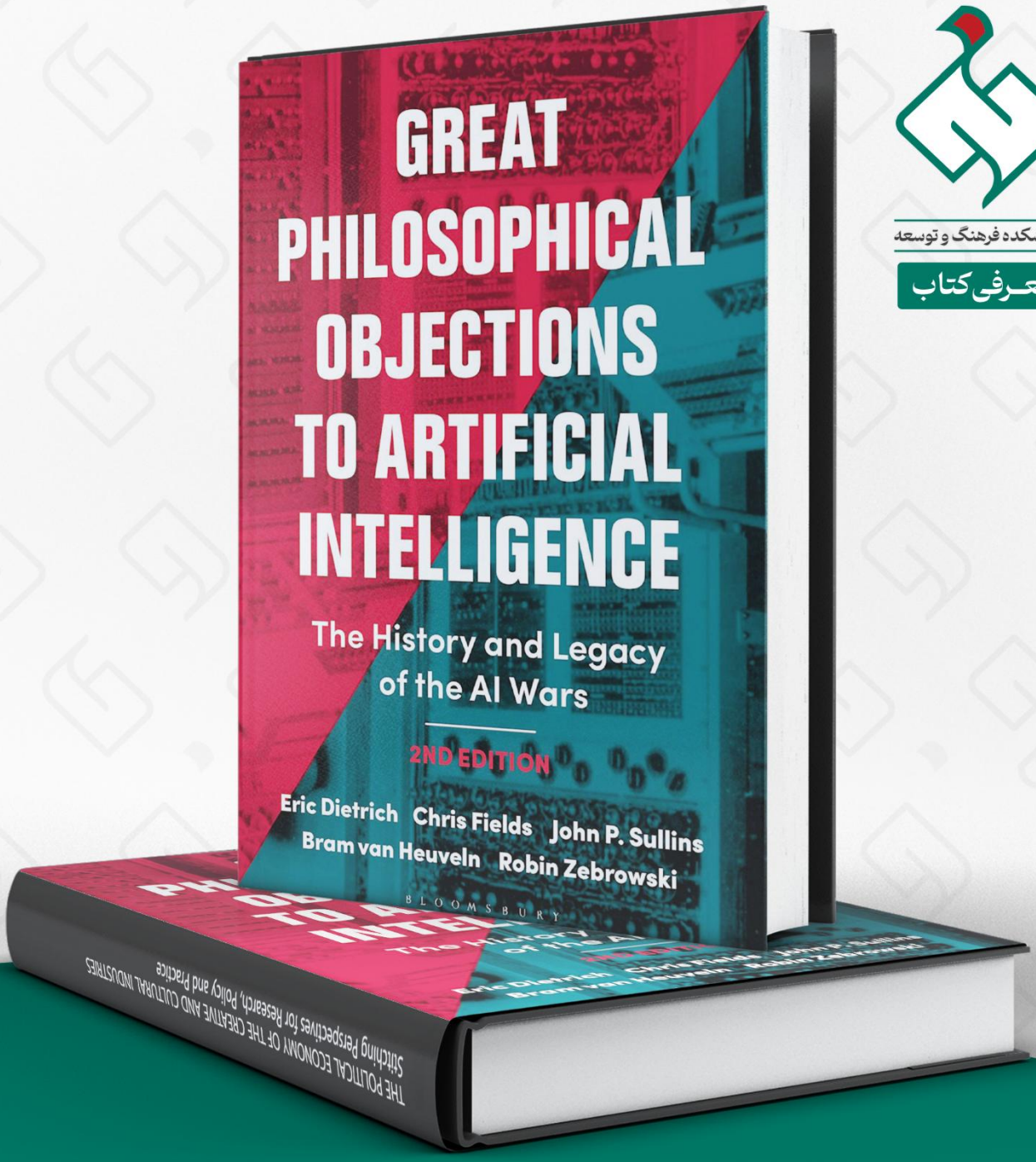
اندیشکده فرهنگ و توسعه

معرفی کتاب



ایرادات فلسفی بزرگ به هوش مصنوعی

تاریخچه و میراث جنگ‌های هوش مصنوعی

زبروسکی، هولن، سالینز، فیلدز، دیتریش | Bloomsbury | ۲۰۲۶

چکیده:

این کتاب مقدمه‌ای سرگرم‌کننده بر رویارویی‌های میان فلسفه و هوش مصنوعی در طول ۷۰ سال گذشته است؛ از ادعاها و ادعاهای متقابل درباره امکان پیاده‌سازی آگاهی گرفته تا استدلال‌هایی درباره معماری شناختی و .ChatGPT این اثر که در ویرایش جدید خود با دربرگرفتن تازه‌ترین تحولات هوش مصنوعی به‌روزرسانی شده، به بررسی مشهورترین استدلال‌های فلسفی علیه ساخت ماشینی با هوش در سطح انسان می‌پردازد و این استدلال‌ها را در قالب چهار جنگ اصلی هوش مصنوعی سازمان‌دهی می‌کند تا نشان دهد چگونه این مناظره میان فیلسوفان، دانشمندان هوش مصنوعی و مهندسان سازنده سامانه‌های هوش مصنوعی شکل گرفته است. کتاب همچنین راهنمایی برای درک پرسش‌های فلسفی مهم و نقدها و حملاتی است که هوش مصنوعی در طول تاریخ خود با آن‌ها مواجه شده و ویرایش دوم آن، با ارائه بینش‌های تازه و مطالب پشتیبان، محتوای جدیدی درباره مدل‌های یادگیری زبان LLM)ها (و وجود هوش مصنوعی مولد، هوش مصنوعی پایدار و توانایی آن در تنظیم اقلیم، مسائل نظری، اخلاقی و قانون‌گذاری پیرامون «خلاقیت محاسباتی»، و اثر دره وهم‌آلود و پیامدهای بالقوه آن برای هوش مصنوعی ارائه می‌کند. آیا در آستانه یک جنگ جدید هوش مصنوعی قرار داریم؟ این مقدمه برای هر کسی است که می‌خواهد مناظراتی را که فلسفه هوش مصنوعی را شکل داده‌اند و استدلال‌هایی را که آینده آن را تعریف خواهند کرد، درک کند؛ اثری که نشان می‌دهد هوش مصنوعی از زمان اختراعش در دهه ۱۹۵۰ چه مسیری را طی کرده است و چگونه بارها و بارها ما را به پرسش‌های فلسفی‌ای بازمی‌گرداند که انسان‌ها همواره با آن‌ها مواجه بوده‌اند: پرسش‌هایی درباره دانش، معنا و اینکه چگونه باید با یکدیگر و با باقی جهان رفتار کنیم.

