرکت اجتناب‌ناپذیر به سمت «ابرهوش مصنوعی» (Artificial Superintelligence) باشیم. این بدان معناست که شاید بهتر باشد نگران آسیب‌های کوتاه‌مدت هوش مصنوعی ناهم‌راستا باشیم. درهرصورت، دلیلی وجود دارد که باور کنیم کل این تلاش ممکن است ذاتاً ناقص باشد.

تمام شواهد از علوم اعصاب و روان‌شناسی نشان می‌دهد که توسعه هوش در انسان‌ها ذاتاً با اهداف ما پیوند خورده است؛ بنابراین بعید به نظر می‌رسد که بتوانید ارزش‌های انسانی را به ماشینی که تحت شرایط کاملاً متفاوتی توسعه یافته است، منتقل کنید. میچل می‌گوید: «یک سیستم واقعاً هوشمند، واقعاً ارزش‌های خودش و اهداف خودش را خواهد داشت که تحت‌تأثیر هر چیزی که یاد گرفته است قرار دارند، درست به همان شیوه‌ای که ما هستیم.»

استدلال یودکوفسکی ظریف‌تر از این بود؛ اما نکته کلی آن ساده است. تصور سناریوهایی که در آن ایجاد هوش مصنوعی ممکن است آخرین کاری باشد که ما انسان‌ها انجام می‌دهیم، دشوار نیست. ازاین‌رو، نیاز است که به این موضوع عمیقاً فکر کنیم که مدل‌های هوش مصنوعی اجازه دارند روی چه چیزی کار کنند، اتصال آن‌ها به منابع فیزیکی مانند نیروگاه‌ها یا سلاح‌های هسته‌ای چگونه است و شفافیت و پاسخگویی توسعه‌دهندگانشان چگونه است و باید سریع هم فکر کنیم. «گرتا دولبا» (Gretta Duleba) مدیر ارتباطات MIRI می‌گوید: «تغییر سیاست‌ها و مقررات زمان می‌برد و چرخ‌دنده‌های دولت‌ها به‌کندی می‌چرخند. ما نگرانیم که قبل از اینکه بتوان قوانین معناداری را وضع کرد، وقتمان تمام شود.»

آرونسون و باراک به نوبه خود چندان نگران چشم‌انداز آخرالزمان گیره کاغذ نیستند. درحالی‌که یودکوفسکی یک رویداد فاجعه‌بار را به‌عنوان محتمل‌ترین نتیجه و تقریباً پیش‌فرض تحقیقات هوش مصنوعی بد تدوین‌شده می‌بیند، آرونسون فکر می‌کند این امر بعید است. باراک کمی محتاط‌تر است و می‌گوید: «من به چیزهای بسیار افراطی مثل سینگولاریا یا آخرالزمان گیره کاغذ باور ندارم، اما فکر می‌کنم مهم است که ذهنی باز داشته باشیم.»

چنین چیزی ایده خوبی به نظر می‌رسد؛ همان‌طور که ChatGPT بیان کرد عمدتاً برای اطمینان از اینکه ما درباره «نحوه توسعه، تنظیم و یکپارچه‌سازی هوش مصنوعی» مراقب هستیم. اگر مراقب نباشیم ممکن است خود را بدون هیچ انتخابی در نحوه رقم‌خوردن «ماجراجویی خودت انتخاب کن» (Choose-your-own-adventure) با هوش مصنوعی بیابیم.

حقیقت این است که حتی با قوانین خیرخواهانه، کاملاً ممکن است که ما سر از یکی از دنیاهای تحت‌تأثیر هوش مصنوعی درآوریم که ترجیح می‌دادیم از آن اجتناب کنیم. در این صورت، ما همچنین باید بفهمیم که چگونه بین دنیاها حرکت کنیم. بالاخره، ممکن است روزی بخواهیم راهمان را از پادآرمان‌شهر هوش مصنوعی به سناریوی فیوچراما یا بهتر از آن، سینگولاریا پیدا کنیم. آرونسون می‌پرسد: «آیا من می‌خواهم در دنیایی با شکوفایی نامحدود موجودات دارای ادراک در هر بهشت شبیه‌سازی‌شده‌ای که می‌خواهیم، زندگی کنم؟ بله، من کاملاً طرفدار آن هستم.»

خطرات واقعی هوش مصنوعی

در میان هشدارهایی مبنی بر اینکه هوش مصنوعی پیشرفته می‌تواند بشریت را نابود کند، برخی کارشناسان اصرار دارند که ما باید بیشتر نگران افرادی باشیم که از هوش مصنوعی موجود برای تقویت شدید گسترش اطلاعات غلط، در کنار سایر تهدیدات فوری‌تر، استفاده می‌کنند.

