هیچ ندیده‌ای هنوز

انتشارات
راه پرداخت

برای سفارش اینترنتی این کتاب به وبسایت انتشارات راه‌پرداخت مراجعه کنید
way2pay.shop

نسخه نمونه

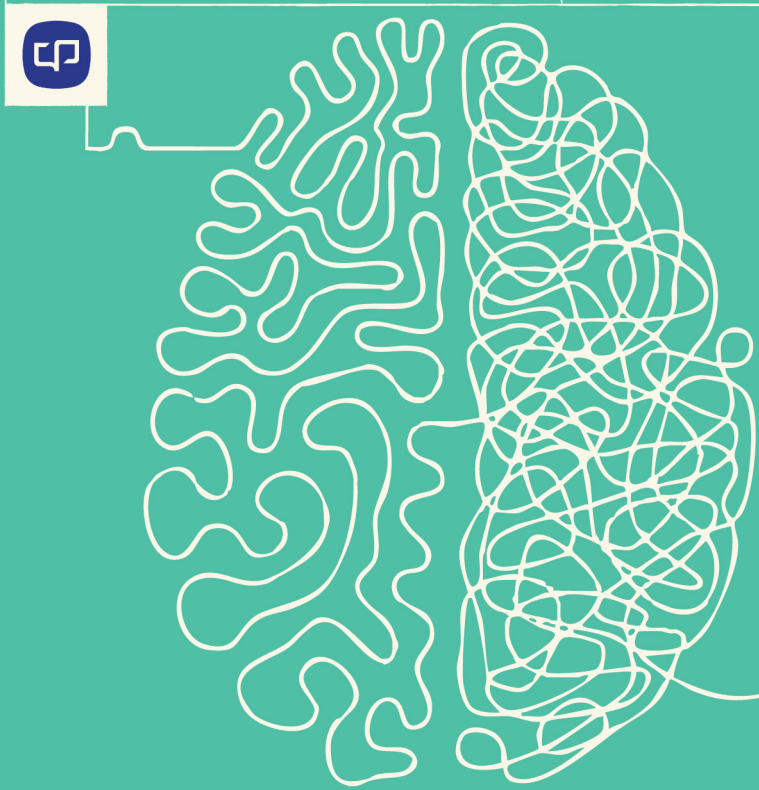
برای دانلود نسخه کامل به وبسایت فروشگاه
انتشارات راه پرداخت مراجعه کنید.

راهنمایی برای عبور از ابهام

چگونه دربارهٔ هوش مصنوعی بیندیشیم

ریچارد ساسکیند

مهدی مهدلو، محمدعلی فائضی‌راد



سرشناسه: ساسکیند، ریچارد ای.، ۱۹۶۱- م. Susskind, Richard E.
عنوان و نام پدیدآور: چگونه درباره هوش مصنوعی بیندیشیم/ریچارد ساسکیند ؛ مترجمان مهدی مهدلو،محمدعلی فائضی‌راد.
مشخصات نشر: تهران: راه پرداخت، ۱۴۰۴.
مشخصات ظاهری: ۲۲۹ص.؛ ۱۴/۵×۲۱/۵س‌م.
شابک: ۹۷۸-۶۲۲-۴۹۵۸-۰۳-۷
وضعیت فهرست نویسی: فیپا
یادداشت: عنوان اصلی: How To Think About AI: A Guide For The Perplexed, 2025.
یادداشت: بالای عنوان: راهنمایی برای عبور از ابهام.
یادداشت: کتابنامه: ص.۲۲۸-۲۲۹.
عنوان دیگر: راهنمایی برای عبور از ابهام.
موضوع: هوش مصنوعی- Artificial intelligence
موضوع: سواد فناوری- Technological literacy
شناسه افزوده: مهدلو، مهدی، ۱۳۵۹-، مترجم
شناسه افزوده: فایضی‌راد، محمدعلی، ۱۳۶۹-، مترجم
رده بندی کنگره: Q۳۳۵
رده بندی دیویی: ۰۰۶/۳
شماره کتابشناسی ملی: ۱۰۳۱۷۶۰۱

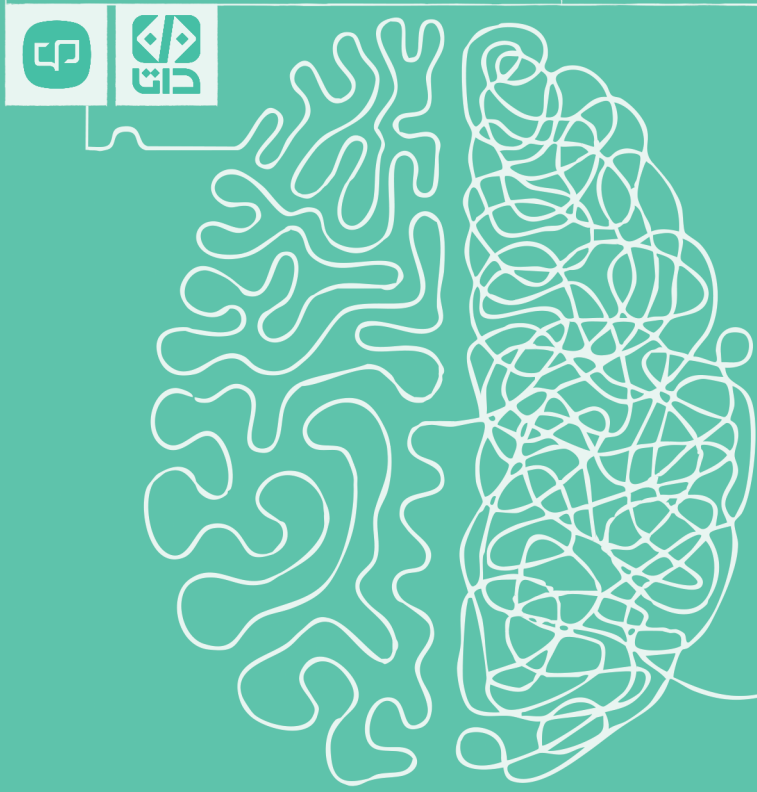
FSC www.fsc.org
The mark of responsible forestry
FSC® C009732

راهنمایی برای عبور از ابهام

چگونه دربارهٔ هوش مصنوعی بیندیشیم

ریچارد ساسکیند

مهدی مهدلو، محمدعلی فائضی‌راد



عنوان:	چگونه دربارهٔ هوش مصنوعی بیندیشیم
ناشر:	راه‌پرداخت
نویسنده:	ریچارد ساسکیند
مترجمان:	مهدی مهدلو - محمدعلی فائضی‌راد
ویراستار ارشد:	مینا والی
ویراستار:	بهار سرلک
بازبینی نهایی متن:	رضا قربانی
صفحه‌آرا:	علیرضا کیوان
ناظر چاپ:	قادر شهبازی
نوبت چاپ:	اول ۱۴۰۴
شمارگان:	۱۰۰۰
شابک:	۹۷۸-۶۲۲-۴۹۵۸-۰۳-۷
تلفن:	۰۲۱-۴۴۴۴۳۹۶۶
ایمیل:	book@way2work.ir
وب‌سایت:	w2ppress.ir
لیتوگرافی:	هنر اشکان
چاپ و صحافی:	واژه

همه حقوق چاپ و نشر این اثر برای «**انتشارات راه‌پرداخت**» محفوظ است. هرگونه تکثیر، انتشار و بازنویسی این اثر یا قسمتی از آن به هر شکل و شیوه ــ (چاپی، صوتی، ویدئویی، دیجیتال و ...) ــ بدون اجازه کتبی ناشر ممنوع است.

نشانی فروشگاه انتشارات راه‌پرداخت:

تهران، جنت‌آباد جنوبی، خیابان لاله غربی، روبه‌روی پاساژ سمرقند، خیابان حدیث، کوچه حدیث دوم، پلاک ۸

way2pay.press way2paypress way2paypress

فهرست

بخش اول: درک هوش مصنوعی ۲۸

فصل اول: تابستان هوش مصنوعی ۲۹

هوش مصنوعی چه‌کاری می‌تواند انجام دهد؟ ۳۰

تاریخچه‌ای کوتاه دربارهٔ هوش مصنوعی ۳۵

پیشرفت‌های بعدی ۴۵

فصل دوم: در باب فناوری ۴۸

فناوری دیجیتال ۴۹

چهار، پنج یا حتی شش مرحله پیشرفت بشری ۵۶

برخی سوگیری‌ها ۵۷

قابلیت‌های آینده ۵۹

بخش دوم: متفاوت اندیشیدن ۶۸

فصل سوم: تفکر فرایندگرا و تفکر نتیجه‌گرا ۶۹

دو فرهنگ هوش مصنوعی ۷۰

متفاوت اندیشیدن ۷۲

تأملات و افکار درونی ۷۵

فصل چهارم: سردرگمی‌ها ۷۷

اندیشهٔ پریشان پیرامون تهدید هوش مصنوعی ۷۸

رویکرد «ما استثنا هستیم» ۸۰

تأثیر هوش مصنوعی بر مشاغل تخصصی ۸۲

مغالطهٔ هوش مصنوعی ۸۴

آنچه ماشین‌ها نمی‌توانند انجام دهند ۸۷

فصل پنجم: واژه‌های مناسب در اختیار نداریم ۹۰

زبان ۹۳

قضاوت و شبهِ‌قضاوت ۹۴

همدلی و شبه‌همدلی ۹۷

خلاقیت و شبه‌خلاقیت ۹۹

بخش سوم: عملیاتی‌سازی هوش مصنوعی ۱۰۴
فصل ششم: خودکارسازی، نوآوری، حذف ۱۰۵
خودکارسازی ۱۰۷
نوآوری ۱۰۸
حذف ۱۰۸
چند نمونه و مثال ۱۰۹
دعاوی حقوقی و دادگاه‌ها ۱۰۹
پول نقد ۱۱۱
کارهای خانه ۱۱۲
خودروهای خودران ۱۱۳

فصل هفتم: تغییر ساختاری بنیادین ۱۱۵
تحول ۱۱۶
بازسازی مبتنی بر چشم‌انداز ۱۱۸
وسایل و ابزارهای نوین ۱۲۰
پژوهش و توسعه ۱۲۱

بخش چهارم: رویارویی با ریسک‌ها ۱۲۳
فصل هشتم: دسته‌بندی انواع ریسک ۱۲۴
دسته‌بندی ریسک‌ها ۱۲۵
دستهٔ ۱: ریسک‌های وجودی ۱۲۶
دستهٔ ۲: ریسک‌های فاجعه‌بار ۱۲۸
دستهٔ ۳: ریسک‌های سیاسی ۱۲۹
دستهٔ ۴: ریسک‌های اجتماعی ــ اقتصادی ۱۳۲
دستهٔ ۵: ریسک‌های ناشی از عملکرد غیرقابل اعتماد ۱۳۴
دسته ۶: ریسک‌های ناشی از وابستگی ۱۳۵
دستهٔ ۷: ریسک‌های ناشی از عدم اقدام ۱۳۷
تأملات اولیه ۱۳۸
یک هیولا؟ ۱۴۰
بیکاری ناشی از فناوری ۱۴۲

مواجهه با نابرابری‌ها ۱۴۳
فرصت‌های ازدست‌رفته ۱۴۳

فصل نهم: مهار هوش مصنوعی ۱۴۵
چرا این‌همه مدت تعلل کرده‌ایم؟ ۱۴۶
آمادگی برای بدترین‌ها ۱۴۸
در باب اخلاق هوش مصنوعی ۱۵۲
در باب قانون و مقررات ۱۵۷
ارتقای نظام عدالت ما ۱۶۳
رویکردی چندرشته‌ای ۱۶۷

بخش پنجم: تأمل در آینده ۱۷۲
فصل دهم: ماشین‌های هوشیار؟ ۱۷۳
هوشیاری انسان ۱۷۴
هوش مصنوعی و هوشیاری ۱۸۰
خاطره‌ای از شبی در اواسط دههٔ ۱۹۸۰ ۱۸۷

فصل یازدهم: به‌زودی ۱۸۸
رابط مغز و کامپیوتر (BCI) ۱۸۹
واقعیت مجازی ۱۹۲
واقعیت؟ ۱۹۸

فصل دوازدهم: شکاف عمیق ۲۰۵
فرگشت و هوش مصنوعی ۲۰۶
گزینه‌های پیش روی آیندهٔ ما ۲۱۰
وظایف ما در قبال آینده ۲۱۵
حکم نهایی ۲۱۸

نتیجه‌گیری ۲۲۴

[یادداشت ناشر]

رضا قربانی / مدیرعامل کارخانه نوآوری راه‌کار

این کتاب دربارهٔ آینده نیست؛ دربارهٔ اکنون است. دربارهٔ لحظه‌ای که در آن ایستاده‌ایم، جایی میان شگفتی و ترس. هوش مصنوعی دیگر رؤیا یا ابزار نیست؛ واقعیتی است که به زندگی ما خزیده، از موبایل تا تصمیم‌های اقتصادی و اخلاقی‌مان را تحت‌تأثیر قرار داده است. در چنین جهانی، روشن اندیشیدن دربارهٔ آن مزیت نیست، بلکه ضرورت است.

در راه پرداخت، هر روز شاهد موجی تازه از هوش مصنوعی هستیم؛ در رسانه‌ها، شرکت‌ها، بانک‌ها، دانشگاه‌ها و حتی خانه‌ها. فناوری با وعدهٔ کارآمدی و سرعت می‌آید، اما پرسش‌های عمیق‌تری هم با خود می‌آورد: دربارهٔ انسان، معنا، آگاهی و عدالت. کتاب *چگونه دربارهٔ هوش مصنوعی بیندیشیم* نوشتهٔ ریچارد ساسکیند، تلاشی است برای بازگرداندن این پرسش‌ها به مرکز گفت‌وگو. ساسکیند از معدود متفکرانی است که چهار دهه از عمرش را صرف رابطهٔ میان انسان و فناوری کرده است. او نه از زاویهٔ مهندسی یا اقتصاد، بلکه از چشم‌انداز فلسفه، عدالت و آیندهٔ بشر به موضوع نگاه می‌کند. این کتاب «راهنمای فنی» نیست؛ دعوتی است به اندیشیدن دوباره دربارهٔ آنچه هوش مصنوعی هست و آنچه می‌تواند بشود.

ساسکیند یادآوری می‌کند که اغلب ما، از مدیران تا مردم عادی، در ابهام زندگی می‌کنیم. نمی‌دانیم هوش مصنوعی قرار است ناجی باشد یا عامل نابودی. او می‌گوید راه فهم این پدیده در «چگونه اندیشیدن» است، نه در پاسخ‌های آماده.

خطرات، از امروز

ساسکیند میان خوش‌بینی افراطی و ترس آخرالزمانی راهی میانه می‌جوید. او می‌گوید خطر اصلی نـه در آیندهٔ دور، بلکـه در بی‌توجهی امروز ماسـت. در جهانـی که هر هفته فناوری تازه‌ای با برچسـب هوش مصنوعی می‌آید، بزرگ‌ترین خطر شاید همین باشد: از دست دادن فرصت اندیشیدن. ساسـکیند از ما می‌خواهد به دههٔ ۲۰۳۰ و ۲۰۴۰ فکر کنیم، به زمانی که شاید با هوش عمومی و فراهوش روبه‌رو شویم. می‌گوید هنوز در دامنهٔ کوه ایستاده‌ایم، اما مسیر صعود یا سقوط را همین امروز انتخاب می‌کنیم.

برای ایران، همین حالا

ما این کتاب را در زمانی منتشر می‌کنیم که گفت‌وگوی هوش مصنوعی در ایران یا در سطح هیجان فناوری می‌ماند، یا در ترس از بیکاری. اما آنچه نیاز داریم، گفت‌وگویی ملی است دربارهٔ معنا، اخلاق و آیندهٔ انسـان ایرانـی در عصر هوش مصنوعی. ایـن کتاب می‌تواند آغاز چنین گفت‌وگویی باشد. ایران به نسلی نیاز دارد که بتواند دربارهٔ فناوری فکر کند، نه‌فقط از آن اسـتفاده کند. جامعه‌ای که هوش مصنوعی را در خدمت عدالت، آموزش، سلامت و توسعه به کار گیرد، نه قربانی ناآگاهی و شتاب‌زدگی شود. انتشار این کتاب در راه پرداخت، بخشی از تعهد ماست به فهم آینده‌ای که در آن زندگی می‌کنیم؛ آینده‌ای که اگر بدون شناخت در آن قدم بگذاریم، ما را می‌بلعد. می‌خواهیم با این کتاب به پژوهشگران، مدیران، سیاست‌گذاران و دانشجویان بگوییم: پیش از تصمیم دربارهٔ هوش مصنوعی، یاد بگیریم چگونه دربارهٔ آن بیندیشیم.

[یادداشت حامی]

مرتضی ترک تبریزی / رئیس هیئت‌مدیرهٔ بانک تجارت

طی سال های اخیر، کمتر واژه‌ای به‌اندازهٔ «هوش مصنوعی» توانسته است چنین حضور پررنگ و همه‌جانبه‌ای در محافل تخصصی، کسب‌وکارها و حتی گفت‌وگوهای روزمرهٔ ما داشته باشد. همگان از تحول‌آفرینی و ضرورت بهره‌گیری از این فناوری سخن می‌گویند، اما کمتر سخنران یا نویسنده‌ای هست که با بیانی روشن، کاربردی و به‌دور از حاشیه‌پردازی‌های پیچیده، مسیر واقعی و عملی مواجهه با این پدیدهٔ شگفت‌انگیز را به ما نشان دهد. در اکثر آثار و مباحث مطرح‌شده، با انبوهی از اصطلاحات فنی و تخصصی روبه‌رو هستیم که باعث سردرگمی، خستگی و عدم درک صحیح می‌شود و در نهایت نیز مسیری مشخص برای مواجهه با هوش مصنوعی ارائه نمی‌دهند. البته این کتاب نیز متمرکز بر نحوهٔ استفاده از هوش مصنوعی نیست اما با رویکردی متفاوت تلاش دارد تا به‌جای ارائهٔ طیفی از واژه‌های فنی و الگوریتم‌های ریاضی، مخاطب را به تأملی فلسفی و انسانی دربارهٔ هوش مصنوعی دعوت کند؛ تأملی که علاوه‌بر کمک به فهم بهتر این فناوری، ما را به تفکر اساسی و بنیادین دربارهٔ نقش انسان، ضرورت اخلاق و خلق تصویری از آینده سوق می‌دهد.

بانک تجارت با درک ضرورت توانمندسازی فناورانهٔ جامع خود ـ اعم از کارکنان، مشتریان و فعالان اقتصادی ـ ترجمهٔ این کتاب را به همگان تقدیم می‌کند. ما اطمینان داریم که این اقدام، در دوره‌ای که «هوش مصنوعی عاملی» (Agentic AI) در حال گسترش است و «هوش عمومی مصنوعی» (AGI) مرزهای آینده را جابجا می‌کند، گامی ارزشمند برای تجهیز فکری مخاطبان به شمار می‌رود. هدف از این کار فراهم آوردن بستری برای مشارکت فعال و مسئولانه در ساختن فردایی است که در آن هوش مصنوعی به‌عنوان نیرویی پیش‌برنده در خدمت تعالی نظام بانکی، توسعهٔ پایدار اقتصادی و رفاه همگانی قرار گیرد.

[یادداشت دوم حامی]

علی گل‌زاده کرمانی / مدیرعامل داتا

هوش مصنوعی و فناوری‌های همراه آن فرصت‌هایی بی‌سابقه پدید آورده‌اند و هم‌زمان مسئولیت‌های جدیدی بر دوش همهٔ ما نهاده‌اند. پرسش‌هایی که هوش مصنوعی برمی‌انگیزد، از جنبه‌های فنی تا اخلاقی و اقتصادی، بنیادین و واقعی‌اند. این کتاب با زبانی روشن و کاربردی، خواننده را به تأمل دربارهٔ این پرسش‌ها دعوت می‌کند؛ نه‌تنها دربارهٔ توانایی‌های فناورانه، بلکه دربارهٔ ارزش‌ها، وظایف و انتخاب‌هایی که در برابر جامعه و نهادها قرار می‌گیرد.

شرکت داتا، به‌عنوان بخشی از خانوادهٔ بانک تجارت، این اثر را با این نگاه منتشر می‌کند که گفت‌وگوی منسجم و کاربردی دربارهٔ داده و هوش مصنوعی پیش‌شرط تصمیم‌گیری درست و بهره‌مندی مسئولانه از فرصت‌هاست. هدف ما از این همراهی فراتر از معرفی یک کتاب است؛ ما می‌خواهیم زمینه‌ای برای ظرفیت‌سازی، گفت‌وگوهای میان‌رشته‌ای و اقدام‌محور شدن در مواجهه با تحولات سریع فناوری فراهم آوریم.

فناوری به‌خودی‌خود خیر یا شر نیست؛ انتخاب‌ها و ساختارهایی که برای هدایت و

مدیریت آن می‌سازیم، تعیین‌کننده است. هم‌سو با نگاه ریچارد ساسکیند، که یادآور می‌شود برقراری توازن میان مضرات و مزایا ـ یعنی خدمت و خیانت ـ بزرگ‌ترین چالش اساسیِ عصرِ امروز در مواجهه با فناوری‌هاست، باور داریم برای تفکر و گفت‌وگوی منطقی دربارهٔ تأثیر هوش مصنوعی، باید شیوهٔ اندیشیدنمان را تغییر دهیم. لازم نیست متخصص فناوری باشیم؛ باید به‌جای غرق شدن در جزئیات فنی، با چارچوبی روشن و گفت‌وگویی میان‌رشته‌ای، پیامدها و انتخاب‌های اجتماعی این فناوری را پیشاپیش بسنجیم. امیدواریم این کتاب نقطهٔ شروعی باشد برای پرسیدنِ پرسش‌های بهتر، گفت‌وگوی گسترده‌تر میان متخصصان و جامعه و اقدام‌هایی که منافع عمومی را در نظر می‌گیرند.

از شما خوانندهٔ گرامی نیز دعوت می‌کنیم این متن را فرصتی برای تأمل و حرکت بدانید؛ حرکتی به‌سوی استفاده‌ای مسئولانه و انسانی از داده و هوش مصنوعی که به پیشرفت پایدار و رفاه همگانی کمک کند.

تمجیدها از کتاب *چگونه دربارهٔ هوش مصنوعی بیندیشیم*

گوردون براون[1]، نخست‌وزیر پیشین انگلستان (۲۰۱۰ـ۲۰۰۷):
کتاب بسیار خواندنی *چگونه دربارهٔ هوش مصنوعی بیندیشیم* دو طیف مخاطب را جذب می‌کند: متخصصانی با دانش هوش مصنوعی که نگران پیامدهای آن هستند و خوانندگان عادی که به درک عظمت تحولات فناوری نیاز دارند. مطالعهٔ این کتاب به تمام علاقه‌مندان به انقلاب هوش مصنوعی به‌شدت توصیه می‌شود.

لرد ویلیام هیگ[2]، رئیس دانشگاه آکسفورد:
این کتاب با وضوحی مثال‌زدنی بسیاری از سوءتفاهم‌ها دربارهٔ هوش مصنوعی را برطرف می‌کند. در این دنیای به‌شدت متغیر، مطالعهٔ این کتاب حیاتی است.

آلستر کمپل[3] و روری استوارت[4]، مجریان پادکست «همه‌چیز زیرِ سرِ سیاست است»:
این کتاب از مهم‌ترین و جذاب‌ترین منابع در حوزهٔ هوش مصنوعی است و برای هر کسی که می‌خواهد تغییرات دنیای پیرامون خود را درک کند، خواندنش ضروری است.

پروفسور دیم وندی هال[5]، استاد علوم کامپیوتر دانشگاه ساوت‌همپتون:
این کتاب، اثری جذاب دربارهٔ هوش مصنوعی است که فردی کاملاً متخصص در این

---

1. Gordon Brown
2. Lord William Hague
3. Alastair Campbell
4. Rory Stewart
5. Dame Wendy Hall

زمینه آن را نوشته است. دیدگاه ارائه‌شده در آن، رویکردی بسیار تازه و عمل‌گرایانه به هوش مصنوعی است که به غیرمتخصص‌ها کمک می‌کند درک درستی از رخدادهای درحال وقوع داشته باشند. مطالعهٔ این کتاب به‌شدت توصیه می‌شود.

ای. سی. گری‌لینگ[1]، نویسندهٔ *فلسفه و زندگی*[2] و *برای خیر جهان*[3]:
نام ساسکیند تقریباً مترادف با عبارت «هوش مصنوعی و تأثیر آن» است. برای درک بهتر پیچیدگی‌های نوظهور در این زمینه، هیچ راهنمایی بهتر از این بررسی قوی، روشن، آموزنده و عمیقاً تفکربرانگیز وجود ندارد. مطالعهٔ آن کاملاً ضروری است.

بارونس اسکاتلند[4]، دبیرکل اتحادیهٔ کشورهای مشترک‌المنافع:
این کتاب قرائتی ضروری برای کسانی است که می‌خواهند دربارهٔ هوش مصنوعی بیندیشند، وعده‌هایش را کشف کنند و جهان را از طریق هوش مصنوعی تغییر دهند.

ویکتور مایر-شونبرگر[5]، استاد دانشگاه آکسفورد و یکی از نویسندگان *چارچوب‌سازان*[6]:
ریچارد ساسکیند با جسارتی مثال‌زدنی، تعریفی تازه از انقلاب هوش مصنوعی ارائه می‌دهد. او با طرح پرسش‌های بنیادین «چه می‌شود اگر»، ما را ـ فارغ از هیاهو و تبلیغات ـ به‌سمتی هدایت می‌کند که بتوانیم با وضوح بیشتری دربارهٔ تصمیمات مهم و حتی حیاتی پیشِ رو بیندیشیم.

---

1. A.C. Grayling
2. *Philosophy and Life: Exploring the Great Questions of How to Live*
3. *For the Good of the World: Why Our Planet's Crises Need Global Agreement Now*
4. Baroness Scotland
5. Viktor Mayer-SchÖnberger
6. *Framers: Human Advantage in an Age of Technology and Turmoil*

دیم جولی ویکاستون[1]، رئیس مؤسسهٔ اِیدا لاولِیس[2]:

ریچارد ساسـکیند، راهنمایی ظریف و قابل فهم دربارهٔ فرصت‌هـا، چالش‌ها، خطرات و مزایای هوش مصنوعی مکتوب کرده است. مطالعهٔ این کتاب برای هر کسی که به دنبال مقدمه‌ای روشن دربارهٔ آینده است، ضروری است.

جان دنتون[3]، دبیرکل اتاق بازرگانی بین‌المللی:

در کتاب *چگونه دربارهٔ هوش مصنوعی بیندیشیم* ریچارد ساسکیند اثرات عمیق هوش مصنوعی بر مـا را بررسـی می‌کند. در ایـن راهنمـای واضـح و قابل فهم، ساسـکیند هم پتانسیل‌ها و هم چالش‌های هوش مصنوعی را می‌سنجد و به خوانندگان کمک می‌کند دربارهٔ پیامدهای اخلاقی، اقتصادی و اجتماعی آن فکر کنند و مسیر مناسبی برای درک آیندهٔ فناوری بیابند. این کتاب برای هر کسـی که به آیندهٔ فناوری‌ها علاقه دارد، ضروری اسـت؛ زیرا راهی را برای تعامل با هـوش مصنوعی ارائه می‌دهد که تفکـر عمیق و تازه‌ای دربارهٔ آینده و چگونگی آمادگی برای آن است.

مایکل ماینلی[4]، شهردار سابق و رئیس اسبق شورای شهر لندن:

ریچارد همواره در تفکر پیرامون فناوری و حرفه‌های ما پیشـرو بوده اسـت. متخصصان ما سال‌هاست که به کیفیت و کمیت تولیدات نوشـتاری خود بالیده‌اند. اما حالا، هوش مصنوعی این حرفه‌ها را با سیلی روزافزون از محتوای نوشتاری با کیفیت روبرو کرده است.

---

1. Dame Julie Maxton
2. Ada Lovelace
3. John Denton
4. Michael Mainelli

کتاب حکیمانه و تفکربرانگیز ریچارد به بررسی این چالش‌ها می‌پردازد و نکتهٔ حیاتی این است که ما را به تفکر دربارهٔ معنای حقیقی انسان‌بودن فرا می‌خواند.

تام ملهام[1]، استاد علوم کامپیوتر دانشگاه آکسفورد:
ریچارد ساسکیند استاد فن بیان است و کتاب *چگونه دربارهٔ هوش مصنوعی بیندیشیم* چون هشداری حیاتی بر فوریت توجه به نحوهٔ همزیستی انسان با هوش مصنوعی بسیار پیشرفته‌تر آینده تأکید دارد. این اثر ارزشمند و به‌هنگام، نوشتهٔ پژوهشگر بااستعدادی است که دهه‌هاست در حوزه هوش مصنوعی فعالیت می‌کند. صرف نظر از اینکه با مقاصدی که ساسکیند در این حوزه به آن‌ها می‌رسد یا حتی با روشی که برای رسیدن به آن‌ها به کار می‌برد موافق باشید یا نه، این کتاب راهنمایی ضروری برای اندیشیدن دربارهٔ هوش مصنوعی است.

پروفسور سر نایجل شادبولت[2]، رئیس کالج جیزس[3] در دانشگاه آکسفورد و رئیس مؤسسهٔ اوپن دیتا[4]:
این کتاب مروری جامع و قابل فهم بر تاریخچه، وضعیت فعلی و مسیرهای آتی احتمالی هوش مصنوعی و پیامدهای آن برای همگان ارائه می‌دهد. ساسکیند از ما دعوت می‌کند تا فرصت‌ها و چالش‌های هوش مصنوعی را در نسبت با بشریت و در نهایت، شاید در ارتباط با جهان گسترده‌تر، مورد تأمل قرار دهیم.

---

1. Tom Melham
2. Sir Nigel Shadbolt
3. Jesus College
4. Open Data Institute

به یاد پدر عزیز و بی‌نظیرم

ما موفق شدیم برنامه‌ای کامپیوتری ابداع کنیم که قادر به تفکر غیرعددی بود. با این دستاورد، ادعا کردیم که مسئلهٔ دیرینهٔ ذهن[1] و جسم[2] را حل کرده‌ایم و توضیح داده‌ایم چگونه یک سیستم متشکل از ماده می‌تواند ویژگی‌های ذهنی از خود بروز دهد. این اختراع، راه را برای خودکارسازی طیف وسیعی از وظایف که پیش از این نیازمند هوش انسانی بود هموار کرد و شبیه‌سازی کامپیوتری را به‌عنوان رویکردی جدید برای مطالعهٔ تفکر ارائه داد. در پی این موفقیت، به شهرت قابل توجهی دست یافتیم، اما در عین حال منتقدانی را نیز برانگیختیم که در اعماق قلبشان می‌دانستند ماشین‌ها نمی‌توانند فکر کنند و قصد داشتند جهان را از پذیرش ادعاهای ما باز دارند.

هربرت اِی. سایمون[3]، برندهٔ جایزهٔ نوبل،
با اشاره به عملیات موفقیت‌آمیزش در ۵۶ـ۱۹۵۵

همگان اصرار دارند کارهایی وجود دارد که ماشین قادر به انجامشان نیست؛ ولی اگر دقیقاً برایم توضیح دهید آن کارها چیست، برایتان ماشینی خواهم ساخت که دقیقاً همان کارها را انجام دهد.

جان فون نویمان[4] (۱۹۴۸)

در فریادِ هر انسان، فغان کودک گریان
ز هر آوا و هر فرمان، چه آسان
به گوش آید صدای بندهای ذهن تو انسان
ویلیام بلیک[5] شعر لندن (۱۷۹۴)

---

۱. به جنبه‌های غیرمادی وجود ما اشاره دارد؛ مانند افکار، احساسات، آگاهی، باورها، ادراکات ذهنی و تجربه‌های درونی. ـم.
۲. به جنبه‌های مادی و فیزیکی وجود ما اشاره دارد؛ مانند اندام‌ها و کل ساختار جسمانی. ـم.

3. Herbert A. Simon
4. John Von Neumann
5. William Blake

این کتاب را به کمک هوش مصنوعی ننوشته‌ایم (چه خوب، چه بد!).

[یادداشت نویسنده]

ریچارد ساسکیند

از ۱۹۸۱ و زمانی که دانشجوی حقوق در دانشگاه گلاسکو بودم و تصادفاً به حوزهٔ هوش مصنوعی کشیده شدم، این مقوله ذهنم را درگیر خود کرده است. در آن زمان رساله‌ای در باب امکان جایگزینی قضات با کامپیوترها می‌نوشتم که تمام مسیرهای تحقیقم را به‌سوی هوش مصنوعی رهنمون کرد. خیلی زود شیفتهٔ این موضوع شدم و مشتاقانه تصمیم گرفتم سطح کاوشم را عمیق‌تر کنم. بنابراین، در اواسط دههٔ ۱۹۸۰ اسکاتلند، سرزمین مادری‌ام، را ترک کردم و برای تحصیل در مقطع دکتری در زمینهٔ هوش مصنوعی و حقوق به دانشگاه آکسفورد رفتم. در اواخر آن دهه، ضمن اینکه تیم هوش مصنوعی بریتانیا را در یک شرکت بین‌المللی حسابداری و مشاوره سرپرستی می‌کردم، در توسعهٔ نخستین سامانهٔ هوش مصنوعی کاربردی برای وکلا مشارکت داشتم. از آن زمان تاکنون، علاوه‌بر مقولهٔ حقوق، چند کتاب مرتبط با هوش مصنوعی نیز با همکاری دانیل ساسکیند دربارهٔ تأثیر هوش مصنوعی بر مشاغل گوناگون ازجمله بهداشت، آموزش، مالیات، حسابرسی، معماری، روزنامه‌نگاری و مواردی ازاین‌دست نوشته‌ام. این روزها، عمدهٔ زمانم به ارائهٔ مشاوره به شرکت‌ها و دولت‌ها دربارهٔ نقش و اهمیت هوش مصنوعی در آینده‌شان می‌گذرد.

در مجموع، هوش مصنوعی همواره برایم موضوعی جذاب و دغدغه‌انگیز بوده است. این کتاب، راهنمایی مقدماتی در این زمینه است و اغلب از تجربیات شخصی‌ام در این حوزه نشئت می‌گیرد. در نگارش این کتاب، از دانش و آگاهی‌ام در چهار حوزهٔ مرتبط با هوش مصنوعی، شامل فلسفه، فناوری دیجیتال، حقوق و برنامه‌ریزی راهبردی بلندمدت و نیز تجربیاتم در طول این سال‌ها در بخش‌های گوناگون خصوصی، دولتی و آموزشی بهره برده‌ام. شاید برای خیلی‌ها عجیب باشد، اما این موضوع در خانوادهٔ ما ریشه دوانده است؛ دو پسرم هم بیشتر وقتشان را صرف تفکر و نگارش دربارهٔ هوش مصنوعی می‌کنند. دانیل که اقتصاددان است، تأثیر فناوری بر اشتغال و رشد اقتصادی را بررسی می‌کند و جیمی که وکیل و نظریه‌پرداز سیاسی است، به مقررات هوش مصنوعی و اثر آن بر سیاست می‌پردازد. در این کتاب، بسیاری از دیدگاه‌های آن‌ها را هم گنجانده‌ام که می‌توانید ردپایشان را در ارجاعات مکرر و آشکار من به کارهایشان ببینید. در حین آماده‌سازی این کتاب و تفکر دربارهٔ آخرین تحولات هوش مصنوعی، دیدگاهم دستخوش تغییر شد. من مدت‌ها خوش‌بین بوده‌ام که هوش مصنوعی می‌تواند دستاوردهای بزرگ اجتماعی و اقتصادی به ارمغان آورد. اکنون حتی خوش‌بین‌تر هم هستم؛ اما هرچه بیشتر در پیشرفت‌های اخیر و شگفت‌انگیز تأمل کرده‌ام، به همان میزان از دامنه و مقیاس تهدیداتِ محتمل هوش مصنوعی نیز آگاه‌تر شده‌ام.

به باور من، برقراری توازن میان مزایا و مضرات هوش مصنوعی ـ یعنی خدمت و خیانت به انسان ـ اساسی‌ترین چالش عصر ماست.

مقدمه

ماروین مینسکی[1] استاد علوم کامپیوتر در ام‌آی‌تی و یکی از پیشگامان اولیهٔ هوش مصنوعی بود که موفق به کسب جایزهٔ معتبر تورینگ در ۱۹۶۹ شد. یک سال بعد در مجلهٔ *لایف* نقل‌قولی از او آورده شد: «وقتی کامپیوترها کنترل را به دست بگیرند، ممکن است دیگر هرگز آن را پس نگیریم. ما به‌لطف آن‌ها زنده خواهیم ماند. اگر خوش‌شانس باشیم، شاید تصمیم بگیرند ما را به‌عنوان حیوان خانگی‌شان نگه دارند.»

قبلاً این حرف برایم خنده‌دار بود، اما حالا خیلی مطمئن نیستم که خنده‌دار باشد.