مقدمه

بخش نخست کتاب با عنوان «جنگ‌های هوش مصنوعی» به تاریخ مناقشات فلسفی درباره امکان ساخت ماشین‌هایی با هوش در سطح انسان، از دهه ۱۹۵۰ تا حدود سال ۲۰۰۰ می‌پردازد. نویسندگان توضیح می‌دهند که نخستین کنفرانس هوش مصنوعی در سال ۱۹۵۶ در دارتموث برگزار شد و در همان سال یکی از نخستین برنامه‌های هوش مصنوعی، یعنی Logic Theorist، ساخته شد. با این حال، تنها چند سال بعد، پروژه ساخت هوش مصنوعی در سطح انسان با مجموعه‌ای از حملات فلسفی روبه‌رو شد. این حملات محدود به یک مسئله نبودند، بلکه محدودیت‌های محاسبات، امکان اندیشیدن ماشین به اشیای مشخص، آگاهی، عقلانیت، اخلاق و معماری شناختی را دربرمی‌گرفتند. نویسندگان تأکید می‌کنند که این مناقشه میان فیلسوفان مخالف و موافق هوش مصنوعی و پژوهشگران این حوزه شکل گرفت و هر دو طرف در بسیاری موارد از پیش فرض‌های نادرستی درباره حوزه مقابل برخوردار بودند.

نویسندگان این مناقشات را در قالب چهار جنگ اصلی سازمان‌دهی می‌کنند. جنگ نخست درباره محدودیت‌های منطقی رایانه‌ها و این پرسش است که آیا اساساً هوش مصنوعی ممکن است یا نه. جنگ دوم به این مسئله می‌پردازد که اگر ساخت هوش ماشینی ممکن باشد، چه معماری‌ای می‌تواند آن را محقق کند. جنگ سوم درباره امکان معنا داشتن و اندیشیدن رایانه به چیزی است و جنگ چهارم مسئله عقلانیت، ارتباط میان امور و خلاقیت را بررسی می‌کند. نکته مهم این است که هیچ یک از این جنگ‌ها با پیروزی قطعی یک طرف پایان نیافتند و بیشتر به نوعی بن بست رسیدند. در عین حال، شکست موج نخست هوش مصنوعی نشان داد پیچیدگی مغز و فرایندهای فکری انسان بسیار بیشتر از آن چیزی است که پژوهشگران اولیه تصور می‌کردند. در همین زمان، هوش مصنوعی به تدریج در قالبی متفاوت پیشرفت کرد و پرسش‌های فلسفی تازه‌ای را پدید آورد.

فصل ۱: گودل و یک اعتراض بنیادین به هوش مصنوعی

فصل نخست به نخستین حمله فلسفی مهم علیه امکان هوش مصنوعی اختصاص دارد؛ حمله‌ای که جان راندولف لوکاس در سال ۱۹۵۹ و بر اساس قضیه ناتمامیت کورت گودل مطرح کرد. گودل نشان داده بود که در هر نظام صوری سازگار و به اندازه کافی قدرتمند برای بیان حساب ساده، گزاره‌هایی

وجود دارند که درون همان نظام قابل اثبات نیستند، هرچند از بیرون نظام می‌توان درستی آن‌ها را تشخیص داد. لوکاس از این نتیجه برای دفاع از این ادعا استفاده کرد که ذهن انسان نمی‌تواند صرفاً یک ماشین یا نظام محاسباتی باشد. استدلال او این بود که رایانه‌ها، به عنوان نظام‌های صوری، با محدودیت‌های گودلی مواجه‌اند، در حالی که انسان می‌تواند حقیقت برخی گزاره‌هایی را که برای آن نظام اثبات‌ناپذیرند دریابد. بنابراین، اگر ذهن انسان بتواند چیزی را بفهمد که هیچ ماشین صوری قادر به اثبات آن نیست، ذهن را نمی‌توان با یک ماشین توضیح داد.

نویسندگان سپس نشان می‌دهند که این استدلال به سادگی قابل پذیرش نیست و باید میان آنچه قضیه گودل واقعاً اثبات می‌کند و نتیجه‌ای که لوکاس از آن می‌گیرد تمایز گذاشت. اثبات گودل درباره یک نظام صوری از سطح فرازبان انجام می‌شود و خودِ این اثبات لزوماً در همان نظام صوری قرار ندارد. بنابراین، این واقعیت که ما از بیرون یک نظام می‌توانیم درباره گزاره‌ای گودلی داوری کنیم، به خودی خود نشان نمی‌دهد که انسان‌ها واجد توانایی شناختی‌ای هستند که هیچ ماشین محاسباتی نمی‌تواند داشته باشد. مسئله اساسی فصل این است که قضیه ناتمامیت گودل محدودیت‌هایی واقعی برای نظام‌های صوری ایجاد می‌کند، اما از این محدودیت‌ها نمی‌توان مستقیماً ناممکن بودن هوش مصنوعی را نتیجه گرفت. به این ترتیب، فصل نخست نشان می‌دهد که یکی از بنیادی‌ترین اعتراض‌های فلسفی به هوش مصنوعی چگونه از یک نتیجه مهم ریاضی آغاز شد، اما برای تبدیل شدن به استدلالی علیه ماشین باید چندین گام فلسفی اضافی برداشته شود.

فصل ۲: چگونه بفهمیم یک رایانه هوشمند است؟ آزمون تورینگ پاسخ نیست

فصل دوم به یکی از مشهورترین ایده‌های تاریخ هوش مصنوعی، یعنی آزمون تورینگ، می‌پردازد و می‌پرسد آیا این آزمون واقعاً می‌تواند مشخص کند که یک رایانه هوشمند است یا نه. نویسندگان استدلال می‌کنند که برداشت رایج از آزمون تورینگ با آنچه آلن تورینگ در نظر داشت تفاوت دارد. تورینگ لزوماً آزمون خود را به عنوان یک ابزار علمی دقیق برای اندازه‌گیری هوش ارائه نکرده بود، بلکه بازی تقلید را بیشتر برای مقابله با پیش فرض‌هایی مطرح کرد که انسان‌ها درباره ماهیت هوش و امکان هوشمند بودن ماشین‌ها دارند. از این دیدگاه، این نکته که یک ماشین از فلز و پلاستیک ساخته شده است نباید به خودی خود دلیلی برای انکار هوش آن باشد، همان گونه که پرواز هواپیما را به دلیل نداشتن بال‌های متحرک، پرواز واقعی نمی‌دانیم. مسئله اصلی رفتار و ویژگی‌هایی است که برای نسبت دادن هوش اهمیت دارند، نه ماده‌ای که یک موجود از آن ساخته شده است.

نویسندگان همچنین نشان می‌دهند که آزمون تورینگ با مشکل ابهام مفهوم هوش مواجه است. اگر برای قبولی در آزمون معیارهای بسیار دقیق تعیین کنیم، این معیارها تا حدی قراردادی و دلخواه خواهند

بود، زیرا هوش انسانی خود مفهومی دقیق و کاملاً تعریف شده نیست. از سوی دیگر، اگر ماشینی بتواند انسان‌ها را در یک گفت‌وگو فریب دهد، هنوز می‌توان درباره اینکه آیا واقعاً هوشمند است یا صرفاً رفتاری شبیه انسان نشان می‌دهد بحث کرد. به همین دلیل، نویسندگان آزمون تورینگ را بیشتر یک آزمایش فکری می‌دانند که پیش داوری‌های ما درباره هوش ماشینی را آشکار می‌کند. تورینگ همچنین به اعتراض‌هایی درباره آگاهی، خلاقیت، یادگیری، خطا کردن و انجام امور جدید پاسخ می‌دهد و تأکید می‌کند که نباید توانایی‌های ماشین‌های موجود را به همه ماشین‌های ممکن تعمیم داد. در نهایت، اهمیت آزمون تورینگ نه در ارائه تعریفی قطعی از هوش، بلکه در واداشتن ما به داوری منصفانه و سازگار درباره هوش ماشین‌هاست.

