اخلاقِ هوش مصنوعی

سرشناسه: کوکلبرگ، مارک، ۱۹۵۷ ـ م. Coeckelbergh, Mark, 1957
عنوان و نام پدیدآور: اخلاقِ هوش مصنوعی/مارک کوکلبرگ؛ ترجمهٔ احسان عارفی‌فر.
مشخصات نشر: تهران: ققنوس، ۱۴۰۳.
مشخصات ظاهری: ۲۲۴ ص.
شابک: ۹۷۸-۶۲۲-۰۴-۰۵۳۱-۳
وضعیت فهرست‌نویسی: فیپا
یادداشت: عنوان اصلی: .AI Ethics [2020]
یادداشت: کتاب حاضر توسط زرنوشت در سال ۱۴۰۳ (۱۵۲ ص.) فیپا دریافت کرده است.
یادداشت: کتابنامه.
یادداشت: نمایه.
موضوع: هوش مصنوعی ــ جنبه‌های اخلاقی
موضوع: Artificial intelligence -- Moral and ethical aspects
شناسهٔ افزوده: عارفی‌فر، احسان، ۱۳۶۶ ـ، مترجم
رده‌بندی کنگره: Q۳۳۴/۷
رده‌بندی دیویی: ۱۷۰
شمارهٔ کتاب‌شناسی ملی: ۹۶۲۲۰۲۰

اخلاقِ هوش مصنوعی

مارک کوکلبرگ
ترجمهٔ احسان عارفی‌فر

این کتاب ترجمه‌ای است از:

AI Ethics
Mark Coeckelbergh
The MIT Press, 2020

انتشارات ققنوس
تهران، خیابان انقلاب، خیابان شهدای ژاندارمری،
شمارهٔ ۱۱۱، تلفن ۴۰ ۸۶ ۴۰ ۶۶
ویرایش، آماده‌سازی و امور فنی:
تحریریهٔ انتشارات ققنوس

***

مارک کوکلبرگ
اخلاقِ هوش مصنوعی
ترجمهٔ احسان عارفی‌فر
چاپ اول
۱۱۰۰ نسخه
۱۴۰۳
چاپ سروش

شابک: ۳-۰۵۳۱-۰۴-۶۲۲-۹۷۸
ISBN: 978-622-04-0531-3
www.qoqnoos.ir
Printed in Iran

تقدیم به مینای مهربان و نیک‌سرشت

مترجم

فهرست

یادداشت نویسنده برای خوانندگان ایرانی

از زمان انتشار کتاب **اخلاق هوش مصنوعی** در سال ۲۰۲۰ به زبان انگلیسی، دنیای هوش مصنوعی با سرعت عجیبی تغییر کرده است. امروز هوش مصنوعیِ مولد[1] قادر است محتوا و تصویر تولید کند، آن هم به گونه‌ای که بسیاری را شگفت‌زده و مبهوت می‌کند. بحث‌های بسیاری در حوزهٔ اخلاق هوش مصنوعی در این مدت درگرفته است، مثلاً در باب خطرهای اگزیستانسیالیستی که گمان می‌رود فناوری به همراه داشته باشد. اما مطالب طرح‌شده در این کتاب همچنان موضوعیت دارند، و شاید حتی بیشتر از قبل. مثلاً پرسش‌هایی نظیر شأن اخلاقی سیستم‌های هوش مصنوعی یا پیامدهای زیست‌محیطی هوش مصنوعی از جمله مواردی‌اند که امروزه اهمیت بیشتری یافته‌اند.

در فصل آخر همین کتاب، دیدگاه‌هایی از نظرتان خواهد گذشت که نشان می‌دهد اخلاقِ هوش مصنوعی روزبه‌روز در حال تبدیل شدن به موضوعی جهانی است. در حال حاضر سازمان‌های بین‌المللی مانند سازمان ملل متحد تلاش‌های پیگیرانه‌ای در زمینهٔ سیاست‌گذاری ناظر به هوش مصنوعی به‌ویژه کار بر روی مدلی جهانی برای راهبری

1. generative AI

آن در پیش گرفته‌اند. درک اهمیت و ضرورت چنین تلاش‌هایی با توجه به ماهیت فرامرزی مسائل اخلاقی در باب هوش مصنوعی چندان دشوار نخواهد بود. با این حال، همچنان چالش‌هایی نیز به چشم می‌خورد: چگونه می‌توان مدل جهانی راهبری هوش مصنوعی را پی‌ریزی کرد که جهانِ جنوب[۱] را نیز در بر گیرد؟ آیا کشورهای غیرغربی نیز سهمی در شکل دادن به این مدل جهانی خواهند داشت؟ تکلیف تفاوت‌های فرهنگی در بحث‌های اخلاق فناوری چیست؟ و چگونه می‌توان شکاف‌ها را پر کرد؟