در ۱۹۸۶ دیدگاه متفاوتی در مجلهٔ *نیوساینتیست* مطرح شد: «در تاریخ پژوهش‌های دانشگاهی، هیچ موضوعی به‌اندازهٔ هوش مصنوعی مردم را فریب نداده است.» در دهه‌های پس‌از آن، بسیاری از محققان و صاحب‌نظران این اتهام گسترده را تکرار کرده‌اند و ادعا می‌کنند هوش مصنوعی چیزی جز یک جنجال تبلیغاتی بزرگ نیست.

چگونه می‌توانیم شکاف میان نظر مینسکی و *نیوساینتیست* را درک کنیم؟

هوش مصنوعی همواره انسان‌ها را به دو دسته تقسیم کرده است. از اواسط دههٔ ۱۹۵۰، زمانی که اصطلاح «هوش مصنوعی» ابداع شد، طرفداران و منتقدان به یک اندازه درباره‌اش اظهار نظر کرده‌اند. تنش میان موافقان و مخالفان بسیار فراتر از اختلافات معمولی میان متفکران و دانشمندان مکاتب مختلف است. در اینجا، مسائل به‌مراتب مهم‌تری مطرح است، زیرا، به بیان ساده، هوش مصنوعی آیندهٔ بشر و جامعه را به‌شدت تحت‌تأثیر قرار خواهد داد.

آیا هوش مصنوعی به ما کمک خواهد کرد به بالاترین سطح پیشرفت و کمال برسیم یا برعکس، منجر به نابودی یا انحطاط انسانیت و تمدن بشری خواهد شد؟ آیا هوش مصنوعی راه‌حل‌های اساسی‌ترین چالش‌های بشر، از تغییرات آب‌وهوایی گرفته تا سلامت جهانی را ارائه خواهد داد یا به تضعیف، وابستگی یا حتی انحطاط ویژگی‌های طبیعی و بیولوژیکی انسان منجر خواهد شد؟[2]

من برای بسیاری از پرسش‌های مربوط به هوش مصنوعی که اکنون مطرح می‌شوند،

---

1. Marvin Minsky

۲. اگر انسان برای انجام وظایف فکری و جسمی بیش‌از حد به هوش مصنوعی تکیه کند، ممکن است به‌تدریج توانایی‌هایش را از دست بدهد و حتی ممکن است ارزش و جایگاه ویژگی‌های منحصربه‌فرد انسانی (مانند هوش هیجانی، خلاقیت به شیوهٔ انسانی، همدلی) کم‌رنگ شود.ـم.

پاسخ قطعی ندارم. همگی ما هنوز در مسیری دشوار و ناآشنا، درحال پیشروی و آزمون و خطا هستیم. همچنین، آگاهم که در سال‌های پیشِ رو، پرسش‌هایی دربارهٔ هوش مصنوعی پیش خواهد آمد که امروز حتی نمی‌توانیم تصورشان کنیم؛ پرسش‌هایی مربوط به مشکلاتی که از دل سیستم‌هایی که هنوز اختراع نشده‌اند، بیرون می‌آیند. درحال حاضر، دیدگاه ما به همان روشنایی محدود است که دانش اندک کنونی‌مان دربارهٔ فناوری‌های موجود و نوظهور بر ما تابانده است.

در این گرداب عدم قطعیت، امیدوارم بتوانم شیوه‌های متنوعی از تفکر دربارهٔ هوش مصنوعی را در اختیار خوانندگان قرار دهم؛ شیوه‌هایی که انگیزه‌ای برای ژرف‌اندیشی و گفت‌وگوی آگاهانه‌تر باشند. این کتاب، اثری فنی نیست؛ بلکه رویکرد و محتوایی مقدماتی دارد. فرض را بر این گذاشته‌ام که خوانندگان دانشی فنی در زمینهٔ هوش مصنوعی و سیستم‌های دیجیتال ندارند. در واقع، یکی از اهدافم، رمزگشایی از هوش مصنوعی و بحث‌های عمومی دربارهٔ این مفهوم شگفت‌انگیز است و می‌خواهم اطمینان دهم که برای تفکر و تکلم منطقی دربارهٔ تأثیر هوش مصنوعی، نیازی نیست متخصص فناوری باشید. به نظرم وقتی از مسائل بااهمیت و پیچیدگی فنی چشم‌پوشی می‌کنم، بسیاری از دانشمندان ناخشنود می‌شوند. این کار را به پای سهل‌انگاری‌ام نگذارید. بااین‌حال، اصرار دارم که بحث دربارهٔ هوش مصنوعی و پیامدهای آن نباید همیشه بر جزئیات فنی متمرکز باشد.

انگیزهٔ نوشتن این کتاب، هیجانی بود که پس‌از عرضهٔ سیستم هوش مصنوعی خارق‌العاده‌ای به نام ChatGPT در اواخر نوامبر ۲۰۲۲ به وجود آمد. این سیستم متعلق به شاخه‌ای از هوش مصنوعی است که عموماً «هوش مصنوعی مولد» (Generative AI) نامیده می‌شود (فعلاً نگران معنای آن نباشید). چت‌جی‌پی‌تی در عرض دو ماه صد میلیون کاربر و طی یک سال پس‌از عرضه، تقریباً دو برابر این رقم را به خود جذب کرد. پس‌از آن، سیل عظیمی از تفسیرها و گمانه‌زنی‌ها به راه افتاد. بااین‌حال، همان‌طور که خواهم گفت، بسیاری از ادعاهای کنونی و پیش‌بینی‌های رایج دربارهٔ هوش مصنوعی مبهم‌اند. برای مثال، بسیاری از تازه‌واردان به این حوزه تصور می‌کنند هوش مصنوعی مولد، نقطهٔ پایان و مقصد نهایی دنیای هوش مصنوعی است؛ اما در حقیقت، ما تازه اول راهیم. هنوز

روزهای نخستین هوش مصنوعی است.

نگرانم بحث‌ها و سیاست‌گذاری‌ها در زمینهٔ هوش مصنوعی بیش‌ازحد تحت تأثیر متخصصان و کارآفرینان حوزهٔ فناوری قرار گرفته باشد. هرچند نمی‌توان نبوغ استعدادهای برتر این عرصه را انکار کرد، اما برخی از آنان با اطمینانی بیش از حد دربارهٔ موضوعاتی مانند اخلاقیات، تأثیرات اجتماعی، روابط بین‌الملل و مقررات‌گذاری اظهار نظر می‌کنند. افزون‌بر این، بسیاری از آن‌ها به‌دلیل منافع شخصی در نحوه و انگیزه‌های فعالیتشان در حوزهٔ هوش مصنوعی، قادر به ارائهٔ تحلیلی بی‌طرفانه یا قابل اعتماد دربارهٔ آینده و پیامدهای این فناوری نیستند.

هدف من در این برههٔ حساس از تاریخ بشر، ایجاد وضوح بیشتر در گفت‌وگوهای روزمره‌مان دربارهٔ هوش مصنوعی است. به همین دلیل، تمرکزم بر این است که ما انسان‌ها چگونه می‌توانیم به‌شکلی مؤثر و ثمربخش دربارهٔ هوش مصنوعی فکر کنیم.

این کتاب در پنج بخش تنظیم شده است که هرکدام متناظر یکی از سنگ‌بناهای اساسی تفکر روشنگرانه‌ای است که نسبت به هوش مصنوعی دارم.

بخش اول به درک ماهیت هوش مصنوعی اختصاص دارد. فصل ۱ نشان می‌دهد یک راه خوب برای بالا رفتن از منحنی یادگیری هوش مصنوعی[۱]، نگاهی به تاریخچه و مسیر آن در هفتادوپنج سال گذشته است. برای روشن‌تر شدن موضوع، فصل ۲ با نگاهی کلی‌تر، خواننده را ترغیب می‌کند که هوش مصنوعی را در بستر پیشرفت‌های گسترده‌تر در فناوری دیجیتال ببیند.

بخش دوم خواننده را به تفکری متفاوت دربارهٔ هوش مصنوعی دعوت می‌کند. فصل ۳ دو روش کلی، اما کاملاً متضاد برای نگاه به هوش مصنوعی را معرفی می‌کند؛ یکی بر نتایج سیستم‌ها تمرکز دارد و دیگری به جنبه‌های مختلف فرایندهای زیربنایی می‌پردازد. فصل ۴ ابهاماتی را روشن می‌سازد که از بی‌توجهی به تفاوت این دو دیدگاه ناشی می‌شوند؛ به‌ویژه دربارهٔ این پرسش که آیا هوش مصنوعی تهدیدی برای بقای انسان است؟ فصل ۵ به توانایی‌هایی در ماشین‌ها می‌پردازد که فعلاً واژگان مناسبی برای توصیفشان در اختیار نداریم و در کنار آن، برخی محدودیت‌های ظاهری این سیستم‌ها را نیز بررسی می‌کند.

---

۱. منظور، شروع یادگیری و طی کردن تدریجی مسیر یادگیری هوش مصنوعی است.ـم.

بخش سوم، جنبه‌های عملی هوش مصنوعی و نحوهٔ پیاده‌سازی آن را بررسی می‌کند و رویکردی کاربردی‌تر دارد. فصل ۶ استدلال می‌کند که برای درک تأثیر هوش مصنوعی بر نیروی کار، نه‌تنها باید بسیار فراتر از اتوماسیون و جایگزینی کارهای فعلی انسان اندیشید، بلکه باید نحوهٔ دگرگونی شیوهٔ کار ما توسط هوش مصنوعی و همچنین توانایی آن در حذف مسائلی را فهمید که اساس بسیاری از مشاغل کنونی هستند. فصل ۷ هشدار می‌دهد که نباید هوش مصنوعی را به‌سادگی به ساختارهای موجود سازمانی افزود، بلکه باید به دنبال تحول بنیادین در ساختار نهادهایمان ازجمله سیستم‌های سلامت، عدالت، آموزش و دانشگاه‌ها بود.

بخش چهارم به جنبهٔ منفی هوش مصنوعی و به خطرات ناشی از هوش مصنوعی می‌پردازد. فصل ۸ طیف وسیعی از تهدیدات مرتبط را دسته‌بندی می‌کند و فصل ۹ ایده‌هایی برای مقابله با آن‌ها ارائه می‌دهد.

بخش پنجم به بررسی آیندهٔ نوع بشر می‌پردازد که در آن، ماشین‌هایمان بسیار تواناتر می‌شوند. این بخش بیشتر فلسفی، گمانه‌زنانه ـ و به نظر من ـ نگران‌کننده است. در فصل ۱۰، دربارهٔ این موضوع بحث می‌کنم که آیا ماشین‌ها می‌توانند دارای هوشیاری باشند. فصل ۱۱ خاطرنشان می‌کند که نباید در نگاه به آینده صرفاً هوش مصنوعی را در نظر گرفت، بلکه باید به فناوری‌های حیاتی دیگری مانند رابط‌های مغز و کامپیوتر[۱] (BCI) و واقعیت مجازی نیز توجه داشت. این امر به بررسی برخی احتمالات فلسفی ناخوشایند مربوط به ماهیت واقعیت روزمره منجر می‌شود. سرانجام، و با نگاهی دوباره به تصویر کلان ترسیم‌شده در فصول قبل، فصل ۱۲ پیشنهاد می‌کند که برای درک تأثیر هوش مصنوعی بر آیندهٔ بشر، باید از منظر تکامل و کیهان‌شناسی با آن روبرو شویم.

هر یک از فصل‌ها مسائل بسیار پیچیده‌ای را مطرح می‌کنند. امیدوار نیستم به سؤالات مطرح‌شده پاسخ قطعی داده شود، بلکه می‌خواهم به خوانندگان کمک کنم بحث‌های خود را ـ با خانواده، دوستان، همکاران و دولت‌هایشان ـ در مورد تأثیر مهم‌ترین مجموعه فناوری‌های زمانه‌مان آغاز کنند.

---

۱. این فناوری‌ها (با نام اختصاری BCI) امکان ارتباط مستقیم بین مغز انسان و دستگاه‌های خارجی مانند کامپیوتر را فراهم می‌کنند. این فناوری می‌تواند پیامدهای گسترده‌ای برای حوزه‌های ارتباطات، درمان بیماری‌های عصبی، افزایش توانایی‌های انسانی و حتی همگرایی انسان و ماشین داشته باشد. ـ م.

بخش اول

۱

درک هوش مصنوعی

فصل اول

۱

تابستان هوش مصنوعی

انبوه تعاریف و تفسیرها در مورد چیستی هوش مصنوعی، اولین مانع در مسیر تفکر روشن دربارهٔ آن است. در نتیجه، رویکردم در این بحث، بیشتر عملی است تا نظری.

برای تعریف هوش مصنوعی دو رویکرد کلی وجود دارد؛ اول، نحوهٔ عملکردش را توصیف کنیم. دوم، توانایی‌هایش را شرح دهیم. متخصصان فناوری تمایل دارند مسیر اول را در پیش بگیرند. آن‌ها با اشاره به روش‌ها و فناوری‌های زیربنایی، هوش مصنوعی را تعریف و اغلب از اصطلاحات تخصصی استفاده می‌کنند. برای مثال، از «شبکه‌های عصبی» (neural networks)، «مدل‌های زبانی بزرگ» (Large language models)، «پس‌انتشار» (Backpropagation)، «واحدهای پردازش گرافیکی» (Graphics processing unit (GPU))، «یادگیری ماشینی خصمانه» (Adversarial machine learning) یا «مدل‌های بنیادی» (Foundation models) سخن می‌گویند. نتیجهٔ این کار، سردرگمی بخش زیادی از مخاطبان است. اغلب مردم ــ از مدیران و سیاست‌گذاران گرفته تا عامهٔ مردم ــ از این اصطلاحات سر در نمی‌آورند و دانش فنی لازم برای فهم راحت چنین مفاهیم پایه‌ای را ندارند.

به همین دلیل ترجیح می‌دهم مسیر دوم را نقطهٔ شروع قرار دهم. بهترین راه برای درک اولیهٔ اینکه هوش مصنوعی چیست، ارائهٔ مثال‌های عملی از کارهایی است که سیستم‌های هوش مصنوعی درحال‌حاضر انجام می‌دهند و به احتمال زیاد در آیندهٔ نزدیک انجام خواهند داد.

هوش مصنوعی چه‌کاری می‌تواند انجام دهد؟

در روزهای اولیهٔ هوش مصنوعی، در دوران آنچه گاهی «هوش مصنوعی قدیمی خوب (Good old-fashioned AI (GOFAI)» نامیده می‌شود ــ در دهه‌های ۱۹۶۰، ۱۹۷۰ و ۱۹۸۰ ــ هوش مصنوعی را به مجموعه‌ای زیرشاخه تقسیم می‌کردیم. از برچسب‌های کلی استفاده می‌کردیم و تمرکزمان بر دسته‌های گسترده‌ای از وظایف بود که هوش مصنوعی برایشان به کار می‌رفت. یکی از حوزه‌ها، تشخیص گفتار (Speech recognition) نام داشت که هدفش توسعهٔ سیستم‌هایی بود که بتوانند کلمات گفته‌شده توسط انسان‌ها را بفهمند؛ سیستم‌هایی که بتوانند، مثلاً، بین عبارت انگلیسی

«abominable» و «a bomb in a bull» تمایز قائل شوند. حوزهٔ دیگر، ترجمه از یک زبان به زبان دیگر بود، با پرهیز از اشتباهات فاحشی مانند ترجمهٔ عبارت انگلیسی «out of sight, out of mind» (از دل برود هر آنکه از دیده برفت) به روسی به‌صورت «invisible insanity» (جنون نامرئی)[1]. حوزهٔ مرتبط دیگر، پردازش زبان طبیعی (Natural language processing (NLP)) بود که این امکان را فراهم می‌کرد تا انسان‌ها بتوانند با استفاده از زبان روزمرهٔ خود ـ با تمام ابهامات و عدم صراحتی که دارد ـ بدون استفاده از کدنویسی، با کامپیوترها ارتباط برقرار کنند و به آن‌ها دستور دهند.

شاخهٔ دیگر هوش مصنوعی، سیستم‌های خبره (Expert systems) بود. محبوب‌ترین رویکرد در این حوزه، تلاش برای شبیه‌سازی و مدل‌سازی فرایندهای استدلال و حل مسئلهٔ برترین متخصصان و گنجاندنشان در سیستم‌ها بود. این کار به افراد غیرمتخصص امکان می‌داد در زمینه‌های مختلف در سطح برجسته‌ترین متخصصان آن زمینه عمل کنند.

کارهای انجام‌شده در چهار زیرشاخهٔ دیگر نیز اساسی بود: یکی بازی‌سازی بود که همچنان زمینه‌ای حاصلخیز برای آزمایش هوش مصنوعی محسوب می‌شود و کانون برخی از برجسته‌ترین پیشرفت‌ها، به‌ویژه در شطرنج و بازی GO، است. دیگری رباتیک بود که با وجود چالش‌های بسیار، اخیراً پیشرفت‌های بزرگی در آن حاصل شده است. سومی، تحقیق در زمینهٔ بینایی و درک بصری بود، یعنی سیستم‌هایی که می‌توانستند اشیا (ازجمله چهره‌ها) را در دنیای فیزیکی تشخیص دهند. نهایتاً، برنامه‌نویسی خودکار مطرح شد که در آن سیستم‌ها خودشان می‌توانستند سیستم‌های دیگر را تولید کنند. این مورد آخر شاید مهم‌ترینشان باشد، زیرا نشان‌دهندهٔ گذار از ماشین‌هایی است که صرفاً توسط انسان‌ها توسعه داده می‌شوند به ماشین‌هایی که می‌توانند کدهای خود را بنویسند و به‌این‌ترتیب، خود را بهبود ببخشند و حتی تکثیر کنند.

دوران GOFAI را پشت سر گذاشته‌ایم و وقتی امروز به توانمندی‌های هوش مصنوعی می‌پردازیم، بیشتر بر وظایف، فعالیت‌ها و زمینه‌های خاص‌تر تمرکز داریم. علاقهٔ اصلی

---

۱. سیستم احتمالاً عبارت «out of sight» را به چیزی مرتبط با «نامرئی» و عبارت «out of mind» را به «جنون» ترجمه کرده است. بنابراین، «جنون نامرئی» به دست می‌آید که عبارتی کاملاً بی‌معنی و بی‌ربط به متن اصلی است. ـ م.

خودم به استفاده از هوش مصنوعی در مشاغل تخصصی است. برای مثال، سیستم‌های هوش مصنوعی در حوزهٔ سلامت می‌توانند بیماری‌ها را تشخیص دهند، اسکن‌ها را تفسیر کنند، دارو تجویز کنند، جراحی رباتیک انجام دهند، به بیماران مشاوره دهند و در توسعهٔ داروها و واکسن‌ها کمک کنند (هوش مصنوعی، تولید واکسن mRNA مدرنا برای کووید۱۹ را تسریع کرد). سیستم‌های حوزهٔ حقوق قادرند پیش‌نویس اسناد حقوقی را تهیه کنند، آن‌ها را با یکدیگر مقایسه، خلاصه و ساده‌سازی کنند. این سیستم‌ها همچنین می‌توانند راهنمایی ارائه دهند، به پرسش‌ها پاسخ گویند، استدلال‌های حقوقی تولید و نتایج احتمالی آرای دادگاه را پیش‌بینی کنند. در مالیات، سیستم‌ها می‌توانند اظهارنامه‌های مالیاتی را تهیه و ارسال کنند و طرح‌ها و استراتژی‌هایی ایجاد کنند که مبلغ مالیات را از لحاظ قانونی به حداقل برسانند. در آموزش، هوش مصنوعی یادگیری شخصی‌سازی‌شده را ممکن می‌سازد و دسترسی فوری به بهترین منابع آموزشی در سراسر جهان را فراهم می‌کند. در معماری، ساختمان‌های قابل توجهی توسط سیستم‌های هوش مصنوعی طراحی می‌شوند (و حتی با استفاده از چاپ سه‌بعدی ساخته می‌شوند).

پیشرفت‌های مشابهی در مشاورهٔ مدیریت، حسابرسی، نقشه‌برداری، محاسبات بیمه‌ای (Actuarial work) و همچنین در بیشتر خدمات تخصصی و مرتبط دیگر مانند بیمه، بانکداری و مدیریت سرمایه‌گذاری درحال انجام است. بسیاری از این سیستم‌ها هنوز در ابتدای راهند. برخی هنوز کاملاً قابل اعتماد نیستند، اگرچه روزبه‌روز بهبود می‌یابند. برخی دیگر درحال‌حاضر می‌توانند با عملکردِ متخصصان انسانی به‌صورت فردی برابری کنند و برخی‌ها حتی از تیم‌هایی متشکل از درخشان‌ترین متخصصان نیز بهتر عمل می‌کنند.

استفاده از هوش مصنوعی فقط به مشاغل تخصصی و اداری محدود نیست، بلکه دامنه‌ای بسیار وسیع‌تر دارد؛ از شرکت در مسابقات تا مدیریت شبکه‌های برق، از خودروهای خودران تا کمک‌خلبانی هواپیماهای نظامی، از بهینه‌سازی فرایندهای کسب‌وکار تا زمان‌بندی مسیرهای پرواز برای کاهش آلایندگی، از تعیین حالات عاطفی انسان تا ایجاد چهره‌های سخنگو که شبیه به نسخهٔ واقعی‌اند. امروزه، در هر عرصه‌ای

که پیش‌تر به هوش انسانی متکی بودیم، شاهد کاربرد روزافزون هوش مصنوعی هستیم.

بسیاری از سیستم‌های هوش مصنوعی به‌صورت مجزا و تک‌منظوره طراحی شده‌اند (مانند سیستم‌های تشخیص پزشکی)، اما امروزه شاهد استفادهٔ فزاینده از هوش مصنوعی در دل سیستم‌های دیگر هستیم. به‌عنوان مثال، تولیدکنندگان تلفن همراه اپل و سامسونگ ادعا می‌کنند که استفاده از گوشی‌هایشان راحت‌تر و بهتر شده، زیرا توسط هوش مصنوعی بهبود یافته‌اند. از سوی دیگر، هوش مصنوعی در نرم‌افزارهای روزمره ادغام و گنجانده می‌شود. مایکروسافت سیستم هوش مصنوعی خود (کوپایلت Copilot) را همراه با نرم‌افزار عمومی آفیس ارائه می‌دهد. این سیستم می‌تواند کارهایی چون تهیهٔ پیش‌نویس ایمیل و اسناد، خلاصه کردن جلسات و مکاتبات ایمیلی، ساخت اسلاید نمایش، ایجاد فهرست کارهای روزمره و بسیاری وظایف دیگر را انجام دهد.

در عمل، به این معناست که اغلب ما به‌زودی دستیاران شخصی و پژوهشی خود را در دسترس خواهیم داشت که به سازماندهی زندگی‌مان کمک می‌کنند و ما را کاملاً مطلع نگه می‌دارند. به همین ترتیب، معلمان خصوصی خودمان را خواهیم داشت که آمادگی دارند هر موضوعی را به ما آموزش دهند. علاوه‌بر این، به‌زودی سیستم‌های هوش مصنوعی خواهیم داشت که دائماً بهتر و بهتر می‌شوند و می‌توانند به هر سؤالی جواب بدهند و قادر به تولید متن (نامهٔ رسمی، شعر، مقاله، ترانه) و همچنین موسیقی، کارتون، طراحی، مناظر طبیعی و ویدیو و موارد دیگر هستند. هوش مصنوعی با این قابلیت‌ها می‌تواند بشر را توانمندتر کند، نه اینکه صرفاً بازدهی و کارآمدی کارمندان اداری را بهبود ببخشد.

بی‌آنکه متوجه باشیم بسیاری از کاربردهای هوش مصنوعی امروز در زندگی روزمرهٔ ما به کار می‌روند. برای مثال، وقتی فروشگاه‌های آنلاین یا تبلیغات ایمیلی، پیشنهادهای فیلم، کتاب، لباس یا غذا به ما می‌دهند، فناوری پنهان در پس این کار، که معمولاً به آن «الگوریتم» می‌گویند، در واقع نوعی هوش مصنوعی است. همین‌طور، هنگامی که برای کار، وام یا بیمه درخواست می‌دهیم، هوش مصنوعی اغلب در پشت‌صحنه مشغول پردازش است. در تمامی سازمان‌های کلیدی که به آن‌ها وابسته‌ایم، از صنایع تولیدی و خرده‌فروشی گرفته تا حمل‌ونقل، بیمارستان، بانک و نهادهای دولتی، سیستم‌های هوش

مصنوعی در مراحل گوناگون به کار گرفته شده‌اند.

این سیستم‌ها همچنین برای خلق دنیاهای نو در واقعیت مجازی و حتی اتصال مغز انسان به کامپیوترها به کار می‌روند (برای جزئیات بیشتر فصل ۱۱ را بخوانید). جاه‌طلبی از این فراتر رفته و برخی از علاقه‌مندان به هوش مصنوعی بر این باورند که هوش مصنوعی ناجی بشریت خواهد بود و در مقابله با جدی‌ترین چالش‌ها، مانند تغییرات اقلیمی، سرطان، فقر، منازعات، اکتشافات فضایی، آموزش همگانی، معضل شهرهای پرجمعیت و زوال دموکراسی، به ما یاری خواهد رساند.

امیدوارم توضیحاتم تا اینجا برای ارائهٔ خلاصه‌ای اولیه از چیستی هوش مصنوعی کفایت کرده باشد. یک نکتهٔ بسیار گویا در مورد هوش مصنوعی را به خاطر داشته باشید که اختلاف نظرهای خود دانشمندان هوش مصنوعی نیز در مورد تعریف آن بی‌پایان است؛ تا جایی که وقتی از چهارصد متخصص هوش مصنوعی خواسته شد از میان تعاریف پیشنهادی یکی را انتخاب کنند، محبوب‌ترین تعریف را فقط ۵۶ درصد از آنان پذیرفتند.

بنابراین، با این تعریف شروع می‌کنم که هوش مصنوعی شاخه‌ای از فناوری دیجیتال به شمار می‌آید و هدفش توسعه و به‌کارگیری سیستم‌هایی است که قادر به انجام وظایف و فعالیت‌هایی هستند که به‌طور سنتی تصور می‌کردیم فقط توسط انسان با ذهن فوق‌العاده‌اش قابل انجام است. اکنون باید این نکتهٔ مهم را اضافه کنیم که سیستم‌های هوش مصنوعی روزبه‌روز بهتر از انسان عمل می‌کنند و کارهایی انجام می‌دهند که فراتر از توانایی‌های انسانی است. برخی مفسران و متخصصان هستند که می‌خواهند حتی از این هم فراتر روند و بگویند که یک سیستم فقط در صورتی هوش مصنوعی واقعی است که بتوان آن را هوشیار (Conscious) نامید. بااین‌حال، به دلایلی که بعداً توضیح خواهم داد، فکر می‌کنم این ویژگی لازم نیست و می‌تواند موجب سردرگمی شود.

یکی از پیامدهای سیستم‌های هوش مصنوعی، همان‌طور که اشاره شد، این است که پرسش‌های تازه‌ای را در عرصه‌های سنتی مطرح می‌کنند. برای نمونه، در آموزش این سؤال مطرح شده که جوانان را برای چه آینده‌ای تربیت می‌کنیم. فعالان حوزهٔ سرمایه‌گذاری خطرپذیر و سهام خصوصی، امروزه بیش از گذشته نگران‌اند صنایعی که سرمایه‌شان

را در آن‌ها به کار انداخته‌اند، مبادا تا زمان سوددهی، دستخوش تحولات اساسی ناشی از هوش مصنوعی شوند. در بخش نظامی نیز با ظهور تسلیحات آتش‌به‌اختیار، استراتژیست‌ها وادار شده‌اند اصول اولیهٔ بازدارندگی و دفاع را از نو بررسی کنند. همچنین، زیرساخت‌های هوش مصنوعی کشورها اکنون منبعی بالقوه برای برتری نظامی و ایجاد مزیت اقتصادی به شمار می‌آیند.

پیشرفت هوش مصنوعی باید در همهٔ حوزه‌هایی که نگاه بلندمدت دارند نیز مورد توجه قرار گیرد؛ اما اغلب به آن بی‌توجهی می‌شود. به‌عنوان مثال، در بحث‌های تغییرات اقلیمی که اغلب بر افق ۲۰۵۰ متمرکز است، کمتر شنیده می‌شود که به ارتباط هوش مصنوعی با این موضوع پرداخته شود. البته نگرانی‌هایی دربارهٔ افزایش مصرف انرژی هوش مصنوعی وجود دارد، اما شاید سؤال مهم‌تر این باشد که سیستم‌های هوش مصنوعی تا ۲۰۵۰ تا چه حد بشر و تمدن را دگرگون خواهند کرد. حداقل از نگاه این کتاب، وضعیت انسان در آن زمان ممکن است تفاوت چشمگیری داشته باشد و شاید سیستم‌های ما آن‌قدر توانمندتر شده باشند که از درک ما خارج باشد. این دیدگاه، اهمیت چالش آب‌وهوایی را کم نمی‌کند، بلکه ترغیبمان می‌کند افق وسیع‌تری برای آنچه تا ۲۰۵۰ می‌خواهیم، در نظر بگیریم.

بااین‌حال، هوش مصنوعی طیف وسیعی از خطرات، تهدیدها و آسیب‌پذیری‌های نگران‌کننده را نیز به وجود می‌آورد. این موارد در فصل‌های ۸ و ۹ بررسی شده‌اند.

تا اینجا تمرکز بحث ما بر قابلیت‌های فعلی و احتمالی هوش مصنوعی بوده است. در گام بعدی، به این پرسش می‌پردازیم که هوش مصنوعی چگونه کار می‌کند. این موضوع را به زبانی غیرتخصصی و از طریق بررسی تکامل هوش مصنوعی شرح خواهم داد.

تاریخچه‌ای کوتاه دربارهٔ هوش مصنوعی

ابتدا باید توضیح دهم چرا از پرداختن به جزئیات عملکرد سیستم‌های هوش مصنوعی پرهیز می‌کنم. بخشی از این تصمیم به این دلیل است که این کتاب برای خوانندگان عمومی و غیرمتخصص نوشته شده، اما مهم‌تر از آن، به این دلیل است که روش‌ها و فناوری‌های توانمندساز به‌سرعت درحال تغییرند. به عقیدهٔ من، برای عموم مردم درک

عمیق جزئیات جدیدترین فناوری‌ها در بلندمدت کمتر از این اهمیت دارد که بدانند این سیستم‌ها در سال‌های پیش رو ممکن است چه‌کارهایی انجام دهند. سیستم‌ها و روش‌های جدیدی پیوسته ظهور کرده‌اند و استفاده از آن‌ها متداول خواهند شد. به‌جای اینکه سعی کنیم تمام جزئیات فنی روش‌های کنونی را بفهمیم، بهتر است درکی کلی از چگونگی تکامل هوش مصنوعی در طول زمان به دست آوریم؛ در حقیقت، درک تاریخچه و مسیر پیشرفت هوش مصنوعی مهم‌تر است. برای دستیابی به این درک و نشان دادن پیشرفت‌های هوش مصنوعی، در ادامه به مرور دستاوردهای برجستهٔ این حوزه در ۷۵ سال گذشته خواهم پرداخت. این روایت تا حد زیادی دیدگاه شخصی‌ام است و داستان هوش مصنوعی را از نگاه نقاط عطفی که از نظر من اهمیت دارند، بیان می‌کند.

بیشتر مردم از طریق داستان‌های علمی ــ تخیلی هوش مصنوعی را شناخته‌اند. برخی از برجسته‌ترین نویسندگان ـ مانند آیزاک آسیموف[1]، نیل استیونسن[2]، ویلیام گیبسون[3]، فیلیپ کی. دیک[4]، آرتور سی. کلارک[5] ـ بسیاری از کاربردهای فعلی و نوظهور هوش مصنوعی را پیش‌بینی کرده بودند. آن‌ها خطراتش را نیز پیش‌بینی کرده بودند. این نویسندگان، در کنار برنامه‌های تلویزیونی و فیلم‌های مختلف ـ به‌ویژه فیلم *۲۰۰۱: ادیسه فضایی* (۱۹۶۸)، *بلید رانر* (۱۹۸۲)، *ترمیناتور* (۱۹۸۴)، و *ماتریکس* (۱۹۹۹) ـ تا حد زیادی انتظارات و ترس‌های ما را از ماشین‌ها و ربات‌های آینده شکل داده‌اند.

هوش مصنوعی به فیلم‌ها، برنامه‌های تلویزیونی و داستان‌های تخیلی معاصر محدود نیست. نشانه‌هایی از تصور هوش مصنوعی را می‌توان در دوران باستان کلاسیک یافت؛ برای نمونه، در *ایلیاد* هومر آنجا که در کارگاه هفائستوس (سازندهٔ سلاح خدایان) از تندیس طلایی کنیزکانی که شبیه به دوشیزگان زنده هستند، سخن گفته می‌شود. بیش از ۲۵۰۰ سال بعد، در قرن نوزدهم نیز تلاش‌های قابل توجهی در زمینهٔ هوش مصنوعی در ادبیات دیده شد که مشهورترینشان در *فرانکنشتاین* اثر مری شلی[6] (۱۸۱۸) و *اروهون* اثر ساموئل

1. Isaac Asimov
2. Neal Stephenson
3. William Gibson
4. Philip K. Dick
5. Arthur C. Clarke
6. Mary Shelley

باتلر[1] (۱۸۷۲) بود. در اوایل قرن بیستم نیز داستان‌های مرتبط وجود داشتند ـ برای مثال، در آثار اچ. جی. ولز[2] که به‌یادماندنی‌ترین آن‌ها در *جنگ دنیاها* (۱۸۹۵ تا ۱۸۹۷) است. چند دهه بعد، در رمان *۱۹۸۴* (منتشرشده در ۱۹۴۹)، جورج اورول[3] زمینه‌ساز بخش عمده‌ای از بحث‌های کنونی دربارهٔ نظارت الکترونیکی گسترده بر شهروندان جامعه شد. در ۱۹۵۰ با انتشار مقالهٔ کلیدی آلن تورینگ[4] با عنوان «ماشین‌های محاسباتی و هوش»، هوش مصنوعی وارد دنیای واقعیت شد. در این مقاله، او به‌شکلی روش‌مندانه به بررسی این پرسش پرداخت که «آیا ماشین‌ها می‌توانند فکر کنند؟» (در فصل ۱۰ دوباره به این مقاله بازخواهم گشت). چند سال بعد ـ در ۱۹۵۶. گروهی از دانشمندان برای برگزاری یک کارگاه تابستانی در کالج دارتموث در نیوهمپشایر آمریکا، گرد هم آمدند. اصطلاح «هوش مصنوعی» در درخواست بودجه‌ای که برای آن رویداد نوشته شده بود، ابداع شد. جان مک‌کارتی، که نویسندهٔ اصلی آن درخواست بود، امروزه خالق اصطلاح «هوش مصنوعی» شناخته می‌شود. گردهمایی دارتموث، محفلی از استعدادهای آمریکایی بود. اولین کنفرانس بین‌المللی هوش مصنوعی، سمپوزیومی با عنوان «مکانیزه کردن فرایندهای تفکر» بود که اواخر ۱۹۵۸ در آزمایشگاه ملی فیزیک در تدینگتونِ انگلستان برگزار شد.

اولین سیستم عملیاتی هوش مصنوعی، پیش از آنکه اصطلاح «هوش مصنوعی» پدید آید، تا حد زیادی طراحی و توسعه یافته بود. این سیستم با نام «نظریه‌پرداز منطقی» (Logic Theorist) شناخته می‌شود[5] که آلن نیوول[6]، دانشمند کامپیوتر، و هربرت سایمون[7]، اقتصاددان، در ۱۹۵۵ طراحی و در ۱۹۵۶ تکمیل کردند[8]. نوآوری این

1. Samuel Butler
2. H. G. Wells
3. George Orwell
4. Alan Turing

۵. «نظریه‌پرداز منطقی» یا «منطق‌دان» برنامه‌ای کامپیوتری است که در ۱۹۵۶ توسط آلن نیوئل، هربرت ای. سایمون و کلیف شا نوشته شد. این برنامه اولین برنامه‌ای بود که برای انجام استدلال خودکار طراحی شده بود و «اولین برنامهٔ هوش مصنوعی» توصیف شده است. «منطق‌دان» ۳۸ قضیه از ۵۲ قضیهٔ اول در فصل دوم کتاب مبادی ریاضیات نوشتهٔ وایت‌هد و برتراند راسل را اثبات کرد و برای برخی از آن‌ها اثبات‌های جدید و کوتاه‌تری پیدا کرد. ـ م.