فصل ۳: چگونه علوم رایانه ذهن را نجات داد

فصل سوم با تاریخچه‌ای از حذف مفهوم ذهن در روان‌شناسی و بازگشت دوباره آن آغاز می‌شود. نویسندگان توضیح می‌دهند که تجربه گرایی منطقی و سپس رفتارگرایی، ذهن و حالات درونی آن را از موضوع اصلی علم کنار گذاشتند. رفتارگرایی، به ویژه، روان‌شناسی را علم رفتار می‌دانست و معتقد بود رفتار را می‌توان بدون ارجاع بنیادی به رویدادهای ذهنی یا فرایندهای درونی توضیح داد. در این چارچوب، مفاهیمی مانند باور، میل، احساس، فکر و تجربه ذهنی از توضیح علمی رفتار کنار گذاشته شدند، زیرا مستقیماً قابل اندازه‌گیری نبودند. اما این وضعیت با ظهور رایانه تغییر کرد. رایانه‌ها نشان دادند که یک سامانه می‌تواند رفتاری پیچیده داشته باشد و این رفتار بر فرایندهای درونی آن متکی باشد. بنابراین، رفتار دیگر الزاماً تنها نتیجه رابطه میان محرک بیرونی و پاسخ قابل مشاهده نبود. رایانه امکان داد دوباره درباره فرایندهای درونی و نقش آن‌ها در تولید رفتار به شیوه‌ای علمی‌تر فکر شود.

در ادامه، فصل مفهوم الگوریتم و ماشین مجازی را برای توضیح این بازگشت ذهن به علم بررسی می‌کند. الگوریتم مجموعه‌ای متناهی از قواعد دقیق و بدون ابهام برای انجام یک فرایند است و علوم شناختی و هوش مصنوعی اولیه به این ایده متکی شدند که تفکر را می‌توان به اجرای الگوریتم‌ها مربوط دانست. مفهوم ماشین مجازی نیز نشان می‌دهد که یک فرایند محاسباتی را می‌توان در سطحی متفاوت از سخت‌افزاری که آن را اجرا می‌کند توصیف کرد؛ برای مثال، یک برنامه مانند Word می‌تواند مستقل از جزئیات سخت‌افزار رایانه‌ای که روی آن اجرا می‌شود بررسی شود. از این منظر، علوم رایانه امکان داد ذهن نه به عنوان یک امر اسرارآمیز، بلکه به عنوان سطحی از فرایندهای سازمان یافته و قابل مطالعه دوباره وارد بحث علمی شود. این تحول زمینه را برای شکل‌گیری دیدگاه محاسباتی درباره ذهن و سپس برای تلاش جهت ساخت ماشین‌هایی با هوش انسانی فراهم کرد.

فصل ۴: پیاده‌سازی یک هوش

فصل چهارم به پرسش محوری جنگ دوم می‌پردازد: اگر هوش انسانی را بتوان از نظر محاسباتی توضیح داد، چه نوع معماری‌ای می‌تواند آن را در یک ماشین پیاده کند؟ نویسندگان در تاریخ هوش مصنوعی چهار معماری اصلی را از یکدیگر متمایز می‌کنند: پردازشگرهای نمادین، شبکه‌های اتصال‌گرا و شبکه‌های عصبی مصنوعی، شناخت تجسم‌یافته و موقعیت‌مند، و سامانه‌های پویا. پردازشگرهای نمادین نخستین معماری غالب بودند و بر این ایده استوار بودند که رایانه و ذهن هر دو با نمادها کار می‌کنند. این دیدگاه در شکل‌گیری هوش مصنوعی اولیه نقشی اساسی داشت و به این تصور انجامید که می‌توان فرایندهای ذهنی را با قواعد صوری و دستکاری نمادها بازسازی کرد. با این حال، مشکل اصلی این معماری آن بود که نمادهای مورد استفاده ماشین عمدتاً به وسیله برنامه‌نویس و از پیش در سیستم وارد می‌شدند. چنین سیستمی در موقعیت‌های از پیش تعریف نشده، یادگیری، انعطاف پذیری و مواجهه با شرایط جدید با دشواری روبه رو می‌شد.

در ادامه، نویسندگان سه رویکرد رقیب را بررسی می‌کنند که هر یک کوشیدند محدودیت‌های معماری نمادین را برطرف کنند. شبکه‌های عصبی مصنوعی به جای اتکا به قواعد صریح، بر یادگیری الگوها از طریق شبکه‌ای از واحدهای به هم پیوسته تکیه دارند و به این ترتیب، بازنمایی‌ها می‌توانند از فرایند یادگیری پدیدار شوند. رویکرد شناخت تجسم‌یافته و موقعیت‌مند نیز بر نقش بدن، ادراک و تعامل مستقیم با محیط تأکید می‌کند و در برخی صورت‌بندی‌های آن، مرز میان ادراک و عمل بسیار کمرنگ می‌شود. سامانه‌های پویا نیز شناخت را بر اساس تغییرات زمانی و روابط میان متغیرهای یک سیستم توضیح می‌دهند. نویسندگان نشان می‌دهند که هیچ یک از این معماری‌ها به تنهایی مسئله هوش را به طور کامل حل نکرده است. در نتیجه، مسئله اصلی همچنان باقی می‌ماند: برای ساخت ماشینی واقعاً هوشمند، چگونه باید نمادها، یادگیری، ادراک، بدن، محیط و استدلال را در یک معماری واحد به هم پیوند داد؟

فصل ۵: مورد عجیب معنای گمشده: آیا رایانه‌ها می‌توانند درباره چیزها فکر کنند؟

فصل پنجم وارد سومین جنگ هوش مصنوعی می‌شود و مسئله «درباره‌بودگی» یا توانایی ذهن برای معطوف بودن به چیزها را در مرکز بحث قرار می‌دهد. مسئله این است که انسان‌ها فقط اطلاعات را پردازش نمی‌کنند، بلکه افکار و باورهایشان درباره اشیا، افراد، رویدادها و جهان هستند. رایانه نیز می‌تواند نمادها را طبق قواعدی مشخص دستکاری کند، اما آیا نمادهای آن برای خود رایانه نیز معنایی دارند؟ نویسندگان این مسئله را با بررسی استدلال اتاق چینی جان سرل پیگیری می‌کنند. در این استدلال، فردی که زبان چینی نمی‌داند می‌تواند با پیروی از قواعدی صوری، پاسخ‌هایی مناسب به پرسش‌های چینی تولید کند، بدون آنکه بداند این پاسخ‌ها چه معنایی دارند. از دیدگاه سرل، این

وضعیت نشان می‌دهد که دستکاری نحوی نمادها به خودی خود برای ایجاد معنا یا فهم کافی نیست. بنابراین، ممکن است یک ماشین رفتاری هوشمندانه داشته باشد، اما همچنان فاقد معنای ذهنی و فهم واقعی باشد.