ترجمه‌هایی که از این کتاب انجام شده، از جمله همین ترجمهٔ فارسی، برای به جریان انداختن گفتگوهای فرامرزی و بین‌قاره‌ای در مورد اخلاق هوش مصنوعی حایز اهمیت‌اند. امروز ایران در زمینهٔ تحقیقات هوش مصنوعی بسیار فعال است. چند سال پیش شاخص مجلهٔ **نیچر**[۲] ایران را از نظر انتشارات پژوهشیِ مرتبط با هوش مصنوعی در رتبهٔ ۱۳ دنیا قرار داده بود، و در سمت دیگر شاهد برخی تلاش‌های موفق در حوزهٔ کاربرد هوش مصنوعی در ایران هستیم، از جمله داستان موفقیت مسیریاب هوشمند «بلد»، که بهتر از من از آن باخبرید! فناوری‌های نوظهور خوش درخشیده‌اند، اما تلاش‌ها در زمینهٔ اخلاق هوش مصنوعی و اخلاق ربات‌ها نیز در حال رشد است. این تلاش‌ها هم باید در بافتارهای بومی معنادار باشند و هم برای رشد از عناصر آن‌ها تغذیه کنند. مثلاً، دیدن نسخه‌های اسلامی اخلاق هوش مصنوعی در نوع خود می‌تواند جالب توجه باشد، و بحث در مورد چگونگی تأثیر مثبت اخلاق هوش مصنوعی

---

۱. Global South: واژه‌ای که امروزه به جای جهان سوم به کار می‌رود و به معنای کشورهای کمترتوسعه‌یافته است که اکثراً در نیمکرهٔ جنوبی قرار دارند.

2. *Nature*

در صنعت هوش مصنوعی و رباتیک ایران و نحوهٔ گنجاندن چنین موضوعاتی در سرفصل‌های آموزشیِ تعداد زیادی از دانش‌آموختگان رشته‌های علمی ــ که هر ساله در ایران فارغ‌التحصیل می‌شوند ــ موضوعی جذاب است.

فناوری هوش مصنوعی روزبه‌روز در حال تکامل است، و همین امر پرداختن به اخلاق هوش مصنوعی را تبدیل به موضوعی بسیار فوری می‌کند. شاید در لحظه‌ای تاریخی قرار گرفته باشیم؛ بی‌شک پیشرفت‌های بیشتری در راه است، از این رو، ضروری است که دولت‌ها و سایر بازیگران این عرصه در سراسر جهان برای استفاده و ادغام هوش مصنوعی در جوامع و اجتماعات متبوعشان به شیوه‌ای معنادار و مسئولانه، و با در نظر گرفتن بافت‌های فرهنگی و تاریخی مرتبط (مثلاً، خاورمیانه که در آن اخلاق غیرسکولار و ارزش‌های سنتی نقش مهمی ایفا می‌کنند)، سیاست‌های مؤثری پی بریزند، که در عین حال نیز با اهداف و ارزش‌های مشترک جهانی، گفتگوهای بین‌فرهنگی و تلاش‌ها برای همکاری‌های بین‌المللی و فراملی همسویی داشته باشد. وجود این همسویی به‌ویژه به این دلیل مهم است که بخواهیم اطمینان یابیم آیندهٔ هوش مصنوعی و رباتیک را فقط چند شرکت جهانی رقم نمی‌زنند و شکاف‌های دیجیتال عمیق‌تر نمی‌شوند، بلکه همهٔ شهروندان، همهٔ کشورها و همهٔ بخش‌های جامعهٔ جهانی از آن بهره‌مند خواهند شد. در این بین قشر جوان سزاوار بهره‌مندی از چنان آیندهٔ فناورانه‌ای هستند که علایق و خواسته‌هایشان نادیده گرفته نشود و قادر باشند خلاقانه در ساخت آن آینده سهیم باشند. هرچند آیندهٔ هوش مصنوعی به اخلاق نیاز دارد، از طرفی نیازمند نیروی کار و استعدادهای آن‌ها نیز هست، و ایران، با جمعیت نسبتاً بزرگش، قابلیت‌های زیادی در این زمینه دارد.

امیدوارم این ترجمهٔ فارسی، که افتخار بزرگی برای من به شمار می‌رود، به این گفتگوها و تلاش‌ها در زمینهٔ اخلاق و سیاست‌گذاری ناظر به هوش مصنوعی در ایران و فراتر از آن یاری رساند.