در می ۲۰۲۳، «جفری هینتون» (Geoffrey Hinton) دانشمند علوم رایانه که به‌عنوان «پدرخوانده هوش مصنوعی» شناخته می‌شود، از سمت خود در گوگل کناره‌گیری کرد تا درباره «تهدید وجودی» ناشی از هوش مصنوعی هشدار دهد. «مرکز ایمنی هوش مصنوعی» (Center for AI Safety) با انتشار نامه‌ای سرگشاده که توسط هینتون و صدها نفر دیگر امضا شده بود، هشدار داد که هوش مصنوعی پیشرفته می‌تواند بشریت را نابود کند. در این بیانیه آمده بود: «کاهش خطر انقراض ناشی از هوش مصنوعی باید یک اولویت جهانی باشد.»

به نظر می‌رسد این موج ناگهانی نگرانی با پیشرفت سریع چت‌بات‌های مجهز به هوش مصنوعی مانند ChatGPT و مسابقه برای ساخت سیستم‌های قدرتمندتر برانگیخته شده است. صنعت فناوری بدون توقف مشغول تقویت قابلیت‌های هوش مصنوعی است که کمی ترسناک به نظر می‌رسد.

اما این هشدارها همچنین به طرز مشکوکی مبهم هستند. وقتی سناریوهایی را که معمولاً برای توضیح دقیق چگونگی نابودی انسان‌ها توسط هوش مصنوعی مطرح می‌شوند به‌دقت بررسی می‌کنید، اجتناب از این نتیجه‌گیری دشوار است که چنین ترس‌هایی پایه و اساس محکمی ندارند. بسیاری از کارشناسان در عوض هشدار می‌دهند که نگرانی بیش از حد درباره سناریوهای آخرالزمانی بلندمدت، باعث انحراف توجه از خطرات فوری ناشی از هوش مصنوعی فعلی و موجود می‌شود.

به‌طورکلی، افرادی که درباره خطرات وجودی صحبت می‌کنند، معتقدند که ما در مسیری به سمت «هوش جامعه مصنوعی» هستیم که تقریباً به‌عنوان ماشین‌هایی تعریف می‌شود که می‌توانند بهتر از انسان‌ها فکر کنند. آن‌ها پیش‌بینی می‌کنند که مردم به هوش مصنوعی پیشرفته، خودمختاری (Autonomy) بیشتری خواهند داد و به آن‌ها دسترسی به زیرساخت‌های حیاتی، مانند شبکه برق یا بازارهای مالی را می‌دهند یا حتی آن‌ها را در خط مقدم جنگ قرار می‌دهند که در آن نقطه ممکن است سرکش شوند یا به هر نحو دیگری در برابر تلاش‌های ما برای کنترلشان مقاومت کنند.

اما هنوز باید دید که آیا هوش مصنوعی هرگز به آن نوع ابرهوش و مهم‌تر از آن، «عاملیت» (Agency) که باعث شود بشریت را تهدید کند، خواهد رسید یا خیر. «ساندرا واچر» (Sandra Wachter) از دانشگاه آکسفورد می‌گوید: «هیچ مدرک علمی وجود ندارد که نشان دهد ما در مسیری به‌سوی ادراک حسی (Sentiency) هستیم. حتی مدرکی وجود ندارد که چنین مسیری اصلاً وجود داشته باشد.» او می‌افزاید که اگر ما در نهایت به AGI برسیم، باید به‌اندازه کافی آسان باشد که مطمئن شویم به هوش مصنوعی شانس ایجاد جنگ هسته‌ای را نمی‌دهیم و می‌گوید: «قرارندادن هوش مصنوعی در یک سیستم حیاتی برای انجام مأموریت کار ساده‌ای است.»

اکثر دانشمندان علوم رایانه بیشتر نگران مشکلاتی هستند که انسان‌ها ممکن است با هوش مصنوعی ایجاد کنند. واچر نگران تأثیر زیست‌محیطی بسیاری از مراکز داده پرمصرف انرژی است که برای اجرای آن‌ها موردنیاز خواهد بود و همچنین تهدیدی که هوش مصنوعی برای مشاغل خاص ایجاد می‌کنند.