نویسندگان سپس راه‌های مختلفی را برای حل مسئله معنا بررسی می‌کنند و به ویژه به ایده ساخت یک ربات متحرک و یادگیرنده می‌پردازند که از طریق حسگرها و عملگرهای خود مستقیماً با جهان تعامل کند. در این نگاه، معنا دیگر چیزی نیست که صرفاً از طریق برنامه نویسی نمادها به ماشین داده شود، بلکه می‌تواند از تعامل مستمر ماشین با محیط شکل بگیرد. با این حال، فصل نشان می‌دهد که حتی این راه حل نیز پرسش‌های تازه‌ای ایجاد می‌کند، زیرا باید توضیح داد چگونه تعامل با جهان به «درباره چیزی بودن» تبدیل می‌شود و چگونه درباره‌بودگی با آگاهی ارتباط پیدا می‌کند. نویسندگان در نهایت تأکید می‌کنند که یکی از دشواری‌های اصلی این جنگ آن است که خود مفهوم درباره‌بودگی تعریف مورد توافقی ندارد. در نتیجه، بخشی از مناقشه درباره اینکه آیا رایانه می‌تواند فکر کند، بدون تعریف روشن از معنایی که قرار است ماشین داشته باشد، پیش رفته است.

فصل ۶: چه چیزی به چه چیزی مربوط است؟ مسئله چارچوب

فصل ششم به چهارمین جنگ هوش مصنوعی و مسئله چارچوب اختصاص دارد؛ مسئله‌ای که در ظاهر فنی است، اما پیامدهای عمیق فلسفی درباره عقلانیت، ارتباط و خلاقیت دارد. یک عامل هوشمند در جهان واقعی باید بتواند تشخیص دهد کدام بخش از اطلاعات موجود برای یک موقعیت مشخص اهمیت دارد و کدام بخش را می‌توان نادیده گرفت. برای مثال، وقتی عاملی عملی انجام می‌دهد، لازم نیست همه چیزهایی را که در جهان تغییر نکرده‌اند دوباره بررسی کند. اما توضیح اینکه ماشین چگونه می‌تواند به شکلی اقتصادی و اصولی تشخیص دهد چه چیزهایی مرتبط‌اند و چه چیزهایی نیستند، بسیار دشوار است. نسخه فنی مسئله چارچوب در هوش مصنوعی عمدتاً به مسئله نمایش تغییرات ناشی از اعمال در سامانه‌های منطقی مربوط می‌شود و پژوهشگران راه حل‌هایی برای آن پیشنهاد کرده‌اند. با این حال، نویسندگان معتقدند مسئله عمیق‌تر، یعنی مسئله فلسفی مربوط به تشخیص ارتباط و اهمیت، همچنان باقی است.

از دیدگاه نویسندگان، مسئله چارچوب فقط مشکل یک شیوه خاص از برنامه نویسی نیست، بلکه به یکی از ویژگی‌های اساسی هوش انسانی مربوط می‌شود. انسان‌ها در موقعیت‌های مختلف، بدون اینکه لازم باشد فهرستی از تمام امور نامرتبط را بررسی کنند، معمولاً تشخیص می‌دهند چه چیزی اهمیت دارد. این توانایی با عقلانیت، یادگیری، استدلال، قیاس و حتی خلاقیت ارتباط پیدا می‌کند. به همین دلیل، نویسندگان استدلال می‌کنند که نادیده گرفتن مسئله فلسفی چارچوب می‌تواند هوش مصنوعی

را به نوعی «تک هوشی» محدود کند که در یک حوزه مشخص عملکرد خوبی دارد اما توانایی انتقال هوشمندانه دانش میان موقعیت‌های متفاوت را ندارد. مسئله چارچوب از این منظر نشان می‌دهد که هوش صرفاً توانایی محاسبه یا پیش بینی نیست، بلکه مستلزم فهم این است که در هر موقعیت چه چیزی اهمیت دارد. انسان‌ها نیز همیشه در حل این مسئله موفق نیستند، اما ظاهراً سازوکارهایی برای مدیریت آن در اختیار دارند.

فصل ۷: پس تکلیف آگاهی چه می‌شود؟

فصل هفتم پس از چهار جنگ تاریخی هوش مصنوعی، به یکی از دشوارترین مسائل فلسفه ذهن می‌پردازد: آیا یک ماشین می‌تواند واقعاً آگاه باشد؟ نویسندگان ابتدا میان توانایی انجام فرایندهای شناختی و تجربه آگاهانه تمایز می‌گذارند. ممکن است یک سامانه بتواند اطلاعات را پردازش کند، زبان تولید کند، مسئله حل کند یا رفتاری شبیه انسان داشته باشد، بدون اینکه تجربه‌ای درونی داشته باشد. در اینجا مسئله «زامبی فلسفی» اهمیت پیدا می‌کند، یعنی موجودی که از بیرون کاملاً شبیه انسان آگاه رفتار می‌کند، اما هیچ تجربه ذهنی و پدیداری ندارد. این بحث نشان می‌دهد که حتی اگر بتوانیم تمام کارکردهای ذهن را به صورت محاسباتی توضیح دهیم، هنوز پرسش دشوار درباره اینکه چرا و چگونه این فرایندها با تجربه آگاهانه همراه می‌شوند، حل نشده است. از همین رو، موفقیت هوش مصنوعی در شبیه سازی رفتار هوشمند الزاماً به معنای حل مسئله آگاهی نیست.

نویسندگان سپس به مشکلات مفهومی زبان و نظریه‌های مختلف آگاهی می‌پردازند و نشان می‌دهند که حتی درباره خود مفهوم آگاهی نیز توافق روشنی وجود ندارد. بخشی از مناقشه به این پرسش مربوط است که آیا آگاهی چیزی جدا از فرایندهای شناختی است یا اینکه می‌توان آن را در قالب همان فرایندها توضیح داد. در این میان، خودکاربودن بسیاری از رفتارهای انسانی نیز مسئله را پیچیده‌تر می‌کند، زیرا انسان همیشه برای انجام یک کار آگاهانه تصمیم نمی‌گیرد و بسیاری از فعالیت‌های ذهنی بدون توجه آگاهانه انجام می‌شوند. فصل در نهایت نشان می‌دهد که «جنگ آگاهی» صرفاً درباره این نیست که آیا می‌توان آگاهی را در یک ماشین ایجاد کرد، بلکه درباره این است که اساساً چه چیزی را آگاهی می‌نامیم و چه شواهدی می‌تواند وجود آن را در یک موجود دیگر، چه انسان و چه ماشین، نشان دهد. به همین دلیل، حتی ماشینی که با دقت بسیار زیادی مانند موجودی آگاه رفتار کند، لزوماً ما را به نتیجه قطعی درباره وجود تجربه درونی آن نمی‌رساند.