6. Alan Newell
7. Herbert Simon

۸. پاسخ برتراند راسل به «منطق‌دان» بسیار آینده‌نگرانه بود. او در نامه‌ای به هربرت سایمون در ۲۱ سپتامبر ۱۹۵۷ نوشت: «دلایل شما را برای پنهان کردن [جزئیات توانایی‌های این سیستم] از دانش‌آموزان درک می‌کنم. چطور می‌توان انتظار داشت آن‌ها یاد بگیرند مسائل ریاضی را حل کنند، وقتی می‌دانند ماشین‌ها این کار را بهتر انجام می‌دهند؟»

سیستم در این بود که فراتر از کاربردهای معمول کامپیوترهای آن زمان ـ مثل محاسبات عددی، علم حساب و رمزگشایی ـ عمل می‌کرد. در عوض، این سیستم نوعی استدلال غیرعددی و نمادین انجام می‌داد که تا اندازه‌ای شبیه‌سازیِ فرایند حل مسئله در انسان بود. این همان سیستمی بود که سایمون در نقل‌قول ابتدای این کتاب به توصیف آن پرداخته بود؛ توصیفی که به اعتراف خودش، «تکبرآمیز» بود.

در دوران آغازین، بذری دیگر نیز کاشته شد. جان فون نویمان، در مجموعه‌ای از سخنرانی‌های مکتوب که عمرش برای تکمیل و ارائه‌شان کفاف نداد، اولین تحلیل ریاضی پایدار را از روابط بین مغز بیولوژیکی و کامپیوتر انجام داد. شاید نام فون نویمان چندان شناخته‌شده نباشد، اما او غولی در ریاضیات و فیزیک قرن بیستم بود و تأثیری بی‌حدوحصر بر توسعهٔ کامپیوترها و هوش مصنوعی داشت. تفکر او در این کتاب و بیشتر کتاب‌های دیگر در این زمینه نفوذ کرده است. کامپیوترهای امروزی هنوز براساس چیزی ساخته می‌شوند که به «معماری فون نویمان» معروف است. در دههٔ ۱۹۵۰، در مؤسسهٔ مطالعات پیشرفتهٔ پرینستون، فون نویمان در همان جایی کار می‌کرد که آلبرت اینشتین، کورت گودل[1] و رابرت اوپنهایمر[2] حضور داشتند. در میان این جمع، فون نویمان بود که به او لقب «مرد بزرگ» داده شده بود. یکی از داستان‌های مورد علاقه‌ام دربارهٔ فون نویمان ـ که شاید ساختگی باشد، اما خنده‌دار است ـ این است که وقتی چهار یا پنج‌ساله بود، از مادرش که سیگار می‌کشید و از پنجره به بیرون خیره شده بود پرسید: «مامان، داری چی رو حساب می‌کنی؟»

دههٔ ۱۹۶۰ پیشرفت‌های متعددی در هوش مصنوعی به همراه داشت، اما از نظر من دو مورد برجسته‌تر بود. اولین مورد، پروژه «دِندرال» (DENDRAL) بود که ادوارد فایگنبام[3] در دانشگاه استنفورد رهبری‌اش می‌کرد. او دانش تخصصی جاشوا لدربرگ[4]، برندهٔ جایزهٔ نوبل پزشکی، و کارل جراسی[5]، مخترع قرص ضدبارداری، را در سیستم‌های مختلفی ثبت

1. Kurt Gödel
2. Robert Oppenheimer
3. Edward Feigenbaum
4. Joshua Lederber
5. Carl Djerassi

و بازنمایی کرد[۱]. این سیستم توانست در سطحی فراتر از هر انسان، ساختار مولکول‌های ناشناخته را با دریافت داده‌های طیف‌سنجی جرمی تعیین کند. دِندرال با موفقیت تکنیک‌های نوظهور یادگیری ماشین (در ادامه توضیح داده می‌شود) و روش‌های اکتشافی را (قواعد سرانگشتی که در ترکیب با دانش رسمی، عملکردی در سطح متخصص ارائه می‌دهند) پیاده‌سازی کرد.

دومین نقطهٔ عطف از نظر من در دههٔ ۱۹۶۰، توسعهٔ سیستمی به نام «الیزا»[۲] توسط جوزف وایزنباوم[۳] ــ استاد هوش مصنوعی دانشگاه ام‌آی‌تی در آن زمان ــ بود. این برنامه که به نام شخصیت الیزا دولیتل در نمایشنامهٔ *پیگمالیون* اثر جورج برنارد شاو نام‌گذاری شده بود، تا حدی لحنی شوخ‌طبعانه داشت. این سیستم تعامل یک بیمار و یک روان‌درمانگر انسانی را تقلید می‌کرد. وایزنباوم به منشی خود پیشنهاد داد از این برنامه استفاده کند، اما وقتی منشی درخواست کرد وایزنباوم اتاق را ترک کند تا بتواند به‌تنهایی با برنامه کار کند، او به‌شدت شوکه شد. در واقع، این اتفاق چنان بر او تأثیر گذاشت که ترغیبش کرد یکی از بهترین کتاب‌ها را در زمینهٔ رابطهٔ انسان و ماشین با عنوان *قدرت کامپیوتر و خرد انسانی*[۴] (۱۹۸۴) به نگارش درآورد. او با پیش‌بینی بسیاری از بحث‌های اخلاقی کنونی دربارهٔ هوش مصنوعی، در آن کتاب نسبت به خطرات ماشین‌هایی که روزبه‌روز توانمندتر می‌شوند و تأثیراتشان بر بشر هشدار داد.

دههٔ ۱۹۷۰، تا حد زیادی با الهام از دِندرال، دهه‌ای سرشار از خوش‌بینی برای دانشمندان هوش مصنوعی بود. آن‌ها سرگرم تبدیل نظریه‌ها به سیل عظیمی از کاربردهای عملی شدند که نخستین نمونه‌های «سیستم‌های خبره» بودند. یکی از این سیستم‌ها، PROSPECTOR بود که به زمین‌شناسان در یافتن کانسارهای معدنی کمک می‌کرد؛ این سیستم چنان موفق عمل کرد که گفته می‌شود کشف یک معدن مولیبدن به ارزش صد میلیون دلار را ممکن کرد. شاخص‌ترین سیستم‌های آن دوره، در زمینهٔ تشخیص

---

۱. منظور این است که او تلاش کرد دانش و تجربهٔ متخصصان را استخراج و به‌شکلی ساخت‌یافته و قابل فهم برای ماشین تبدیل کند، به‌گونه‌ای که کامپیوتر بتواند براساس آن استدلال و تصمیم‌گیری کند. به بیان دیگر، این پروژه توانست فرایندهای فکری متخصصان را در قالب کدهای کامپیوتری بازآفرینی کند. ــ م.

2. ELIZA

3. Joseph Weizenbaum

4. *Computer Power and Human Reason*

پزشکی و توصیهٔ درمانی فعال بودند. ازجملهٔ این‌ها می‌توان به MYCIN برای بیماری‌های عفونی، CADUCEUS در پزشکی داخلی و CASNET برای گلوکوم (آب‌سیاه) اشاره کرد. گفته می‌شد این سیستم‌ها همگی در سطح متخصص یا حتی بالاتر عمل می‌کردند.

به‌رغم خوش‌بینی و پیشرفت‌های اولیه، هوش مصنوعی در دههٔ ۱۹۷۰ با فرازونشیب‌هایی روبرو شد. اصطلاح «زمستان هوش مصنوعی» معمولاً به دوره‌ای اشاره دارد که کاهش چشمگیری در علاقه و سرمایه‌گذاری در این حوزه رخ داد. منتقدان و مخالفان فناوری همیشه از چنین دوره‌هایی استقبال می‌کنند. اولین زمستان هوش مصنوعی در اواخر دههٔ ۱۹۷۰ به وقوع پیوست. این افول، پس‌از کاهش بودجه‌های دولتی ـ به‌ویژه در ایالات متحده ـ رخ داد و بازتاب‌دهندهٔ این دیدگاه بود که فناوری موجودی جذاب است، اما بعید است نتایج عملی قابل توجهی ارائه دهد و به وعده‌های اولیهٔ پیشگامانش جامه عمل بپوشاند. هوش مصنوعی نتوانسته بود بازار را متقاعد کند.

بااین‌حال، با ظهور فناوری «سیستم‌های خبرهٔ قاعده‌محور» (Rule-based expert systems) در دههٔ ۱۹۸۰، زمستان به بهار تبدیل شد؛ این فناوری در آن زمان شاخهٔ کاربردی و تجاری هوش مصنوعی تلقی می‌شد. من سیستم‌های خبره را نسل اول سیستم‌های عملی هوش مصنوعی می‌دانم. جامعهٔ کسب‌وکار خیلی سریع متوجه شد چگونه می‌توان از آن‌ها به‌طور جدی در بخش‌های مختلف ازجمله حقوق، امور مالی، بانکداری، بیمه، خرده‌فروشی، تولید و غیره استفاده کرد.

ایدهٔ اصلی در این سیستم‌ها، ایجاد درخت‌های تصمیم یا فلوچارت‌هایی بود که به‌صورت کلی برای تقلید از روند استدلال متخصصان انسانی طراحی شده بودند. کاربران غیرمتخصص قادر بودند در این درخت‌های تصمیم حرکت کنند. سیستم‌ها از کاربران سؤال می‌کردند و در پاسخ، جواب‌هایی ارائه می‌دادند و مستنداتی تولید می‌کردند که باز هم در سطح یک متخصص بود.

در دههٔ ۱۹۸۰، تمام فکر و ذکرم این سیستم‌ها بود. پس‌از اینکه در ۱۹۸۶ رسالهٔ دکترایم را در این زمینه نوشتم، از سال ۱۹۸۷ تا ۱۹۸۸ به همراه فیلیپ کپر ـ از متخصصان برجستهٔ حقوق ـ اولین سیستم هوش مصنوعی کاملاً عملیاتی و تجاری جهان را برای وکلا ایجاد کردم. سیستم ما در مواردی مشاوره می‌داد که به دلیل انقضای مهلت قانونی، امکان

پیگیری دعوی در دادگاه وجود نداشت. در توسعهٔ این سیستم، من و فیلیپ یک درخت تصمیم‌گیری عظیم با بیش از دو میلیون مسیر ایجاد کردیم که تمام ترکیبات ممکن از الگوهای واقعی و مسائل حقوقی را در این شاخهٔ پیچیده از قانون پوشش می‌داد. به‌طور کلی، ما تفسیر او از قوانین مربوطه و سوابق قضایی را به یک برنامهٔ کامپیوتری پیچیده تبدیل کردیم؛ به عبارت دیگر، دانش او در قالب کد بیان شد. سیستم ما زمان لازم برای تحقیقات را از چند ساعت به چند دقیقه کاهش داد و دربارهٔ روند استدلال خود نیز توضیحاتی ارائه می‌داد. با گذشت بیش از سه دهه، فیلیپ هنوز مشتاقانه اعتراف می‌کند که سیستم از او بهتر عمل می‌کرد.

بعد، در اواخر دههٔ ۱۹۸۰ و اوایل دههٔ ۱۹۹۰ شاهد آغاز دومین زمستان هوش مصنوعی بودیم. اغلب دلیل این رکود را دشواری و هزینهٔ بالای عملیاتی‌سازی سیستم‌های خبره در مقیاس صنعتی و گاهی نیز سقوط بازار سخت‌افزارهای مخصوص هوش مصنوعی می‌دانند. اما به عقیدهٔ من، ظهور شبکهٔ جهانی وب، عامل تسریع این زمستان بود. در اوایل دههٔ ۱۹۹۰، متخصصان هوش مصنوعی در حوزه‌های کاری خود به‌وضوح دیدند که ارائهٔ محتوا و خدمات به‌صورت آنلاین تا چه حد آسان است. همین باعث شد بسیاری از ما که در زمینهٔ هوش مصنوعی برای کاربردهای حرفه‌ای کار می‌کردیم، تا چند سال به‌سمت توسعهٔ ابزارها و برنامه‌های تحت وب تغییر مسیر بدهیم. قبول داشتیم که این ابزارها به‌اندازهٔ سیستم‌های خبرهٔ قاعده‌محور پیچیده ـ که پیش از آن، برایشان زحمت کشیده بودیم ـ پیشرفته نبودند، اما توسعهٔ وب نوعی از فوریت، جذابیت تجاری و رویکرد عمل‌گرایانه داشت که بخش بزرگی از جامعهٔ هوش مصنوعی را مجذوب خود کرد.

بسیاری از ما از اوایل تا اواسط دههٔ ۱۹۹۰ چندان توجهی به هوش مصنوعی نداشتیم. اما در ۱۹۹۷، وقتی باخبر شدیم که قهرمان شطرنج جهان، گری کاسپاروف، توسط یک سیستم کامپیوتری از شرکت آی‌بی‌ام به نام «دیپ‌بلو» (Deep Blue) شکست خورده است، شگفت‌زده شدیم. در دههٔ ۱۹۸۰، فکر می‌کردیم این اتفاق هرگز ممکن نیست. تنها راه ما در آن زمان برای ارائهٔ هوش مصنوعی عملی، سرهم‌بندی درخت‌های تصمیم بسیار بزرگی بود که روند استدلال ظاهری متخصصان انسانی را بازنمایی می‌کرد، درست مثل کاری که در زمینهٔ حقوق انجام داده بودیم.

بااین‌حال، متخصصان برجسته ـ برای مثال، شطرنج‌بازان و پزشکان ـ معتقد بودند در چالش‌برانگیزترین شرایطی که با آن روبرو می‌شوند، تخصص و تجربه‌شان را نمی‌شود به صرف فلوچارت فروکاست. آن‌ها احساس می‌کردند بهترین حرکات یا تشخیص‌هایشان وابسته به نوعی دانش ناگفتنی و ضمنی ـ شهود یا غریزه ـ است که در هیچ سیستم کامپیوتری نمی‌شود کدگذاری‌اش کرد. همین باعث شد در دههٔ ۱۹۸۰ باور داشته باشیم همواره یک تقسیم کار مشخص بین ماشین‌ها و انسان‌ها وجود خواهد داشت: ماشین‌ها برای انجام کارهای روتین مناسب‌تر خواهند بود، اما اگر تخیل و خلاقیت لازم باشد، انسان مالک بدون رقیب آن‌ها باقی خواهد ماند.

بسیاری از مردم امروز نیز به این تقسیم کار (میان انسان و ماشین) باور دارند، اما اشتباه می‌کنند. این باوری غلط در ۱۹۹۷ بود، چراکه درک درستی از تأثیر افزایش نمایی در قدرت پردازش کامپیوترها نداشتیم (به فصل ۲ مراجعه کنید). هنگامی که دیپ‌بلو، کاسپاروف را شکست داد، سیستم قادر بود بیش از سیصد میلیون حرکت ممکن را در یک ثانیه بررسی کند. درحالی‌که بهترین شطرنج‌بازان در هر لحظه می‌توانند حدود ۱۱۰ حرکت را در نظر بگیرند. کاسپاروف در واقع مغلوب سیستمی نشد که مانند خودش بازی می‌کرد، بلکه بیشتر در برابر توان پردازش غیرهوشمندانهٔ[1] آن شکست خورد.

اما اشتباه ما فقط این نبود؛ ما در این باور که ماشین‌ها همیشه به فعالیت‌های عادی محدود خواهند ماند، اشتباه می‌کردیم. معلوم شد راه متفاوتی برای توسعهٔ سیستم‌هایی که می‌توانند عملکردی در سطح متخصصان انسانی داشته باشند، پیدا شده است. این اتفاق به پدیده‌ای منجر شد که آن را سیستم‌های هوش مصنوعی نسل دوم می‌نامم.

نسل اول، یعنی سیستم‌های خبره، نیازمند «مهندسان دانش» بودند تا گوهرهای دانش را از ذهن «متخصصان هر حوزه» استخراج و دانش آن‌ها را به درخت‌های تصمیم تبدیل کنند. در واقع، جنبهٔ هوشمندانهٔ این سیستم‌ها، تلاش انسانی برای ساخت این فلوچارت‌ها بود. خود سیستم‌های عملیاتی مانند ابزارهای مسیریابی عمل می‌کردند. در مقابل، سیستم‌های هوش مصنوعی نسل دوم خودشان بخش بیشتری از کار سنگین

۱. منظور از پردازش بروت‌فورس (Brute-force processing) که در برخی منابع آن را غیرهوشمندانه یا فراگیر می‌نامند، آزمایشِ تمام راه‌حل‌های ممکن به‌طور نظام‌مند اما بدون استفاده از الگوریتم‌های بهینه یا هوشمندانه است تا پاسخ صحیح پیدا شود. در این مثال، کامپیوتر در بازی شطرنج از روش بروت‌فورس استفاده می‌کند تا تمام حرکات ممکن و نتایج آن‌ها را بررسی کند و در نهایت، بهترین حرکت را پیدا کند. ـ م.

را بر عهده دارند. نسل دوم، به‌جای استخراج مستقیم دانش از انسان‌ها، به استخراج گوهرها از اقیانوس‌های وسیع داده می‌پردازد. جای تعجب نیست که تمام داده‌های دیجیتالی که انباشته کرده‌ایم ــ در وب، در سیستم‌ها و در پایگاه‌های داده‌مان، چه به‌صورت متنی و چه چندرسانه‌ای ـ بسیاری از تجربیات، ایده‌ها، بینش‌ها، تخصص‌ها و احساسات مشترک و انباشتهٔ بشریت را در خود جای داده، مجسم کرده، منعکس کرده و در بر گرفته است. جهش تکنولوژیکی در اینجا، توسعهٔ سیستم‌هایی بود که می‌توانند مستقیماً الگوها، روندها و روابط را در این اقیانوس‌های داده تشخیص دهند. هدف، توسعهٔ سیستم‌های کامپیوتری بوده که بتوانند به شیوه‌های مختلف از حجم عظیمی از داده‌ها «یاد بگیرند»، بدون اینکه به برنامه‌ریزی صریح نیاز داشته باشند؛ ازاین‌رو، عبارت «یادگیری ماشین» پدید آمد. اغلب گفته می‌شود این سیستم‌ها برای انجام وظایف خاصی ـ از تشخیص بیماری گرفته تا نگارش اسناد پیچیده ـ «آموزش دیده‌اند». تکنیک اصلی‌ای که از یادگیری ماشین پشتیبانی می‌کند، با عنوان «شبکه‌های عصبی» (Neural networks / Neural nets) شناخته می‌شود. بخش عمده‌ای از کارهای پیشگامانه در این زمینه توسط جفری هینتون[1] ـ برندهٔ جایزهٔ نوبل ـ در دههٔ ۱۹۸۰ انجام شد، بااین‌حال، مقالهٔ او دربارهٔ «یادگیری عمیق» در سال ۲۰۰۶ به‌عنوان نقطهٔ عطفی خاص در هوش مصنوعی قلمداد می‌شود. شبکه‌های عصبی، همان‌طور که از نامشان پیداست، از ساختار مغز انسان الهام گرفته شده‌اند. البته سیستم‌های کنونی مغز انسان را بازتولید نمی‌کنند، بلکه ایدئال‌سازی (Idealization)[2] از نحوهٔ عملکرد مغز به نظر می‌رسند.

یکی از اولین موفقیت‌های مهم و پرمخاطب در حوزهٔ یادگیری ماشین در ۲۰۱۱ اتفاق افتاد، زمانی که آی‌بی‌ام به دستاورد دیگری رسید. در برنامهٔ زندهٔ تلویزیونی مسابقهٔ معلومات عمومی در آمریکا به نام *جئوپاردی!* (*Jeopardy!*) کامپیوتری تحت عنوان «واتسون» (Watson) ـ که به نام بنیان‌گذار آی‌بی‌ام، توماس جی. واتسون، نام‌گذاری شده بود ـ دو نفر از بهترین شرکت‌کنندگان تاریخ این مسابقه را در هم کوبید. در واقع، این سیستم هوش مصنوعی توانست به هر سؤالی با سرعت و دقتی بالاتر از هر انسانی پاسخ دهد. واتسون

1. Geoffrey Hinton

۲. گرفتن یک سیستم پیچیده (مثل مغز) و ساختن یک مدل ساده‌شده از آن که فقط ویژگی‌های کلیدی و مهم آن را حفظ می‌کند. به عبارت دیگر، ایدئال‌سازی یعنی ساده کردن واقعیت برای ساختن یک مدل قابل فهم و قابل محاسبه که تحلیل مسائل را راحت‌تر کند. ـ م.

ترکیبی از هوش مصنوعی نسل اول بود که در آن دانش به‌صورت صریح کدنویسی شده بود و هوش مصنوعی نسل دوم که متکی بر یادگیری ماشین از حجم وسیعی از داده‌ها بود. آن موقع بود که مشتاقان هوش مصنوعی دوباره گرمای تابستان جدیدی را احساس کردند.

چند سال بعد، آلفاگو (AlphaGo) از راه رسید که هنوز یکی از مثال‌های مورد علاقه‌ام از یک سیستم یادگیری ماشین با عملکرد بالاست (هرچند عناصری از هوش مصنوعی نسل اول نیز در آن وجود دارد). این سیستم در ابتدا توسط شرکت دیپ‌مایند (DeepMind) ـ که در آن زمان یکی از زیرمجموعه‌های گوگل بود ـ برای بازی پیچیدهٔ گو (Go) طراحی شد؛ بازی‌ای چنان پیچیده که تعداد حرکات ممکن در آن بیشتر از تعداد اتم‌ها در جهان است. پیش از این، متخصصان هوش مصنوعی تردید داشتند که هرگز سیستمی بتواند گو را به‌صورت حرفه‌ای بازی کند. اما در اوایل ۲۰۱۶، آلفاگو توانست لی سدول ـ بازیکن کلاس جهانی گو ـ را با نتیجهٔ ۴ بر یک شکست دهد. این سیستم با بهره‌گیری از «شبکه‌های عصبی عمیق» (Deep neural networks) آموزش دیده بود. این آموزش هم از طریق «یادگیری با نظارت» (Supervised learning) (با استفاده از بازی‌های قبلی انجام‌شده توسط بازیکنان خبرهٔ انسانی) و هم از طریق «یادگیری تقویتی» (Reinforcement learning) (که در آن با انجام میلیون‌ها بازی در برابر خودش، خود را بهبود می‌بخشید) صورت گرفته بود.

علاقه‌مندان به هوش مصنوعی دربارهٔ حرکت سی‌وهفتم طی بازی دوم مسابقهٔ آلفاگو در ۲۰۱۶ زیاد صحبت می‌کنند. این حرکتِ آلفاگو بود و هیچ انسانی پیش از آن به فکر چنین حرکتی نیفتاده بود. یکی از بازیکنان برجستهٔ انسانی آن را «زیبا» توصیف کرد و گفت از دیدنش اشک در چشمانش جمع شد. اگر این حرکت را یک انسان انجام داده بود، آن را «خلاقانه» یا «نوآورانه» یا حتی «نبوغ‌آمیز» می‌نامیدیم؛ اما هیچ‌کدام از این‌ها نبود. این حرکت حاصل ترکیبی از قدرت پردازش عظیم و الگوریتم‌های خارق‌العاده‌ای بود که روی حجم وسیعی از داده‌ها عمل می‌کردند. نکتهٔ قابل توجه این است که برخی از حرکات آلفاگو، فراتر از تصور کسانی بود که خود سیستم را طراحی کرده بودند.

طبق مشاهدهٔ برخی مفسران، آلفاگو سیستم قابل توجهی بود، اما عملکرد سیستم عمدتاً مبتنی بر یادگیری از روی بازی‌های گذشتهٔ اساتید این حوزه بود؛ به عبارت دیگر،

بدون داده‌های آموزشی اولیه که محصول استعداد انسانی بودند، نمی‌توانست چنین عملکردی داشته باشد. در ۲۰۱۷ آلفاگو زیرو (AlphaGo Zero) آمد. به این سیستم هیچ‌یک از بازی‌های انسانی گذشته داده نشد. در عوض، فقط قواعد بازی را به آن آموزش دادند و خود سیستم باید استراتژی‌ها و تاکتیک‌های بهینه را کشف می‌کرد. آلفاگو زیرو توانست آلفاگوی اصلی را با نتیجهٔ صد بر صفر شکست دهد. نسخه‌ای از آلفاگو زیرو به نام آلفازیرو (AlphaZero) حتی پا را فراتر گذاشت و برترین برنامه‌های جهان در بازی‌های شطرنج، شوگی و گو را شکست داد.

در ۲۰۱۸، از همان شرکت، آلفافولد (AlphaFold) معرفی شد. به گفتهٔ ونکی راماکریشنان، دانشمند برندهٔ جایزهٔ نوبل، این سیستم «پیشرفتی خیره‌کننده در مسئلهٔ تاشدگی پروتئین[1] بود؛ مسئله‌ای که پنجاه سال یکی از چالش‌های بزرگ زیست‌شناسی محسوب می‌شد. این اتفاق دهه‌ها زودتر از آنچه بسیاری در این رشته پیش‌بینی می‌کردند، رخ داد». این دستاورد در ۲۰۲۴ ثبت شد؛ دمیس هاسابیس و جان جامپر، دو دانشمند هوش مصنوعی که رهبری توسعهٔ آلفافولد را بر عهده داشتند، مشترکاً جایزهٔ نوبل شیمی را به پاس «پیش‌بینی ساختار پروتئین» دریافت کردند. انتظار می‌رود آلفافولد و نسخه‌های پس‌از آن تأثیر عمیقی بر درمان بیماری‌ها و مدیریت پسماندهای صنعتی داشته باشند.

پیشرفت‌های بعدی

زمانی که دیپ‌مایند روی اولین نسخهٔ آلفافولد کار می‌کرد، گروهی دیگر از دانشمندان هوش مصنوعی در گوگل مقاله‌ای منتشر کردند که گامی مهم در توسعهٔ هوش مصنوعی به شمار می‌رفت. این مقاله به تولد یک معماری جدید یادگیری عمیق برای شبکه‌های عصبی به نام «ترنسفورمرها» (Transformers) منجر شد. این همان معماری‌ای است که «مدل‌های زبانی بزرگ» بر پایهٔ آن ساخته می‌شوند. این سیستم‌ها روی حجم عظیمی از داده‌ها «آموزش» می‌بینند. باز هم تأکید می‌کنم که زیاد نگران اصطلاحات تخصصی و نحوهٔ کارکردشان نخواهیم بود. کافی است که به گفتهٔ استیون ولفرام[2] دربارهٔ این

---

۱. به فرایند تبدیل زنجیرهٔ خطی اسیدهای آمینه به ساختار سه‌بعدی منحصربه‌فرد، تاشدگی پروتئین می‌گویند.ـ م.

2. Stephen Wolfram

معماری جدید اشاره کنیم که: «به‌طور شگفت‌انگیزی برای همگان، یک شبکهٔ عصبی... قادر به انجام همان کاری است که پیش از این منحصراً از عهدهٔ انسان برمی‌آمد و آن هم تولید زبان انسانی قابل فهم است.» با استفاده از معماری ترنسفورمر، در سی‌ام نوامبر ۲۰۲۲، شرکت اوپن اِی‌آی ـ که در آن زمان یک آزمایشگاه تحقیقاتی آمریکایی بود ـ چت‌جی‌پی‌تی را راه‌اندازی کرد که شناخته‌شده‌ترین مدل زبانی بزرگ است. من که نسخه‌های قبلی خانوادهٔ جی‌پی‌تی را دنبال می‌کردم و درباره‌شان نوشته بودم، از این سیستم جدید حیرت کردم[۱]. چشمگیرترین پیشرفتی بود که طی چهل سال فعالیتم در زمینهٔ هوش مصنوعی دیده بودم. معرفی چت‌جی‌پی‌تی، سرآغازی برای شناخت گستردهٔ ظرفیت‌های هوش مصنوعی در میان عموم مردم بود. ناگهان، به نظر می‌رسید سیستمی داریم که می‌تواند تقریباً به هر سؤالی که با زبان عادی پرسیده می‌شود، پاسخ‌های مشخص و ملموس ارائه دهد. شگفت‌انگیزتر اینکه وقتی از آن خواسته می‌شد متنی را بنویسد، پاسخ در عرض چند ثانیه ظاهر می‌شد؛ خواه یک غزل، خواه یادداشت تسلیت یا حتی مقاله‌ای مفصل دربارهٔ هر موضوعی. دامنه و خروجی آن بی‌حدومرز به نظر می‌رسید.

مانند بسیاری از سیستم‌های مشابه دیگری که پس‌از آن عرضه شدند (مثل جِمنای (Gemini) از گوگل، بینگ (Bing) از مایکروسافت، کِلاد (Claude) از انتروپیک (Anthropic)، چت‌جی‌پی‌تی یک چت‌بات است؛ یعنی مکالمهٔ انسانی را تقلید می‌کند. همچنین، به‌عنوان سیستم هوش مصنوعی «مولد» (Generative) طبقه‌بندی می‌شود، به این معنی که محتوا را براساس درخواست تولید می‌کند. این محتوا فقط به متن محدود نیست، این سیستم‌ها می‌توانند آثار هنری، موسیقی و ویدیوهای چشمگیری تولید کنند. آن‌ها حتی این قابلیت را دارند که کدهای برنامه‌نویسی با کیفیت بالا تولید کنند؛ به بیان ساده، قادر به نوشتن نرم‌افزار هستند.

یک نکتهٔ اساسی در مورد چت‌جی‌پی‌تی که اغلب از نظر پنهان می‌ماند این است که نسخهٔ اولیهٔ آن فقط یک گام آزمایشی بود. اهمیت چت‌جی‌پی‌تی و هوش مصنوعی مولد را باید در نقش پیشگامی‌شان برای نسل‌های آیندهٔ این سیستم‌ها جستجو کرد، نه در

۱. GPT مخفف Generative Pre-trained Transformer به معنی ترنسفورمر مولد از پیش آموزش‌دیده است. ـ م.

آنچه اواخر ۲۰۲۲ بودند یا حتی امروز هستند؛ زیرا با وجود توانایی‌های چشمگیر، همچنان کاستی‌هایی دارند. هنوز در ابتدای راهیم. بااین‌حال، نکتهٔ کلیدی این است که سرعت تغییر و پذیرش درحال افزایش است و همان‌طور که در فصل ۲ خواهد آمد، می‌توان به‌طور منطقی انتظار داشت این سیستم‌ها به‌مراتب توانمندتر و قابل اعتمادتر شوند.

چت‌جی‌پی‌تی و هوش مصنوعی مولد، جدیدترین فصل از داستانی ادامه‌دارند که ریشه در کارهای پیشگامانهٔ دههٔ ۱۹۵۰ دارد. این سیستم‌های جدید جایگزین سیستم‌های قدیمی‌تر هوش مصنوعی مانند سیستم‌های خبره و سیستم‌های پیش‌بینی اولیه نمی‌شوند. آن‌ها اساساً نقطهٔ پایان این داستان نیز نیستند. حوزهٔ هوش مصنوعی مجموعه‌ای درهم‌تنیده از تکنیک‌ها بوده و خواهد بود که جدیدترینشان هوش مصنوعی مولد است. اکنون حدود ۷۵ سال است که هوش مصنوعی وجود دارد. در چند دههٔ اول، هر پنج یا ده سال یک بار پیشرفت‌های بزرگ رخ می‌داد. اما امروز، به نظر می‌رسد هر شش تا دوازده ماه یک بار شاهد پیشرفت‌های مختلف هستیم که برخی به‌عنوان نوآوری‌های تکنولوژیکی و برخی دیگر به‌عنوان کاربردهای جدید و چشمگیر ظهور می‌کنند. همین حالا که در ۲۰۲۴ می‌نویسم، همگام شدن با تحولات ـ هم در فناوری‌های زیربنایی و هم در دامنهٔ کاربردها ـ دشوار است. اکنون داریم از تابستان داغ هوش مصنوعی لذت می‌بریم. برخی معتقدند هیجان و مقیاس سرمایه‌گذاری کنونی به یک گرمایش بیش از حد مصنوعی و موقت منجر شده است، اما هیچ‌کس نمی‌تواند سطح فوق‌العادهٔ علاقه و سرمایه‌گذاری را به روشی معقول انکار کند. دیدگاه من این است که تا آیندهٔ نزدیک در تابستان هوش مصنوعی به سر خواهیم برد.

به‌سختی می‌توان حوزه‌ای از زندگی بشر را یافت که تحت‌تأثیر هوش مصنوعی قرار نگرفته باشد. این مسیر به کجا ختم خواهد شد؟ سیستم‌های هوش مصنوعی تا چه حد می‌توانند پیشرفته‌تر شوند؟ آیا ما به خلق نوعی «ابرهوشمندی» (Superintelligence) نزدیک می‌شویم؟ آیا در آستانهٔ «تکینگی» (Singularity) هستیم؟ این اصطلاحات دقیقاً به چه معنا هستند؟ در فصل ۲ پس از بررسی جایگاه هوش مصنوعی در بستر کلی‌تر فناوری دیجیتال، آن‌ها را توضیح می‌دهم.

فصل دوم

۲

درباب فناوری

برای درک اهمیت بلندمدت هوش مصنوعی، چه مولد باشد و چه نباشد، لازم است با دیدگاهی جامع به تحولات جاری در عرصهٔ فناوری، به‌ویژه فناوری دیجیتال بنگریم. به باور من، درک جایگاه پیشرفت‌های هوش مصنوعی در چارچوب گسترده‌ترِ مراحل تکامل جوامع و تمدن بشری نیز بینش ارزشمندی ارائه می‌دهد. همچنین، این رویکرد کمک می‌کند سوگیری‌های احتمالی را که ممکن است درک ما را از رشد فناوری تحت‌تأثیر قرار دهند، تشخیص دهیم. با تکیه بر این چارچوب تحلیلی، می‌توانیم به این پرسش پاسخ دهیم: سیستم‌های هوش مصنوعی در نهایت تا چه حد می‌توانند پیشرفت کنند؟

فناوری دیجیتال

متخصصان فناوری روش‌های گوناگونی برای تبیین دوران خارق‌العاده‌ای دارند که در آن زیست می‌کنیم. من این موضوع را در قالب پنج عنوان بررسی می‌کنم.

نخستین و مهم‌ترین مفهومی که باید درک شود این است که ماشین‌های ما هرروزه توانمندتر می‌شوند. برای لحظه‌ای، یادگیری ماشین، محاسبات کوانتومی و سایر فناوری‌ها و مفاهیم توانمندساز رایج را فراموش کنید. به‌جای آن، بر این واقعیت تمرکز کنید که سیستم‌ها و ماشین‌های ما به‌سادگی کارهای بیشتر و بیشتری انجام می‌دهند. این موضوع جانِ کلام است. به نظر می‌رسد تقریباً هر روز خبرهایی از پیشرفت‌های شگفت‌انگیز جدید، سیستم‌ها، ابزارها، فناوری‌ها، اپلیکیشن‌ها و استارتاپ‌ها می‌شنویم. بخش عمده‌ای از این اخبار به سیستم‌هایی مربوط است که وظایفی را انجام می‌دهند که تا همین اواخر تصور می‌کردیم تنها در حوزهٔ توانایی‌های انسان قرار دارند؛ از ابزارهای تشخیص پزشکی گرفته تا سیستم‌هایی که می‌توانند نقاشی بکشند، از سیستم‌هایی که می‌توانند تشخیص دهند لبخند یک انسان واقعی یا ساختگی است تا سیستم‌هایی که می‌توانند پیش‌بینی کنند چند سال دیگر زنده خواهید ماند.

مدت‌هاست پذیرفته‌ایم که بیل مکانیکی، اره‌برقی و جرثقیل می‌توانند در انجام کارهای یدی که در باغ‌ها و ساخت‌وسازهایمان انجام می‌دهیم، از ما بهتر عمل کنند. پذیرفته‌ایم که خودرو سریع‌تر از ما حرکت می‌کند و نسل‌هاست که می‌دانیم کمباین می‌تواند محصولات را سریع‌تر از کارگر انسانی جمع کند. در کارخانه‌ها دهه‌هاست آگاهیم ماشین

برای بسیاری از فعالیت‌ها بسیار پربازده‌تر و ایمن‌تر از انسان‌هاست.

در زمینهٔ فعالیت‌ها و کارهایی که به توانایی‌های ذهنی انسان مربوط می‌شوند ـ مثل تفکر، استدلال، حل مسئله، فرضیه‌سازی و برنامه‌ریزی ـ هنوز به‌سختی می‌پذیریم که ماشین بتواند با ما رقابت کند یا پیشی بگیرد. بااین‌حال، در عصر دیجیتال، روند کلی روشن است؛ احتمالاً با گذشت زمان، بیشتر فعالیت‌ها و وظایفی که تصور می‌کردیم فقط توسط انسان قابل انجام هستند، به‌طور خودکار توسط سیستم‌های توانمند یا توسط انسان‌های کم‌تجربه‌تر با پشتیبانی این سیستم‌ها انجام خواهند شد. برخی نمونه‌های مرتبط را در فصل ۱ مطرح کردم. دستاوردهای سیستم‌های فعلی و نوظهور نه‌تنها نشان می‌دهد که سیستم‌های ما روزانه توانمندتر می‌شوند، بلکه حاکی از آن است که با سرعتی ظاهراً فزاینده درحال پیشرفت هستند. در بسیاری از زمینه‌ها، این سیستم‌ها همین حالا نیز از انسان بهتر عمل می‌کنند.