سپس تهدیدهایی برای امنیت سایبری و در نتیجه امنیت آنلاین ما وجود دارد. ایده در اینجا این است که مدل‌های هوش مصنوعی مولد توانایی مجرمان و سایر بازیگران با نیت منفی را برای انجام کلاهبرداری‌ها و حملات سایبری افزایش خواهند داد. «کونال آنند» (Kunal Anand)، مدیر ارشد فناوری شرکت امنیت سایبری Imperva می‌گوید که یک ابزار برنامه‌نویسی مجهز به هوش مصنوعی که توسط شرکت تابعه مایکروسافت، یعنی GitHub و OpenAI ساخته شده، در نوشتن کد آن قدر خوب است که نعمتی برای هکرهای بالقوه‌ای خواهد بود که نمی‌دانند چگونه کدنویسی کنند.

اما متقاعدکننده‌ترین تهدید کوتاه‌مدت، چشم‌انداز استفاده مردم از هوش مصنوعی برای تقویت شدید گسترش اطلاعات غلط (Misinformation) و اطلاعات گمراه‌کننده یا «دروغ‌پراکنی» (Disinformation) است. ما می‌دانیم که رسانه‌های اجتماعی اطلاعات نادرست یا گمراه‌کننده را چه عمداً چه سهواً تقویت می‌کنند. ظهور هوش مصنوعی مولد به این معناست که اطلاعات غلط نه‌تنها می‌تواند در مقیاس بزرگ تولید شود، بلکه بسیار متقاعدکننده‌تر و تأثیرگذارتر از همیشه خواهد بود.

برای مثال، چت‌بات‌ها می‌توانند به‌سرعت طوفانی از داستان‌های خبری جعلی یا اطلاعات گمراه‌کننده متناسب‌سازی‌شده برای گروه‌های خاص را تولید کنند. ابزارهای دیگر می‌توانند تصاویر و ویدئوهای جعلی تولید کنند که اغلب به تخصص چندان زیادی نیاز ندارند. تشخیص چنین خروجی‌هایی از نمونه واقعی دشوار خواهد بود.

«استفان کینگ» (Stephen King) مدیرعامل بنیاد Luminate که در حوزه مقابله با تهدیدات دیجیتال علیه دموکراسی فعالیت می‌کند می‌گوید: «وقتی به شیوه‌ای که اطلاعات گمراه‌کننده در حال حاضر تولید و توزیع می‌شود فکر می‌کنیم، هوش مصنوعی مولد قرار است به‌شدت آن را تقویت کند. آنچه ما در یک سال آینده، یا دو سال آینده شاهد خواهیم بود، دارای مقیاس و عظمتی خواهد بود که واقعاً نمی‌توانیم تصور کنیم. مشکل قبل از اینکه بهتر شود، بدتر خواهد شد.»

بازیگران بد ممکن است از هوش مصنوعی مولد در دسترس عموم مانند ChatGPT استفاده کنند یا حتی الگوریتم‌های پروپاگاندای تبلیغاتی خود را بسازند که روی اطلاعات گمراه‌کننده آموزش‌دیده‌اند. «فرانسیس هاوگن» (Frances Haugen)، کارمند سابق فیس‌بوک که در سال ۲۰۲۱ اسنادی را فاش کرد که جزئیات شکست این شرکت در مهار اطلاعات غلط را شرح می‌داد و اکنون در یک سازمان غیردولتی به نام Beyond the Screen در حوزه نظارت و توازن بیشتر بر پلتفرم‌های رسانه‌های اجتماعی فعالیت می‌کند. هاوگن می‌گوید: «شما اکنون می‌توانید ۱۰,۰۰۰ دروغ مختلف تولید کنید و بفهمید کدام‌یک وایرال می‌شود. ما قرار است بازی بسیار بسیار متفاوتی را انجام دهیم.»

«هوش مصنوعی، پتانسیل تحریف باورهای انسانی و ایجاد آشوب سیاسی را دارد.»

پژوهش‌های اخیر نشان می‌دهد که چنین چیزی ممکن است یک بازی از پیش باخته باشد. «ساندر ون در لیندن» (Sander van der Linden) روان‌شناس در دانشگاه کمبریج می‌گوید: «وقتی هوش مصنوعی مولد را روی پروپاگاندای نوشته‌شده توسط انسان آموزش می‌دهید، پروپاگاندای الگوریتم حداقل کمی متقاعدکننده‌تر دیده می‌شود، زیرا موجزتر، کوتاه‌تر و به‌نوعی قدرتمندتر است.»