فصل ۸: مسائل اخلاقی پیرامون کاربردهای هوش مصنوعی

فصل هشتم بحث را از این پرسش که «آیا ماشین می‌تواند ذهن داشته باشد؟» به پرسشی عملی‌تر منتقل می‌کند: «وقتی هوش مصنوعی در جهان واقعی به کار گرفته می‌شود، چه مسائل اخلاقی ایجاد می‌کند؟» نویسندگان توضیح می‌دهند که گسترش داده‌های بزرگ و یادگیری ماشین باعث شده هوش مصنوعی دیگر محدود به بازی، حل معما و اثبات قضیه نباشد و در حوزه‌هایی وارد شود که تصمیم‌های آن مستقیماً بر زندگی انسان‌ها اثر می‌گذارند. سامانه‌های توصیه گر، کاربردهای پزشکی، خودروهای خودران و سلاح‌های نیمه خودمختار و خودمختار نمونه‌هایی هستند که در آن‌ها خروجی یک سامانه هوش مصنوعی می‌تواند پیامدهای جدی انسانی داشته باشد. بنابراین، اخلاق هوش مصنوعی دیگر صرفاً مسئله‌ای نظری درباره آینده ماشین‌های هوشمند نیست، بلکه به مسئله‌ای مربوط به طراحی، کاربرد و مسئولیت اجتماعی سامانه‌هایی تبدیل شده است که همین امروز در جهان واقعی فعالیت می‌کنند. در چنین شرایطی، پرسش اخلاقی دیگر نمی‌تواند از تصمیم‌های فنی جدا باشد.

نویسندگان در ادامه میان «اخلاق درباره هوش مصنوعی» و مسئله اینکه خود هوش مصنوعی چگونه باید اخلاقی عمل کند تمایز می‌گذارند. یکی از رویکردهای سنتی این بوده است که قواعد اخلاقی مستقیماً در سامانه‌ها برنامه نویسی شوند، گویی می‌توان مجموعه‌ای از قوانین روشن و از پیش تعیین شده را در ماشین قرار داد و مسئله اخلاق را حل کرد. اما این دیدگاه با مشکلات جدی روبه رو است، زیرا اخلاق انسانی فقط پیروی مکانیکی از مجموعه‌ای ثابت از قواعد نیست و در موقعیت‌های واقعی، تشخیص زمینه، پیامد، تعارض ارزش‌ها و شرایط خاص اهمیت زیادی دارد. از سوی دیگر، بسیاری از کاربردهای هوش مصنوعی مسئله مسئولیت را نیز پیچیده می‌کنند، زیرا تصمیم نهایی حاصل تعامل میان طراحان، شرکت‌ها، کاربران، داده‌ها و خود سامانه است. بنابراین، مسئله اخلاق هوش مصنوعی صرفاً این نیست که چگونه یک ماشین را اخلاقی کنیم، بلکه باید بررسی کرد چه کسانی مسئول طراحی و استفاده از آن هستند و چگونه می‌توان از پیامدهای اخلاقی و اجتماعی تصمیم‌های الگوریتمی جلوگیری کرد.

فصل ۹: آیا هوش مصنوعی‌های تجسم یافته می‌توانند عاملان اخلاقی باشند؟

فصل نهم از این ایده آغاز می‌کند که برای بررسی امکان عامل اخلاقی بودن یک ماشین، باید فراتر از سامانه‌های صرفاً نرم افزاری رفت و نقش بدن و محیط را در شناخت و کنش در نظر گرفت. نویسندگان این بحث را به سنت مطالعات شناختی چهارگانه مرتبط می‌کنند که بر ذهن تجسم یافته، درگیرشده با محیط، کنشگر و امتداد یافته تأکید دارد. پرسش اصلی این است که اگر اخلاق انسانی به ذهنی وابسته است که در بدن و محیط اجتماعی قرار دارد، آیا می‌توان انتظار داشت یک هوش مصنوعی نیز بدون

چنین شرایطی عامل اخلاقی باشد؟ نویسندگان همچنین میان استفاده از ربات به عنوان صرفاً یک پوسته فیزیکی برای اجرای برنامه و رویکردی که بدن را بخشی واقعی از فرایند شناخت و تصمیم گیری می‌داند تفاوت می‌گذارند. در نگاه نخست، بدن صرفاً ظاهر سامانه را تغییر می‌دهد، اما در رویکرد تجسم یافته، تعامل بدنی با محیط می‌تواند بخشی از خود فرایند شناخت و در نتیجه بخشی از امکان کنش اخلاقی باشد.

فصل سپس به این پرسش می‌رسد که آیا آگاهی برای عامل اخلاقی بودن یک ماشین ضروری است یا خیر. اگر آگاهی را شرط لازم بدانیم، دشواری اثبات آگاهی ماشین تقریباً امکان نسبت دادن عاملیت اخلاقی کامل به آن را از میان می‌برد. اما نویسندگان با بررسی دیدگاه‌هایی که عاملیت اخلاقی را بدون اتکا به آگاهی نیز ممکن می‌دانند، نشان می‌دهند که می‌توان مسئله را از زاویه دیگری بررسی کرد. در این نگاه، عاملیت می‌تواند ویژگی‌ای برآمده از تعامل یک سامانه با محیط، توانایی سازگاری و سطح مناسب استقلال آن باشد. بنابراین، ممکن است موجودی در سطحی از توصیف یک سازوکار مکانیکی به نظر برسد، اما در سطحی دیگر، به دلیل تعامل و سازگاری با محیط، بتوان آن را عامل تلقی کرد. با این حال، نویسندگان تأکید می‌کنند که امکان نسبت دادن عاملیت اخلاقی به ماشین با مطلوب بودن واگذاری مسئولیت اخلاقی به آن یکی نیست. در نهایت، بدن، تعامل اجتماعی، سازگاری با محیط و نوعی خرد عملی برای اعتماد اخلاقی به سامانه‌های هوش مصنوعی اهمیت پیدا می‌کنند.

فصل ۱۰: بازنمایی و معناشناسی در مدل‌های بزرگ زبانی

فصل دهم با ظهور «هوش مصنوعی جدید» و به طور مشخص مدل‌های بزرگ زبانی، بحث قدیمی درباره بازنمایی و معنا را دوباره مطرح می‌کند. مسئله اصلی این است که مدل‌های زبانی بزرگ چگونه می‌توانند با استفاده از حجم عظیمی از داده‌های زبانی، الگوهای پیچیده میان واژه‌ها و عبارات را یاد بگیرند و متن‌هایی تولید کنند که از نظر انسانی معنادار به نظر می‌رسند. این موضوع مستقیماً به یکی از مناقشات قدیمی کتاب بازمی‌گردد: آیا دستکاری ساختارهای نمادین و روابط میان آن‌ها برای داشتن معنا کافی است یا اینکه معنا به نوعی ارتباط با جهان نیاز دارد؟ در رویکردهای نمادین کلاسیک، بازنمایی‌ها معمولاً به صورت نمادهایی در نظر گرفته می‌شدند که معنای آن‌ها از پیش مشخص شده بود، اما مدل‌های زبانی بزرگ از طریق یادگیری آماری از داده‌ها، بازنمایی‌هایی ایجاد می‌کنند که رابطه میان عناصر زبانی را در فضای محاسباتی خود منعکس می‌کنند. همین تحول باعث می‌شود پرسش قدیمی درباره «معنای گمشده» بار دیگر در برابر هوش مصنوعی قرار گیرد.