نکتهٔ اساسی دوم من دربارهٔ فناوری این است که تکنیک‌ها و فناوری‌های قابل توجهی که به ماشین‌های ما قابلیت می‌بخشند، با سرعتی تصاعدی درحال پیشرفت هستند. چه در مورد قدرت پردازش صحبت کنیم، چه ظرفیت حافظه، پهنای باند شبکه، فضای ذخیره‌سازی داده یا هر قابلیت دیگر، شاهد رشدی انفجاری هستیم که گاهی درک‌نشدنی است. شبکه‌های عصبی را در نظر بگیرید که فناوری زیربنایی عمدهٔ سیستم‌های هوش مصنوعی کنونی هستند. توان محاسباتی موجود برای «آموزشِ» این سیستم‌ها بین سال‌های ۲۰۱۰ تا ۲۰۲۲ تقریباً هر شش ماه، دو برابر شد[1]. پیشرفت در فناوری تراشه‌ها خیره‌کننده بوده است. در نتیجهٔ این روند، تقریباً هر ده سال، سرعت محاسبات حدود یک میلیون برابر بهبود خواهد یافت. درحالی‌که می‌توان و باید از پیشرفت‌های اخیر مانند چت‌جی‌پی‌تی و دیگر سیستم‌های هوش مصنوعی مولد شگفت‌زده شویم، شاید بیشتر از آن باید از این واقعیت متحیر شویم که خودِ فناوری زیربنایی این سیستم‌ها ممکن است در دههٔ آینده یک میلیون برابر پیشرفت کند.

ما در دورانی به سر می‌بریم که دگرگونی‌های فناورانه‌اش در تاریخ بشر بی‌سابقه بوده‌اند.

---

۱. شواهدی نیز وجود دارد که نشان می‌دهد فاصلهٔ دوبرابر شدن از هر شش ماه ممکن است به هر ده ماه افزایش یابد، اما هنوز در مورد روند بلندمدت آن (مثلاً، در افق ده‌ساله) دیدگاهی وجود ندارد.

افزون‌بر این، همان‌طور که پیش‌تر گفتم، سرعت خودِ این تغییرات نیز روبه‌افزایش است. رِی کورزوِیل[1]ـ مخترع و آینده‌پژوه برجسته ـ این شتاب فزاینده را در قالب «قانون بازده شتاب‌دهنده» (Law of accelerating returns) تبیین کرده است. این قانون بیان می‌کند که: «فناوری‌های اطلاعاتی مانند رایانش، به‌صورت تصاعدی ارزان‌تر می‌شوند، زیرا هر پیشرفت، طراحی مرحلهٔ بعدی تکامل آن‌ها را آسان‌تر می‌کند.» سیستم‌های نوظهور، خود در طراحی و توسعهٔ نسل‌های بعدی فناوری، نقش ایفا می‌کنند و این چرخه ادامه می‌یابد. واضح‌ترین نمونهٔ اخیر این پدیده را در سیستم‌های هوش مصنوعی مولد می‌بینیم که می‌توانند کد بنویسند؛ یعنی سیستم‌هایی که سیستم‌های دیگر را توسعه می‌دهند. این توانایی برای آیندهٔ فناوری و بشریت اهمیتی بنیادین دارد: وقتی سیستم‌ها قادر به خوداصلاحی، خودبهبودی و خودگسترشی باشند، باید شاهد پیشرفتی حتی سریع‌تر از آنچه تاکنون دیده‌ایم باشیم. تأثیر این تحول هم بر سیستم‌های موجود و هم در ظهور سیستم‌های کاملاً جدید مشهود خواهد بود. برای مثال، این فناوری‌ها می‌توانند به خلق جهان‌های مجازی عظیم و پویایی منجر شوند که به‌طور مستمر تکامل می‌یابند. این جهش فناورانه نه‌تنها ابزارهای ما، بلکه خود ماهیت پیشرفت را متحول خواهد کرد.

در مجموع، در پیشرفت فناوری‌های توانمندساز، ما از نقطهٔ زانویی[2] (knee point) منحنی نمایی عبور کرده‌ایم و باید انتظار پیشرفت‌های انفجاری بیشتری را در کل حوزهٔ فناوری دیجیتال داشته باشیم.

سومین نکته این است که هیچ «خط پایان آشکاری» در رقابت جهانی برای توسعهٔ فناوری‌های دیجیتال وجود ندارد. هیچ‌کس در شرکت‌های پیشرو فناوری جهان یا در آزمایشگاه‌های تحقیقاتی در ایالات متحده، چین یا کرهٔ جنوبی انتظار ندارد که کار فناوری به این زودی‌ها تمام شود. مقصدی نهایی در چشم‌انداز فناوری نیست؛ بلکه سیر آن، مسابقه‌ای بی‌امان و بی‌پایان است.

هوش مصنوعی، بازار و سرمایه‌گذاران را شیفته و تحت‌تأثیر خود قرار داده است. سرمایه‌گذاری‌های چشمگیری در این حوزه در جریان است. شرکت‌های بزرگ فناوری و

1. Ray Kurzweil

۲. در ریاضیات، نقطه زانویی نقطه‌ای از منحنی است که در آن، منحنی به‌وضوح خم می‌شود یا به عبارت دیگر، تغییر شیب از زیاد به کم (یا برعکس) به‌طور محسوس رخ می‌دهد.ـم.

فعالان حوزهٔ هوش مصنوعی، خود مبالغ سرسام‌آوری را صرف تحقیق و توسعه می‌کنند، دولت‌ها کمک‌های مالی کلانی اعطا می‌کنند، اکثر کسب‌وکارهای بزرگ بودجه‌هایی گزاف به پروژه‌های پایلوت و توسعهٔ راهکارهای مبتنی بر این فناوری اختصاص می‌دهند و از سوی دیگر، جامعهٔ استارتاپی پویا که از استعدادهای درخشان بهره می‌برد، شاهد جذب حجم گسترده‌ای از سرمایه‌گذاری‌های خطرپذیر است. حتی شرکت‌های سرمایه‌گذاری خصوصی نیز به این عرصه وارد شده‌اند تا فرصت‌های بزرگ را شکار کنند. بررسی‌های مؤسسه حسابداری PwC در ۲۰۱۷ (پیش از ظهور هوش مصنوعی مولد) نشان داد که سهم بالقوهٔ هوش مصنوعی در اقتصاد جهانی تا ۲۰۳۰ می‌تواند به ۱۵٫۷ تریلیون دلار برسد. در ۲۰۲۳ نیز شرکت مشاورهٔ مدیریت مک‌کینزی اعلام کرد هوش مصنوعی مولد فقط می‌تواند سالانه تا ۴٫۴ تریلیون پوند به اقتصاد جهانی بیفزاید. شاید در مورد رقم دقیق این اعداد بحث باشد، اما جهت‌گیری بازار شفاف است: هوش مصنوعی پدیده‌ای با ارزش اقتصادی فوق‌العاده است. در نتیجه، بسیاری از فارغ‌التحصیلان برتر و دانشمندان برجسته نتوانسته‌اند در برابر جاذبهٔ بودجه‌های تحقیقاتی، حقوق‌های کلان و سرمایه‌گذاری‌های سنگینی که کسب‌وکارها و پروژه‌های هوش مصنوعی به همراه آورده‌اند، مقاومت کنند.

ظهور سریع سیستم‌های جدید، گواهی بر گستردگی فعالیت‌ها در این حوزه است. به‌عنوان نمونه می‌توان به چت‌جی‌پی‌تی اشاره کرد که پرچم‌دار آخرین نسل از سیستم‌های هوش مصنوعی شناخته می‌شود. تنها طی چهار ماه پس از عرضه، نسخهٔ ارتقایافتهٔ آن (GPT-4) منتشر شد و هم‌زمان، رقبای متعددی نیز پدید آمدند. پیش‌بینی می‌کنم وقتی شما این کتاب را می‌خوانید، نسخه GPT-5 نیز در دسترس قرار گرفته باشد[1].

هم‌زمان، شاهد رشد و گسترش روش‌های نوینی در این حوزه هستیم. ازجمله می‌توان به تحقیقات در زمینهٔ «داده‌های مصنوعی» (Synthetic data) اشاره کرد. هرچند این مفهوم در نگاه اول ممکن است تکراری یا دور از ذهن به نظر برسد، اما در واقع به توانایی سیستم‌ها در تولید حجم انبوهی از داده‌های مصنوعی اشاره دارد که می‌توانند به‌عنوان منبع تغذیهٔ

۱. شرکت OpenAI، نسخهٔ GPT-5 را در هفتم اوت ۲۰۲۵ (۱۶ مرداد ۱۴۰۴) عرضه کرد.ـم.

سیستم‌های آیندهٔ هوش مصنوعی مورد استفاده قرار گیرند. این رویکرد می‌تواند وابستگی به فرایند جمع‌آوری داده از وب و پایگاه‌های دادهٔ موجود را که از نظر حقوقی چالش‌برانگیز است، کاهش دهد[۱]. نمونهٔ دیگر، توسعهٔ سیستم‌هایی است که دیگر نیازی به تخصص کاربران در زمینهٔ مهندسی پرامپت (Prompt engineering) ندارند. در عوض، خودِ این سیستم‌ها به‌صورت ساختاریافته قادرند با پرسش‌های هدفمند از کاربران، به تعریف دقیق، دسته‌بندی و طبقه‌بندی مسائل یا موضوعاتی بپردازند که نیاز به راهنمایی دارند. همچنین، می‌توان به ظهور سیستم‌های هوش مصنوعی خودمختار اشاره کرد که قادر به برقراری ارتباط با یکدیگر، اتخاذ تصمیمات مستقل یا دسترسی به منابع محاسباتی و فیزیکی و استفاده از آن‌ها بدون نیاز به دخالت مستقیم انسان هستند که گاهی از آن‌ها با عبارت «هوش مصنوعی عاملی» (Agentic AI) یاد می‌شود.

در گام بعدی، به چهارمین نکته می‌رسیم که هم شورانگیز است و هم به‌راستی دلهره‌آور. با فرض اینکه بشر در اثر یک فاجعهٔ جهانی (خواسته یا ناخواسته) از میان نرود، به نظر می‌رسد که تأثیرگذارترین و پیشرفته‌ترین فناوری‌های دیجیتال در راهند و در سال‌های پیش رو ظهور خواهند کرد. این دستاوردها را «فناوری‌های هنوز اختراع‌نشده» می‌نامم. لازم به یادآوری است که فناوری دیجیتال تنها حدود هشتادوپنج سال است که در زندگی ما حضور دارد و ما هنوز در مراحل ابتدایی بهره‌برداری از آن هستیم. دور از ذهن نیست که در آینده‌ای نزدیک، فناوری‌های امروزی ما همان‌قدر ابتدایی به نظر برسند که اکنون سخت‌افزارهای دههٔ ۱۹۸۰ را قدیمی می‌پنداریم.

در آینده، زندگی انسان‌ها با فناوری‌هایی متحول خواهد شد که هنوز اختراع نشده‌اند و حتی قادر به تصورشان نیستیم. پیشرفته‌ترین و اثرگذارترین سیستم‌ها هنوز در پردهٔ ابهام آینده قرار دارند و منتظر کشف و توسعه به دست بشر هستند. اگرچه این پیش‌بینی را نمی‌توان به‌صورت قطعی اثبات کرد، اما سیر تحولات دهه‌های اخیر ـ از پیدایش اینترنت تا ظهور شبکه‌های اجتماعی و هوش مصنوعی مولد ـ گواه روشنی بر صحت این ادعاست.

در این زمینه، باید حداقل از نظر مفهومی، هم سیستم‌های هوش مصنوعی قابل تصور

۱. بااین‌حال، در اینجا مشکلات فنی بالقوه‌ای مانند فروپاشی مدل (Model collapse) نیز وجود دارد.

و هم انواع غیرقابل تصور آن را مدنظر قرار دهیم. یاد گفتهٔ ادوارد دوبونو[1]، نویسنده و متفکر بزرگ، می‌افتم که تمایزی مشابه را به زبانی ساده توضیح داده است. گفته‌اش را نقل به مضمون در اینجا می‌آورم: از زمانی که انسان توانست به آسمان نگاه کند، دست‌کم توانسته است قدم گذاشتن بر ماه را تصور کند. این امر قابل تصور بود. اما آنچه را تا ربع دوم قرن بیستم اصلاً نمی‌شد تصور کرد، این بود که در روزی که انسان‌ها واقعاً بر ماه قدم می‌گذارند، بیش از ۶۵۰ میلیون نفر این رویداد را از جعبه‌های کوچکی در گوشهٔ اتاق نشیمن خود تماشا خواهند کرد. هوش مصنوعی نیز همین‌گونه است؛ چیزهای بسیاری وجود دارند که درحال‌حاضر می‌توانیم تصورشان کنیم، اما چیزهای بسیار بیشتری وجود دارند که کاملاً خارج از تصورات کنونی ما هستند.

بسیاری از تازه‌واردان به عرصهٔ هوش مصنوعی گمان می‌کنند فناوری‌های اخیر مانند چت‌جی‌پی‌تی، نقطهٔ پایانی پیشرفت‌های این حوزه هستند. اما در واقع، این‌ها تنها نخستین گام‌های لرزان کودکانهٔ ما در این مسیر به شمار می‌آیند. به باور من، اکثر پیش‌بینی‌ها دربارهٔ آینده اشکال اساسی دارند، چراکه فناوری‌هایی را که هنوز اختراع نشده‌اند، نادیده می‌گیرند.

پنجمین و آخرین نکته‌ای که شاید از همهٔ موارد جمع‌بندی‌ام دربارهٔ دوران فناوری دیجیتال نگران‌کننده‌تر باشد، این است که ما حتی درحال‌حاضر نیز درک جامعی از چگونگی کارکرد پیشرفته‌ترین سیستم‌های هوش مصنوعی خود نداریم و به نظر می‌رسد این عدم درک روزبه‌روز افزایش خواهد یافت. این واقعیت زمانی برایم آشکار شد که مقالهٔ «چت‌جی‌پی‌تی چه می‌کند... و چرا کار می‌کند؟»[2] نوشتهٔ استیون ولفرام[3] از برجسته‌ترین دانشمندان علوم کامپیوتر معاصر را مطالعه کردم. نکتهٔ حیرت‌آور این بود که ولفرام بارها اعتراف می‌کند که نمی‌تواند توضیح دهد چت‌جی‌پی‌تی (نسخه GPT-3.5 در آن زمان) چگونه کار می‌کند. ولفرام مکرراً عملکرد چت‌جی‌پی‌تی را «شگفت‌انگیز» و خروجی آن را «غیرمنتظره» توصیف می‌کند. او حتی فراتر می‌رود و می‌گوید درک ما از عملکرد این

1. Edward de Bono
2. What Is ChatGPT Doing...and Why Does It Work?
3. Stephen Wolfram

سیستم‌ها هنوز پیشانظری (Pre-theoretical) است؛ یعنی ما فاقد روش‌های علمی نظام‌مند و کلی برای توضیح نحوهٔ عملکرد چت‌جی‌پی‌تی و سیستم‌های مشابه هستیم. ولفرام حتی به شرایطی اشاره دارد که در آن «کمی جادو و خرافات نیز در کار رخنه کرده است». او از «هنر» آموزش شبکه‌های عصبی می‌گوید و دربارهٔ یکی از تکنیک‌های کلیدی اظهار می‌کند: «فکر نمی‌کنم هیچ علم خاصی پشت این موضوع باشد. فقط روش‌های مختلفی امتحان شده و این یکی ظاهراً جواب می‌دهد». او همچنین می‌گوید: «واقعاً به درک در سطح مهندسی نیاز نیست تا بدانیم که شبکهٔ عصبی چطور خود را پیکربندی کرده است.» در جمع‌بندی، ولفرام می‌گوید ما «درحال کشف این واقعیت هستیم که بله، درون این سیستم‌ها پیچیدگی زیادی وجود دارد و ما آن را درک نمی‌کنیم؛ اگرچه در نهایت چیزی تولید می‌کنند که به زبان انسانی قابل تشخیص است.»

این حقیقت باید به تمام کسانی که فکر می‌کنند هنگام کار با چت‌جی‌پی‌تی و ابزارهای مشابه می‌فهمند چه اتفاقی در جریان است، گوشزد کند که چقدر از آنچه واقعاً می‌گذرد بی‌خبرند. وقتی سیستم‌ها به سطح خاصی از پیچیدگی می‌رسند، عملکرد روزمره‌شان (که متفاوت از طراحی کلی‌شان است) اغلب از توضیح و درک انسانی فراتر می‌رود. ما این را در آستانهٔ هزارهٔ جدید با مشکل Y2K مشاهده کردیم[۱]. افراد غیرمتخصص به‌درستی می‌پرسیدند چرا متخصصان کامپیوتر نمی‌توانند به ما بگویند با رسیدن به ساعت ۲۴ چه اتفاقی خواهد افتاد. حتی در آن زمان نیز درک کامل سیستم‌ها برایمان بسیار پیچیده بود. اکنون، سیستم‌های هوش مصنوعی این ناتوانی در درک را به سطحی کاملاً جدید ارتقا داده‌اند.

برای جمع‌بندی دربارهٔ عصر دیجیتال باید گفت که توانایی سیستم‌ها و ماشین‌های ما روزبه‌روز درحال افزایش است. فناوری‌های زیربنایی با سرعت سرسام‌آور و رشد نمایی درحال تکامل‌اند. این مسیر پیشرفت هیچ افق مشخصی ندارد و ما هنوز بسیاری از فناوری‌های تحول‌آفرین آینده را حتی طراحی نکرده‌ایم. جالب‌تر آنکه امروز نیز سازوکار برخی از سامانه‌های هوش مصنوعی خود را کاملاً نمی‌فهمیم.

---

۱. مسئلهٔ Y2K که به مشکل ۲۰۰۰ هم معروف است، یک چالش فنی جهانی بود که به دلیل روش کدنویسی قدیمی (ذخیرهٔ سال به‌صورت دو رقمی مثل ۹۸ به‌جای ۱۹۹۸) ایجاد شد و نگرانی‌هایی دربارهٔ باگی در سیستم‌های کامپیوتری در آغاز سال ۲۰۰۰ به وجود آورد که باعث می‌شد سال ۲۰۰۰ از سال ۱۹۰۰ غیرقابل تشخیص باشد.ـم.

چهار، پنج یا حتی شش مرحله پیشرفت بشری

یکی دیگر از روش‌های بررسی تأثیر فناوری دیجیتال، نگاهی به شیوه‌های اصلی ثبت، اشتراک و انتقال اطلاعات در جامعه است[۱]. من از انسان‌شناسانی همچون والتر اُنگ[۲] پیروی می‌کنم که معتقدند بشر از دوره‌های گوناگونی گذر کرده است: ابتدا عصر شفاهی که گفتار ابزار اصلی ارتباط بود؛ سپس دورهٔ خط؛ پس‌از آن دوران چاپ؛ و اکنون دنیایی که در آن بیشتر ارتباطات به‌کمک انواع فناوری‌های دیجیتال میسر شده است. با در نظر گرفتن مسیر تحولات، احتمالاً مرحلهٔ پنجم نیز در راه است که چه بسا عصر فراانسان‌گرایی (Transhumanism) (که در فصل‌های ۱۱ و ۱۲ به آن پرداخته شده است) باشد؛ یعنی زمانی که انسان، فناوری، رباتیک، نانوفناوری مولکولی، واقعیت مجازی، زیست‌شناسی مصنوعی و مهندسی ژنتیک در هم ادغام شوند. دست‌کم به‌صورت فرضی، می‌توان مرحلهٔ ششمی را نیز تصور کرد؛ دورانی که در آن سیستم‌های هوش مصنوعی احتمالاً کاملاً جایگزین ما خواهند شد (توضیحات بیشتر در فصل ۱۲ آمده است).

به نظر من، ما درحال‌حاضر به‌آرامی به پایان دورهٔ گذار بین مرحلهٔ سوم و چهارم توسعه، یعنی بین یک جامعهٔ صنعتی مبتنی بر چاپ و یک جامعهٔ دیجیتال مبتنی بر هوش مصنوعی، نزدیک می‌شویم. این اتفاق تغییری جزئی نیست. برخی تأثیر آن را با مقیاس تغییری که با ظهور دستگاه چاپ ایجاد شد، مقایسه می‌کنند؛ فناوری‌ای که گذار از مرحلهٔ دوم به سوم پیشرفت بشری را ممکن کرد. من این قیاس را در این می‌بینم که ماشین‌های چاپ به انقلابی در آموزش، در گسترش ایده‌ها و ایجاد بازار جهانی برای مطالعه منجر شدند. همان‌طور که به یک جامعهٔ دیجیتال تمام‌عیار مبتنی بر هوش مصنوعی که زیربنای آن اینترنت و وب است و از طریق دستگاه‌های همراه فراگیر (و به‌زودی از طریق رابط‌های مغز و کامپیوتر نیز ممکن است که در فصل ۱۱ به آن می‌پردازیم) ارائه می‌شود، می‌رویم، در واقع تغییر عظیم دیگری در نحوهٔ ایجاد، ذخیره، انتشار و مصرف اطلاعات رخ

---

۱. من سال‌هاست که این مفهوم را «زیرساخت اطلاعاتی» جامعه می‌نامم؛ مفهومی که آن را نخستین‌بار در کتابم با عنوان آیندهٔ حقوق (انتشارات آکسفورد، ۱۹۹۶، صفحات ۹۱ تا ۹۶) معرفی کردم.

2. Walter Ong

خواهد داد. اما ماهیت هوش مصنوعی بسیار فراتر از داده و اطلاعات است. سیستم‌های هوش مصنوعی صرفاً دانش را ذخیره و توزیع نمی‌کنند، بلکه دانش را به کار می‌گیرند.

در گذار از جامعهٔ مبتنی بر چاپ به جامعهٔ دیجیتال، تحول دیگری نیز رخ داده است. در حدود هشتادوپنج سال گذشته، دانشمندان علوم کامپیوتر و طراحان سیستم، لایه‌های متعددی بین انسان‌ها و صفر و یک‌های ماشین‌های دیجیتال قرار داده‌اند. در ابتدا، کامپیوترها فقط توسط مهندسانی که با زبان‌های برنامه‌نویسی سطح پایین کد می‌نوشتند، قابل استفاده بودند. با گذشت زمان، زبان‌های برنامه‌نویسی سطح بالاتری توسعه یافتند که استفاده و اجرای آن‌ها آسان‌تر بود. برای کاربران، پیشرفت‌هایی مانند رابط‌های کاربری گرافیکی و تشخیص صدا، افرادی را که هیچ دانش فنی نداشتند قادر ساخت تا از کامپیوترها ـ ازجمله تبلت‌ها و دستگاه‌های همراه ـ استفاده کنند. بخش عمده‌ای از صنعت کامپیوتر به این اختصاص یافته است که فناوری را برای افراد عادی ــ و نه‌فقط برای متخصصان علوم کامپیوتر ــ قابل دسترس و قابل استفاده کند. با پیشرفت‌های هوش مصنوعی و توانایی برقراری ارتباط با ماشین‌هایمان به زبان طبیعی، شکاف بین انسان‌ها و کدهای صفر و یک تقریباً از بین رفته است. با ظهور واقعیت مجازی و رابط‌های مغز و کامپیوتر، این شکاف کاملاً بسته خواهد شد. این نشان‌دهندهٔ یک نقطهٔ عطف اساسی و برگشت‌ناپذیر در داستان تمدن بشری خواهد بود.

برخی سوگیری‌ها

به هنگام اندیشه دربارهٔ فناوری دیجیتال و هوش مصنوعی، دو سوگیری رایج اما به‌شدت گمراه‌کننده وجود دارد که معمولاً دید ما را از آنچه در آینده امکان دارد اتفاق بیفتد، مخدوش می‌کنند.

من اولی را «نزدیک‌بینی فناورانه» می‌نامم؛ یعنی تمایل به اینکه هنگام ارزیابی ظرفیت بلندمدت یک تکنیک یا فناوری جدید، بیش از حد به نسخهٔ فعلی و محدودیت‌های کنونی آن توجه شود. به‌عنوان نمونه، برخی از منتقدان هوش مصنوعی مولد به این نکته اشاره می‌کنند که سیستم‌های کنونی «توهم‌زایی» (Hallucinate) دارند. این وضعیت زمانی رخ می‌دهد که این سیستم‌ها اطلاعات یا دیدگاه‌هایی را ارائه می‌دهند که در ظاهر

موثق به نظر می‌رسند، اما در واقع نادرست یا گمراه‌کننده‌اند. بدون شک این مسئله یکی از محدودیت‌های موجود در ۲۰۲۴ است و درک همگانی از این کاستی‌ها اهمیت دارد. کسانی که هشدارهای بجا می‌دهند، دیدگاه‌های ارزشمندی را مطرح می‌کنند.

بااین‌حال، باید با دقت بسیار به بدبینان افراطی و دانشمندان علوم اجتماعی عبوس نگریست که پیوسته بر ضعف‌ها و ایرادات سیستم‌های گذشته تأکید می‌ورزند. به نظر می‌رسد آن‌ها از روندهای رشد نمایی و دوبرابرشدن‌های مداوم در پیشرفت فناوری غافل مانده‌اند. القای این حس که مشکلات کنونی برای همیشه باقی خواهند ماند، اقدامی نادرست و اغلب غیرمسئولانه است؛ رفتاری که متأسفانه از برخی منتقدان هوش مصنوعی سر می‌زند. اگرچه چنین امکانی منتفی نیست، اما بسیار محتمل است که این صنعت با کوشش فراوان بر «توهم‌زایی» و دیگر مشکلات فعلی فائق آید. در مجموع، نسخه‌های امروزی بدترین نسخه‌هایی هستند که این سیستم‌ها خواهند داشت.

این تصور که سیستم‌های فعلی ـ به‌رغم تمام کاستی‌هایشان ـ در آیندهٔ نزدیک بدون تغییر باقی بمانند، بسیار بعید است؛ خصوصاً با توجه به اینکه فناوری‌های موردنظر، توجه بازار را به خود جلب کرده‌اند. برنامه‌ریزی استراتژیک بلندمدت و سیاست‌گذاری طبیعتاً باید واقع‌بینانه باشند، اما در حوزهٔ هوش مصنوعی، اگر بیش از حد روی سیستم‌های کنونی و محدودیت‌های آن‌ها متمرکز شود، بی‌نتیجه خواهد بود. دومین سوگیری موضوعی است که آن را «انکارگرایی غیرمنطقی» (Irrational rejectionism) می‌نامم. این سوگیری زمانی بروز می‌کند که منتقدان بدون اینکه زحمت دیدن عملکرد سیستم‌های خاص را به خود بدهند، اهمیت، کاربرد یا پتانسیل آن‌ها را متعصبانه رد می‌کنند. به‌عنوان مثال، بارها شنیده‌ام که افراد مغرضانه دربارهٔ چت‌جی‌پی‌تی سخن می‌گویند و با کمال تعجب پس‌از اندکی پرس‌وجو ـ یا به تعبیری، بازجویی ـ متوجه می‌شوم که اصلاً از این سیستم استفاده نکرده‌اند. انکارگرایی غیرمنطقی، اختلالی است که بیشتر در میان رهبران و متخصصان پرمشغله خارج از حوزهٔ هوش مصنوعی مشاهده می‌شود و امری غیرقابل تحمل است. چنانچه آن‌قدر مشغول هستید که فرصت مشاهدهٔ عملکرد یک سیستم را ندارید، ارزیابی‌های شما نه‌تنها فاقد اعتبار، بلکه عمیقاً توهین‌آمیز است.

قابلیت‌های آینده

حال، به پرسش مطرح‌شده در فصل نخست دربارهٔ اینکه سرانجام این تحولات به کجا خواهد انجامید، بازمی‌گردیم. پاسخ کلی من ـ همان‌طور که پیش‌تر نیز گفتم ـ این است که هیچ خط پایان مشخصی در پیش نیست و هیچ وضعیت نهایی قابل تصوری وجود ندارد. این تصور نادرست است که گمان کنیم نسخه‌ای نهایی از هوش مصنوعی ظهور خواهد کرد. باوجوداین، در پسِ این پرسش، نگرانی عمیقی نهفته است که در گفت‌وگوهای رایج خود را نشان می‌دهد: ماشین‌های ما تا چه میزان توانمند خواهند شد؟ آیا حتی ممکن است به‌نوعی زمام امور را به دست بگیرند؟

برای پاسخ به این پرسش‌ها، اکنون می‌خواهم پنج فرضیه دربارهٔ وضعیت احتمالی آیندهٔ هوش مصنوعی را بیان و بعد درباره‌شان تأمل کنم. در اصطلاح‌شناسی استراتژی بلندمدت، این‌ها پنج سناریوی محتمل هستند، همان‌طور که در جدول ۱.۲ خلاصه شده‌اند:

جدول ۲-۱. پنج فرضیهٔ (کنونی) دربارهٔ وضعیت آتی هوش مصنوعی

۱. فرضیهٔ هیاهو
۲. فرضیهٔ GenAI+ (هوش مصنوعی مولد پیشرفته)
۳. فرضیهٔ AGI (هوش عمومی مصنوعی)
۴. فرضیهٔ ابرهوشمندی
۵. فرضیهٔ تکینگی

این فرضیه‌ها را هم از آثار حوزهٔ هوش مصنوعی و هم از گفت‌وگوهای بی‌شماری با متخصصانی که در زمینهٔ بهره‌برداری تجاری از هوش مصنوعی فعالیت می‌کنند، استخراج کرده‌ام. در فصل ۱۲، فرضیهٔ ششم را اضافه می‌کنم که آن را «فرضیهٔ تکامل هوش مصنوعی» می‌نامم، اما قبل از پرداختن به آن، کارهای زیادی باید انجام بدهیم.

«فرضیهٔ هیاهو» (Hype) معتقد است ما به پایان راه هوش مصنوعی رسیده‌ایم. این

فرضیه اذعان می‌کند که پیشرفت‌های چشمگیری صورت گرفته است، اما اصرار دارد که محدودیت‌های ذاتی در هوش مصنوعی وجود دارد و این محدودیت‌ها به موانع دائمی تبدیل می‌شوند. این فرضیه مدعی است که ما شاهد بهبودهای اندک اما نه‌چندان اساسی در هوش مصنوعی فعلی خود خواهیم بود، اما این سیستم‌ها هیچ‌گاه کاملاً قابل اعتماد نخواهند بود و داده‌های زیربنایی آن‌ها به‌طور جبران‌ناپذیری ناقص‌اند و سوگیری دارند. در پسِ این نگاه، شک و بی‌اعتمادی عمیق نسبت به جنجال و هیجانی وجود دارد که تحولات اخیر را در بر گرفته است.

فرضیهٔ بعدی، +GenAI یا «هوش مصنوعی مولد پیشرفته» است. این فرضیه بیان می‌کند که موج کنونی سیستم‌های هوش مصنوعی مولد به‌طور چشمگیری ارتقا و بهبود خواهند یافت، اما بعید است شاهد پیشرفت‌های بزرگ دیگری در هوش مصنوعی باشیم. بنابراین، ما از هیاهو و آزمایش فراتر رفته و به استفادهٔ روزانهٔ گسترده و قابل اعتماد خواهیم رسید، به‌طوری که بیشتر توهمات و سوگیری‌ها برطرف خواهند شد. براساس این دیدگاه، نسخهٔ بهبودیافتهٔ هوش مصنوعی مولد امروزی کاملاً در عملیات اغلب سازمان‌ها ادغام خواهد شد و دستاوردهای قابل توجهی در کارایی و بهره‌وری به ارمغان خواهد آورد. همچنین بسیاری از شهروندان را توانمند خواهد ساخت و ابزارهایی در اختیارشان قرار می‌دهد که در امور اداری، تحقیق، یادگیری و حل مسئله به کارشان می‌آید. این فرضیه به‌سرعت توجه جامعهٔ کسب‌وکار را به خود جلب کرده است.

در ادامه به فرضیهٔ AGI می‌رسیم. در اینجا، AGI به «هوش عمومی مصنوعی» (Artificial General Intelligence) اشاره دارد. تعریفی را که مایکل وولدریج[1]، پژوهشگر باسابقهٔ هوش مصنوعی، ارائه کرده بود در نظر بگیرید که می‌گوید: «AGI تقریباً معادل داشتن کامپیوتری است که طیف کاملی از توانایی‌های فکری یک انسان را داشته باشد؛ این‌ها شامل توانایی مکالمه به زبان طبیعی، حل مسئله، استدلال، درک محیط و غیره در سطحی برابر یا فراتر از سطح یک انسان معمولی خواهد بود.» این مورد، در اصل به معنای عملکردی تمام‌عیار و هم‌سطح انسان توسط ماشین‌هاست. شاید دقیق‌تر باشد که از آن با عنوان «فرضیهٔ شبه AGI» یاد کنیم، زیرا می‌توانیم سیستم‌های هوش

1. Michael Wooldridge

مصنوعی را تصور کنیم که می‌توانند در تقریباً تمام وظایف شناختی با سطح انسان برابری داشته باشند یا از آن فراتر روند، اما در تعداد کمی از آن‌ها همچنان نقص داشته باشند. سیستم‌های AGI و شبه AGI به‌مراتب پیشرفته‌تر از سیستم‌های هوش مصنوعی مولد امروزی خواهند بود که دامنهٔ آن‌ها هنوز محدود است. AGI نشان‌دهندهٔ یک جهش بزرگ در عملکرد نسبت به هوش مصنوعی مولد فعلی است.

فرضیهٔ دیگر این است که ماشین‌های ما در برهه‌ای به سطح «ابرهوشمندی» خواهند رسید؛ اصطلاحی که نیک بوستروم[1] در ۲۰۱۴ آن را رایج کرد. او ابرهوشمندی را این‌گونه تعریف می‌کند: «هر نوع هوشی که عملکرد شناختی انسان‌ها را در تقریباً تمامی زمینه‌های مورد توجه، به‌طور چشمگیری پشت سر بگذارد.» اما میزان این «چشمگیر بودن» چقدر است؟ جیمز لاولاک[2] که ترجیح می‌دهد از اصطلاح «فراهوشمندی» (Hyperintelligence) استفاده کند، معتقد است که سیستم‌های ما «هزاران و سپس میلیون‌ها بار باهوش‌تر از ما خواهند شد». «فرضیهٔ ابرهوشمندی» که من در نظر دارم، پیش‌بینی می‌کند که سیستم‌های هوش مصنوعی در نهایت به‌شکلی غیرقابل تصور از انسان‌ها توانمندتر خواهند شد.

در نهایت، «فرضیهٔ تکینگی» وجود دارد که برجسته‌ترین طرفدار آن ری کورزویل است. او در آخرین دیدگاه خود پیش‌بینی می‌کند که: «ما با هوش مصنوعی یکی خواهیم شد و خودمان را با نوعی قدرت محاسباتی میلیون‌ها برابر بیشتر از آنچه زیست‌شناسی در اختیارمان گذاشته، تقویت خواهیم کرد. این امر هوش و آگاهی ما را چنان گسترش خواهد داد که درک آن دشوار است.» وی اصطلاح «تکینگی» را از حوزه‌های ریاضیات و فیزیک اقتباس می‌کند، اما تأکید دارد که کاربرد آن در اینجا استعاری است؛ چراکه به‌زعم او «این واژه ناتوانی ما را در فهم چنین تحول بنیادینی با سطح کنونی هوشمان نشان می‌دهد». او در ادامه پیش‌بینی می‌کند: «فرایندی از هم‌آفرینی (Co-creation) رخ خواهد داد؛ تکامل ذهن‌هایمان برای دستیابی به بینش عمیق‌تر و بهره‌گیری از آن توانایی‌ها برای تولید ایده‌های نو و متعالی که ذهن‌های آتی ما به کاوش آن‌ها خواهند

1. Nick Bostrom
2. James Lovelock

پرداخت.»

کدام‌یک را باید باور کنیم؟ «فرضیهٔ هیاهو»، «فرضیهٔ هوش مصنوعی مولد پیشرفته»، «فرضیهٔ هوش عمومی مصنوعی»، «فرضیهٔ ابرهوشمندی» یا «فرضیهٔ تکینگی»؟ خودِ جامعهٔ هوش مصنوعی ـ متشکل از دانشمندان علوم کامپیوتر، علوم شناختی، علوم اجتماعی و فلسفه ـ در این مورد اختلاف نظر دارند. به نظر می‌رسد برخی انتظار کند شدن پیشرفت فناوری را دارند، اما برخی دیگر پیش‌بینی می‌کنند که این روند پیوسته و بدون توقف شتاب گیرد. گروهی فرضیه‌های ابرهوشمندی و تکینگی را چیزی بیش از داستان علمی‌ـتخیلی نمی‌دانند؛ اما عده‌ای دیگر جهش‌های بزرگ فناورانه را تقریباً اجتناب‌ناپذیر می‌دانند. کاش می‌توانستم خوانندگانی را که می‌خواهند زندگی، کسب‌وکار یا مشاغل فرزندان خود را برای آینده آماده کنند، مطمئن کنم؛ اما آینده‌ای مقرر و قطعی وجود ندارد. پس چگونه باید به این موضوع نگاه کنیم؟ اول از همه، باید بین پیشرفت فناوری و پذیرش آن تمایز قائل شویم. اگرچه پیشرفت فنون و فناوری‌های توانمندساز (Enabling technologies)[1] درحال‌حاضر به‌صورت یک منحنی نمایی ظاهراً هموار و انفجاری پیش می‌رود، اما سرعت پذیرش و استفاده از آن‌ها همیشه مسیری متفاوت را دنبال می‌کند؛ پذیرش فناوری‌های جدید توسط انسان به‌صورت مجموعه‌ای ناهموار و پر از نوسان، بهتر توصیف می‌شود[2]. به عبارت دیگر، حتی اگر برخی کاربردهای هوش مصنوعی از بُعد فناورانه امکان‌پذیر باشند، ممکن است به دلیل عواملی تجاری، فرهنگی، قانونی، سیاسی یا اخلاقی، الان یا حتی در آینده مورد استقبال قرار نگیرد. درحال‌حاضر، انسان‌ها هنوز این انتخاب را دارند که هر سیستم هوش مصنوعی مشخص را بپذیرند و از آن استفاده کنند یا خیر؛ به این معنا که قدرت نسبی هر فرضیه را نمی‌توان صرفاً براساس مبانی فناورانه قضاوت کرد.