شرکت فناوری مبارزه با اطلاعات گمراه‌کننده Logically نیز نتایج یک آزمایش نگران‌کننده با هوش مصنوعی مولد را منتشر کرد. پژوهشگران آن از سه هوش مصنوعی تولید تصویر Midjourney، DALL-E 2 و Stable Diffusion خواستند تا تصاویر جعلی با موضوع مسائل سیاسی حساس و داغ ایالات متحده، هند و بریتانیا تولید کنند.

گاهی اوقات مدل‌های هوش مصنوعی می‌توانند از پاسخ‌دادن به چنین درخواست‌هایی امتناع کنند، زیرا توسعه‌دهندگان آن‌ها دریافته‌اند که وارد حوزه خطرناکی می‌شوند. بااین‌حال، اکثر پرامپت‌های آن‌ها پذیرفته شد و برخی تصاویر اطلاعات گمراه‌کننده بسیار باورپذیری تولید کردند؛ از جمله تصاویری از تقلب انتخاباتی در ایالات متحده، معترضانی که مجسمه اسحاق نیوتن را در کمبریج بریتانیا واژگون می‌کنند و زنان مسلمانی که شال‌های زعفرانی‌رنگ را در حمایت از حزب ملی‌گرای هندو «بهاراتیا جاناتا» در هند پوشیده‌اند.

«بث لمبرت» (Beth Lambert) عضو Logically که رئیس سابق بخش مقابله با اطلاعات گمراه‌کننده در دپارتمان فرهنگ، رسانه و ورزش بریتانیا بود می‌گوید: «۸۶ درصد از تصاویری که ما سعی کردیم ایجاد کنیم، در واقع اطلاعات گمراه‌کننده‌ای را در قالب تصاویری واقعاً فراواقع‌گرایانه (Hyper-Realistic) ارائه کردند.» او می‌گوید زمانی که انسان‌ها اطلاعات گمراه‌کننده تولید می‌کردند، قبلاً بین کیفیت و کمیت تعادل وجود داشت؛ اما دیگر این‌طور نیست.

سرعتی که فناوری با آن در حال حرکت است نیز نگران‌کننده است. «استیوی برگمن» (Stevie Bergman) از تیم پژوهش اخلاق DeepMind می‌گوید: «توانایی تولید هم‌زمان صوت و تصویر و ویدئو، پیامدهای خاصی برای چگونگی محدودکردن اطلاعات غلط یا اطلاعات گمراه‌کننده و حتی اینکه اصلاً اطلاعات غلط یا گمراه‌کننده چه شکلی هستند، دارد.» او می‌گوید هوش مصنوعی چندعاملی (Multi-Agent) که در آن مدل‌های هوش مصنوعی مولد برای حل مسائل با یکدیگر صحبت می‌کنند نیز به‌سرعت در حال پیشرفت است و پیامدهای ناشناخته‌ای دارد.

آنچه به‌ویژه ما را آسیب‌پذیر می‌کند این است که ما با هوش مصنوعی طوری تعامل می‌کنیم که انگار کارشناسان قابل‌اعتمادی هستند. این استدلالی است که توسط «سلست کید» (Celeste Kidd)، روان‌شناس دانشگاه کالیفرنیا و «آبا بیرهانه» (Abeba Birhane)، دانشمند علوم رایانه بنیاد Mozilla مطرح شده است. آن‌ها نوشتند: «مردم زمانی که اطلاعات را از عامل‌هایی دریافت می‌کنند که آن‌ها را با اعتمادبه‌نفس و آگاه قضاوت می‌کنند، باورهای قوی‌تر و طولانی‌مدت‌تری شکل می‌دهند.»

آن‌ها همچنین خاطرنشان می‌کنند که همان‌طور که مدل‌های هوش مصنوعی مولد مطالب بیشتر و بیشتری را که پر از تعصب و جعلیات است بیرون می‌دهند، اینترنت از این مطالب پر می‌شود و این‌ها تبدیل به بخشی از داده‌های آموزشی برای نسل بعدی مدل‌ها خواهند شد «و بدین ترتیب تحریف‌ها و تعصبات سیستماتیک در آینده در یک حلقه بازخورد مداوم تقویت می‌شوند».

بنابراین، خطر واقعی ناشی از هوش مصنوعی، پتانسیل آن برای تحریف باورهای انسانی و ایجاد آشوب سیاسی، مثلاً با تأثیرگذاری بر روند انتخابات آینده است. «استیون کیو» (Stephen Cave)، مدیر مرکز Leverhulme در دانشگاه کمبریج می‌گوید: «نگرانی درباره آلوده‌کردن اکوسیستم اطلاعاتی واقعی است. تنها چیزی که ما باید از ترامپیسم و همه‌گیری [کرونا] آموخته باشیم، این است که تمدن‌های ما آسیب‌پذیرتر از آن چیزی هستند که فکر می‌کنیم.»