فصل سپس این پرسش را دنبال می‌کند که آیا چنین بازنمایی‌هایی واقعاً واجد معنا هستند یا صرفاً روابط آماری پیچیده میان نشانه‌های زبانی را ثبت می‌کنند. اگر یک مدل زبانی بتواند درباره اشیا، افراد،

رویدادها و روابط میان آن‌ها متن تولید کند، آیا می‌توان گفت که این مدل به همان چیزهایی اشاره می‌کند که انسان هنگام استفاده از زبان به آن‌ها اشاره می‌کند؟ این پرسش به مسئله «درباره‌بودگی» بازمی‌گردد که در جنگ سوم هوش مصنوعی مطرح شده بود. تفاوت مهم مدل‌های جدید با سامانه‌های نمادین کلاسیک این است که بازنمایی‌های آن‌ها به صورت دستی و صریح در برنامه وارد نمی‌شوند، بلکه از فرایند آموزش و ارتباط میان داده‌ها پدید می‌آیند. با این حال، نویسندگان نشان می‌دهند که همین موفقیت می‌تواند پرسش فلسفی را دشوارتر کند: اگر یک سامانه بدون آنکه به معنای انسانی کلمه چیزی را «بداند»، بتواند زبان را با چنین سطحی از پیچیدگی تولید کند، آیا باید معنا را به خودِ ساختارهای محاسباتی نسبت داد یا همچنان باید میان پردازش نمادها و فهم معنا تمایز گذاشت؟

فصل ۱۱: آیاLLM ها بحث درباره آگاهی را تغییر داده‌اند؟

فصل یازدهم مناقشه درباره آگاهی را در پرتو ظهور مدل‌های بزرگ زبانی از نو بررسی می‌کند. مدل‌هایی مانند ChatGPT نشان داده‌اند که یک سامانه مصنوعی می‌تواند در گفت‌وگوهای طولانی، درباره موضوعات متنوع پاسخ تولید کند، درباره حالات ذهنی صحبت کند، از زبان اول شخص استفاده کند و حتی درباره آگاهی و تجربه ذهنی اظهارنظر کند. همین توانایی‌ها پرسش قدیمی را دوباره زنده کرده‌اند که آیا رفتار زبانی پیچیده می‌تواند نشانه‌ای از وجود ذهن و آگاهی باشد. مسئله از اینجا دشوار می‌شود که ما نیز در مورد آگاهی موجودات دیگر مستقیماً به تجربه درونی آن‌ها دسترسی نداریم و معمولاً از رفتار و گزارش‌های آن‌ها نتیجه گیری می‌کنیم. بنابراین، اگر یک مدل زبانی رفتاری شبیه موجودی آگاه نشان دهد، باید مشخص شود چه چیزی می‌تواند وجود آگاهی را در آن تأیید یا رد کند. این موضوع همان شکاف قدیمی میان رفتار قابل مشاهده و تجربه درونی را دوباره به مرکز فلسفه هوش مصنوعی بازمی‌گرداند.

با این حال، فصل تأکید می‌کند که توانایی تولید زبان به خودی خود مسئله آگاهی را حل نمی‌کند. یک مدل زبانی می‌تواند درباره احساس، درد، خودآگاهی یا تجربه ذهنی صحبت کند، اما از این واقعیت نمی‌توان مستقیماً نتیجه گرفت که خود مدل چنین تجربه‌ای دارد. این وضعیت یادآور همان مسئله‌ای است که در بحث‌های قدیمی‌تر درباره آزمون تورینگ، اتاق چینی و تمایز میان رفتار هوشمندانه و فهم واقعی مطرح شده بود. از سوی دیگر، پیشرفت مدل‌های زبانی باعث شده است که برخی اعتراض‌های سنتی دیگر به سادگی قابل تکرار نباشند، زیرا سامانه‌های جدید توانایی‌هایی را نشان می‌دهند که زمانی برای ماشین‌ها بسیار دور از دسترس تصور می‌شد. بنابراین، LLMها الزاماً پاسخ قطعی به مسئله آگاهی ارائه نمی‌کنند، اما معیارهای ما برای بحث درباره ذهن مصنوعی را به چالش می‌کشند. مسئله همچنان

این است که آیا آگاهی را باید بر اساس عملکرد و رفتار تعریف کرد یا اینکه وجود تجربه پدیداری چیزی فراتر از آنچه از عملکرد یک سامانه قابل مشاهده است.

فصل ۱۲: پایداری، مسئله اخلاقی جدید مطرح شده توسط هوش مصنوعی مولد

فصل دوازدهم یکی از مسائل متفاوت اما بسیار معاصر هوش مصنوعی مولد را بررسی می‌کند: پیامدهای زیست محیطی و مسئله پایداری. نویسندگان توجه را از پرسش‌هایی مانند اینکه آیا ماشین می‌تواند فکر کند یا آگاه باشد، به شرایط مادی تولید و استفاده از سامانه‌های هوش مصنوعی منتقل می‌کنند. مدل‌های مولد برای آموزش و اجرای خود به زیرساخت‌های محاسباتی گسترده، مراکز داده، تجهیزات سخت افزاری و مقدار قابل توجهی انرژی نیاز دارند. بنابراین، گسترش سریع این فناوری فقط یک تحول علمی یا اقتصادی نیست، بلکه مسئله‌ای اخلاقی درباره هزینه‌هایی است که برای تولید و استفاده از آن پرداخت می‌شود. از این منظر، ارزیابی هوش مصنوعی نمی‌تواند تنها بر قابلیت‌های آن متمرکز باشد، بلکه باید منابعی را که برای ایجاد و نگهداری این قابلیت‌ها مصرف می‌شوند و پیامدهای زیست محیطی آن‌ها نیز در نظر بگیرد. قرار گرفتن پایداری در کنار آگاهی و اخلاق در بخش جدید کتاب نشان می‌دهد که دامنه فلسفه هوش مصنوعی از مسئله «آیا ماشین می‌تواند مانند انسان فکر کند؟» به مسئله گسترده‌تر «این فناوری چه اثری بر جهان انسانی و طبیعی دارد؟» گسترش یافته است.