دانشمندان علوم اجتماعی، محققان بازار و سیاست‌گذاران از ما می‌خواهند دربارهٔ آینده‌های احتمالی هوش مصنوعی براساس شواهد بحث کنیم؛ اما این رویکرد، فقط

۱. به فناوری‌هایی گفته می‌شود که به وجود آمدن یا پیشرفت سایر فناوری‌ها را ممکن می‌سازند. این فناوری‌ها به‌تنهایی ممکن است به چشم نیایند، اما نقشی حیاتی در پشتیبانی و تسهیل نوآوری‌های بزرگ‌تر دارند. آن‌ها زیرساخت لازم را فراهم می‌کنند تا چیزهای جدید و پیچیده‌تر ساخته شوند.ـم.

۲. در اینجا، نویسنده قصد دارد بگوید که پیشرفت فناوری خیلی سریع، روان و پیوسته همراه با رشد انفجاری به نظر می‌آید، اما وقتی پای پذیرش و استفادهٔ انسان از فناوری وسط می‌آید، روندها دیگر آن‌قدر هموار نیستند، بلکه مسیری پر فرازونشیب خواهند بود.ـم.

بخشی از داستانِ درحال شکل‌گیری را نشان خواهد داد. بی‌تردید، باید داده‌ها را در مورد سیستم‌های کنونی جمع‌آوری و ارزیابی کنیم؛ بااین‌حال، اگر این تصویر فوری از فناوری امروز و محدودیت‌های آن را در بستر گسترده‌تر مسیر پیشرفت در هوش مصنوعی نبینیم تصویری ناقص از واقعیت در اختیار خواهیم داشت. تمرکز بر یک نقطه از منحنی (مثلاً، انجام تحقیق روی یک سیستم در یک زمان خاص) بینش محدودی به دست می‌دهد و اغلبْ تحلیل‌گران را دچار کوته‌بینی فناورانه می‌کند.

مهم‌تر و واضح‌تر اینکه هیچ شواهدی از آینده وجود ندارد. به این معنا که علاوه‌بر برون‌یابی، باید به گمانه‌زنی نیز متوسل شویم. در همین راستا و با توجه به بسیاری از یافته‌های این فصل، در اینجا برخی از گمانه‌زنی‌های شخصی خود را ارائه می‌دهم: مسیر پیشرفت‌ها، افزایش توانایی سیستم‌های ما، رشد نمایی فناوری‌های زیربنایی، قانون بازده فزاینده، تأثیر احتمالی سیستم‌های هنوزاختراع‌نشده، مقیاس فعلی و پیش‌بینی‌شدهٔ سرمایه‌گذاری، تقاضای آشکار بازار، مقیاس و پویایی جامعهٔ استارتاپ‌های هوش مصنوعی و مشارکت بسیاری از انسان‌های توانمند و بلندپرواز.

این عوامل باعث می‌شوند فرضیهٔ هیاهو را نامحتمل بدانم. این فرضیه آینده‌ای بسیار بعید را ترسیم می‌کند. به اعتقاد من، کسانی که از زمستان قریب‌الوقوع هوش مصنوعی یا ترکیدن کامل حباب هوش مصنوعی سخن می‌گویند، در اشتباهند. گمان می‌کنم اشتباه آن‌ها در درک نکردن اشتهای فعلی بازار است. هر چقدر هم در مورد اینکه هوش مصنوعی به‌زودی منجر به اختلال فراگیر و تحول برگشت‌ناپذیر خواهد شد، اغراق شده باشد، ما همین حالا هم شاهد دستیابی هوش مصنوعی به دستاوردهای قابل توجهی در کارایی و بهره‌وری هستیم. این موضوع صرفاً از طریق کاربرد هوش مصنوعی در خودکارسازیِ (Automation) شیوهٔ عملکرد فعلی کسب‌وکارها و دولت‌ها حاصل شده است (در فصل ۶ به تأثیر هوش مصنوعی فراتر از خودکارسازی خواهم پرداخت). ادعای اینکه هوش مصنوعی درحال‌حاضر به ساده‌سازی و بهینه‌سازی کارهای روزمره، عادی، تکراری، اداری و مبتنی بر فرایند می‌پردازد ــ وعده‌های بسیار بیشتری نیز در این زمینه می‌دهد ــ خیال‌پردازانه یا اغراق‌آمیز نیست. در این امر ارزش اقتصادی بسیار بالایی نهفته است، چنان‌که شرکت‌های بزرگ مشاوره با حرص‌وولع آن را برای هر سازمانی که بودجه‌ای در

اختیار دارد، توضیح می‌دهند. بااین‌حال، فقط قطره‌ای از این دریا را چشیده‌ایم.

در زمستان‌های قبلی، برخلاف وضعیت کنونی، هوش مصنوعی ارزش اقتصادی آن‌چنانی خلق نکرد. سرمایه‌گذاری عمدتاً در تحقیق و توسعه‌های بهره‌برداری‌نشده و در پروژه‌های آزمایشی و تجربی بود. اما دلیل تجاری قانع‌کننده‌ای برای ادامهٔ مسیر وجود نداشت. بازار بازده تجاری پرباری را از هوش مصنوعی کسب نمی‌کرد. من آن زمان از نزدیک شاهد این اتفاق بودم. امروزه، بسیار متفاوت‌تر از آن زمان، پیاده‌سازی هوش مصنوعی به‌وضوح قرار است بر سود و زیان تأثیر بگذارد. این موضوع درحال نفوذ به قلب برنامه‌های بهبود فرایند بسیاری از کسب‌وکارها در بخش‌های مختلف است. استفادهٔ عملی و به‌کارگیری هوش مصنوعی درحال‌حاضر در برنامه‌های بلندمدت خدمات عمومی در سراسر جهان گنجانده می‌شود. در کار روزانه‌ام به‌عنوان مشاور استراتژی، این روند را به‌وضوح می‌بینم، حتی اگر مستقیماً در معیارهای استاندارد کاملاً مشهود نباشد. بازار با تقاضای بیشتری مواجه است که بی‌تردید به نوآوری و پیشرفت بیشتر منجر خواهد شد. ممکن است هوش مصنوعی آهسته‌تر از آنچه بسیاری انتظار دارند توسعه یابد ـ باز هم همان تمایز بین پیشرفت و پذیرش مطرح است ـ اما تصور اینکه دوباره به اعماق زمستانی سرد فرو خواهد رفت، به‌معنای نادیده گرفتن نیروهای بازار است.

با توجه به عوامل ذکر شده، کاملاً انتظار دارم که فرضیهٔ GenAI+ در همین دهه تأیید شود. در این زمینه، پیشرفت فناوری چند سال جلوتر از پذیرش آن خواهد بود. باید کاملاً واضح بگوییم که حتی GenAI+ که از AGI پایین‌تر است، باز هم می‌تواند هوش مصنوعی‌ای ارائه دهد که زندگی انسان را دستخوش تغییر کند، بهبود دهد، نجات بخشد یا تهدید کند.

فرضیه AGI چطور؟ اینجا ماجرا جالب می‌شود. دیدگاه جامعهٔ هوش مصنوعی در این مورد از زمان عرضهٔ چت‌جی‌پی‌تی تغییر کرده است. تا اواخر ۲۰۲۲، وقتی دربارهٔ امکان توسعهٔ سیستم‌های AGI سؤال می‌شد، به نظر می‌رسید بیشتر متخصصان هوش مصنوعی می‌گفتند تحقق آن به‌شرط امکان، بین بیست تا چهل سال طول می‌کشد. ری کورزویل جزو گروه جسورتر بود. او از ۱۹۹۹ پیش‌بینی کرده بود که AGI تا ۲۰۲۹ به وجود خواهد آمد.

از زمـان ظهور سیسـتم‌های هوش مصنوعـی مولد، متوجه شـده‌ام که بازه بیسـت تا چهل سال اغلب به پنج تا ده سال کاهش یافته است. نکتهٔ جالب این است: کسانی که از «فرضیهٔ مقیاس‌پذیری» (Scaling hypothesis)[۱] ـ حمایت می‌کنند، معتقدند برای دستیابی به AGI پیشرفت فناورانهٔ جدیدی لازم نخواهد بود؛ کافی است «محاسبات» و داده‌های بیشـتری به این چالش اختصاص داده شـود. منظور متخصصان فناوری از «محاسبات»، منابع محاسباتی اسـت که به‌طور کلی به معنای «توان پردازش» است. جالب است که متخصصان فناوری فعل‌ها را به‌راحتی به اسم تبدیل می‌کنند![۲]

در مورد داده‌ها ـ سـوخت سیسـتم‌های هوش مصنوعـی امروزی ـ چالش این اسـت که داده‌های بیشـتر و بیشـتری جمع‌آوری شـود؛ چه از منابـع موجود (چندرسـانه‌ای و نه‌فقط متن)، چه از طریق تولید مصنوعی و چه از نو به‌صورت مستقیم جمع‌آوری شده باشـند. دیگر متخصصان بر این باورند که روش‌های فعلی هوش مصنوعی ما را به AGI نمی‌رساند، بلکه ابزارها و روش‌های کاملاً جدیدی مورد نیاز است. برخی نیز معتقدند که AGI، در آیندهٔ نزدیک، چیزی جز داستان نیست.

در مقیاس بزرگ تاریخ بشریت، اینکه AGI در پنج، ده یا پنجاه سال آینده محقق شود، چندان تفاوتی ندارد. با توجه به مسیری که هوش مصنوعی در پیش گرفته و نیز با توجه به تقاضای بالای بازار و نوآوری‌های چشمگیر این صنعت، معتقدم باید برای دستیابی به AGI (به‌طور دقیق‌تر، شـبه AGI) در سـال‌های ۲۰۳۰ تا ۲۰۳۵ آماده باشیم. این همان چیزی است که به مشتریان بخش خصوصی و دولتی خود توصیه می‌کنم. مصرانه ادعا نمی‌کنم که قطعاً تا آن زمان AGI خواهیم داشت، اما فکر می‌کنم این سناریو به‌اندازه‌ای محتمل است که باید به پیامدهای آن برای بشر، جامعه و بازار فکر کنیم. افراد کمی خارج از دنیای هوش مصنوعی هستند که به‌طور نظام‌مند و جدی در مورد AGI فکر می‌کنند. بیشـتر برنامه‌ریزی‌های بلندمدت دولت‌ها و کسب‌وکارها، ناخواسـته براساس فرضیهٔ GenAI+ بنا شـده است. این نگرش ممکن اسـت کوته‌بینانه باشـد و بدتر از آن، ما را در

۱. نظریه‌ای بنیادین و بسیار تأثیرگذار در حوزهٔ یادگیری عمیق و هوش مصنوعی است که مسیر توسعهٔ مدل‌های زبانی بزرگ (مانند GPT) را هدایت کرده است. به بیان ساده، این فرضیه معتقد است که افزایش ساده و مستمر منابع ورودی به مدل‌های هوش مصنوعی در نهایت منجر به افزایش هوش و ظهور توانایی‌های جدید در آن‌ها خواهد شد. ـ م.

۲. همان‌طور که «Compute» در اصل به معنی «محاسبه کردن» است و یک فعل به شمار می‌آید، امروزه در گفت‌وگوها و مستندات فنی در قالب یک اسم به معنای «پردازش» یا «قدرت پردازش» به کار می‌رود. ـ م.

معرض ریسک‌های بزرگ قرار دهد؛ به همین دلیل است که در فصل ۹ خواستار تمرکز بسیار بیشتر بر این سؤالم که «چه می‌شود اگر AGI محقق شود؟»

در گفت‌وگوهای روزمره دربارهٔ هوش مصنوعی، اغلب می‌بینم که افراد غیرمتخصص اما اندیشمند، تصور می‌کنند مقصد نهایی هوش مصنوعی ـ که معمولاً در آینده‌ای دور است ـ ماشین‌هایی خواهند بود که تقریباً از بسیاری جهات به‌اندازهٔ انسان‌های باهوش توانمند هستند. به‌نوعی این‌طور فرض می‌شود که آن‌ها نهایتاً در همان سطح متوقف خواهند شد. به‌ندرت به‌طور جدی تصور می‌شود که سیستم‌های هوش مصنوعی ممکن است به‌شکل غیرقابل محاسبه، غیرقابل درک و غیرقابل وصف، از حتی بهترین انسان‌ها نیز توانمندتر شوند. اینجاست که به سرزمین فرضیهٔ ابرهوشمندی وارد می‌شویم.

از یک منظر، اگر و زمانی که به AGI دست یابیم، آن‌گاه در سراشیبی لغزنده‌ای قرار می‌گیریم که ناگزیر ـ و شاید خیلی سریع ـ به ابرهوشمندی منتهی می‌شود. بهتر است به اواسط دههٔ ۱۹۶۰ بازگردیم؛ زمانی که ایروینگ جان گود[1] در مقاله خود با عنوان «گمانه‌زنی‌هایی دربارهٔ اولین ماشین فوق‌هوشمند» این بحث را به راه انداخت. گود ریاضیدانی بود که با آلن تورینگ در زمینهٔ رمزگشایی کار کرده بود. با مرور و بازنویسی یک بخش برجسته از آن مقاله می‌توان گفت که به‌محض دستیابی به ماشین‌هایی که خودشان بتوانند ماشین‌های دیگری طراحی کنند، این ماشین‌ها «ماشین‌های بهتر و بهتر» طراحی خواهند کرد و «آن‌گاه بدون شک نوعی انفجار هوش رخ خواهد داد... هوش انسان بسیار عقب خواهد ماند... بنابراین اولین ماشین فوق‌هوشمند تنها اختراعی است که انسان به ساخت آن نیاز خواهد داشت». فرضیهٔ ابرهوشمندی، با الهام از گود، انسان‌ها را خالق پیشرفته‌ترین ماشین‌های فردا نمی‌داند. بلکه به زبان ساده، یک چرخهٔ رفت‌وبرگشتی حیرت‌انگیز به وجود می‌آید: سیستم‌ها، سیستم‌های جدید و بهتری پدید می‌آورند و این سیستم‌های جدید نیز خودشان را بهبود می‌دهند و این روند ادامه می‌یابد تا نقطه‌ای که هنوز برایمان نامشخص است.

ترجیح می‌دهم آنچه را در اینجا تصور می‌شود به‌جای «انفجار هوش»، «انفجار قابلیت» بنامم؛ اما درهرصورت، این مفهوم حیرت‌انگیز است. هرچند جهش ناگهانی

1. Irving John Good

به ابرهوشمندی را دشوار می‌دانم، اما منطق حرکت از AGI به ابرهوشمندی قانع‌کننده‌تر است. سؤال بزرگ این است که برای بازگرداندن این روند خیلی دیر شده است یا خیر.

به باور من، شرایط هنوز کاملاً دستخوش تغییر نشده است. بازار بی‌ثبات است و احتمالاً در چند سال آینده، تمرکز بیشتر بر بهره‌برداری تجاری خواهد بود تا پیشرفت‌های فناورانه. ما هنوز تا رسیدن به سیستم‌های هوش مصنوعی مستقل و خودتکثیرشونده که قادر به تعیین و تغییر اهداف خود هستند، فاصلهٔ زیادی داریم. اما نکتهٔ حیاتی این است که باید از همین حالا به پیامدهای AGI بیندیشیم. به همین دلیل است که در فصل ۹ خواستار تفکر پیرامون سناریوی «چه می‌شود اگر AGI محقق شود؟» هستم.

دربارهٔ فرضیهٔ تکینگی، در به چالش کشیدن کورزویل مرددم، چراکه او در گذشته، پیش‌بینی‌های فوق‌العاده دقیقی دربارهٔ پیشرفت فناوری‌های دیجیتال گوناگون داشته است. او بیش از دو دهه است که پیش‌بینی می‌کند تکینگی تا ۲۰۴۵ به وقوع خواهد پیوست. باوجوداین، در این مورد بر تمایز میان پیشرفت و پذیرش فناوری تأکید می‌کنم. تردید اصلی‌ام در مورد ادغام انسان و هوش مصنوعی ـ که هستهٔ اصلی فرضیهٔ تکینگی است ـ نه از بابت امکان‌پذیری فنی آن، بلکه بیشتر به‌جهت مطلوبیت و ملاحظات اخلاقی است. در فصل ۱۲ به این پرسش بازخواهم گشت.

در پایان، به نکته‌ای دربارهٔ اصطلاحات اشاره می‌کنم. همان‌طور که در فصل ۵ توضیح می‌دهم، واژهٔ «هوش»، چه انسانی و چه مصنوعی، برایم مبهم و گیج‌کننده است. از این پس ـ این ابهام را عامدانه حفظ می‌کنم ـ از عبارت «سیستم‌های بسیار توانمند» برای اشاره به سیستم‌هایی بهره می‌گیرم که به‌راحتی و گاهی به‌شکلی چشمگیر عملکرد بهتری از انسان‌ها دارند.

بخش دوم

۲

متفاوت اندیشیدن

فصل سوم

۳

تفکر فرایندگرا و تفکر نتیجه‌گرا

پس‌از عرضهٔ چت‌جی‌پی‌تی، واکنش‌های دو شخصیت برجسته و روشنفکر، هنری کیسینجر[1] و نوام چامسکی[2]، توجهم را جلب کرد. در آن زمان، هر دو از روشنفکران آمریکایی پرتلاش، جنجالی و سنت‌شکن بودند که سنشان از نود گذشته بود. هریک خیل عظیمی از طرفداران پرشور و منتقدان سرسخت داشتند.

دو فرهنگ هوش مصنوعی

در مورد هوش مصنوعی اما به نظر می‌رسد کیسینجر و چامسکی کاملاً با هم اختلاف نظر دارند. کیسینجر در مقاله‌ای که اواخر فوریهٔ ۲۰۲۳ به همراه همکاران برجسته‌اش در *وال استریت ژورنال* دربارهٔ چت‌جی‌پی‌تی نوشت، با شور و هیجان اعلام کرد که هوش مصنوعی مولد «دانش بشری را بازتعریف خواهد کرد، به تغییرات در تاروپود واقعیت ما (جهانِ مادی و اجتماعی ما) شتاب خواهد بخشید و سیاست و جامعه را سازماندهی مجدد خواهد کرد... مسیرهای تحول‌آفرینی را برای عقلانیت انسان و افق‌های جدیدی را برای دانش منسجم (Consolidated knowledge)[3] خواهد گشود.» از نظر کیسینجر و همکارانش، موج جدید هوش مصنوعی، رویدادی به اهمیت اختراع چاپ برای بشریت بود: «یک دستاورد بدیع و شگفت‌انگیز که به‌عنوان اوج افتخار ذهن انسان در کنار هوش مصنوعی قرار می‌گیرد». در عمل، برای کاربری که با فناوری چندان آشنا نیست، این سیستم مانند یک کتابدار و دانشمندِ به‌غایت سریع و فصیح به نظر می‌رسد که هم‌زمان نقش یک استاد را نیز ایفا می‌کند. این فناوری، فرایند خلاصه‌سازی و پرس‌وجوی دانش جهانی را مؤثرتر از هر رابط فناورانه یا انسانی موجود ممکن می‌سازد.

در مقابل، نوام چامسکی با حمایت همکاران برجسته‌اش، موضعی کاملاً متفاوت اتخاذ کرد. او در اوایل مارس ۲۰۲۳ در *نیویورک تایمز* اظهار داشت: «هم مضحک است و هم غم‌انگیز... که این حجم از سرمایه و توجه صرف چیزی چنین ناچیز می‌شود...

---

1. Henry Kissinger

2. Noam Chomsky

۳. عبارت «دانش منسجم/ تلفیق‌شده» به توانایی هوش مصنوعی مولد (مانند چت‌جی‌پی‌تی) در گردآوری، سازماندهی و فشرده‌سازی حجم عظیمی از داده‌ها و اطلاعات پراکنده اشاره دارد. وقتی کیسینجر می‌گوید هوش مصنوعی مولد «افق‌های جدیدی را برای دانش منسجم خواهد گشود»، منظورش این است که این فناوری به ما اجازه می‌دهد: اولاً، تمام اطلاعات بشری را به یک پایگاه دانش واحد، یکپارچه و عظیم تبدیل کنیم. ثانیاً، این دانش فشرده‌شده، بستر جدیدی برای تفکر، نوآوری و پیشرفت ایجاد خواهد کرد، زیرا دیگر نیازی نیست انسان برای یافتن و ارتباط دادن اطلاعات پراکنده در کتاب‌ها و منابع متعدد زمان زیادی صرف کند.ـم.

با توجه به بی‌اخلاقی، علم‌نمایی و ناتوانی زبانی این سیستم‌ها، فقط می‌توان در برابر محبوبیتشان خندید یا گریست.» علاوه‌بر این، چامسکی و همکارانش بیم داشتند که یادگیری ماشینی «با وارد کردن تصوری اساساً معیوب از زبان و دانش در فناوری ما، علم ما را تنزل بدهد و اخلاق ما را فاسد کند.» آن‌ها چت‌جی‌پی‌تی را به «یک موتور آماری ناشی برای تطبیق الگوها» تشبیه کردند. آن‌ها این سیستم را در مقایسه با ذهن انسان «چیزی بسیار بی‌اهمیت» دانستند و توصیه کردند که «پیش‌بینی‌های سیستم‌های یادگیری ماشینی همیشه سطحی و مشکوک خواهند بود».

کدام‌یک درست است: آیا هوش مصنوعی واقعاً انقلابی در ساختار واقعیت ایجاد کرده یا صرفاً ابزاری کم‌اهمیت در مقایسه با توانایی‌های انسانی است؟ این تقابل دیدگاه‌ها چگونه قابل درک است و چرا دو اندیشمند برجسته به چنین ارزیابی‌های متضادی دربارهٔ ماشین‌ها رسیده‌اند؟

پاسخ من این است که کیسینجر و چامسکی، نمایندهٔ دو دیدگاه و فرهنگ بسیار متفاوت در دنیای هوش مصنوعی هستند. من نوع تفکر این دو را تفکر فرایندگرا (Process-thinking) و تفکر نتیجه‌گرا (Outcome-thinking) می‌نامم. این تمایز پیامدهای مهمی برای نحوهٔ تفکر ما دربارهٔ هوش مصنوعی دارد. من از اصطلاح «دو فرهنگ» که سی.پی اسنو[1] مطرح کرد، الهام گرفته‌ام؛ او معتقد بود شکاف عمیق میان هنر (علوم انسانی و ادبیات) و علم (رشته‌های علمی و مهندسی)، جریان فکری غرب را به «دو قطب مجزا» تقسیم کرده است. به نظر من، شکافی مشابه در حوزهٔ هوش مصنوعی نیز وجود دارد که تفکر فرایندگرا و تفکر نتیجه‌گرا را شامل می‌شود.

متفکران فرایندگرا به چگونگی کارکرد سیستم‌های پیچیده علاقه‌مندند؛ متفکران نتیجه‌گرا به نتایجی که آن سیستم‌ها به بار می‌آورند اشتیاق دارند. متفکران فرایندگرا به معماری سیستم‌ها تمایل دارند؛ متفکران نتیجه‌گرا بر عملکرد آن‌ها تمرکز می‌کنند. متفکران فرایندگرا معمولاً تفکر «پایین به بالا» دارند و مجذوب جزئیات عملیاتی می‌شوند؛ متفکران نتیجه‌گرا معمولاً به‌شیوهٔ «بالا به پایین» فکر می‌کنند و دغدغهٔ تأثیر کلی را دارند. طرز فکر هریک بسیار متفاوت است.

1. C. P. Snow

متفاوت اندیشیدن

متفکران فرایندگرا عموماً در سه دستهٔ اصلی جای می‌گیرند. دستهٔ اول عمدتاً به چگونگی عملکرد سیستم‌های هوش مصنوعی توجه دارند. این گروه به موضوعاتی مانند الگوریتم‌ها، داده‌ها، معماری ماشین‌ها، پردازش کامپیوتری، شبکه‌های عصبی، واحدهای پردازش گرافیکی (GPU)، یادگیری ماشین، وزن‌ها، توکن‌ها و سایر جزئیات فنی می‌پردازند. این دسته معمولاً شامل فناوران، برنامه‌نویسان و دانشمندان داده می‌شود. اگرچه ممکن است این مهندسان در زیرگروه‌های متفاوتی فعالیت کنند و تکنیک‌ها یا فناوری‌های خاصی را ترجیح دهند، اما نقطهٔ اشتراک همهٔ آن‌ها درگیری عمیق در طراحی و عملیات واقعی ماشین‌هاست.

دستهٔ دوم از متفکران فرایندگرا در هوش مصنوعی، عمدتاً به شیوهٔ عملکرد ذهن و مغز انسان توجه دارند. مطالعهٔ این حوزه نه‌تنها می‌تواند الهام‌بخش طراحی و بهبود سیستم‌های هوش مصنوعی باشد، بلکه می‌تواند محدودیت‌های ماشین‌هایی را که می‌سازیم نیز آشکار کند. این گروه عموماً شامل پژوهشگران حوزه‌های علوم اعصاب، روان‌شناسی شناختی، علوم شناختی و فلسفه می‌شود.

دستهٔ سوم از متفکران فرایندگرا مجذوب توانایی سیستم‌های هوش مصنوعی در روشن کردن ذهن انسان هستند. برای بسیاری از این افراد، سیستم‌های هوش مصنوعی خود به‌تنهایی موضوع اصلی نیست. نمونهٔ بارز این نگرش را می‌توان در گفته‌های مارگارت بودِن[1] ـ پیشگام حوزهٔ هوش مصنوعیِ بریتانیا ـ مشاهده کرد که در ۱۹۹۲ به مجلهٔ *نیوساینتیست* گفته بود: «اساساً به کامپیوترها کاری ندارم. دغدغهٔ من ذهن انسان است، نه کامپیوترها... شاید امروز خودم را یک روان‌شناس محاسباتی (Computational psychologist)[2] می‌دانم.» این اظهارات نمونهٔ بارزی از دستهٔ سوم اندیشهٔ فرایندگرا است.

در مقابل متفکران فرایندگرا، متفکران نتیجه‌گرا بر خروجی سیستم‌های هوش مصنوعی تمرکز می‌کنند: پیامدهای اجتماعی، تأثیر اقتصادی، کاربردهای عملی،

---

1. Margaret Boden

۲. روان‌شناس محاسباتی کسی است که از قدرت کامپیوتر و مدل‌سازی ریاضی برای درک عمیق‌تر سازوکار ذهن انسان استفاده می‌کند. ـ م.

چالش‌های نظارتی، نتایج کسب‌وکاری، پیامدهای اخلاقی، تحقق سیستم‌های عملیاتی و موارد دیگر. در این دسته می‌توان سیاست‌گذاران، فعالان کسب‌وکار، متخصصان اخلاق، حقوق‌دانان، اقتصاددانان و همچنین انواع کاربران فناوری و هوش مصنوعی را جای داد.

این تمایز بین تفکر فرایندگرا و تفکر نتیجه‌گرا کمک می‌کند شکاف میان چامسکی و کیسینجر را درک کنیم. چامسکی عمدتاً متفکری فرایندگرا است، اما کیسینجر اغلب متفکری نتیجه‌گرا محسوب می‌شود. از دیدگاه چامسکی، فرایندهای زیربنایی چت‌جی‌پی‌تی جالب نیستند. او در زمینهٔ چگونگی کسب توانایی‌های زبانی توسط انسان تخصص دارد و استدلال می‌کند که این سازوکار انسانی به‌مراتب برتر از فرایندهای زیربنایی چت‌جی‌پی‌تی است. چامسکی به‌عنوان متفکر فرایندگرا از دستهٔ دوم، معتقد است ذهن انسان بی‌نهایت جذاب‌تر و توانمندتر از سیستم‌های هوش مصنوعی است. از دیدگاه او، ذهن آدمی به‌مراتب پیچیده‌تر و پیشرفته‌تر از آن است که بتوان با مدل‌های ماشینی آن را بازتولید کرد. چامسکی را نمی‌توان متفکر نتیجه‌گرا دانست. اگر چنین بود، قطعاً دست‌کم از عملکرد فعلی سیستم‌های هوش مصنوعی قدردانی می‌کرد. او می‌توانست به ظرفیت بالای آن‌ها در زمینه‌هایی مانند تشخیص پزشکی، درمان و پیشگیری از بیماری‌ها اشاره کند یا به تهدیدهای جدی آن‌ها برای بشریت (مانند هوش مصنوعی در دستان بازیگران شرور) بپردازد. در نقطهٔ مقابل، کیسینجر بر تأثیرات و پیامدها تمرکز دارد. اگرچه او متخصص عملکرد درونی ماشین‌ها یا ذهن‌ها نیست، اما به‌خوبی می‌تواند ببیند هوش مصنوعی ممکن است چه تأثیری داشته باشد. برای کیسینجر، آنچه این سیستم‌ها انجام می‌دهند اهمیت دارد، نه نحوهٔ انجام آن‌ها.

چامسکی و کیسینجر دو نمونهٔ افراطی از تفکر فرایندگرا و تفکر نتیجه‌گرا محسوب می‌شوند. اکثر افرادی که در زمینهٔ هوش مصنوعی کار می‌کنند، ترکیبی از هر دو نوع تفکر را دارند؛ اما معمولاً وقتی فردی به پتانسیل هوش مصنوعی شک می‌کند، اغلب به یکی از دسته‌های تفکر فرایندگرا روی می‌آورد.

در دههٔ ۱۹۸۰، کتاب *گودل، اشر، باخ*[1] اثر داگلاس هافستادر[2]، برندهٔ جایزهٔ پولیتزر،

1. *Gödel, Escher, Bach*
2. Douglas Hofstadter

به‌عنوان کتاب مقدس متخصصان مشتاق هوش مصنوعی شناخته می‌شد. در آن کتاب و سایر آثار، هافستادر خود را یکی از بزرگ‌ترین متفکران هوش مصنوعی نشان داد؛ به همین دلیل وقتی در اوایل ۲۰۱۸ مشاهدات او را در مورد ترجمهٔ ماشینی زبان‌ها خواندم، شگفت‌زده شدم:

> **به‌هیچ‌وجه تمایل ندارم ماشین‌های بی‌جان جایگزین مترجمان انسانی شوند. در واقع، این ایده مرا می‌ترساند و منزجر می‌کند. به نظر من، ترجمه هنری فوق‌العاده ظریف است که همیشه از تجربهٔ زندگی و تخیل خلاق فرد بهره می‌برد. اگر آن روز به‌اصطلاح خوش بیاید که مترجمان انسانی فقط یادگاری از گذشته باشند، آن‌گاه احترام من به ذهن انسانی درهم خواهد شکست و این شوک مرا در سردرگمی وحشتناک و اندوه عظیم و دائمی فرو خواهد برد.**

اگر مقصود هوفستادر از این سخنان، ترجمهٔ خودکار و بلادرنگ است که به مردم کشورهای مختلف امکان ارتباط آزادانه در امور روزمره را بدهد، باید بگویم که من با انزجار او موافق نیستم. این پیشرفت از نظر اجتماعی و اقتصادی منافع بی‌شماری خواهد داشت. شاید او نگران افول هنر اصیل ترجمه است؛ مثلاً ترجمهٔ آثار فلاسفهٔ یونان باستان به انگلیسی یا تبدیل متون برجستهٔ روسی به فرانسوی. درهرصورت، آنچه جلب توجه می‌کند، احساسات شدید و تمرکز فولادین او بر فرایند به‌جای نتیجهٔ نهایی است. این نوع تفکر فرایندگرا در میان برخی از بزرگان اولیهٔ هوش مصنوعی، به‌ویژه کسانی که نسبت به پتانسیل آن تردید داشتند و دارند، غیرعادی نیست.

ناب‌گرایی متفکران فرایندگرا و عمل‌گراییِ برخی از متفکران نتیجه‌گرا را مقایسه کنید. نمونهٔ بارز این تفاوت را می‌توان در دو سیاستمدار پیشین یافت که زمانی رقیب یکدیگر بودند: تونی بلر[1] و ویلیام هیگ. شاید این دو، ترکیبی غیرمنتظره به نظر برسند، اما در ۲۰۲۳ دست به کار شدند و با همکاری یکدیگر گزارشی تأثیرگذار را با

1. Tony Blair

عنوان «هدف ملی جدید: هوش مصنوعی نویدبخش آینده‌ای پیشرو در جهان برای بریتانیاست» منتشر کردند. آن‌ها هوش مصنوعی را این‌گونه توصیف می‌کنند:

مهم‌ترین فناوری نسل ما محسوب می‌شود... این فناوری فرصت‌های بالقوهٔ بسیار زیادی برای دگرگونی ساختار دولت، ماهیت علم و ارتقای توانمندی شهروندان ارائه می‌دهد... بااین‌حال، ریسک‌های آن نیز عمیق است و اکنون همان زمانی است که باید این فناوری را به‌شکلی سازنده شکل دهیم.

علاوه‌بر این، آن‌ها با اشاره به عرصهٔ بین‌المللی استدلال می‌کنند: «کشورهایی که توانایی بازآفرینی دولت‌های خود حول فناوری را داشته باشند، همان‌هایی خواهند بود که آینده را تعریف می‌کنند».

نمونهٔ قدرتمند دیگر از اندیشهٔ نتیجه‌گرا، بیل گیتس است، کسی که درصورت نیاز به‌راحتی قادر است به تفکر فرایندگرا در سطح جهانی روی آورد. او نسخه‌ای از GPT را «انقلابی» خوانده و این تجربه را «کاملاً خیره‌کننده» توصیف می‌کند. او پیش‌بینی می‌کند که هوش مصنوعی «نحوهٔ کار، یادگیری، سفر، دریافت خدمات درمانی و ارتباطات مردم را تغییر خواهد داد. کلیهٔ صنایع حول آن بازتعریف خواهند شد. کسب‌وکارها به میزان استفادهٔ خوبشان از آن، خود را متمایز می‌کنند». او همچنین توضیح می‌دهد که هوش مصنوعی «می‌تواند برخی از وخیم‌ترین نابرابری‌های جهان را کاهش دهد»، به‌ویژه در حوزه‌های سلامت، آموزش و تغییرات آب‌وهوایی.

باور اینکه هافستادر و گیتس دربارهٔ یک پدیدهٔ واحد صحبت می‌کنند، دشوار است. دیدگاه‌های آن‌ها همانند چامسکی و کسینجر، به دلیل نگرش متفاوتشان به هوش مصنوعی از هم فاصله می‌گیرد؛ یکی با نگاه فرایندگرا و دیگری با رویکرد نتیجه‌گرا.

تأملات و افکار درونی

وقتی روی رسالهٔ دکترایم در زمینهٔ هوش مصنوعی و حقوق کار می‌کردم، خود را به تمام معنا در زمرهٔ متفکران فرایندگرا قرار می‌دادم. تمرکز من عمدتاً بر شیوه‌های حل مسئله

توسط وکلا و قضات بود و اینکه آیا می‌توان شیوه‌های استدلالی آن‌ها را در آن نوع اولیه از سیستم‌های هوش مصنوعی ـ که به سیستم‌های خبره معروف بودند ـ بازآفرینی کرد؟ به‌عنوان متفکری فرایندگرا، مثلاً، به‌شدت درگیر این پرسش بودم که چگونه می‌توان دانش حقوقی را در سیستم‌های کامپیوتری بازنمایی و ذخیره کرد و این سیستم‌ها چگونه می‌توانند به نتایج حقوقی برسند. آیا «منطق گزاره‌ای» برای این کار کافی بود یا به «حساب محمولات مرتبهٔ اول» یا حتی شکلی از «منطق وظیفه‌گرا» نیاز داشتیم؟ واقعاً این‌گونه صحبت می‌کردم! این‌ها سؤالات پیچیده اما مهم دربارهٔ فرایند بودند که سال‌ها ذهنم را مشغول کردند.