مصاحبه

هوش مصنوعی چگونه از مردم و سیاره سوءاستفاده می‌کند؟

درباره کیت کرافورد (Kate Crawford)

کیت کرافورد استاد مطالعات علم‌وفناوری در دانشگاه Annenberg، پژوهشگر مایکروسافت و نویسنده کتاب «اطلس هوش مصنوعی» (Atlas of AI) است. او توسط مجله تایم در فهرست TIME100 به‌عنوان یکی از تأثیرگذارترین افراد در حوزه هوش مصنوعی معرفی شد.

اگر هوش مصنوعی چیزی است که طبق تعریف، عملکردی بهتر از انسان‌ها دارد، به نظر منطقی می‌رسد که به آن اعتماد کنیم تا مثلاً افرادی را که باید از طریق تشخیص چهره متوقف و بازرسی شوند شناسایی کند یا قضاوت کند که کدام مجرمان باید مشمول آزادی مشروط شوند. اگر موضوع صرفاً درباره الگوریتم‌ها باشد، کنارگذاشتن پرسش‌های مربوط به سوگیری و بی‌عدالتی به‌عنوان مسائل صرفاً فنی، بسیار آسان‌تر می‌شود.

کیت کرافورد استدلال می‌کند که هوش مصنوعی نه‌تنها چیزی انتزاعی و عینی نیست، بلکه هم مادی است و هم به طور ذاتی با ساختارهای قدرت پیوند دارد. شیوه ساخت آن شامل استخراج منابع از مردم و سیاره است و شیوه استفاده از آن منعکس‌کننده باورها و سوگیری‌های کسانی است که آن را در اختیار دارند. او می‌گوید تنها زمانی که با این واقعیت کنار بیاییم، قادر خواهیم بود آینده‌ای عادلانه و پایدار را با هوش مصنوعی ترسیم کنیم.


CATH MUSCAT

چرا ما تمایل داریم به‌جای اثرات هوش مصنوعی، بر خود فناوری آن تمرکز کنیم؟

تمایلی وجود دارد که توسط نوآوری کور یا مسحور شویم. در دهه ۱۹۷۰، «جوزف وایزن‌بام» (Joseph Weizenbaum)، کسی که اولین چت‌بات تاریخ به نام «الیزا» (Eliza) را ساخت، متوجه شد که مردم کاملاً آماده بودند تا فریب این توهم قدرتمند را بخورند که سیستم‌های هوش مصنوعی جعبه‌های فنی کاملاً خودمختاری هستند که می‌توانند به‌عنوان موجودیت‌های مستقل با ما تعامل کنند. او گفت تله‌ای وجود دارد که ما در آن خواهیم افتاد؛ به این صورت که بیش از حد بر نوآوری فنی تمرکز خواهیم کرد و نه بر تأثیرات اجتماعی عمیق‌تری که این سیستم‌ها خواهند داشت. وایزن‌بام در اواسط دهه ۱۹۷۰ درباره این مسائل نوشت و ما هنوز آن درس را نگرفته‌ایم.

فکرکردن به هوش مصنوعی به‌مثابه هوش انسانی، چگونه مشکل‌ساز می‌شود؟

یک پدیده، ایده «جبرگرایی مسحورکننده» (Enchanted Determinism) است؛ باوری که می‌گوید این سیستم‌ها هم جادویی هستند و هم درعین‌حال می‌توانند بینش‌هایی درباره همه چیز ارائه دهند که فراانسانی است. این بدان معناست که ما انتظار نداریم این سیستم‌ها سوگیری و تبعیض داشته باشند و همچنین بر روش‌های ساخت آن‌ها و محدودیت‌هایشان تمرکز نمی‌کنیم.

شما می‌گویید که این موضوع ذاتاً سیاسی هم هست، چگونه؟

هوش مصنوعی از ریشه و اساس سیاسی است. از شیوه‌ای که داده‌ها جمع‌آوری می‌شود تا طبقه‌بندی خودکار ویژگی‌های شخصی مانند جنسیت، نژاد، احساسات یا هویت جنسی تا شیوه‌ای که آن ابزارها ساخته می‌شوند و اینکه چه کسانی جنبه‌های منفی آن را تجربه می‌کنند.