در این چارچوب، مسئله پایداری به یکی از تازه‌ترین پرسش‌های اخلاقی هوش مصنوعی مولد تبدیل می‌شود. اگر یک فناوری بتواند تولید محتوا، پژوهش، ارتباط و بسیاری از فعالیت‌های انسانی را دگرگون کند، باید همزمان درباره هزینه‌های مادی و زیست محیطی آن نیز تصمیم گیری شود. این مسئله فقط به میزان انرژی مصرف شده محدود نیست و به چرخه گسترده‌تری از زیرساخت‌ها و منابع مورد نیاز برای توسعه و استفاده از سامانه‌های مولد مربوط می‌شود. از همین رو، فصل پایانی بخش مربوط به «هوش مصنوعی جدید» نشان می‌دهد که ظهور مدل‌های مولد نه تنها برخی مناقشات قدیمی درباره معنا، بازنمایی و آگاهی را دوباره فعال کرده، بلکه مسائل اخلاقی تازه‌ای نیز به وجود آورده است که در زمان شکل گیری جنگ‌های کلاسیک هوش مصنوعی وجود نداشتند. در نتیجه، فلسفه هوش مصنوعی اکنون ناچار است علاوه بر پرسش درباره ماهیت ذهن و هوش، درباره مسئولیت انسان در قبال منابع، محیط زیست و پیامدهای گسترده توسعه این فناوری نیز بحث کند.

جمع‌بندی کتاب:

در پایان کتاب، نویسندگان تأکید می‌کنند که «جنگ‌های هوش مصنوعی» در اواخر قرن بیستم به سکوت رسیده‌اند. فیلسوفان دیگر، به‌طور کلی، مانند گذشته با اصل امکان هوش مصنوعی مخالفت

نمی‌کنند؛ بحث درباره امکان منطقی ساخت هوش مصنوعی تا حد زیادی کنار گذاشته شده و مباحث مربوط به معماری هوش نیز بیشتر به متخصصان و پژوهشگران این حوزه واگذار شده است. در مقابل، جنگ مربوط به «درباره‌بودگی» یا معنا به تدریج به جنگ درباره آگاهی تبدیل شده است؛ بحثی که همچنان فیلسوفان بسیاری را درگیر کرده، هرچند انگیزه اصلی آن بیش از گذشته از علوم اعصاب می‌آید. مسئله چارچوب نیز همچنان حل نشده باقی مانده است، اما در بسیاری از موقعیت‌ها بدون آنکه به صراحت شناخته شود، حضور دارد. بنابراین، جنگ‌های قرن بیستم نه با پیروزی فیلسوفان پایان یافتند و نه با پیروزی پژوهشگران هوش مصنوعی؛ هیچ‌کس برنده جنگ‌های هوش مصنوعی قرن بیستم نشد.

با این حال، نویسندگان می‌پرسند که آیا مناقشات قرن بیست‌ویکم درباره آگاهی و اخلاق را باید ادامه همان «جنگ‌های هوش مصنوعی» دانست یا خیر. پاسخ نهایی را به آینده واگذار می‌کنند. پژوهشگران هوش مصنوعی همچنان به توسعه سامانه‌های جدید ادامه می‌دهند و دیگر مانند گذشته فیلسوفان در پی آنان نیستند تا به پروژه‌هایشان اعتراض کنند، اما پرسش‌های فلسفی از میان نرفته‌اند. ظهور هوش مصنوعی مولد و مدل‌های بزرگ زبانی نیز برخی از مناقشات قدیمی درباره معنا، بازنمایی و آگاهی را دوباره فعال کرده و در کنار آن‌ها مسائل تازه‌ای مانند پایداری و اخلاق کاربردهای هوش مصنوعی را مطرح کرده است. از این منظر، پایان کتاب نه اعلام پایان مناقشه، بلکه تأکیدی بر تداوم آن است: جنگ‌های قدیمی خاموش شده‌اند، اما پرسش‌هایی که آن‌ها ایجاد کردند همچنان باقی‌اند و ممکن است در قالب جنگ‌های تازه‌ای بازگردند.

جمع‌بندی و یافته‌های سیاستی:

این کتاب از منظر سیاستگذاری، یک عینک مهم در برابر نگاه صرفاً فنی به هوش مصنوعی قرار می‌دهد: سیاستگذار نباید پیشرفت هوش مصنوعی را فقط بر اساس افزایش توان محاسباتی و کاربردهای آن ارزیابی کند، بلکه باید همزمان به پرسش‌های مربوط به معنا، آگاهی، عاملیت، مسئولیت و پیامدهای اجتماعی و زیست محیطی آن توجه داشته باشد. سیر چهار جنگ هوش مصنوعی و مباحث جدید درباره آگاهی، اخلاق، مدل‌های زبانی بزرگ و پایداری نشان می‌دهد که تصمیم‌گیری درباره AI همواره با پرسش‌هایی فراتر از فناوری درهم تنیده است.

1) سیاستگذاری هوش مصنوعی باید از نگاه صرفاً فناورانه فراتر برود

تاریخ جنگ‌های هوش مصنوعی نشان می‌دهد که بسیاری از مسائل بنیادی این حوزه، فنی صرف نیستند و به مفاهیمی مانند هوش، معنا، آگاهی، عقلانیت و عاملیت مربوط می‌شوند. بنابراین سیاستگذاری AI نیازمند مشارکت همزمان متخصصان فنی، فیلسوفان و علوم انسانی است.

2) نباید توانایی عملکردی ماشین را با هوش یا آگاهی یکی دانست

توانایی یک سامانه در تولید پاسخ‌های هوشمندانه، لزوماً به معنای برخورداری آن از آگاهی یا فهم نیست. این تمایز برای سیاستگذار اهمیت دارد، زیرا نسبت دادن بیش از حد توانایی‌های انسانی به ماشین می‌تواند به ارزیابی نادرست میزان استقلال، مسئولیت و قابلیت اعتماد سامانه‌های هوش مصنوعی منجر شود.

3) مسئولیت اخلاقی نباید صرفاً به خود سامانه هوش مصنوعی واگذار شود

کتاب نشان می‌دهد که حتی اگر بتوان سامانه‌هایی با نوعی عاملیت اخلاقی ساخت، این پرسش همچنان باقی است که آیا واگذاری مسئولیت‌های اخلاقی به ماشین‌ها اساساً مطلوب است یا خیر. بنابراین سیاستگذاری باید مسئولیت طراحان، تولیدکنندگان، بهره برداران و نهادهای استفاده کننده از AI را در کنار نقش خود سامانه مورد توجه قرار دهد.

4) بدن و زمینه اجتماعی در طراحی و ارزیابی هوش مصنوعی اهمیت دارند

بحث هوش مصنوعی تجسم یافته نشان می‌دهد که شناخت و عاملیت را نمی‌توان همیشه از بدن و محیط جدا کرد. از منظر سیاستگذاری، این موضوع ضرورت توجه به زمینه واقعی استفاده از AI، تعامل آن با انسان‌ها و محیط اجتماعی و نه فقط عملکرد آن در شرایط آزمایشگاهی را برجسته می‌کند.