در اواخر همان دهه، هنگامی که مدیریت یک پروژهٔ هوش مصنوعی را در یکی از بزرگ‌ترین شرکت‌های بین‌المللی حسابداری و مشاوره برعهده داشتم، نگاهم تغییر کرد و توجهم به نتایج این سیستم‌ها معطوف شد. ما نیازمند ابزاری بودیم که به حسابرسان عمومی در ارزیابی امور مالیات بر ارزش افزودهٔ شرکت‌ها کمک کند. همچنین، به دنبال سیستمی بودیم که بتواند مشخص کند چه زمانی به‌علت سپری شدن مهلت قانونی، دیگر امکان طرح دعوی قانونی وجود ندارد. این‌ها پرسش‌های کاربردی دربارهٔ نتایج بودند که اهمیت فراوانی داشتند. برای کاربران سیستم‌های ما اهمیتی نداشت که سیستم ما مبتنی بر منطق گزاره‌ای بود یا منطق وظیفه‌گرا. حال که به یک متفکر کاملاً نتیجه‌گرا تبدیل شده بودم، مأموریت تیم خود را این‌گونه تعریف کردم: «به‌کارگیری فناوری کامپیوتر برای دسترسی گسترده‌تر و آسان‌تر به تخصص و دانش کمیاب». این تعریف، عمل‌گرایانه بود تا نظری و نسبت به روش‌های مورد استفادهٔ ما بی‌طرف بود.

در فعالیت‌های مشاوره‌ای کنونی‌ام همچنان رویکردی نتیجه‌گرا دارم، اما هنوز شیفتهٔ گسترهٔ وسیعی از مسائل فرایندگرا هستم که اساس هوش مصنوعی را تشکیل می‌دهند؛ به‌ویژه پرسش‌های فلسفی، روان‌شناختی و فناورانه. با این حال، آگاهم که تفکر فرایندگرا و تفکر نتیجه‌گرا دو نگرش کاملاً متفاوت هستند که از فرهنگ‌های بسیار متفاوتی نشئت می‌گیرند. نادیده گرفتن یا درهم آمیختن این دو ممکن است به انواع گوناگونی از سردرگمی‌ها، برداشت‌های نادرست و خطاها منجر شود که به برجسته‌ترین آن‌ها خواهم پرداخت.

فصل چهارم

۱۴

سردرگمی‌ها

وقتی چامسکی و دیگران ظاهراً هوش مصنوعی را ناچیز می‌شمارند، در واقع فرصت‌ها و تهدیدهای بزرگ آن را نادیده می‌گیرند. از سوی دیگر، وقتی کیسینجر و هم‌فکرانش ادعاهای جسورانه و همه‌جانبه مطرح می‌کنند، پیشرفت‌های عظیم فناوری، تحولات فرهنگی و تغییرات قانونی را که برای تحقق پیش‌بینی‌هایشان لازم است، دست‌کم می‌گیرند. اما مشکلات ناشی از این دیدگاه‌های کاملاً متضاد، بسیار فراتر از این است که صرفاً در مجادلهٔ دو اندیشمند برجسته در مقالاتشان خلاصه شود.

اندیشهٔ پریشان پیرامون تهدید هوش مصنوعی

در وهلهٔ اول، متخصصان هوش مصنوعی که ادعاها و پیش‌بینی‌های کاملاً متفاوتی ارائه می‌دهند، باعث سردرگمی گسترده در جامعه شده‌اند. وقتی کسانی که به‌عنوان متخصص شناخته می‌شوند، با آن شور، اطمینان و بینش ظاهراً یکسان، مواضعی کاملاً متضاد اتخاذ می‌کنند، مردم عادی چه برداشتی باید داشته باشند؟

آیا ما محکوم به نابودی هستیم و به‌سوی نوعی ویران‌شهر (Dystopia) پیش می‌رویم که در آن انسان‌ها یا نابود می‌شوند یا در بهترین حالت، تنها با اجازه و میل سیستم‌های هوش مصنوعی فوق‌پیشرفته به حیات خود ادامه می‌دهند؟ یا اینکه این رویداد صرفاً تکرار پیچیده‌تری از همان جنجال‌هایی است که روزگاری حول ماشین‌های ریسندگی و کمباین‌های درو ـ ماشین‌هایی که در نهایت فقط زندگی ما را بهتر کردند ـ به پا شد؟

بحث دربارهٔ تهدیدهای وجودی هوش مصنوعی برای بشر را به فصل ۸ موکول می‌کنم. در این بخش می‌خواهم به ارائهٔ نمونه‌های بیشتری از ابهامات و آشفتگی‌های موجود در بحث‌های عمومی این حوزه بپردازم. پادکست پرطرفدار بریتانیایی «همه‌چیز زیر سر سیاست است» (The Rest is Politics) با اجرای آلستیر کمپل و روری استوارت را در نظر بگیرید. این پادکست با میلیون‌ها دانلود، مخاطبان گسترده‌ای دارد. در اواسط ۲۰۲۳، بسیاری از شنوندگان این برنامه، اپیزودی را شنیدند که در آن پل نِرس[1]، برندهٔ جایزهٔ نوبل، اظهار داشت: «هوش مصنوعی بیش از حد بزرگنمایی شده است.» بعد پیشنهاد داد که

1. Paul Nurse

«بخشی از مشکل به نام آن برمی‌گردد... بهتر است آن را یادگیری ماشین بنامیم تا مردم این‌قدر نترسند.» چند هفته بعد در همین مجموعه پادکست، یووال هراری[1] ـ مورخی با آثار پرفروش ـ موضعی کاملاً متفاوت اتخاذ کرد و پیشرفت‌های اخیر در هوش مصنوعی را «پیشرفتی فوق‌العاده خطرناک» خواند. او در مورد احتمال شکل‌گیری «گونه‌های جدیدی از نظام‌های تمامیت‌خواه» هشدار داد و حتی فراتر رفت و از «پایان تاریخ بشر» سخن گفت: «تاریخ ادامه پیدا خواهد کرد، اما این بار کس دیگری [هوش مصنوعی] پشت فرمان است».

باز هم می‌پرسم: یک فرد غیرمتخصص در برابر این تناقض‌ها چه برداشتی باید داشته باشد؟ بار دیگر تأکید می‌کنم که اینجا شاهد تفاوتِ نگرش و فرهنگِ «اندیشمندان فرایندگرا» و «اندیشمندان نتیجه‌گرا» هستیم. هراری دغدغه‌اش تأثیر هوش مصنوعی بر بشر است. واکنش پل نرس به‌عنوان دانشمندی برجسته، نمونه‌ای بارز از تفکر فرایندگراست. این باور که اگر مردم بفهمند سیستم‌های هوش مصنوعی فقط داده‌ها را جمع‌آوری و پردازش می‌کنند آن‌گاه خطراتش محو خواهد شد، به‌نوعی غفلت از تأثیرات یادگیری ماشین است. مثل این است که بگوییم کامپیوترها نمی‌توانند آسیب برسانند، چون در نهایت فقط با صفر و یک سروکار دارند. حدس می‌زنم که شاید پل نرس با نگاه تخصصی‌اش به زیست‌شناسی انسانی، سیستم‌های یادگیری ماشین را در مقایسه با ساختار زیستی انسان، چیزهایی ساده و ابتدایی ببیند. این نگرش تا حدی شبیه استدلال چامسکی است که مصرانه معتقد است کامپیوترها به‌وضوح در یادگیری زبان از ذهن انسان عقب‌ترند. شاید این ادعا نادرست نباشد، اما چنین تفکر فرایندگرایانه‌ای از نوع دوم می‌تواند مردم را ترغیب کند تا پیامدهای عملی گسترده سیستم‌های هوش مصنوعی فعلی و آینده را نادیده بگیرند.

پل نرس در این دیدگاه تنها نیست. برای مثال، مایکل وولدریج، استاد علوم کامپیوتر دانشگاه آکسفورد، در نطق برجسته و ممتاز خود در سخنرانی تورینگ در ۲۰۲۳، هنگام صحبت دربارهٔ چت‌جی‌پی‌تی و هوش مصنوعی مولد گفت: «شما با یک ذهن در تعامل نیستید. این سیستم دربارهٔ اینکه چه بگوید فکر نمی‌کند... در واقع فقط

1. Yuval Noah Harari

نسخهٔ پیشرفته‌ای از قابلیت تکمیل خودکار متن روی گوشی شماست». من این نوع تفکر فرایندگرا (دستهٔ اول) را بارها شنیده‌ام. به‌گمانم، وولدریج و دیگران با چنین بیاناتی می‌خواهند این اطمینان‌خاطر را بدهند که سیستم‌های هوش مصنوعی چیزی شبیه به ذهن انسان در اختیار ندارند. اما من معتقدم با کم‌اهمیت جلوه دادن فرآیندهای زیربنایی، این واقعیت را پنهان می‌کنیم که همین سیستم‌های خالی از ذهن، قادرند طیف گسترده‌ای از وظایف را در سطح انسانی و حتی فراتر از آن انجام دهند.

رویکرد «ما استثنا هستیم»

یکی دیگر از ابهامات موجود در گفتمان هوش مصنوعی برای من زمانی روشن شد که من و دانیل ساسکیند درحال مصاحبه با طیف گسترده‌ای از متخصصان برای آماده‌سازی کتابمان با عنوان آیندهٔ مشاغل بودیم. بارها و بارها متوجه شدیم که هر گروه از متخصصان، ظرفیت بسیار بیشتری برای هوش مصنوعی در حوزه‌هایی غیر از حوزهٔ تخصصی خودشان می‌بینند. پزشکان به‌راحتی اشاره می‌کنند که هوش مصنوعی در حوزه‌هایی مانند حقوق، حسابداری و معماری پتانسیل بالایی دارد، اما به‌طور غریزی از به‌کارگیری آن در زمینهٔ سلامت مقاومت می‌کنند. وکلا با اطمینان ادعا می‌کنند که حسابرسی، روزنامه‌نگاری و مشاورهٔ مدیریت آمادهٔ جایگزینی با هوش مصنوعی هستند، اما کاملاً معتقدند که کاربرد بسیار محدودی در حرفهٔ وکالت و نظام دادرسی دارد.

حتی در درون مشاغل نیز این نگرش دیده می‌شود. مثلاً در حقوق، وکلای شرکتی منکر هرگونه تغییر احتمالی در کار خود هستند، اما بلافاصله به وکلای ملکی اشاره می‌کنند و معتقدند تحولی اساسی در آن حوزه ضروری بوده است. در پزشکی هم شاهد همین الگو هستیم: پزشکان عمومی معتقدند روش سنتی ارائهٔ خدمات اولیهٔ درمانی تقریباً تنها راه ممکن است اما در همان حال، رادیولوژیست‌ها را صرفاً پردازشگر تصویر می‌دانند که به‌شدت نیازمند تحول هستند.

بسیاری از متخصصان ـ اگر نگوییم اکثر آن‌ها ــ خود را در حرفهٔ خویش هنرمند می‌پندارند. آن‌ها کار خود را مظهر چیزی می‌بینند که ماشین‌ها هرگز قادر به انجامش نخواهند بود. انواع سوگیری‌ها و ناهماهنگی‌ها در اینجا درحال وقوع است، اما واقعیت

انکارنشدنی این است که دنیای مشاغل تخصصی مملو از افرادی است که معتقدند هوش مصنوعی از پتانسیل بالایی برخوردار است، اما «نه برای ما» زیرا «ما استثنا هستیم».

این نگرشِ «ما استثنا هستیم» را تا حدی می‌توان درک کرد، اما عموماً فاقد انسجام و غیرقابل دفاع است. افراد متخصص، سرمایه‌های عظیمی نظیر سال‌ها تحصیل، آموزش و تلاش‌های بی‌وقفه را در تخصص خود هزینه کرده‌اند. موقعیت اجتماعی و منافع اقتصادی آن‌ها پیوندی ناگسستنی با تخصصشان دارد. این گروه چنان در چارچوب روش‌های سنتی محصور شده‌اند که نمی‌توانند جایگزینی قابل قبول برای شیوهٔ کاری خود تصور کنند. آن‌ها حتی قادر به درک پارادایمی متفاوت از الگوی حاکم بر کار خود نیستند. بنابراین، تفکر «ما استثنا هستیم» نشان‌دهندهٔ محافظه‌کاری روان‌شناختی و فرهنگی است؛ اما اغلب با تعهد اخلاقی اصیل نسبت به روش‌های کاری قدیمی نیز همراه می‌شود. شاید ترسی نهفته از «بیکاری فناورانه» (دربارهٔ جایگزینی نیروی کار توسط ماشین‌ها در فصل ۸ صحبت کرده‌ام) نیز در این میان وجود داشته باشد.

چند سال پیش، در مواجهه با گروهی قدرتمند از متخصصان محاسبات بیمه‌ای (Actuaries)، تجربه‌ای عجیب کسب کردم که نمونه‌ای بارز از همین تفکر «ما استثنا هستیم» بود. وقتی دربارهٔ پیشرفت‌های هوش مصنوعی صحبت کردم، قاطعانه گفتند صحبت‌هایم چندان ارتباطی با کار آن‌ها ندارد. در پاسخ گفتم که به‌سختی می‌توانم حرفه‌ای را تصور کنم که بیش از حرفهٔ آن‌ها در معرض تأثیر هوش مصنوعی باشد. بااین‌حال، آن روز بحثم به جایی نرسید. واقعاً شگفت‌انگیز است که چه اندازه انرژی صرف می‌کنیم تا خود را متقاعد کنیم که جایگزین‌ناپذیر هستیم.

تفکر «ما استثنا هستیم» نیز همواره شکلی از تفکر فرایندگراست. نکتهٔ کلیدی اینجاست که مردم تهدید شدن فرایندهای انسانی را که برایشان ارزشمند است دوست ندارند، به‌ویژه فرایندهایی که در آن سرمایه‌گذاری شخصی زیادی کرده‌اند و به آن می‌بالند. اما در عمل مشخص می‌شود که خود فرایند برای دریافت‌کنندگان خدمات تخصصی چندان اهمیتی ندارد. بیماران پزشک نمی‌خواهند، سلامتی می‌خواهند. آن‌ها روان‌درمان‌گر هم نمی‌خواهند، آرامش ذهنی می‌خواهند. مشتریان، معمار نمی‌خواهند؛

ساختمان‌های ایمن، کاربردی و جذاب می‌خواهند. همچنین افراد خواهان حل اختلاف در دادگاه نیستند، اساساً می‌خواهند از مشکلات حقوقی اجتناب کنند. وقتی پرده‌ها را کنار بزنیم و هنجارهای هر حرفه را مورد بررسی قرار دهیم، غالباً مشخص می‌شود که روش فعلی کار متخصصان ـ یعنی فرایندشان ـ ذاتاً چندان اهمیتی ندارد، به این معنا که به‌خودی‌خود ارزشمند نیست. در عوض، اهمیت در نتایج به‌دست‌آمده و ارزش‌ها و منافعی است که حاصل می‌شود. مشتریان دریل برقی نمی‌خواهند، سوراخی در دیوار می‌خواهند.

برای همه‌مان آزاردهنده است اگر بپرسند کدام جنبه‌های کارمان ذاتاً ارزشمند هستند، یعنی ارزش پرداختن و انجام دادن را برای خودِ آن کار دارند؛ زیرا بخش زیادی از زمان ما صرف فعالیت‌هایی می‌شود که صرفاً ارزش ابزاری دارند. به این معنا که اهمیت آن فعالیت‌ها تنها در نتایجی است که به ارمغان می‌آورند و بنابراین ممکن است در برابر هوش مصنوعی آسیب‌پذیر باشند.

تأثیر هوش مصنوعی بر مشاغل تخصصی

با توجه به همهٔ این ملاحظات، طبیعتاً در مورد تأثیر هوش مصنوعی بر کار شخصی خودم بارها فکر کرده‌ام. اندکی پس‌از راه‌اندازی چت‌جی‌پی‌تی، آزمایش کوچکی ترتیب دادم. به‌عنوان نویسنده‌ای در روزنامهٔ *تایمز*، مقالات متعددی برای این نشریه نوشته‌ام. اوایل ۲۰۲۳ از من خواسته شد مقاله‌ای دربارهٔ تأثیر هوش مصنوعی بر حرفهٔ وکالت بنویسم، از چت‌جی‌پی‌تی (نسخه ۳٫۵) خواستم متنی هشتصدکلمه‌ای دربارهٔ همین موضوع تهیه کند، آن هم برای *تایمز* و به سبک ریچارد ساسکیند که با الهام از نوشته‌های او باشد. با نگاهی تحقیرآمیز توأم با دلسوزی به متنی که در چند ثانیه تولید شده بود، آن را چندان هم بد نیافتم. اما با همان نگرش کلاسیک «ما استثنا هستیم»، نتیجه گرفتم که اصلاً قابل مقایسه با کار واقعی نیست. سپس از نسخهٔ تازه‌منتشرشدهٔ GPT-4 خواستم همین کار را انجام دهد. اینجا بود که مو بر تنم سیخ شد. پیشرفت محسوسی بود؛ بی‌شک این متن می‌توانست پیش‌نویس اولیهٔ یکی از مقالات خودم باشد. اولین فکری که به ذهنم خطور کرد این بود: وقتی به GPT-6 یا GPT-7 برسیم،

دیگر نیازی به من به‌عنوان ستون‌نویس روزنامه نخواهد بود. همچنین به این فکر افتادم که پرامپت بهتر می‌توانست این باشد که «تحلیلی هشتصدکلمه‌ای شامل تلفیق نظرات بیست کارشناس برتر جهانی در زمینهٔ هوش مصنوعی و حقوق را به سبک نویسنده‌ای صاحب‌سبک مانند مارگارت اتوود[1] بنویسد». به همه توصیه می‌کنم به این بیندیشند که رشتهٔ کاری یا تخصصی‌شان در عصر GPT-6 یا GPT-7 چه شکلی خواهد بود. (این نوعی از تفکر «چه می‌شود اگر هوش عمومی مصنوعی (AGI) محقق شود؟» محسوب می‌شود).

تد لِویت[2]، نویسندهٔ محبوبم در حوزهٔ مدیریت، تشویقمان می‌کند این باور رایج را به چالش بکشیم که هیچ «جایگزین رقابتی» (Competitive substitute)[3] برای «محصول اصلی» یک صنعت وجود ندارد[4]. همچنین باید این احتمال را بپذیریم که ممکن است به‌راحتی قابلِ جایگزینی باشیم، یعنی آنچه امروز انجام می‌دهیم و خدمات و خروجی‌هایی که ارائه می‌کنیم، ممکن است به شیوه‌ای کاملاً متفاوت عرضه شود. اگر، برای مثال، استارتاپ‌های کارآفرین ابزارها یا روش‌های کاملاً جدیدی برای پاسخگویی به نیازهای مصرف‌کنندگان یا درخواست‌های مشتریان بیابند، آن‌گاه ــ همان‌طور که پیوسته به مراجعان خود می‌گویم ــ بازار و نسلی که دنیای قبل از هوش مصنوعی را به یاد نمی‌آورند، پایبندی‌ای به شیوه‌های قدیمی کار نخواهند داشت.

بااین‌حال و با وجود تمام این بحث‌ها، به نظر من بیشتر ادعاهای کوتاه‌مدتی که این روزها دربارهٔ تأثیر هوش مصنوعی بر مشاغل تخصصی و کارهای اداری مطرح می‌شود، به‌شدت مبالغه‌آمیز است. اما از آن مهم‌تر، معتقدم اغلب ادعاهای بلندمدت نیز تأثیر آن را تا حد زیادی دست‌کم می‌گیرند. هوش مصنوعی در دو سه سال آینده زندگی اقتصادی و اجتماعی ما را زیرورو نخواهد کرد، اما به‌گمانم، تحول واقعی از اواخر دههٔ ۲۰۲۰ به بعد آغاز خواهد شد. در این مدت کارهای زیادی باید انجام بدهیم، اما باید بدانیم این تغییر

---

1. Margaret Atwood

2. Ted Levitt

۳. عبارت «جایگزین رقابتی» به هر محصول، خدمت، فناوری یا روشی اشاره دارد که می‌تواند همان نیاز یا خواسته را برآورده کند، اما به‌شکلی بهتر، ارزان‌تر یا کارآمدتر از محصول یا خدمت موجود.ــم.

۴. لویت استدلال می‌کند که بسیاری از شرکت‌ها یا صنایع شکست می‌خورند، زیرا تمرکز خود را فقط بر ویژگی‌های فیزیکی یا ظاهری محصول اصلی خود محدود می‌کنند، درصورتی‌که باید بر نیاز اصلی مشتری که توسط آن محصول برآورده می‌شود، تمرکز داشته باشند.ــم.

برای متخصصـان و کارمندان تدریجی خواهد بود، نه به‌شـکل یـک دگرگونی ناگهانی و انفجاری.

بـه اعتقـاد مـن، در کوتاه‌مـدت، عمده‌تریـن تأثیر فنـاوری بر مشـاغل، خودکارسـازی فعالیت‌های فعلی خواهد بود که موجب افزایش کارایی و بازدهی می‌شود. در میان‌مدت، هـوش مصنوعی بـا امکان‌دهـی بـه ارائـهٔ خدمـات چندرشـته‌ای (Multidisciplinary service)، مرزبندی‌هـای رایـج بیـن تخصص‌هـای سـنتی را از میـان خواهد بـرد. اما در بلندمدت، تهدید اصلی برای شرکت‌های تخصصی، مشتریانی خواهند بود که به هوش مصنوعی مجهز شده‌اند و سازمان‌هایی که بدون نیاز به مراجعه به کارشناسان بیرونی، طیف وسیعی از امور را خودشان انجام می‌دهند. این یک هشدار جدی به متخصصان است که «رقیبی که شما را از میدان به در می‌کند، شبیه شما نخواهد بود».

مغالطهٔ هوش مصنوعی

وقتـی این اسـتدلال را بـا متخصصان در میـان می‌گـذارم، اغلب پاسـخی شـبیه به این می‌دهند: «یک لحظـه صبر کن. من فکـر می‌کنم، اسـتدلال می‌کنم، نـوآورم، متفکرم، خوش‌برخوردم، اما یک ماشین هیچ‌یک از این‌ها نیسـت، بنابراین... کاری که من انجام می‌دهم از خطر هوش مصنوعی در امان اسـت.» این نوع نگرشِ «ما اسـتثنا هستیم» مرتکب آن چیزی می‌شـود که من و دانیل ساسـکیند آن را «مغالطـهٔ هوش مصنوعی» می‌نامیم؛ فرض نادرسـتی کـه می‌پندارد تنها راه رسـاندن ماشـین‌ها به سـطح عملکرد انسان‌های برتر، تقلید از شیوهٔ کار انسان‌هاست. معمولاً وقتی افراد امکان محول کردن بخشـی از کارشـان به هوش مصنوعی را رد می‌کنند، به چگونگی انجام کار خود ـ یعنی فرایند فعلی ـ استناد می‌کنند و اصرار دارند یک ماشین نمی‌تواند و هرگز نخواهد توانست مانند آن‌ها کار کند. درستی یا نادرستی این ادعا محل بحث است. اما نکته‌ای اساسی‌تر در این میان نادیده گرفته می‌شود: برای دستیابی به نتایج و خروجی‌های مطلوب مشتریان و کاربران، لازم نیست ماشین‌ها از ما تقلید کنند.

اشتباه اساسی این است که درک نکنیم سیستم‌های هوش مصنوعی اساساً به تقلید یا بازسازی استدلال انسـانی نمی‌پردازند. فکر کردن خلاف این موضوع، نگاهی بیش از

حد انسان‌محور نسبت به هوش مصنوعی است. ما در دام این نگرش گرفتاریم؛ همان «بندهای ذهنی» که ویلیام بلیک از آن‌ها سخن گفته است[۱].

خودروهای خودران را در نظر بگیرید. گمان نمی‌کنم کسی عملاً باور داشته باشد که بهترین راه توسعهٔ این فناوری، ساخت ربات‌هایی است که پشت فرمان خودروهای معمولی بنشینند و شیوهٔ رانندگی انسان را تقلید کنند. بااین‌حال، این دقیقاً مشابه آن است که تصور کنیم ربات‌ها در بیمارستان، دادگاه یا محیط اداری به‌عنوان جایگزین متخصص انسانی حضور می‌یابند. در عوض، سیستم‌های هوشمند ما در آینده با بهره‌گیری از قابلیت‌های منحصربه‌فرد خود و نه تقلید از ما، به ایفای نقش خواهند پرداخت. شواهد این ادعا را می‌توان در پیروزی «دیپ‌بلو» بر گری کاسپاروف در ۱۹۹۷ و شکست لی سدول از «آلفاگو» در ۲۰۱۶ مشاهده کرد. این انسان‌ها مغلوب الگوریتم‌های پیشرفته‌ای شدند که با پردازش حجم بی‌سابقه‌ای از داده‌ها و به‌کارگیری قدرت پردازشی بسیار بالا رفتار می‌کردند، نه ماشین‌هایی که صرفاً فرایندهای کاری این استادان را کپی کرده باشند.

مصداق دیگری از مغالطهٔ هوش مصنوعی در مباحث مربوط به «شفافیت» سیستم‌های هوش مصنوعی مشاهده می‌شود. برای درک این موضوع، لازم است یک پیش‌زمینه بگوییم. سیستم‌های خبره یا نسل اول هوش مصنوعی، توانایی تشریح مسیر استدلال خود را داشتند. آن‌ها این کار را با نشان دادن مسیرهای پیموده‌شده در ساختار درخت‌گونهٔ تصمیم‌گیری‌های خود، شفاف می‌کردند. در مقابل، سیستم‌های نسل دوم مبتنی بر یادگیری ماشین نمی‌توانند خروجی خود را به شیوه‌ای مشابه توضیح دهند. در این سیستم‌ها، هیچ مسیر استدلالی ساده و مشخصی وجود ندارد. عملکرد آن‌ها به‌مراتب پیچیده‌تر و غیرشفاف‌تر است؛ همان‌گونه که استیون ولفرام اشاره می‌کند، این فرایندها اغلب خارج از درک ماست. باید تأکید کرد که سیستم‌های هوش مصنوعی مولد، مانند انسان‌ها استدلال نمی‌کنند.

تحلیلگران و کارشناسان بی‌شماری این وضعیت را ناپذیرفتنی می‌دانند. وقتی در

---

۱. ویلیام بلیک شاعر و هنرمند انگلیسی است که در یکی از شعرهایش با عنوان «لندن» از عبارتِ «بندهای ذهنی» (mind-forg'd manacles) استفاده می‌کند که استعاره‌ای است برای محدودیت‌هایی که ذهن انسان برای خود می‌سازد.ـم.

موقعیت‌های حیاتی به سیستم‌های هوش مصنوعی اتکا می‌کنیم، باید بتوانیم نگاهی به درون آن‌ها بیندازیم تا بفهمیم چه می‌گذرد. در نتیجه، این استدلال مطرح می‌شود که اگر انسان‌ها می‌توانند اعمال خود را توضیح دهند، ماشین‌ها نیز باید چنین قابلیتی داشته باشند. اما این انتظار برای تقلید از عملکرد انسانی، خود نمونه‌ای از همان مغالطهٔ هوش مصنوعی است، چراکه از هوش مصنوعی می‌خواهد مانند انسان عمل کند.

در این زمینه مسائل پیچیده و مورد مناقشهٔ بسیاری وجود دارد. دو نکتهٔ اساسی در این بحث حائز اهمیت است: نخست آنکه در زندگی روزمره برای بسیاری از سیستم‌های فراگیر و حیاتی مورد استفاده‌مان ـ مانند ساعت یا ماشین‌حساب ـ تقاضای توضیح نمی‌کنیم. اعتماد ما ناشی از اطمینان به توسعهٔ مسئولانه و آزمایش‌های جامع این سیستم‌هاست (موضوعی که در فصل ۹ تفصیل بیشتری خواهد یافت) و نیز می‌دانیم که روش‌های توسعه و آزمون آن‌ها در معرض بررسی و ارزیابی قرار دارند. دوم آنکه نباید از سیستم‌هایمان بیش از قابلیت‌های انسانی انتظار داشته باشیم. مثال گویای این موضوع، عملکرد قضات و پزشکان است. اگرچه هنگام تصمیم‌گیری‌های این متخصصان انتظار دریافت توضیح داریم، اما ماهیت این توضیحات اغلب مبهم است. حتی باهوش‌ترین قاضی‌ها یا پزشکان هم قادر نیستند فرایندهای روان‌شناسی عصبی یا فیزیولوژی عصبی مؤثر در تصمیمات خود را تشریح کنند. حتی عمیق‌ترین تحلیل‌های درونی افراد نیز به چنین سطحی از تبیین دست نمی‌یابد. در عوض، این افراد توضیح‌هایی ارائه می‌دهند که بهتر است آن‌ها را توجیهات عقلانی (Rationalizations) بدانیم که نمایشی منسجم از مقدمات و منطقی است که با نتیجهٔ نهایی‌شان سازگار است؛ اما این توجیهات لزوماً حتی بازتاب‌دهندهٔ فرایندهای فکری آگاهانه خود آن‌ها هم نیستند. قضات ممکن است به‌جای دلایل واقعی، توجیهات ظاهری برای تصمیمات خود ارائه دهند و پزشکان نیز به‌جای توضیح دقیق علمی و پایه‌ای، به بیان ارتباط کلی بین علائم و نام بیماری‌ها بسنده می‌کنند. بااین‌حال، همین سطح از توضیح معمولاً را ما می‌توانیم بپذیریم.

شاید در این صورت، راه‌حل نه در تقلید از شیوه‌های تبیین انسانی، بلکه اطمینان از روش‌های توسعه و آزمون سیستم‌ها باشد و در کنار آن نیز توجیهات عقلانی از سوی

هوش مصنوعی تولید شود که با عملکرد سیستم‌ها سازگار است، اما لزوماً بازتاب مستقیم چگونگی انجام کار توسط آن‌ها نیست.

آنچه ماشین‌ها نمی‌توانند انجام دهند

خب، حالا نوبت مطرح کردن این پرسش معترضانه است: آیا واقعاً کارهایی وجود دارد که ماشین‌ها قادر به انجامشان نباشند؟ مدت‌هاست انسان‌ها با انتخاب وظایفی که می‌توانند انجام دهند، اما ظاهراً ماشین‌ها قادر به انجامشان نیستند، تلاش می‌کنند محدودیت‌های مسلم هوش مصنوعی را شناسایی کنند. البته انگیزه‌ها در اینجا متفاوت است. ما باید میان کسانی که به دلایل اخلاقی معتقدند ماشین‌ها نباید در برخی از زمینه‌ها یا کارها به کار گرفته شوند و کسانی که از منظر فنی ادعا می‌کنند که سیستم‌ها اساساً قادر نیستند وظیفه‌ای خاص را انجام دهند، تمایز قائل شویم؛ البته، باید مراقب کسانی باشیم که صرفاً به دام اندیشهٔ «ما استثنا هستیم» می‌افتند.

معمولاً کسانی که ادعای عدم امکان فنی دارند، معتقدند ماشین‌ها نمی‌توانند خلاق، خودآگاه، همدل، کنجکاو، شجاع یا صمیمی باشند؛ قادر به اعمال قضاوت یا تشخیص بین درست و غلط نیستند؛ نمی‌توانند بخشنده باشند یا عشق بورزند؛ همچنین، هرگز به نبوغی مانند اینشتین دست نخواهند یافت. افراد بدبین به هوش مصنوعی با اشتیاقی خاص به گسترش این فهرست از قابلیت‌ها ادامه می‌دهند که به‌زعم آن‌ها برای ماشین‌ها اساساً دست‌نیافتنی باقی خواهند ماند.

به یک معنا، تمام این ادعاها صحیح‌اند، اما از منظری بدیهی و پیش‌پاافتاده. چراکه این ویژگی‌ها را می‌توان همان خصوصیات ذاتی و متمایزکنندهٔ انسان دانست که تقریباً ما را از گیاهان، حیوانات و ماشین‌ها متمایز می‌کند. این قابلیت‌ها بخشی از مجموعه توانایی‌های انسان، یعنی فرایندهای انسانی هستند. آن‌ها در عمیق‌ترین لایه‌های زندگی روزمره ما جای گرفته‌اند.

در واقع، تأکید صرف بر اینکه بگوییم هوش مصنوعی مثلاً فاقد خلاقیت انسانی است، کمکی به پیشبرد درک ما از این فناوری نمی‌کند. بی‌تردید، این گزاره تفاوت ماهوی عملکرد انسان و ماشین را نشان می‌دهد که یکی مبتنی بر خلاقیت ذاتی است و دیگری

(حداقل درحال‌حاضر) بر پایهٔ هوش مصنوعی مولد عمل می‌کند. اما این ادعا، به پرسش عملی دربارهٔ نتایج نسبی که هریک می‌توانند ارائه دهند پاسخ نمی‌دهد. هرچند یک سیستم هوش مصنوعی ممکن است نتواند متنی را با آن اصالتی که یک نویسندهٔ انسانی خلق می‌کند ارائه دهد، اما قادر است متنی را در قالبی تولید کند که پیش از آن هرگز وجود نداشته است.

بسیاری از این مناقشات دربارهٔ آنچه سیستم‌های هوش مصنوعی نمی‌توانند انجام دهند، برخاسته از ناتوانی در درک تمایز میان «تفکر فرایندگرا» و «تفکر نتیجه‌گرا» یا گاهی از آمیختن نادرست این دو رویکرد است. از منظر تفکر فرایندگرا، وقتی می‌گوییم ماشین نمی‌تواند فلان کار را انجام دهد، در واقع به شیوه‌های عملکردی اشاره می‌کنیم که ماشین قادر نیست آن‌ها را تکرار کند. این دیدگاه بر نحوهٔ عملکرد انسان‌ها تأکید دارد و ادعا می‌کند سیستم‌های هوش مصنوعی نمی‌توانند این فرایندها را کپی کنند. اگرچه این ادعا در نگاه اول بدیهی می‌نماید، زیرا صرفاً تأیید می‌کند که به‌رغم کاربرد گستردهٔ اصطلاحات انسان‌واره‌ای مانند «شبکه‌های عصبی»، سیستم‌های هوش مصنوعی انسان‌های بیولوژیکی نیستند. هرچند موج کنونی هوش مصنوعی تا حدودی از ساختار عصبی انسان الهام گرفته، اما این اصطلاحات (حداقل درحال‌حاضر) عمدتاً جنبهٔ استعاری دارند.

به‌این‌ترتیب، تاریخچهٔ پرباری از بدبینی نسبت به هوش مصنوعی این‌گونه رنگ می‌بازد. نمونهٔ بارز آن استاد فلسفه‌ای به نام هوبرت دریفوس[1] است که از پیشگامان نقد هوش مصنوعی محسوب می‌شود. او در ۱۹۷۲ کتاب *آنچه کامپیوترها نمی‌توانند انجام دهند (What Computers Can't Do)* را منتشر کرد و دو دهه بعد در ۱۹۹۲، با کتاب دیگری با عنوان *آنچه کامپیوترها همچنان نمی‌توانند انجام دهند (What Computers Still Can't Do)* به این بحث ادامه داد. عنوان کتاب دوم، آن‌قدر ساده و قابل پیش‌بینی بود که حتی یک هوش مصنوعی قدیمی و ابتدایی هم می‌توانست تولیدش کند. نقد دریفوس بر «عقلانیت مصنوعی» (Artificial reason) نمونه‌ای کلاسیک از تفکر فرایندگرای نوع دوم است که در دغدغهٔ دائمی او نسبت به

1. Hubert Dreyfuss

تفاوت‌های بنیادین عملکرد ماشین‌ها و انسان‌ها ریشه دارد. نکتهٔ قابل تأمل اینجاست که با پیشرفت فزایندهٔ قابلیت‌های ماشین‌ها، اندیشمندان سنتی هوش مصنوعی مانند دریفوس و هافستادتر حسرتی عمیق از خود بروز داده‌اند که نشان می‌دهد دغدغه‌شان صرفاً به جنبه‌های علمی محدود نبوده، بلکه نگرانی‌هایی دربارهٔ مطلوبیت این فناوری نیز در کار بوده است.

هنگامی که می‌پرسیم ماشین‌ها چه کارهایی نمی‌توانند انجام دهند، معتقدم باید تمرکز خود را بر نتایج معطوف کنیم. در فصل ۵ با همین رویکرد به بررسی امکان قضاوت، همدلی و خلاقیت مبتنی بر هوش مصنوعی می‌پردازم. اما پاسخ به این پرسش را که «آیا سیستم‌های هوش مصنوعی خواهند توانست به بینش و اصالتی در سطح اینشتین دست یابند» به زمانی دیگر و شاید کتابی دیگر موکول می‌کنم. بااین‌حال، با تمام وجود اطمینان دارم ماشین‌ها بالاخره چنین خواهند کرد[۱].

۱. برای یک بحث بسیار خوب و مرتبط دربارهٔ چیزی که نویسندگان آن را توانایی شناختی منحصربه‌فرد می‌دانند ـ یعنی قابلیت «چارچوب‌بندی» ـ به کتاب *Framers: Human Advantage in an Age of Technology and Turmoil* اثر کنت کوکیر و همکاران مراجعه کنید.