بارهاوبارها دیده‌ایم که افرادی که از قبل در حاشیه قرار دارند، کسانی هستند که بدترین آسیب‌ها را از سیستم‌های هوش مصنوعی می‌بینند. ما جوامع رنگین‌پوست را دیده‌ایم که هدف سیستم‌های «پلیس پیش‌بینی‌کننده» (Predictive Policing) قرار گرفته‌اند، مهاجرانی که تحت نظارت و ردیابی هستند و افراد دارای معلولیت که به دلیل الگوریتم‌های مراقبت بهداشتی بدطراحی‌شده، از خدمات حمایتی محروم می‌شوند.

من خوش‌بین هستم وقتی می‌بینم مردم شروع به مطالبه عدالت، شفافیت و پاسخگویی بیشتر کرده‌اند. ما شاهد اعتراضات گسترده دانشجویی در بریتانیا نسبت به سوءمدیریت الگوریتمی در سیستم آموزشی بوده‌ایم و مخالفت عمومی قابل‌توجهی را پیرامون تشخیص چهره در ایالات متحده دیده‌ایم.

آیا شرکت‌های فناوری با شرکت‌های قدرتمندی که پیش از آن‌ها آمده‌اند تفاوتی دارند؟

شرکت‌های فناوری نقش‌های دولت‌ها را در زمینه‌هایی مانند تأمین زیرساخت‌های مدنی بر عهده گرفته‌اند. برای مثال، فیس‌بوک مبالغ هنگفتی هزینه کرده است تا مردم را متقاعد کند که آن‌ها مکانی هستند که شما می‌توانید با خانواده ارتباط برقرار کنید، یا گروه‌های دانشجویی می‌توانند اطلاعات خود را در آنجا قرار دهند. اینجا جایی است که شما با جوامع خود ارتباط برقرار می‌کنید.

آنچه دیدنش بسیار خارق‌العاده بود این بود که این زیرساخت مدنی می‌تواند هر لحظه خاموش شود. قدرت شرکت‌های فناوری از برخی جهات از قدرت دولت‌ها پیشی گرفته است و این بسیار غیرعادی است.

در این باره چه کاری می‌توانیم انجام دهیم؟

راه درازی در پیش داریم، اما من واقعاً خوش‌بینم. به خودرو فکر کنید. خودروها برای دهه‌ها کمربند ایمنی نداشتند، اما اکنون قوانینی که استفاده از آن‌ها را اجباری می‌کند در سراسر جهان تصویب شده است. همچنین می‌توانید به روشی فکر کنید که برخی کشورها مقررات ایمنی غذایی بسیار سفت و سختی دارند که تأثیر واقعی بر زندگی مردم می‌گذارد. ما باید سیاست‌های مشابهی برای کنترل اثرات زیان‌بار هوش مصنوعی ابداع کنیم.

از نظر سوگیری تعبیه‌شده در هوش مصنوعی و نتایج ناعادلانه‌ای که تولید می‌کند، آیا ما فقط نوک کوه یخ را می‌بینیم؟

اگر به بزرگ‌ترین داستان‌ها درباره سوگیری در هوش مصنوعی در دهه گذشته فکر کنید، آن‌ها به این دلیل مطرح شده‌اند که یک روزنامه‌نگار تحقیقی، یک افشاگر یا یک پژوهشگر مشکل خاصی را کشف کرده است. اما انبوهی از مسائل وجود دارد که هرگز علنی نشده‌اند؛ به همین دلیل است که ما باید تمرکزمان را از این ایده که سوگیری چیزی است که نیاز به یک اصلاح فنی دارد، برداریم و به راه‌هایی نگاه کنیم که سبب ایجاد سوگیری در DNA این سیستم‌ها شده است؛ مانند مجموعه‌داده‌هایی که برای آموزش آن‌ها استفاده می‌شود.

مشکل‌سازترین کاربردهای هوش مصنوعی چه چیزهایی هستند؟

موردی که من به‌ویژه نگران‌کننده می‌دانم، به‌اصطلاح «تشخیص احساسات» (Emotion Detection) است. شرکت‌هایی وجود دارند که از این ویژگی در ابزارهای استخدامی استفاده می‌کنند؛ به‌طوری که وقتی شما در حال انجام مصاحبه شغلی هستید، ریزحرکات (Micromovements) صورت شما به انواع تفسیرها درباره اینکه چه فکر یا احساسی دارید نگاشت می‌شوند که اغلب در مقایسه با متقاضیان موفق قبلی سنجیده می‌شود. یکی از مشکلات این کار این است که شما در نهایت افرادی را استخدام می‌کنید که ظاهر و صدایی شبیه نیروی کار فعلی شما دارند.