5) مدل‌های زبانی بزرگ، مسائل فلسفی قدیمی را دوباره به مسئله سیاستی تبدیل کرده‌اند

ظهورLLM ها بحث‌های قدیمی درباره بازنمایی، معنا و آگاهی را دوباره فعال کرده است. بنابراین سیاستگذاری درباره هوش مصنوعی مولد نباید صرفاً بر قابلیت‌های کاربردی آن متمرکز باشد، بلکه باید درباره حدود نسبت دادن فهم، معنا و آگاهی به این سامانه‌ها نیز حساس باشد.

6) پایداری باید به بخشی از ارزیابی اخلاقی هوش مصنوعی تبدیل شود

ویرایش جدید کتاب، پایداری را به عنوان یک مسئله اخلاقی جدید مطرح شده توسط هوش مصنوعی مولد وارد بحث می‌کند. این تغییر نشان می‌دهد که ارزیابی سیاستی AI باید علاوه بر قابلیت‌ها و منافع آن، هزینه‌ها و پیامدهای زیست محیطی توسعه و استفاده از سامانه‌های هوش مصنوعی را نیز در نظر بگیرد.

نقاط قوت کتاب:

۱- روایت تاریخی منسجم از هفت دهه مناقشه:

یکی از مهم‌ترین نقاط قوت کتاب، تبدیل تاریخ پراکنده فلسفه هوش مصنوعی به روایتی منسجم در قالب چهار «جنگ» است. نویسندگان اعتراض‌های فلسفی به هوش مصنوعی را در یک توالی تاریخی قرار می‌دهند و نشان می‌دهند هر مناقشه چگونه شکل گرفت، چه پاسخی دریافت کرد و چه میراثی برای مباحث بعدی باقی گذاشت. یک نقد منتشرشده در Teaching Philosophy نیز کتاب را مروری جامع و در عین حال قابل دسترس بر تاریخ اعتراض‌های فلسفی به هوش مصنوعی ارزیابی می‌کند.

۲- توضیح قابل فهم مباحث پیچیده فلسفی و فنی:

کتاب موفق می‌شود مباحثی دشوار مانند قضیه ناتمامیت گودل، آزمون تورینگ، معماری‌های شناختی، مسئله معنا، مسئله چارچوب و آگاهی را برای خواننده‌ای فراتر از متخصصان فلسفه ذهن و علوم رایانه قابل پیگیری کند. مرور Teaching Philosophy نیز به طور مشخص توضیح روشن و قابل دسترس نویسندگان درباره استدلال گودل و همچنین ارائه جامع معماری‌های اصلی هوش مصنوعی را از نقاط قوت کتاب می‌داند. این ویژگی، کتاب را برای آموزش و ورود به حوزه فلسفه هوش مصنوعی مناسب کرده است.

۳- پیوند دادن فلسفه با تحول واقعی هوش مصنوعی:

نقطه قوت دیگر کتاب این است که فلسفه را از بحثی انتزاعی درباره «آیا ماشین می‌تواند فکر کند؟» جدا نمی‌کند، بلکه آن را با تاریخ واقعی توسعه هوش مصنوعی پیوند می‌دهد. نویسندگان نشان می‌دهند اعتراض‌های فلسفی چگونه با تغییر معماری‌ها و رویکردهای فنی AI ارتباط داشته‌اند و چگونه پرسش‌هایی مانند معنا، آگاهی و عقلانیت همچنان در توسعه فناوری حضور دارند. ناشر نیز یکی از اهداف اصلی کتاب را توضیح همین رابطه میان فلسفه، علم و مسیر تحول هوش مصنوعی معرفی می‌کند.

۴- ارائه دیدگاه مستقل در کنار گزارش تاریخی:

کتاب صرفاً یک تاریخ نگاری خنثی از مناقشات نیست و در بخش‌هایی مانند «درباره‌بودگی»، مسئله چارچوب و عاملیت اخلاقی هوش مصنوعی، نویسندگان موضع تحلیلی خود را نیز مطرح می‌کنند. این ویژگی باعث می‌شود اثر از یک مرور صرف فراتر برود و خواننده با تفسیر مشخص نویسندگان از مسائل بنیادی مواجه شود. مرور Juvshik نیز فصل‌های پنجم، ششم و به ویژه نهم را از بخش‌هایی می‌داند که کتاب در آن‌ها از گزارش تاریخی فراتر می‌رود و استدلال‌های خاص خود نویسندگان را ارائه می‌کند.

نقدهای اصلی کتاب:

۱- پوشش کم‌رنگ اخلاق کاربردی هوش مصنوعی:

مهم‌ترین نقد واردشده به کتاب، عدم تناسب میان اهمیت اخلاق کاربردی AI و حجم اختصاص‌یافته به آن است. فصل هشتم عمدتاً به سلاح‌های خودمختار و خودروهای خودران محدود می‌شود و موضوعاتی مانند الگوریتم‌های پیش‌بینی‌کننده در شبکه‌های اجتماعی و پلیس، اطلاعات نادرست، تشخیص پزشکی و ربات‌های مراقبتی را بررسی نمی‌کند. Juvshik این فصل را «بزرگ‌ترین ناامیدی» کتاب می‌داند و معتقد است با توجه به اهمیت اخلاق کاربردی AI، بحث آن بسیار کوتاه و محدود باقی مانده است.

۲- ابرابری در سطح فنی و لحن فصل‌ها:

کتاب از نظر سطح دشواری و شیوه ارائه کاملاً یکدست نیست. برخی فصل‌ها بسیار روشن و قابل دسترس‌اند، در حالی که فصل‌هایی مانند بحث معنا، مسئله چارچوب و معماری‌های هوش مصنوعی فنی‌تر و برای فهم کامل نیازمند آشنایی قبلی با فلسفه ذهن یا علوم رایانه‌اند. Juvshik نیز کتاب را از نظر لحن و مفهوم تا حدی «ناهمگون» توصیف می‌کند و معتقد است برخی فصل‌ها برای دانشجویان دوره کارشناسی دشوارترند. در نتیجه، کتاب بیشتر برای خوانندگان تحصیلات تکمیلی یا پژوهشگران مناسب است.

۳- اصله گرفتن از تحولات هوش مصنوعی کاربردی و مولد:

یکی از محدودیت‌های مهم نسخه نخست کتاب این است که در حالی که نویسندگان کاربردهای هوش مصنوعی ضعیف را به عنوان «جنگ‌های جدید» معرفی می‌کنند، تنها بخش محدودی از کتاب را به این حوزه اختصاص داده‌اند. این مسئله با انتشار کتاب در سال ۲۰۲۱ تشدید شد، زیرا ChatGPT تنها یک سال بعد منتشر شد و موج جدیدی از پرسش‌های فلسفی و اخلاقی درباره مدل‌های زبانی بزرگ و هوش مصنوعی مولد ایجاد کرد. البته ویرایش دوم در سال ۲۰۲۵ با فصل‌هایی درباره LLMها، آگاهی و پایداری این فاصله را تا حد زیادی جبران کرده است.

اندیشکده فرهنگ و توسعه

باشگاه کتاب فرتو