فصل پنجم

۵

واژه‌های مناسب در اختیار نداریم

در فصل‌های ۳ و ۴، سردرگمی‌هایی در درک عمومی از هوش مصنوعی شناسایی شد. یکی از این موارد، عدم تمایز بین تفکر فرایندگرا و تفکر نتیجه‌گراست. دیگری، گرایش به دیدگاه «ما استثنا هستیم» است که اغلب به مغالطهٔ هوش مصنوعی منجر می‌شود؛ یعنی گمان کنیم برای دستیابی به عملکرد سطح بالا، سیستم‌های هوش مصنوعی لزوماً باید به شیوه‌ای مشابه انسان‌ها عمل کنند. علاوه‌بر این موارد، انتظار یا تمایل به اینکه سیستم‌های هوش مصنوعی توانایی تبیین عملکرد خود را کم‌وبیش به همان روشی که انسان‌ها توضیح می‌دهند، داشته باشند نیز از دیگر سوءتفاهم‌های رایج در این حوزه به شمار می‌رود.

ریشهٔ تمام این سردرگمی‌ها و ابهامات دیگری که به آن‌ها می‌پردازم، یک مشکل اساسی‌تر است: ما فاقد واژگان و مفاهیم مناسب برای درک و بررسی نحوهٔ عملکرد سیستم‌های روبه‌رشد خود هستیم. به‌جای آن، گفتمانمان را در چارچوبی از زبان مرتبطِ با انسان‌ها بنا می‌کنیم. با این کار، ناخواسته انسان‌وارگی می‌آفرینیم، به استعاره متوسل می‌شویم و ایده‌های نو را در واژه‌های کهنه می‌گنجانیم. این چالش به‌ویژه در بحث‌های مربوط به ناتوانی‌های کامپیوترها به‌وضوح نمایان می‌شود. برای مثال، همان‌طور که در فصل ۴ اشاره کردم، اغلب ادعا می‌شود که ماشین‌ها قادر به قضاوت، همدلی یا خلاقیت نیستند. بااین‌حال، تأکید بر اینکه ماشین‌ها مانند انسان‌ها عمل نمی‌کنند می‌تواند بینشی دربارهٔ فرایندهای متفاوتی که زیربنای عملکرد هریک هستند ارائه کند، اما اطلاعات چندانی دربارهٔ قابلیت‌های واقعی‌شان در اختیارمان نمی‌گذارد. هنگامی که ماشین‌ها خروجی‌هایی تولید می‌کنند که پیش‌تر منحصراً محصول انسان‌ها تصور می‌شد، به‌سختی می‌توان از برچسب‌های انسانی برای توصیف عملکردشان چشم پوشید. می‌توانیم آن را استعاری بنامیم، اما تأکید می‌کنم که توصیف ماشین‌ها با ویژگی‌های انسانی، به سردرگمی می‌انجامد.

مدتی بسیاری از شرکت‌های فناوری و مشاوران مدیریت، برای صحبت در مورد هوش مصنوعی از اصطلاح «رایانش شناختی» (cognitive computing) استفاده می‌کردند. شاید این عبارت پیشرو و مدرن به نظر می‌رسید، اما کاملاً نادرست بود. سیستم‌های تحت این عنوان، حتی به‌اندازهٔ یک کدوی پخته هم حالت شناختی نداشتند. این

اصطلاح، گمراه‌کننده و به‌ویژه هیاهوی تبلیغاتی بود، زیرا «رایانش شناختی» به قابلیت‌هایی اشاره داشت که سیستم‌های هوش مصنوعی فاقد آن بودند.

مشابه همین مسئله در مورد اصطلاح «توهمات» (hallucinations) نیز صادق است. این واژه به‌صورت گسترده برای توصیف خطاها و اطلاعات نادرستی که توسط سیستم‌های هوش مصنوعی مولد تولید می‌شوند، مورد استفاده قرار می‌گیرد. در بهترین حالت، «توهم» یک استعاره است و در بدترین حالت، به حالت‌های شناختی اشاره می‌کند که در هوش مصنوعی وجود ندارند. توهمات در اصل، ادراک‌های اشتباه از تجربیات حسی محسوب می‌شوند و این چیزی نیست که وقتی چت‌جی‌پی‌تی خروجی‌های بی‌معنی تولید می‌کند، اتفاق می‌افتد.

حتی مفهوم اصلی و بنیادین «هوش» در حوزهٔ هوش مصنوعی، بدون تعریف دقیق و مشخصی استفاده می‌شود (هر کسی آن را به‌شکلی متفاوت و دلخواه به کار می‌برد). پیش از ظهور هوش مصنوعی، این اصطلاح در رشته‌های مختلفی مانند فلسفه، روان‌شناسی و علوم شناختی، معانی متفاوتی داشت. حتی درون همین رشته‌ها نیز ماهیت هوش به‌شدت مورد بحث و مناقشه بود؛ برای مثال، در روان‌شناسی، بحث‌های داغی در مورد معنای «باهوش‌تر بودن» یک فرد نسبت به دیگری وجود داشت و مردم نسبت به مفهوم «ضریب هوشی» یا IQ حساسیت زیادی نشان می‌دادند.

در حوزهٔ هوش مصنوعی، همان‌طور که استوارت راسل[1] در سخنرانی‌های ریث[2] دربارهٔ هوش مصنوعی اشاره می‌کند: «از همان ابتدا... هوش در ماشین‌ها به یک شکل تعریف شده است: ماشین‌ها به‌اندازه‌ای هوشمند هستند که بتوان انتظار داشت اقداماتشان به تحقق اهدافشان بینجامد.» این تعریف به ظرفیتی از ماشین‌ها اشاره می‌کند که مشابه انسان‌هاست. در همین زمینهٔ فرایندگرایی، شرکت پیشگام هوش مصنوعی دیپ‌مایند در مقطعی تأکید داشت مأموریتشان «حل مسئلهٔ هوش» است. این عبارت مرا سردرگم می‌کرد، زیرا به‌خودی‌خود توضیح چندانی دربارهٔ اینکه سیستم‌هایشان دقیقاً چه‌کاری

1. Stuart Russell

۲. سخنرانی‌های ریث (The Reith Lectures) یک مجموعه سخنرانی رادیویی سالانه و بسیار معتبر است که شبکهٔ بی‌بی‌سی در بریتانیا برگزار و پخش می‌کند. این مجموعه به افتخار جان ریث (John Reith)، اولین مدیر کل بی‌بی‌سی (که نقش تاریخی در شکل‌گیری مفهوم «پخش عمومی» داشت)، نام‌گذاری شده است. ـ م.

انجام خواهند داد، ارائه نمی‌کرد.

در مقابل این کاربرد از «هوش»، رویکرد نتیجه‌گرا قرار می‌گیرد که تا پیش از پیدایش چت‌جی‌پی‌تی رواج داشت. در آن زمان، عموماً این دیدگاه وجود داشت که هوش مصنوعی در آیندهٔ قابل پیش‌بینی همچنان «محدود» (narrow) باقی خواهد ماند. به این معنا که هوش مصنوعی فقط در حوزه‌های مشخصی به ارائهٔ عملکردی در سطح انسانی قادر است. براساس این استدلال، دستیابی هوش مصنوعی به هوش عمومی مصنوعی (AGI) ـ اگر اصلاً چنین اتفاقی بیفتد ـ سال‌های زیادی زمان می‌برد. وقتی صحبت از هوش مصنوعی محدود می‌شود، کاملاً روشن است که منظور نتایج است. برای مثال، در حوزهٔ حقوق، هوش مصنوعی محدود به سیستمی اشاره دارد که می‌تواند فقط در یک شاخهٔ خاص ــ اغلب نسبتاً محدود و دارای مرزهای مشخص ـ از نظام حقوقی فعالیت کند، نه ماشینی که قادر به حل هرگونه مسئلهٔ حقوقی باشد.

به‌این‌ترتیب، حتی مفهوم پایه‌ای «هوش» نیز در حوزهٔ هوش مصنوعی به‌شکلی گیج‌کننده به کار می‌رود؛ گاهی به چگونگی عملکرد انسان‌ها یا سیستم‌ها اشاره دارد و گاهی به آنچه می‌توانند انجام دهند. اینجا نیز، مانند دیگر مباحث هوش مصنوعی، به واژگان غنی‌تری نیاز داریم. ما هنوز واژه‌های لازم را در اختیار نداریم.

زبان

با گذر جوامع بشری از مقطعی از توسعهٔ انسانی و اجتماعی به مرحله‌ای دیگر، واژه‌ها و مفاهیم نیز اغلب معانی جدیدی به خود می‌گیرند یا کاملاً ابداع می‌شوند تا ایده‌های نوظهوری را که واژگان پیشین برای بیانشان ناکافی هستند، دربرگیرند و منعکس کنند. اریک هابزبام[1] این نکته را به‌شکلی ماندگار در صفحهٔ آغازین کتاب *عصر انقلاب: ۱۸۴۸_۱۷۸۹* به تصویر می‌کشد. او از ما می‌خواهد به چند واژهٔ انگلیسی توجه کنیم که در طول آن دورهٔ شصت‌ساله (مورد مطالعه‌اش) ابداع شدند یا معنای مدرن خود را یافتند:

1. Eric Hobsbawm

این واژه‌ها شامل کلماتی مانند «صنعت»، «صنعت‌گر»، «کارخانه»، «طبقهٔ متوسط»، «طبقهٔ کارگر»، «سرمایه‌داری» و «سوسیالیسم» هستند. همچنین، اصطلاحات سیاسی همچون «اشرافیت» و «راه‌آهن»، «لیبرال» و «محافظه‌کار» و مفاهیمی نظیر «ملیت»، «دانشمند» و «مهندس» به همراه واژه‌هایی چون «آمار»، «جامعه‌شناسی» و چند نام دیگر از علوم مدرن و نیز واژگانی نظیر «روزنامه‌نگاری» و «ایدئولوژی» همگی زاییده یا جرح‌وتعدیل‌شده در این دورهٔ تاریخی هستند.

شگفت‌انگیز نیست که این اصطلاحات پیش از انقلاب صنعتی وجود نداشتند یا معنی متفاوتی داشتند؟ چه واژه‌های جدید یا بازتعریف‌شده‌ای در آینده پدیدار خواهند شد تا مفاهیم و مسائل نوظهوری را که در عصر هوش مصنوعی زاده شده‌اند، توصیف کنند؟

در ادامه به این پرسش تا حدودی پاسخ می‌دهم و طی آن به بررسی سه مفهومی که برایشان به واژه‌های جدید نیاز مبرم داریم، می‌پردازم. این مفاهیم به قضاوت، همدلی و خلاقیت مربوط می‌شوند؛ پدیده‌هایی که شاید توسط سیستم‌های هوش مصنوعی بروز داده نمی‌شوند. در هر مورد، پیشنهاد می‌کنم از اصطلاحاتی که برای انسان‌ها به کار می‌بریم، استفاده نکنیم. در عوض، پیشوند «شبهِ» (quasi-) را اضافه می‌کنم که در لاتین به معنای «شبیه به» یا پسوند «ـوار» است[1]. در اینجا هیچ رتبه‌بندی‌ای انجام نمی‌دهم. منظورم این نیست که سیستم‌های دارای شبهِقضاوت، شبهِهمدلی و شبهِخلاقیت عملکرد ضعیف‌تری نسبت به قضاوت، همدلی و خلاقیت انسانی دارند. صرفاً در پی روشی برای تمایز بین فرایندها و نتایج انسانی و ماشینی هستم که به‌ظاهر مشابه به نظر می‌رسند. عذرخواهی می‌کنم که شهامت یا تخیل لازم برای ابداع واژگان کاملاً جدید را ندارم؛ تلاش‌های گذشته‌ام در واژه‌سازی موفق نبوده است.

قضاوت و شبهِقضاوت

بیایید با مفهوم قضاوت شروع کنیم. در حوزهٔ حقوق، وکلا اغلب ذهنیت من دربارهٔ

---

۱. در اینجا نوعی هم‌سویی با عنوان و روح کتاب *چنان‌که گویی انسان است: اخلاق و هوش مصنوعی* نوشتهٔ نایجل شادبولت و راجر همپسن (Roger Hampson) دیده می‌شود.

هوش مصنوعی را به چالش می‌کشند با این ادعا که «ماشین‌ها قادر به اعمال قضاوت نیستند» (و در نتیجه، طبیعتاً شغل آن‌ها در امان است، زیرا قضاوت هستهٔ اصلی حرفهٔ آن‌هاست). پزشکان نیز ادعایی مشابه مطرح می‌کنند و به «قضاوت بالینی» (audit judgement) اشاره می‌کنند که به‌ظاهر جوهرهٔ اصلی خدمات درمانی مطلوب است و البته که خارج از دسترس هوش مصنوعی است. حسابداران نیز بر این نکته تأکید می‌کنند که «قضاوت حسابرسی» (clinical judgement) را، که تاکنون منحصراً در حیطهٔ توانایی انسان بوده، ماشین نمی‌تواند بازآفرینی کند. به گفتهٔ آن‌ها هیچ ماشینی قادر نیست تشخیص دهد کاسه‌ای زیر نیم‌کاسه است، مهارتی که فقط با حضور در محیط کار و بررسی ریزبینانهٔ شرایط به دست می‌آید. این بار نیز با تفکری فرایندگرا روبروییم که قضاوت (مهارتی تخصصی و انسانی) را با نتیجه (سلامت، ایمنی از مخاطرات و اطمینان از وضعیت بازار که خواست جامعه است) اشتباه می‌گیرد. این نگرش همچنین نمونهٔ بارز «ما استثنا هستیم» است؛ نوعی حربهٔ دفاعی و توجیه ویژهٔ کسانی است که تمایل دارند برای حفظ آبرو و روش‌های کاری‌شان، به دور خود دیوار بکشند.

در اینجا مقصودم از «قضاوت» دقیقاً همان است که متخصصان در کار خود اعمال می‌کنند. انواع دیگری از قضاوت وجود دارند ـ نظیر قضاوت زیبایی‌شناختی یا اخلاقی ـ که خارج از حیطهٔ بحث حاضر قرار می‌گیرند. در فضای فعالیت‌های تخصصی، با این پرسش بنیادین مواجه می‌شویم که یک ماشین چگونه می‌تواند قضاوت کند؟ اما به اعتقاد من، طرح این پرسش از پایه نادرست است؛ یا دقیق‌تر بگویم، این پرسشی است فرایندگرا و کم‌ثمر که در دام «مغالطهٔ هوش مصنوعی» گرفتار شده است. چنین پرسشی مثل این است که بپرسیم: چگونه یک ماشین می‌تواند عیناً همان عملکردی را داشته باشد که انسان دارد؟ پاسخ روشن است: نمی‌تواند. اما این پاسخ، مسائل جالب‌توجهی را که در اینجا مطرح است، از بین نمی‌برد. پرسش بهتری که باید پیش از آن مطرح کرد این است: قضاوت حرفه‌ای، پاسخ به کدام مشکل است؟ یا به عبارت دیگر، اصلاً چرا شرایطی پیش می‌آید که انسان‌ها نیاز به اعمال قضاوت پیدا می‌کنند؟ مراجعانی که به دنبال دریافت خدمات تخصصی‌اند، عموماً خطاب به مشاوران نمی‌گویند: «سلام، لطفاً مقداری قضاوت به من بدهید!» قضاوت فی‌نفسه هدف نیست؛ بلکه ابزار یا امکانی است

که انسان‌ها در مواجهه با عدم قطعیت به کار می‌گیرند. به تعبیری کلی‌تر، قضاوت فرایند است و نتیجهٔ نهایی، زدودن عدم قطعیت است.

به‌جای این پرسش که آیا ماشین‌ها می‌توانند قضاوت کنند، پرسش به‌مراتب بهتر این است که آیا ماشین‌ها می‌توانند با عدم قطعیت مواجه شوند؟ به‌طور کلی، با توجه به حجم عظیم داده‌هایی که در اختیار دارند، پاسخ به این سؤال یک «بله»ی قاطعانه است.

پدر مرحومم پزشک متخصص پوست بود. او حامی بزرگ کارهایم بود و ما، به‌ویژه در دههٔ ۱۹۸۰، گفت‌وگوهای بسیاری دربارهٔ هوش مصنوعی داشتیم. من به دلیل برخی از موفقیت‌های اولیهٔ هوش مصنوعی در کاربردهای پزشکی، به این حوزه علاقه‌مند شده بودم. این توفیق را داشتم که با برخی پیشگامان هوش مصنوعی در عرصهٔ پزشکی همکاری کنم؛ همکاری‌هایی که در پژوهش‌های مشترک من و دانیل ساسکیند دربارهٔ هوش مصنوعی در مشاغل تخصصی تثبیت شد و سلامت یکی از محورهای اصلی تحقیق ما بود.

در اواخر دههٔ ۱۹۸۰، من و پدرم به فکر ایجاد یک سیستم هوش مصنوعی افتادیم که بتواند ملانوم بدخیم (خطرناک‌ترین نوع سرطان پوست) را تشخیص دهد. طرح اولیهٔ ما که کاملاً منطبق با جریان اصلی سیستم‌های خبرهٔ مبتنی بر قواعد آن زمان بود، این بود که اساساً یک فلوچارت عظیم بسازیم که افراد غیرمتخصص را طبق قواعدی که متخصصان پوست برای ارزیابی خوش‌خیم یا بدخیم بودن ضایعات پوستی استفاده می‌کنند، راهنمایی کند. اما این کار دشوارتر از آن بود که تصور می‌کردیم. ما در شناسایی تمامی قواعد با مشکل مواجه شدیم و حتی سخت‌تر این بود که وزن نسبی قواعدی که استخراج کرده بودیم و چگونگی تعاملشان با یکدیگر را تعیین کنیم.

در ۱۹۸۹، دومین و آخرین تلاش ما استفاده از یک ابزار اولیهٔ یادگیری ماشین بود که از الگوریتمی به نام ID3 برای تولید درخت‌های تصمیم از دلِ مجموعهٔ داده‌ها استفاده می‌کرد. ایده این بود که به‌جای ساخت دستی درخت تصمیم از طریق «استخراج دانش» (که در واقع سعی می‌کردم گنجینهٔ دانش پدرم را استخراج کنم)، بتوانیم داده‌هایی دربارهٔ علائم و نتایج بافت‌شناسی (Histology) جمع‌آوری کنیم تا سیستم به‌طور خودکار درخت تصمیم را تولید کند. در نهایت، به دلیل عدم دسترسی به داده‌های کافی، پروژه را کنار گذاشتیم. (الان که به آن زمان فکر می‌کنم، شاید جالب و کمی خجالت‌آور باشد

که فکر می‌کردیم کاربرد اصلی یادگیری ماشین تولید درخت‌های تصمیم است؛ اما این تصور ما از هوش مصنوعی در آن زمان بود).

سال‌ها بعد، در ۲۰۱۷ پدرم مقاله‌ای از مجلهٔ *نیچر (Nature)* با عنوان «طبقه‌بندی سرطان پوست در سطح متخصصان پوست با بهره‌گیری از شبکه‌های عصبی عمیق» را برایم ارسال کرد. این پژوهش به‌راستی نقطهٔ عطفی در این حوزه به نظر می‌رسید. وقتی درباره‌اش صحبت کردیم، او گفت در همان نگاه اول دریافته بود که این سیستم بهتر از خودش و همکارانش عمل خواهد کرد. او با وضوح توضیح داد که آنچه در اصطلاحات پزشکی به «شهود»، «غریزه» یا «قضاوت» معروف است، در حقیقت فرایندی شناختی است که پزشک وضعیت بیمار را ـ چه آگاهانه و چه ناخودآگاه ـ با بانک اطلاعات کوچک تجربیات خود مقایسه می‌کند. پدرم اشاره کرد که در طول دهه‌ها فعالیت حرفه‌ای‌اش فقط چند صد مورد ملانوم دیده بود. بنابراین، اذعان داشت که یک سیستم هوش مصنوعی با دسترسی به پایگاه داده‌ای شامل ۱۲۹ هزار و ۴۵۰ مورد ضایعهٔ پوستی مشکوک، احتمالاً این امکان را دارد که از او و دیگر متخصصان پوست پیشی بگیرد؛ حتی اگر فرایندهای زیربنایی آن از روش‌های کار پزشکان باتجربه کاملاً متفاوت باشد. این استدلال به بسیاری از حرفه‌های تخصصی دیگر تعمیم‌پذیر است. تطابق الگوها با تجربیات گستردهٔ قبلی را می‌توان تحت عنوان «تشخیص کلی‌نگر شباهت» (Holistic Similarity Recognition)[1] توصیف و تبیین کرد[2].

همدلی و شبه‌همدلی

یکی دیگر از انتقادات رایج به سیستم‌های هوش مصنوعی، ناتوانی آن‌ها در بروز همدلی است. چگونه یک سیستم می‌تواند حالت‌های عاطفی انسان‌ها را تشخیص دهد و از همه مهم‌تر، چگونه می‌تواند خود را جای آن‌ها بگذارد و در غم و شادی‌شان شریک شود؟

در تأملی که اوایل ۲۰۲۳ داشتم، تصمیم گرفتم این وظیفه را به یک هوش مصنوعی

---

۱. به شیوه‌ای اشاره دارد که در آن شباهت بین دو چیز مثل تصویر، جسم یا مجموعه‌ای از داده‌ها به‌صورت کلی و جامع بررسی می‌شود، نه صرفاً براساس ویژگی‌های مجزا یا بخش‌های خاص. ـ م.

۲. این اصطلاح توسط هوبرت دریفوس و استوارت دریفوس (Stuart Dreyfus) در کتاب *فراتر از ماشین: قدرت شهود و تخصص انسانی در عصر کامپیوتر* (۱۹۸۶) ابداع شده است.

مولد محول کنم. از نسخهٔ اولیهٔ چت‌جی‌پی‌تی پرسیدم آیا می‌تواند «یک ایمیل همدلانه از طرف من برای دوستی که به‌تازگی پدرش را از دست داده» بنویسد؟ پاسخش چنین بود:

> **از شنیدن خبر درگذشت پدرتان بسیار متأسفم. از دست دادن والدین یکی از دشوارترین تجربه‌های زندگی است و در این دوران سخت، همدرد شما و خانواده‌تان هستم. می‌دانم هیچ واژه‌ای قادر به بیان عمق غم شما نیست، اما بدانید که من اینجا هستم تا به هر نحوی که بتوانم از شما حمایت کنم. چه نیاز به هم‌صحبت یا شانه‌ای برای گریستن داشتید و یا حتی به دنبال فرصتی برای خلوت هستید بدون هیچ تردیدی با من در میان بگذارید. در این لحظات غم‌بار، محبت و همدردی بی‌قیدوشرط من را بپذیرید.**

البته که این پاسخ بی‌نقص نبود ـ چند اشتباه داشت ـ اما خیلی بد نبود. می‌شود گفت به‌خوبی بسیاری از پیام‌های تسلیتی بود که پس از درگذشت پدرم از دوستانم دریافت کردم. در واقع، هنگام خواندن این پیام به یاد پدرم افتادم و چشمانم پر از اشک شد. در آن لحظه به این فکر کردم که واژه‌ای برای توصیف آنچه برایم اتفاق افتاده بود نداریم؛ مفهومی که بتواند حالتی را بیان کند که یک ماشین باعث گریهٔ انسان شده است.

این آزمایش من را بر آن می‌دارد که در مورد تعمیم فرضیات گسترده دربارهٔ هوش مصنوعی و همدلی، احتیاط به خرج دهیم. همدلی دو بُعد دارد: شناختی و احساسی. بُعد شناختی به فرایندهای ذهنی اشاره دارد که از طریق آن‌ها حالت عاطفی فرد دیگری درک و دسته‌بندی می‌شود. آیا سیستم‌های هوش مصنوعی می‌توانند حالت‌های عاطفی را تشخیص دهند (یا به‌نوعی شبه‌تشخیص دهند)؟ پیشرفت‌ها در حوزهٔ رایانش احساسی (Affective computing) نشان می‌دهد پاسخ مثبت است (رایانش احساسی به سیستم‌هایی می‌پردازد که هم می‌توانند عواطف را تشخیص دهند و هم ابراز کنند). تحقیقات مرتبط در این زمینه به ما دلیلی می‌دهد تا انتظار داشته باشیم که سیستم‌های هوش مصنوعی در آیندهٔ نزدیک در تشخیص حالات روانی انسان‌ها از خود انسان‌ها بهتر عمل کنند. فراتر از آن، کاملاً قابل پیش‌بینی است که سیستم‌های هوش

مصنوعی به‌زودی خواهند توانست به‌گونه‌ای به کاربران پاسخ دهند که حتی همدلانه‌تر از اکثر انسان‌ها به نظر برسد که ترکیبی از هوش مصنوعی مولد، پروفایلینگ عاطفی و روان‌شناختی و فناوری سنتز گفتار (شاید با ارائهٔ صدایی دلسوزانه) است. در اینجا خوانندگان را به کار جالب جرمی هویک[1] ارجاع می‌دهم که روی ایدهٔ «آزمون تورینگ همدلی» کار می‌کند: «آیا کاربر انسانی می‌تواند بین همدلی ابرازشده توسط یک مراقب رباتیک (artificial carer)[2] و یک پزشک انسانی تمایز قائل شود؟»

صِرف اینکه ماشین‌ها ممکن است واکنش‌های مناسبی از خود نشان دهند، به این معنی نیست که آن‌ها به‌نوعی جهان را از دریچهٔ نگاه کاربران خود می‌بینند یا غم و شادی آن‌ها را احساس می‌کنند و با فرد همدرد می‌شوند. این امر مستلزم بازسازی مؤلفهٔ دوم همدلی، یعنی بخش عاطفی است؛ همان جنبه‌ای که فرض می‌کند سیستم‌های هوش مصنوعی می‌توانند نوعی هوشیاری را تجربه کنند (ادعایی که در فصل ۱۰ به نقد آن خواهم پرداخت و آن را فاقد پایه و اساس علمی می‌دانم).

به‌این‌ترتیب، احتمالاً به مرزهای همدلی ماشینی رسیده‌ایم. واقعیت این است که سیستم‌های هوش مصنوعی فاقد توانایی برای قرار دادن خود در جایگاه انسان‌ها و تجربهٔ غیرمستقیم عواطف آنان هستند. در مواجهه با این مسئله، می‌توان به این نکته اشاره کرد که حتی در مورد انسان‌ها نیز هرگز نمی‌توان با قطعیت دریافت که ابراز همدردی دیگران اصیل است یا خیر. چه بسا آنان ــ آگاهانه یا ناآگاهانه ــ درحال پنهان‌کاری، تظاهر یا وانمود کردن باشند. به بیان صریح‌تر، هیچ‌گاه نمی‌توانیم از وجود احساسات واقعی دیگران نسبت به خود اطمینان حاصل کنیم. هرچند، احتمال اصیل بودن این احساسات زیاد است، اما دسترسی مستقیم به آنچه فلاسفه «ذهن‌های دیگر» (problem of other minds)[3] می‌خوانند، برایمان مقدور نیست.

با توجه به پیشرفت‌های اخیر در حوزهٔ هوش مصنوعی، محتمل است که در آینده‌ای

---

1. Jeremy Howick

۲. مراقب رباتیک به نوعی سیستم یا دستگاه مبتنی بر هوش مصنوعی (مانند یک ربات، یک نرم‌افزار پیشرفته یا یک دستیار صوتی) اشاره دارد که برای ارائهٔ مراقبت و پشتیبانی به افراد، به‌ویژه سالمندان یا بیماران، طراحی شده است. ـ م.

۳. مسئلهٔ ذهن‌های دیگر به این پرسش فلسفی اشاره دارد که از کجا می‌توانیم مطمئن شویم دیگران هم ذهن یا آگاهی‌ای شبیه ما دارند؟ ما فقط می‌توانیم رفتار یا حرف‌هایشان را ببینیم و بشنویم، اما نمی‌توانیم مستقیماً به افکار یا احساسات درونی‌شان دسترسی داشته باشیم؛ بنابراین، همیشه نوعی عدم قطعیت وجود دارد. ـ م.

نه‌چندان دور، سیستم‌های هوش مصنوعی در شبیه‌سازی همدلی (تظاهر به آن) حتی از انسان‌های غیرصادق نیز بهتر عمل کنند و شاید حتی قادر باشند واکنش عاطفی مطلوب‌تری در کاربران خود ایجاد نمایند تا انسان‌های صادق.

همهٔ این‌ها مرا به این نتیجه می‌رساند که تا زمانی که واژهٔ بهتری برای این مفهوم نداشته باشیم، منطقی و مفید است که بگوییم سامانه‌های هوش مصنوعی می‌توانند «شبه‌همدل» باشند.

خلاقیت و شبه‌خلاقیت

مسائل اصلی پیرامون رابطهٔ خلاقیت و هوش مصنوعی در ۱۹۸۴ توسط دونالد میکی[1] (پدر هوش مصنوعی بریتانیا) و روری جانستون[2] در کتاب پیشگویانه‌شان با عنوان *کامپیوتر خلاق* مطرح شد. آن‌ها در این اثر، این پرسش را مطرح کردند که «آیا ماشین‌ها می‌توانند راهکارهایی بیندیشند که از ذهن انسان دور مانده است؟» برای نمونه «آیا ممکن است کامپیوترها آثار هنری جدید خلق کنند؟»

در جریان سخنرانی‌هایم دربارهٔ کتاب *آیندهٔ مشاغل*، پرتکرارترین پرسشی که با آن روبرو می‌شوم این است که «فرزندان ما در آینده چه شغل‌هایی خواهند داشت؟» جالب اینجاست که متخصصان ارشد امروز چندان نگران آیندهٔ خود نیستند، اما به‌روشنی درمی‌یابند که در آینده‌ای نه‌چندان دور، بسیاری از مشاغل سنتی توسط سیستم‌هایی که روزبه‌روز توانمندتر می‌شوند، دستخوش تحولات اساسی خواهند شد؛ اگر نگوییم که به‌طور کامل جایگزین می‌شوند. در پاسخ به این دغدغه‌ها، تا چندی پیش معمولاً صنایع خلاق را به‌عنوان آخرین سنگر مقاومت در برابر پیشرفت هوش مصنوعی معرفی می‌کردیم. همانند اکثر تحلیلگران، ما نیز بر این باور بودیم که دستاوردهای خلاقانهٔ بشر، حداقل در افق‌های کوتاه‌مدت و میان‌مدت برای ماشین‌ها در دسترس نخواهد بود. اما این تصور اشتباه از آب درآمد، چراکه نسل جدید سیستم‌های هوش مصنوعی اکنون توانایی خلق آثاری همچون شعر، نثر، ترانه، موسیقی، کاریکاتور، طراحی‌های خطی،

1. Donald Michie
2. Rory Johnston

مناظر، ویدیو و بسیاری چیزهای دیگر را دارد؛ آن هم نه در سطحی معمولی، بلکه اغلب فراتر از استانداردهای کاری که اکثر انسان‌ها قادر به تولیدش هستند. برای مثال، این واقعیت می‌تواند برای هنرمندان خلاق نگران‌کننده باشد، چراکه اکنون حتی افراد فاقد مهارت‌های هنری نیز می‌توانند به‌سادگی آثار بصری خیره‌کننده‌ای برای مقاصدی مانند ارائهٔ مجموعه اسلایدهای حرفه‌ای تولید کنند.

خروجی سیستم‌های هوش مصنوعی در تولید آثار هنری و ادبی از کیفیت قابل قبولی برخوردار است، اما ممکن است همچنان تمایل داشته باشیم خلاقیت سیستم‌های هوش مصنوعی مولد را انکار کنیم. من این احساس را به‌خوبی درک می‌کنم، اما باید توجه داشته باشیم که چنین نگرشی نمونه‌ای از تفکر فرایندگرا است. هرچند این دیدگاه لزوماً نادرست نیست، اما صرفاً یک دیدگاه است. این نگرش بر شیوهٔ انسانی خلق آثار هنری و ادبی تأکید دارد، نه صرفاً بر نتیجهٔ نهایی. با وجود بدیع بودن خروجی هوش مصنوعی، ظاهراً برایمان اهمیت دارد که آن اثر از «جوشش خلاقانهٔ انسانی» سرچشمه نگرفته است. حتی اگر خروجی هوش مصنوعی به‌نوعی برتر باشد، باز هم ممکن است چنین فکری کنیم. شاید عبارت «شبه‌خلاق» توصیف مناسب‌تری برای این فرایند محاسباتی زیربنایی باشد؛ دست‌کم تا زمانی که واژهٔ دقیق‌تری برای آن ساخته شود.

باوجوداین، چالشی عمیق‌تر از انتخاب صرفِ یک عنوان برای سیستم‌های خالق آثار بدیع در میان است. پرسش مهم‌تر این است که ما انسان‌ها چگونه باید به این محتوای اصیل تولیدشدهٔ سیستم‌های هوش مصنوعی بنگریم؟ آیا وقتی به طرحی نگاه می‌کنیم یا داستان کوتاهی را می‌خوانیم که یک سیستم هوش مصنوعی تولیدش کرده ــ حتی اگر فوق‌العاده باشد ــ چیزی اساسی کم نیست؟

برای روشن‌تر شدن این موضوع، می‌توانیم از تمثیلی ورزشی بهره ببریم. فرض کنید در آینده‌ای نه‌چندان دور، ربات‌هایی طراحی شوند که بتوانند مسیر صد متر را از یوسین بولت سریع‌تر طی کنند یا در زمین گلف، عملکردی برتر از تایگر وودز داشته باشند. اما آیا این موضوع برای ما جذاب خواهد بود؟ شاید اگر به عملکرد ربات‌ها علاقه‌مند باشید، پاسخ شما مثبت باشد؛ اما اکثر ما از تماشای بولت و وودز دقیقاً به این دلیل هیجان‌زده می‌شدیم که آن‌ها نیز انسان‌هایی از جنس گوشت، پوست و استخوان خودمان بودند.

ما از دستاوردهای اسطوره‌های ورزشی سرمست می‌شویم؛ نوعی لذت نیابتی از عملکرد آنان کسب می‌کنیم، تا حدی به این دلیل که آن‌ها ـ البته با فاصله‌ای بسیار زیاد ـ شبیه به خودمان هستند. اگر خودمان دونده یا گلف‌باز باشیم، به‌خوبی درمی‌یابیم که این افراد تا چه حد استثنایی‌اند. ما آرزو داریم حتی ذره‌ای مانند آن‌ها باشیم. از مهارتشان لذت می‌بریم. از تلاش و فداکاری‌هایی که در تمریناتشان متحمل شده‌اند به حیرت می‌آییم. مسئله صرفاً سرعت و دقت نیست. آنچه اهمیت دارد الهامی است که از همنوعان خود در اوج توانایی‌شان می‌گیریم. این موضوع فقط به نتایج مربوط نیست. شیوهٔ دستیابی به این نتایج، بخشی حیاتی و غیرقابل جایگزین از تجربهٔ تماشاگر بودن است.

در هنر نیز همین‌گونه است. وقتی اثر ادبی بزرگی را می‌خوانیم، به موسیقی دل‌نشینی گوش می‌دهیم یا نقاشی خیره‌کننده‌ای را تماشا می‌کنیم، بخشی از لذت ما ناشی از این آگاهی است که انسان دیگری در خلق آن نقش داشته است، کسی که کوشیده، ارتباط برقرار کرده، آفریده و در نهایت به زندگی ما الهام بخشیده، آن را برانگیخته و تعالی بخشیده است. عنصر کلیدی این تجربه، دانستنِ حضور یک انسان در آن سوی اثر است. ما می‌توانیم درد، ناامیدی، شادی یا شوروشوق خالق را در اثر احساس کنیم.

صرف‌نظر از پیشرفت‌های خیره‌کنندهٔ سیستم‌های هوش مصنوعی، به نظر می‌رسد بسیاری از شکل‌های بیان انسانی ـ به‌ویژه اجراهای زنده ـ کماکان به‌خاطر خودشان برای انسان‌ها ارزشمند خواهند بود. این همان جوهره‌ای است که تفاوت بنیادین بین خلاقیت انسان و شبه‌خلاقیت ماشین‌ها را برای ما تعریف می‌کند. این دیدگاه به‌معنای انکار توانایی‌های چشمگیر هوش مصنوعی نیست، بلکه تأکیدی است بر این باور که فرایند خلاقیت انسانی ارزشی فراتر از خود اثر هنری دارد.

بااین‌حال، می‌پذیریم که اکنون و در آینده، گاه در مواجهه با یک نقاشی، شعر یا قطعهٔ موسیقی، تشخیص منشأ انسانی یا ماشینی اثر ناممکن خواهد بود. این ابهام ممکن است تجربهٔ لذت بردن ما را تحت‌تأثیر قرار دهد؛ آیا داریم اثر همنوعان خود را می‌ستاییم یا با شبه‌اثر از یک سیستم هوش مصنوعی طرف هستیم؟ این وضعیت مسبوق به سابقه است، زیرا همواره با آثار تقلبی و جعل‌های هنری دست‌وپنجه نرم کرده‌ایم. بااین‌حال، انتظار دارم این موضوع روزبه‌روز برای ما اهمیت بیشتری پیدا کند و در آینده

به‌طور معمول به دنبال نوعی تأیید اصالت برای آثار هنری باشیم؛ گواهی که منشأ آن‌ها را مشخص کند و اطمینان دهد که خالق اثر یک انسان بوده است. در این زمینه، واسطه‌های مورد اعتماد بـه کمک خواهند آمـد. اقدامـات اولیـه در ته‌نقش‌گـذاری (watermark) مستندات تولیدشده توسط هوش مصنوعی نشان می‌دهد که این موضوع جدی گرفته شده است[۱].