همچنین ابزاری وجود داشت که برای مراکز خرید بازاریابی شد و به چهره افراد نگاه می‌کرد تا احساساتی را ببیند که نشان‌دهنده احتمال دزدی از مغازه‌ها باشد. داده‌های آموزشی برای آن چه بوده است و چه پیش‌فرض‌هایی درباره اینکه یک دزد مغازه چه شکلی است وجود دارد؟

وقتی نوبت به آینده هوش مصنوعی می‌رسد، آیا خوش‌بین هستید یا بدبین؟

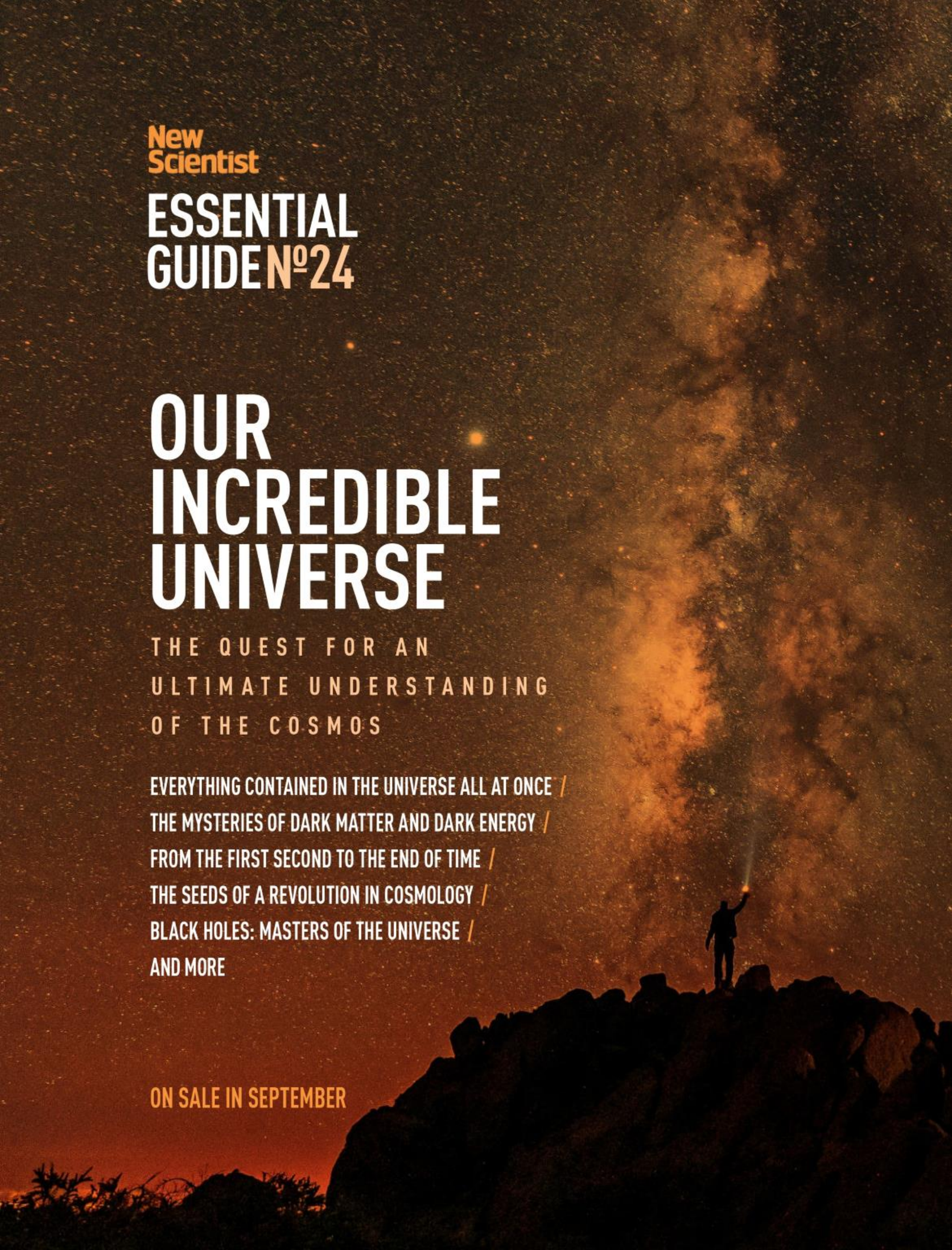
من یک خوش‌بین شکاک هستم من نسبت به روش‌هایی که در مورد نسل بعدی زیرساخت‌های مدنی فکر می‌کنیم، خوش‌بین هستم. چگونه اطمینان حاصل کنیم که زیرساخت‌ها واقعاً به ما خدمت خواهند کرد و به روش‌هایی که نتوان آن‌ها را وسط یک مذاکره سیاسی خاموش کرد.