نسـل‌های آینده ممکن اسـت نگاهی متفاوت داشته باشـند. اگر هوش مصنوعی به مرحله‌ای برسد که آثارش برتری غیرقابل انکاری از تولیدات انسانی داشته باشد، می‌توان تصور کرد که برای برخی از شبه‌اثرها (مثل انیمیشن، بازی یا فیلم) منشأ خلق اثر اهمیت خود را از دست بدهد.

۱. ته‌نقش یا واترمارک در اسناد هوش مصنوعی به نشانهٔ پنهان یا آشکار در محتوای تولیدشده توسط AI اشاره می‌کند که برای شناسایی منشأ آن به محتوا افزوده می‌شود. ـ م.

بخش سوم

۱۳

عملیاتی‌سازی هوش مصنوعی

فصل ششم

خودکارسازی، نوآوری، حذف

در ادامه به موضوعات عملی‌تر می‌پردازیم اما همچنان در مباحث مربوط به مغز می‌مانیم. در ۲۰۱۷، این افتخار را داشتم که در کنفرانسی در بوستون برای جمعی متشکل از دو هزار جراح مغز و اعصاب سخنرانی کنم. مخاطبانی ترسناک بودند! برگزارکنندگان رویداد دعوت کرده بودند دربارهٔ پژوهش‌های اصلی صورت‌گرفته برای تألیف کتاب *آیندهٔ مشاغل* صحبت کنم و به این پرسش پاسخ دهم که در آینده چه سرنوشتی در انتظار حرفهٔ جراحی مغز و اعصاب است؟ از من خواسته بودند با صراحت تمام دربارهٔ تأثیر هوش مصنوعی و رباتیک بر حرفه‌شان صحبت کنم، حتی اگر این بحث‌ها جنجالی باشد.

در همان ابتدای سخنرانی، جسورانه اعلام کردم: «بیماران به جراحان مغز و اعصاب نیاز ندارند.» زمزمه‌های نگران‌کننده‌ای در سالن پیچید. رنگ از چهرهٔ رئیس کنفرانس رخت بست. فوراً به اصل مطلب بازگشتم و توضیح دادم: «آنچه بیماران می‌خواهند، سلامتی است.» تأکید کردم که در شرایط فعلی، برای مشکلات خاص مغز و اعصاب، جراحان انسانی بی‌تردید بهترین گزینه‌ای هستند که اکنون در اختیار داریم و از خدماتشان قدردانی کردم. اما این گمان را مطرح کردم که در آینده، نسل‌های بعد به گذشته نگاه خواهند کرد و باورشان نخواهد شد که پیشینیانشان این‌قدر سر و بدن می‌شکافتند؛ گویی به روزگار جراحی‌های اولیه و ابتدایی نظر می‌افکنند!

سپس به نکتهٔ کلیدی سخنانم پرداختم. به حضار گفتم که آیندهٔ نظام سلامت، صرفاً به جراحی‌های بهتر یا حتی رباتیک محدود نخواهد شد، بلکه بدون شک درمان‌های غیرتهاجمی، کم‌تهاجمی و همچنین پزشکی پیشگیرانه از جایگاه ویژه‌ای برخوردار خواهند بود.

در طول سالیان متمادی، این روایت را نمونه‌ای عینی از آنچه «تفکر نتیجه‌گرا» می‌نامم، مطرح کرده‌ام. این روش، متخصصان را به چالش می‌کشد تا به‌جای تمرکز بر شیوه‌های کنونی دستیابی به این نتایج، روی خودِ نتایج مطلوب متمرکز شوند. این نگرش تصریح می‌کند که ارزش خدمات تخصصی عمدتاً ـ اگر نگوییم کاملاً ـ در نتایج حاصله نهفته است، نه در روش‌ها و تکنیک‌های مورد استفاده که برای دستیابی به آن‌ها به‌کار می‌روند. بااین‌حال، اخیراً متوجه شدم که این روایت جراحان مغز، درس دیگری نیز در بر دارد که

به همان اندازه مهم است. درسی که به یک تمایز اساسی اشاره می‌کند: تفاوت میان «خودکارسازی»، «نوآوری» و «حذف»؛ تفاوتی که اکنون به توضیحش می‌پردازم.

خودکارسازی

منظورم از «خودکارسازی»، کامپیوتری کردن کارهایی است که هم‌اکنون انجام می‌دهیم. این مفهوم به‌معنای به‌کارگیری فناوری برای انجام وظایف، فرایندها یا فعالیت‌های موجود است که معمولاً به‌منظور افزایش کارایی صورت می‌گیرد. به بیان دیگر، زمانی که رویه‌های فعلی خود را ساده‌سازی، بهینه‌سازی یا حتی تقویت می‌کنیم، در واقع داریم آن‌ها را خودکارسازی یا نظام‌مند می‌کنیم. این نوع خودکارسازی تاکنون نقش اصلی را در داستان کامپیوترها ایفا کرده است. طی هشتادوپنج سال گذشته، فناوری را به روش‌های سنتیِ انجام کارها اضافه کرده‌ایم. در واقع، وقتی اکثر مردم به فناوری دیجیتال فکر می‌کنند، اولین چیزی که به ذهنشان می‌رسد همین مفهوم خودکارسازی است.

تقریباً شک ندارم که خودکارسازی، به‌عنوان یک روشِ استفاده از فناوری، به کاهش هزینه، صرفه‌جویی در زمان، بهبود فرایند، افزایش بهره‌وری و سایر مزایای مشابه منجر شده است. علاوه‌بر این، ازآنجاکه خودکارسازی از وضعیت موجود حمایت و آن را تقویت می‌کند، برای بسیاری از کارمندان اداری و کارشناسان، موضوعی آشنا و اطمینان‌بخش به نظر می‌رسد. در واقع، خودکارسازی بسیاری از جنبه‌های روش‌های سنتی کار ـ حتی بخش‌هایی که ناکارآمد و معیوب‌اند ـ را حفظ می‌کند، به همین دلیل است که بعضی از منتقدان با کنایه می‌گویند خودکارسازی به ما «آشفتگی بیشتر با هزینهٔ کمتر» تحویل می‌دهد!

مفهوم و اصطلاح خودکارسازی، محور اصلی گفت‌وگوها دربارهٔ فناوری دیجیتال است. اگر به مثال جراحی مغز برگردیم، جراحی رباتیک به‌وضوح نمونه‌ای از خودکارسازی است که در آن، برخی یا تمام فرایندها و فعالیت‌های جراحی توسط ماشین به‌جای انسان انجام می‌شود. این نمونه همچنین مصداق چیزی است که برخی اقتصاددانان و مشاوران کسب‌وکار آن را «جایگزینی وظایف» می‌نامند، یعنی در برخی وظایف، انسان از چرخه خارج می‌شود و ماشین به‌جای او، آن کار را انجام می‌دهد.

نوآوری

تمرکز صرف بر جایگزینی وظایف، این تصور را ایجاد می‌کند که آیندهٔ جراحی منحصراً به جراحی رباتیک محدود می‌شود؛ یعنی کار جراحان به‌طور کامل خودکار شده و جایگزین می‌شود. اما روایت من از جراحی مغز، در بُعدی مهم‌تر، بر مقولهٔ «نوآوری» تأکید دارد. من این اصطلاح را به‌شکلی بسیار خاص به کار می‌برم تا به استفاده از روش‌ها و فناوری‌هایی اشاره کنم که انجام کارهایی را ممکن می‌سازند که قبلاً غیرممکن یا حتی غیرقابل تصور بودند. روش‌های غیرتهاجمی یا کم‌تهاجمی، نمونه‌های بارز نوآوری محسوب می‌شوند. این روش‌ها به‌گونه‌ای طراحی شده‌اند که نتایج مطلوب بیماران را با درد و اضطراب کمتر، ریسک‌های پایین‌تر و بدون دوره‌های طولانی نقاهت که معمولاً با جراحی‌های سنتی همراه است، تأمین کنند. روش‌های غیرتهاجمی و کم‌تهاجمی به‌هیچ‌وجه جراحی سنتی را خودکار نمی‌کنند؛ بلکه از همان ابتدا، راهکارهای کاملاً جدید و نوآورانه‌ای برای درمان بیماران ارائه داده‌اند که قبلاً در دسترس نبوده‌اند. بر این مبنا، نوآوری را می‌توان چنین تعریف کرد: دستیابی به نتایج مطلوبی که بیماران، مراجعان و کاربران همواره به دنبال آن هستند، اما این بار با استفاده از روش‌ها و فناوری‌هایی کاملاً جدید.

حذف

در روایت من از جراحی مغز و اعصاب، نمونه‌ای روشن از آنچه «حذف» می‌نامم نیز قابل مشاهده است. تأثیر برخی فناوری‌ها صرفاً حل یک مشکل نیست، بلکه اساساً خودِ آن مشکل را از بین می‌برند. مورد مطالعاتی مورد علاقه‌ام در زمینهٔ فناوری‌های حذف‌کننده، «بحران بزرگ فضولات اسب» در دههٔ ۱۸۹۰ است. در آن زمان، حمل‌ونقل در نیویورک کاملاً در اختیار اسب‌ها و کالسکه‌ها بود. اسب‌ها فضولات بسیار زیادی تولید می‌کردند که در خیابان‌ها پراکنده بود. سیاست‌گذاران نمی‌دانستند چگونه با این خطر بی‌سابقه مقابله کنند. چندی نگذشت که خودرو اختراع شد و در اوایل قرن بیستم، دیگر برای جابجایی مردم و کالاها به اسب‌ها نیاز نبود. بنابراین، فناوری بنیادین جدید یعنی موتور احتراق داخلی، منجر به حذف کامل مشکل فضولات شد.

همین امر در مورد هوش مصنوعی نیز صادق است. در راستای پزشکی پیشگیرانه،

به‌راحتی می‌توان تصور کرد که هوش مصنوعی سیستم‌ها و فرایندهایی معرفی خواهد کرد که اساساً باعث می‌شود مشکلی برای حل کردن وجود نداشته باشد. تمرکزم در اینجا بر حذف نقش انسان‌ها توسط هوش مصنوعی نیست (اگرچه این اتفاق ممکن است رخ دهد)، بلکه بر تکنیک‌ها و فناوری‌هایی است که طیف وسیعی از مشکلات و مسائلی را حذف خواهند کرد که امروزه بسیاری از افراد و سازمان‌ها برای حلشان تلاش می‌کنند.

قدرت حذف‌کنندگی فناوری به‌ندرت مورد توجه یا بحث قرار می‌گیرد، اما به نظر من، این نیرو در نهایت ممکن است قدرت‌مندترین نیروی تغییر در میان تمامی نیروهای تحول‌آفرین باشد.

چند نمونه و مثال

مرزهای تمایز میان «خودکارسازی»، «نوآوری» و «حذف» کاملاً دقیق و مشخص نیست. هم‌پوشانی‌ها و مناطق خاکستری‌ای بین آن‌ها وجود دارد. اما امیدوارم این تقسیم‌بندی، چشم‌اندازی گسترده‌تر از تأثیر هوش مصنوعی ارائه دهد که در بسیاری از بحث‌های جاری غایب است. ــ درک عمومی از هوش مصنوعی عمدتاً بر خودکارسازی متمرکز شده و نوآوری و حذف را نادیده گرفته است. این خطایی فاحش است. مهم‌ترین پیامد این نگرش محدود، آن است که وقتی مردم از هوش مصنوعی و آیندهٔ مشاغل سخن می‌گویند، تأثیرات آتی آن را بیش از حد دست‌کم می‌گیرند. چند پروژهٔ تحقیقاتی دانشگاهی و تجاری پراستناد وجود دارد که از همین زاویه می‌توان آن‌ها را به چالش کشید. این باور که تهدیدهای شغلی تنها به «تجاوز وظیفه‌ای» (Task encroachment) محدود می‌ماند (یعنی ماشین‌ها فقط بخشی از وظایفی را که انسان‌ها امروز انجام می‌دهند بر عهده می‌گیرند) به‌کلی امکان دو تحول دیگر را نادیده می‌گیرد: نوآوری (خودِ وظایف دیگر وجود ندارند) و حذف (مشکلاتی که مشاغل برای آن‌ها وجود دارند، خودشان ناپدید می‌شوند).

دعاوی حقوقی و دادگاه‌ها

چند مثال دیگر این تمایز را ملموس‌تر می‌سازد. یکی از این موارد در دنیای حقوق،

دادگاه‌ها و حل اختلافات است. در شصت‌وپنج سالِ نخستِ به‌کارگیری فناوری در این حوزه، تمرکز تقریباً به‌طور کامل بر خودکارسازی بود. وکلا و متخصصان فناوری حقوقی صرفاً بر توسعه و استفاده از فناوری‌هایی متمرکز بودند که روش‌های سنتی کار را تسهیل کرده و بهبود می‌بخشیدند؛ مانند بررسی اسناد، ارائهٔ ادله در دادگاه، ثبت شواهد، مدیریت دادگاه و تحقیقات حقوقی برای وکلا و قضات. هیچ‌یک از این سیستم‌ها ماهیت اساسی دعاوی را تغییر ندادند، بلکه صرفاً برخی فرایندها را کارآمدتر کردند.

در مقابل، در دههٔ اخیر، جامعه‌ای پراکنده در سراسر جهان بر چیزی تمرکز کرده‌اند که به «حل اختلاف آنلاین» معروف شده است. این موضوع یکی از دغدغه‌های اصلی کتاب من با عنوان *دادگاه‌های آنلاین و آیندهٔ عدالت*[1] است که در آن این پرسش را مطرح می‌کنم که «آیا دادگاه یک خدمت است یا یک مکان؟» من این فرض را به چالش می‌کشم که برای حل اختلافات حقوقی، همهٔ طرف‌های ذی‌ربط باید در یک مکان و زمان مشخص گرد هم آیند. من و برخی دیگر، رویکردهای مختلفی برای حل اختلاف به‌صورت آنلاین پیشنهاد می‌دهیم، ازجمله جلسات مجازی (که در دوران کرونا متداول شد)، خدمات ناهم‌زمان (که در آن هیچ جلسه یا شهادت شفاهی وجود ندارد، بلکه ادله و استدلال‌ها به‌صورت الکترونیکی ارائه می‌شوند و قضات نیز تصمیمات خود را به همین صورت صادر می‌کنند) و حتی استفاده از هوش مصنوعی برای صدور احکام یا تصمیماتی الزام‌آور. حل اختلاف آنلاین نمونه‌ای از نوآوری است. نتایجی که طرفین به دنبال آن هستند با استفاده از فرایندهایی کاملاً متفاوت محقق می‌شود؛ این خدمات جدید پیش از ظهور و گسترش فناوری دیجیتال اساساً ممکن نبودند. با توجه به تمایز سه‌گانهٔ من، این مورد معادل درمان غیرتهاجمی در حوزهٔ سلامت است.

وقتی صحبت از تأثیر هوش مصنوعی بر حرفهٔ وکالت می‌شود، اکثر وکلای دادگستری ادعا می‌کنند که حرفهٔ آن‌ها در امان است، زیرا هیچ رباتی نمی‌تواند در دادگاه حاضر شود و دفاعیه ارائه دهد. استدلال آن‌ها این است که اکثر وظایف مربوط به وکالت شفاهی را نمی‌توان به ماشین‌ها سپرد. بیشتر محققان نیز با این دیدگاه موافق‌اند و معتقدند که شغل وکلا نسبتاً ایمن است. اما آن‌ها از این واقعیت غافل می‌شوند که دادگاه‌های آنلاین

۱. کتاب *Online Courts and the Future of Justice* اثر ریچارد ساسکیند، در ۲۰۱۹ توسط انتشارات دانشگاه آکسفورد منتشر شد.ـم.

نیاز به وکالت شفاهی را از بین می‌برند. حتی اگر درست فرض کنند که هیچ ماشینی نمی‌تواند در دادگاه استدلال کند (با این توضیح که در بلندمدت، این فرض نیز ممکن است نادرست از آب درآید)، نکته اینجاست که مجموعه وظایف مربوط به حضور در دادگاه دیگر بخشی از فرایند نیست؛ وکالت شفاهی حذف شده است. ممکن است نعل‌بند ماهری باشید، اما اگر اسب استفاده‌ای نداشته باشد، این مهارت بی‌ارزش خواهد بود.

پزشکی پیشگیرانه را در نظر بگیرید؛ این مفهوم در حوزهٔ حقوق نیز نمونهٔ واضحی دارد. من به این رویکرد «پیشگیری از اختلاف» می‌گویم، نه «حل اختلاف پس‌از وقوع»؛ همان‌طور که بهتر است در بالای پرتگاه حفاظ بگذاریم تا اینکه در پایین آن آمبولانس مستقر کنیم. تقریباً همه ترجیح می‌دهند متحمل هزینه‌ها و ناراحتی‌های ناشی از دعوای حقوقی نشوند. بهترین راه برای تحقق این امر، این است که اصلاً دچار مشکل حقوقی نشویم. امروزه هوش مصنوعی ابزارهای مفیدی در این زمینه ارائه می‌دهد که به افراد کمک می‌کند خطرات و مشکلات حقوقی را زودتر از گذشته شناسایی و همچنین مشکلات حقوقی را قبل از اینکه تشدید شوند، مهار کنند. در دنیای اختلافات، اجتناب از مشکل از همان ابتدا نمونهٔ دیگری از حذف است؛ حذفِ همان مشکلاتی که دادگاه‌ها برای حلشان ایجاد شده‌اند.

پول نقد

اگر فراتر از حوزه‌های پزشکی و حقوق بنگریم، بسیاری و چه بسا اکثر کاربردهای واقعاً متحول‌کنندهٔ فناوری دیجیتال، نمونه‌هایی از نوآوری و حذف هستند تا خودکارسازی. یک مثال ساده در این زمینه، دستگاه خودپرداز (ATM) یا تحویل‌دهندهٔ الکترونیکی پول نقد بود. قبل از ظهور این فناوری، چنین نبود که اگر مردم نیمه‌شب به پول نیاز داشتند، به بانک محلی می‌رفتند و سوراخی در دیوار می‌یافتند که فرد بیچاره‌ای در طرف دیگر آن نشسته و به‌محض درخواست، اسکناس تحویل می‌داد! این‌گونه نبود که فرایندی دستی بابت توزیع پول از قبل وجود داشته باشد، اما بانک‌داران به دلیل ناکارآمد بودن (یا مثلاً، سرمای هوا!) تصمیم به خودکارسازی آن گرفته باشند؛ بدیهی است که چنین

چیزی نبوده است. فناوری راهی کاملاً جدید برای ارائهٔ خدمات بانک‌داری ایجاد کرد. این نوآوری، حتی اگر مبتنی بر هوش مصنوعی هم نبود، بانک‌داری خرد را متحول کرد. بعدها، فناوری کارت‌های اعتباری و سیستم‌های پرداخت الکترونیک پا را فراتر گذاشتند و تا حد زیادی نیاز به پول نقد را از بین بردند.

کارهای خانه

مثال دیگری را بررسی کنیم. از من دعوت شد در پروژه‌ای تحقیقاتی که مؤسسهٔ اینترنت آکسفورد (Oxford Internet Institute) انجامش می‌داد نقشی فرعی ایفا کنم. این پژوهش از متخصصان هوش مصنوعی می‌پرسید چقدر طول می‌کشد تا ربات‌ها برای انجام کارهای روزانهٔ خانه مورد استفاده قرار بگیرند[1]. یکی از پرسش‌گران سؤالی مشابه این را مطرح کرد: «چقدر طول می‌کشد تا ربات‌هایی ایجاد شوند که بتوانند لباس‌ها را اتو کنند؟» با کمی جسارت ـ و شاید گستاخی ـ پاسخ دادم که سؤال اشتباهی پرسیده‌اند. گفتم بهتر است که بپرسیم اساساً اتو کردن چه مشکلی را حل می‌کند. من آن‌ها را تشویق می‌کردم به‌جای «تفکر فرایندگرا» به «تفکر نتیجه‌گرا» رو بیاورند (به فصل ۳ مراجعه کنید). اگر رویکرد فرایندگرا داشته باشیم، ممکن است تصور کنیم رباتی انسان‌نما با اتوی معمولی مشغول اتوکشی است. این مصداقی از خودکارسازی و جایگزینی وظایف خواهد بود. اما اگر نتیجه‌گرا فکر کنیم، قطعاً به این نتیجه می‌رسیم که آنچه واقعاً می‌خواهیم، «پارچه‌ای صاف و بدون چین‌وچروک» است. می‌توان دستگاهی برای صاف کردن پارچه تصور کرد که اصلاً شبیه انسان یا ربات انسان‌نما با اتو نباشد. اما به محققان گفتم که در بلندمدت، حتی نیاز به هرگونه سازوکار حرارتی برای صاف کردن لباس‌ها از بین خواهد رفت، زیرا علم مواد (که بی‌شک با کمک هوش مصنوعی پیشرفت خواهد کرد) پیراهن، شلوار، ملحفه و سایر منسوجات بدون چین‌وچروک تولید خواهد کرد. در واقع، همین حالا هم تا حدی با تولید پیراهن‌های ضدچروک که نیاز به اتو ندارند، به این نقطه رسیده‌ایم.

---

۱. این پژوهش توسط اکاترینا هرتوگ و همکارانش با عنوان «مختصصان هوش مصنوعی می‌گویند که ۴۰ درصد از زمان صرف‌شده برای کارهای خانه می‌تواند ظرف ده سال آینده حذف شود» انجام گرفته که در ۲۲ فوریه ۲۰۲۳ در سایت مؤسسهٔ اینترنت آکسفورد منتشر شد.

خودروهای خودران

آخرین مطالعهٔ موردی این فصل به خودروها اختصاص دارد. براساس تفکر نتیجه‌گرا، باید پرسید: فلسفهٔ وجودی خودروها چیست؟ اگر از لذت رانندگی بگذریم، همان‌طور که اغلب افرادی که علاقه‌ای به خودروها ندارند می‌گویند، خودروها وسیله‌ای هستند برای رسیدن از نقطهٔ الف به نقطهٔ ب. اوبر (Uber) و ویز (Waze) اغلب به‌عنوان نمونه‌های بارزی از قدرت فناوری مطرح می‌شوند. درست است، اما هر دو نمونه‌هایی از خودکارسازی هستند؛ یکی در رزرو تاکسی و دیگری در برنامه‌ریزی مسیر. اولی فرایند یافتن، رزرو و پرداخت برای خدمات تاکسی را کامپیوتری کرده و به‌این‌ترتیب، آن را ساده‌تر می‌کند و دومی فرایندی را بهینه می‌کند که سریع‌ترین مسیر برای هر سفر تعیین شود و جای نقشه‌ها یا حتی دانش رانندگان تاکسی را می‌گیرد.

در مقابل، خودروهای کاملاً خودران به‌هیچ‌وجه نمونه‌ای از خودکارسازی و جایگزینی وظایف نیستند. همان‌طور که در فصل ۴ اشاره شد، تا جایی که من می‌دانم، هیچ‌کس ادعا نمی‌کند که بهترین راه برای توسعهٔ خودروهای خودران این است که ربات‌هایی بسازیم که روی صندلی رانندهٔ یک خودروی معمولی بنشینند و به‌نحوی از رانندگان انسان با دست‌هایی روی فرمان و پاهایی روی پدال‌ها تقلید کنند. خودروهای خودران یک نوآوری خیره‌کننده‌اند (نگاه کنید به اولین تجربهٔ من که در فصل ۸ شرح داده‌ام). آن‌ها نه توسط ربات‌ها، بلکه توسط داده‌ها هدایت می‌شوند. خودروهای خودران به‌صورت بنیادین فرایند دستیابی به نتایج حمل‌ونقل را تغییر خواهند داد. هرچند با کمی اغراق می‌توان گفت که این هم نوعی از خودکارسازی است، اما برای کسانی که امروز نوجوان هستند یا توانایی رانندگی ندارند، چنین تعبیری صادق نیست. برای آن‌ها، این فناوری امکان شکلی از جابجایی را فراهم می‌کند که در گذشته غیرممکن بود. (ذره‌ای تردیدی ندارم که نوه‌های ما روزی حیرت‌زده سر تکان خواهند داد و خواهند گفت: «باورکردنی نیست که انسان‌ها زمانی خودروها را می‌راندند!»)

این فناوری برای بسیاری از افراد به معنای از دست دادن شغل خواهد بود، اما این پیامد، منظور اصلی من از «حذف» نیست. البته همان‌طور که اشاره کردم، برخی

فناوری‌ها واقعاً نیاز به حضور انسان را از بین می‌برند. اما آنچه تحت عنوان «حذف» در مورد خودروهای خودران مدنظرم است، مربوط به رانندگان نیست، بلکه مربوط به هدفی است که خودروها برای آن طراحی شده‌اند، یعنی مسئلهٔ رفتن از نقطهٔ الف به نقطهٔ ب. فناوری حذف‌کننده در این زمینه درحال‌حاضر جلسات ویدیویی است و در آینده چیزی شبیه به حضور سه‌بعدی تولیدشده توسط هوش مصنوعی در نوعی واقعیت مجازی خواهد بود. درآن‌صورت، اصلاً نیازی به سفر کردن وجود نخواهد داشت.

در پایان این بحث لازم می‌دانم دربارهٔ تمایز سه‌گانهٔ بین خودکارسازی، نوآوری و حذف، دو نکته را مطرح کنم. نخست آنکه نوآوری و حذف، وابستگی بسیار بیشتری نسبت به خودکارسازی ایجاد می‌کنند. اگر صرفاً روش قدیمی انجام کارها را خودکار کرده باشیم، معمولاً در صورتِ بروز خرابی یا ازکارافتادگی سیستم، می‌توانیم به سیستم‌های قدیمی که غالباً کاغذی، آنالوگ و دستی هستند، بازگردیم. در مقابل، وقتی فعالیت، خدمت یا سازمان را با نوآوری یا حذف متحول کرده باشیم، به‌ناچار فرایندها یا روش‌های کاملاً جدیدی معرفی کرده‌ایم که هیچ معادل سنتی‌ای برای بازگشت به آن وجود نخواهد داشت. بنابراین، وابستگی ما به نوآوری و حذف، هزینه در پی دارد. درصورتِ از کار افتادن این سیستم‌ها، اگر برنامه‌های جایگزین نداشته باشیم، به‌شدت در معرض خطر خواهیم بود.

نکتهٔ دوم این است که کاربردهای تحول‌آفرین و حذف‌کنندهٔ فناوری، اغلب به دانش و تخصص سنتی انسان‌ها متکی نیستند یا از آن پیروی نمی‌کنند. در عوض، این فناوری‌ها به داده‌ها ـ معمولاً در مقیاس کلان ـ تکیه دارند. سیستم مبتنی بر هوش مصنوعی که اغلب از آن نام برده می‌شود و در تشخیص ملانوم بدخیم عملکرد چشمگیری دارد (به فصل ۵ مراجعه کنید)، هیچ درک یا دانش پایه‌ای از تخصص پوست یا آسیب‌شناسی بافتی ندارد؛ بلکه تنها تصاویر قبلی از ضایعات پوستی را همراه با نتایج آزمایشگاهی در اختیار دارد.

به‌طور خلاصه، اینکه مانند اکثر مردم، هوش مصنوعی را صرفاً ابزاری برای خودکارسازی وظایف و فعالیت‌های امروزی در نظر بگیریم، به این معناست که تأثیرات کنونی و آیندهٔ آن را به‌شدت دست‌کم گرفته‌ایم.

انتشارات
راه پرداخت

[درباره انتشارات راه پرداخت]

ما عاشق فناوری و نوآوری هستیم؛ تلاش می‌کنیم تأثیر فناوری و نوآوری بر زندگی و کار را درک کنیم و یافته‌هایمان را با دیگران به اشتراک می‌گذاریم. در انتشارات راه پرداخت تلاش می‌کنیم ناشر کتاب‌های برتر فناوری و نوآوری جهان باشیم؛ به همین دلیل نیز بهترین کتاب‌های ناشران حرفه‌ای جهان را به‌دقت بررسی، انتخاب و آماده می‌کنیم. با اینکه کتاب‌های ما در حوزه‌های رمزارز و بلاکچین بیشتر شناخته شده و ممکن است بسیاری ما را بیشتر با فعالیت‌های رسانه‌ای و آموزشی در حوزه فین‌تک و فناوری‌های مالی بشناسند، ولی ما به فراتر از فین‌تک می‌اندیشیم و تلاش می‌کنیم بر اقتصاد نوآوری و اقتصاد دیجیتال ایران اثرگذار باشیم.

چرا کتاب؟ چون کتاب کامل‌ترین رسانه برای انتقال پیام است. کتاب کارکرد آموزش، اطلاع‌رسانی و سرگرم‌کنندگی را به‌صورت هم‌زمان دارد. کتاب شروع، اوج و پایان دارد. کتاب خوب مدل ذهنی منسجم دارد و آشفتگی انسان را کاهش می‌دهد. هیچ‌کدام از ابزارهای ارتباطی جدید و سنتی توان بالای کتاب در آموزش را ندارند؛ بنابراین کتاب بهترین ابزار خلق‌شده توسط بشر برای انتقال مفاهیم است.

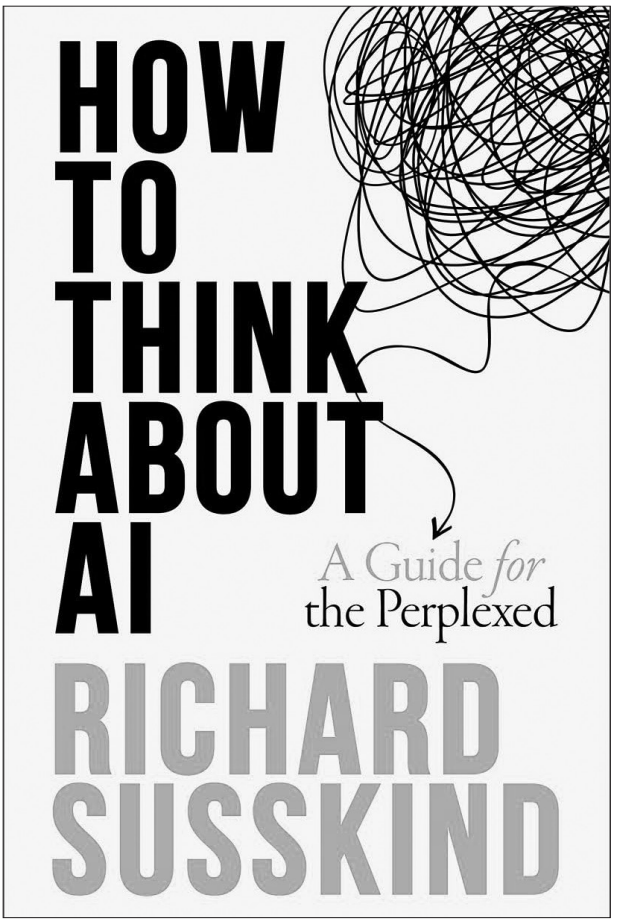
زمانی به این فکر می‌کردیم که چگونه می‌توانیم حداقل ۱۰۰ کتاب در زمینه فناوری و نوآوری منتشر کنیم، اما وقتی از این تعداد گذر کردیم، حواس‌مان نبود! تمرکز ما بر انتشار کتاب‌های باکیفیت و خواندنی در حوزه‌های فناوری و نوآوری بود و حساب کار از دست ما در رفته بود. امروز به این فکر می‌کنیم که در پنج سال آینده بیش از هزار کتاب در زمینه فناوری و نوآوری منتشر کنیم.

ما در انتشارات راه پرداخت تلاش می‌کنیم برای انتشار کتاب‌ها حمایت بخش خصوصی را در کنار خود داشته باشیم. به همین منظور صمیمانه از تمام کسب‌وکارهایی که در این سال‌ها حامی انتشار کتاب‌های راه پرداخت بوده‌اند، تشکر می‌کنیم. ما قدردان کسب‌وکارهایی هستیم که در مسیر خود انتخاب کردند از انتشار دانش حمایت کنند. امیدواریم بتوانیم این راه را قوی‌تر ادامه دهیم.

تمرکز ما در انتشارات انتخاب بهترین کتاب‌های بهترین ناشران دنیا و ترجمه آنهاست. تلاش می‌کنیم بهترین‌ها را انتخاب کنیم و با دقت بالایی متن‌ها را ترجمه و در اختیار فارسی‌زبانان قرار دهیم.

در صورت تمایل برای مشارکت در تولید کتاب‌ها، می‌توانید با ما تماس بگیرید. همچنین می‌توانید هر کدام از کتاب‌های منتشرشده راه پرداخت را انتخاب و حامی انتشار چاپ جدیدی از آن شوید.

انتشارات راه پرداخت بخشی از کارخانه نوآوری رسانه راه‌کار است. رسانه‌های راه‌کار در زمینه فناوری‌های مالی (فین‌تک) شناخته‌شده هستند، ولی ما در راه‌کار به فراتر از فین‌تک می‌اندیشیم. تمرکز راه‌کار تولید و توزیع محتوای اثربخش است. نخ تسبیح تمام فعالیت‌های راه‌کار هم پنج واژه است؛ فناوری، نوآوری، ایده، تفکر طراحی و کسب‌وکار. راه‌کار در این سال‌ها تلاش کرده صدای بخش خصوصی اقتصاد ایران باشد. تلاش کردیم روندها و رویدادهای مرتبط با اقتصاد نوآوری و اقتصاد دیجیتال ایران را پوشش دهیم و امیدواریم در این راه سخت، سخت‌جان بمانیم.



چگونه دربارهٔ هوش مصنوعی بیندیشیم

در سال‌های اخیر، به‌ویژه از زمان راه‌اندازی چت‌جی‌پی‌تی، علاقهٔ عموم مردم و متخصصان به هوش مصنوعی چند برابر شده است. بااین‌حال، بسیاری هنوز دقیقاً نمی‌دانند هوش مصنوعی چیست، چه کارهایی از عهده‌اش برمی‌آید و در چه زمینه‌هایی ناتوان است. آیا هوش مصنوعی پدیده‌ای سودمند برای انسان و تمدن بشری است؟ آیا می‌تواند راه‌حل‌هایی برای مسائل بنیادین ما ارائه دهد یا تهدیدی وجودی برای انسان‌ها به شمار می‌آید؟ در کنار این پرسش‌ها، همچنان دربارهٔ چگونگی قانون‌گذاری برای استفاده از هوش مصنوعی و تعیین مرزهای اخلاقی آن، ابهام‌های فراوانی وجود دارد. ریچارد ساسکیند، نویسنده و سخنران بریتانیایی، در کتاب چگونه دربارهٔ هوش مصنوعی بیندیشیم با تکیه بر تجربهٔ چند دهه‌ای خود در مطالعه و کار بر روی هوش مصنوعی از اوایل دههٔ ۱۹۸۰، می‌کوشد به این پرسش‌ها پاسخ دهد. او در این کتاب، ضمن بررسی تاریخچهٔ هوش مصنوعی و سناریوهای محتمل آیندهٔ آن، تأکید می‌کند که برقراری تعادل میان مزایا و تهدیدهای هوش مصنوعی، چالش بنیادین دوران ماست.

ریچارد ساسکیند با روایت داستانِ غیرقابل پیش‌بینی هوش مصنوعی، در کتاب چگونه دربارهٔ هوش مصنوعی بیندیشیم راهنمایی کوتاه و غیرتخصصی ارائه می‌دهد که ما را به نگاهی تازه و متفاوت به این پدیده دعوت می‌کند. او هوش مصنوعی را از آزمایشگاه‌های رایانه‌ای، شرکت‌های بزرگ فناوری و استارتاپ‌ها بیرون می‌آورد و آن را به قلب زندگی روزمره می‌برد.

ساسکیند در کتاب چگونه دربارهٔ هوش مصنوعی بیندیشیم با تکیه بر تجربهٔ خود در زمینهٔ کار بر هوش مصنوعی از اوایل دههٔ ۱۹۸۰، تاریخچهٔ هوش مصنوعی را مرور می‌کند و به بررسی سناریوهای محتمل آیندهٔ آن می‌پردازد. دیدگاه‌های ساسکیند دربارهٔ هوش مصنوعی همواره مطابق با جریان غالب نیست. او چت‌جی‌پی‌تی و دیگر فناوری‌های هوش مصنوعی مولد را صرفاً فصلی تازه در داستان ادامه‌دارِ هوش مصنوعی می‌داند و معتقد است هنوز در آغاز راهیم. به باور او، برای اندیشیدن مسئولانه دربارهٔ تأثیر هوش مصنوعی باید فراتر از فناوری‌های امروز نگاه کنیم؛ چراکه در دههٔ ۲۰۳۰، فناوری‌هایی که هنوز اختراع نشده‌اند، تأثیری به‌مراتب عمیق‌تر از ابزارهای کنونی بر زندگی انسان خواهند گذاشت.

انتشارات راه پرداخت
ناشر فناوری و نوآوری
way2pay.press

ISBN 978-622-4958-03-7
۴۵۹ هزار تومان