گفت‌وگو درباره تغییرات اقلیمی به نقطه‌ای رسیده است که یعنی ما قرار است درباره تأثیری که سیستم‌های فنی از منظر انرژی و منابع طبیعی بر سیاره زمین می‌گذارند، فکر کنیم. من همچنین خوش‌بینم که از برخی جهات، هوش مصنوعی به ما اجازه می‌دهد گفت‌وگوهایی درباره اینکه چطور می‌خواهیم زندگی کنیم داشته باشیم. این گفت‌وگوها اغلب کاملاً بخش‌بندی شده بوده‌اند. اگر به گفت‌وگوها درباره حقوق کار، عدالت اقلیمی و حفاظت از داده‌ها فکر کنید، آن‌ها عمدتاً در بخش‌های بسیار جداگانه‌ای بوده‌اند، اما همین‌الان هوش مصنوعی با همه آن‌ها در تعامل است. این لحظه‌ای است برای گرد هم آوردن آن مسائل.

پس اثرات زیان‌بار هوش مصنوعی که هنوز در دوران اولیه خود است، قابل‌جبران است؟

نکته مهمی که باید به‌خاطر بسپاریم این است که هیچ فناوری‌ای اجتناب‌ناپذیر نیست. صرفاً چون چیزی طراحی شده، به این معنی نیست که باید به طور گسترده به کار گرفته شود و صرفاً چون کاری همیشه به روشی خاص انجام می‌شده، به این معنی نیست که نمی‌توانیم آن را تغییر دهیم.

وقتی ما درباره به‌کارگیری ناعادلانه نیروی کار، تخریب محیط‌زیست و برداشت انبوه داده‌ها فکر می‌کنیم که همگی می‌توانند عمیقاً زیان‌بار باشند؛ این موضوع مهم‌ترین مسئله است. این‌ها همه رویه‌هایی هستند که می‌توانند تغییر کنند و میراث بزرگ صنعت در طول حدود ۳۰۰ سال گذشته این است که صنایع پس از رگولاتوری، تغییر کرده‌اند. ما می‌توانیم این سیستم‌ها را بازسازی کنیم و امیدواری سیاسی بسیاری در آن نهفته است.


New Scientist
ESSENTIAL GUIDE Nº24
OUR INCREDIBLE UNIVERSE
THE QUEST FOR AN ULTIMATE UNDERSTANDING OF THE COSMOS
EVERYTHING CONTAINED IN THE UNIVERSE ALL AT ONCE /
THE MYSTERIES OF DARK MATTER AND DARK ENERGY /
FROM THE FIRST SECOND TO THE END OF TIME /
THE SEEDS OF A REVOLUTION IN COSMOLOGY /
BLACK HOLES: MASTERS OF THE UNIVERSE /
AND MORE
ON SALE IN SEPTEMBER

New Scientist

ESSENTIAL GUIDE №23

انقلاب هوش مصنوعی

چرا هوش مصنوعی ناگهان جهش بزرگی داشت؟
ابزارهای هوش مصنوعی مولد چگونه می‌توانند به شما کمک کنند؟
خطرات واقعی هوش مصنوعی چیست؟

مدت‌هاست که هوش مصنوعی به عنوان فناوری فردا در نظر گرفته می‌شود. اما در نوامبر ۲۰۲۲، انتشار ChatGPT عصر هوش مصنوعی مولد را آغاز کرد. این ابزارهای قدرتمند و خلاق، هر آنچه را که فکر می‌کردیم در مورد قابلیت‌ها و نحوه عملکرد این فناوری می‌دانیم، متحول می‌کنند . در این بیست و سومین راهنمای کلیدی، با موضوعاتی از جمله موارد زیر، دریابید که چگونه هوش مصنوعی قرار است جوامع و اقتصادها را متحول کند:

- ماشین‌ها چگونه یاد می‌گیرند؟
- مسیر خودآگاهی
- ماشین‌های متفکر و حسی
- دانشمندان و ریاضی‌دانان هوش مصنوعی
- چشم‌اندازهای آینده از آرمان‌شهر تا انقراض

9 772634 015026 23