مارک کوکلبرگ
۲۶ سپتامبر ۲۰۲۳، وین

یادداشت مترجم

این ترجمه دستاورد نخستین تلاش جدی‌ام در زمینهٔ فلسفه است. از همان کودکی که کتاب‌های فلسفی را در کتابخانهٔ پدرم می‌دیدم، تا بعدها که خود نیز کتاب به دست گرفتم، کوشیده‌ام در متن دیدگاه‌های فلسفی حضور داشته باشم و از سرگذشت تفکر سر درآورم. به‌مرور دوستانی اهل فلسفه یافتم که در گذر سال‌ها بیشترین نزدیکی را با ایشان داشتم و همراهی با ایشان نیز بیش از پیش شیفتگی‌ام به فلسفه و فلسفه‌ورزی را فزونی بخشید. با این حال، تحصیل در زمینهٔ علوم مهندسی و اشتغال در حوزهٔ فناوری، که با ضرورت‌های زندگی روزمره نیز همراه بود، امکان آن را از من ربود تا فرصتی کافی را صرف مطالعهٔ اصولی و هدفمند فلسفه سازم. اما مطالعهٔ کتاب‌های فلسفی و گفتگوهای گاه‌وبی‌گاه با دوستان فلسفه‌ورز همچنان مرا در عرصهٔ تفکر فلسفی نگاه داشت که مایهٔ دلخوشی‌ام است. با این حال، چنان‌که هایدگر آموخته است: «اگر ما هم نتوانیم با فلسفه کاری کنیم، فلسفه با ما کاری خواهد کرد،» و با من نیز به سبب دلمشغولی‌ام چنین کرد.

در یکی از گفتگوهای فلسفی با استاد و دوست عزیز و ارجمند جناب آقای دکتر جلال پیکانی، که خود نیز دلمشغول فلسفه بوده

و هستند، دغدغه‌مندی خود را با ایشان در میان گذاشتم و ایشان پیشنهاد ترجمهٔ کتاب در یکی از زمینه‌های مرتبط با فلسفهٔ فناوری را مطرح کردند. این پیشنهاد مبارک بود؛ نه‌فقط به این دلیل که نظم و چارچوبی به فعالیت‌های فلسفی من می‌داد، بلکه به سبب پیوندی که میان دانش و تجربه‌های شغلی‌ام در عرصهٔ مهندسی با دغدغه‌مندی فلسفی‌ام برقرار می‌ساخت. حمایت‌ها و راهنمایی‌هایی ایشان در پیمودن این مسیر هم در گزینش کتابِ مناسبی که درخور فرهیختگان و خوانندگان ایرانی باشد، و هم در نکات ارزنده‌ای ادامه یافت که طی بازخوانی‌های چندبارهٔ متنِ ترجمه یادآور شدند. این حمایت‌ها نه‌تنها برای این نخستین تجربه‌ام در عرصهٔ ترجمه ضرورت داشت، بلکه سبب شور و اشتیاقی شد که مرا برای انجام این امر دلگرم نگه می‌داشت. سپاسگزار تمامی مهربانی‌های ایشان هستم و امید موفقیت روزافزونشان را در عرصهٔ اندیشه‌ورزی دارم.

در خصوص خودِ کتاب و نویسنده نیاز به توضیح بیشتری نیست؛ ترجمه به زبان‌های مختلف و بازخوردهای مناسب خود گواهی بر اقبال اهل فلسفه و فناوری از این اثر است. همچنین توضیحات ناشر اصلی کتاب، انتشارات ام‌آی‌تی، در باب عمق علمی و جامعیت این کتاب، در کنار شهرت نویسنده در حوزهٔ مباحث اخلاق هوش مصنوعی، مخصوصاً در محافل دانشگاهی، مراکز پژوهشی و ارگان‌های سیاست‌گذاری در اروپا، همگی نشان از اهمیت این کتاب در زمینهٔ تخصصی خود داشته و دارد. با نگاه به بازار کتاب در ایران، چه‌بسا این کتاب اگر نه تنها کتاب، اما جزء آثار انتشاریافتهٔ انگشت‌شماری است که مسائل اساسیِ پیرامون اخلاقِ هوش مصنوعی را با رویکردی فلسفی تحلیل کرده است، و در کنار پرداختن به جنبه‌های مختلف نظری و کاربردی موضوع با ارائهٔ مثال‌ها و ارجاعات متعدد به جامعه و صنعت، پرسش‌های روزآمد و جدیِ این حوزه را نیز به شکلی

منسجم و ساختارمند پیش روی خواننده قرار می‌دهد. امیدوارم این ترجمه روزنی باشد به دنیای پرهیاهو و پرشتاب اخلاق هوش مصنوعی تا علاقه‌مندان به مباحث فلسفه و فناوری، چه اهل دانش در مراکز دانشگاهی و چه خوانندگان پرشمار کتاب‌های عمومی فلسفی در ایران، به این عرصه نظری افکنند یا راهی بگشایند. در پایان، این تأکید را لازم می‌دانم که هر کاستی و خطایی در ترجمهٔ فارسی بر عهدهٔ مترجمی است که به عذر نخستین بودنِ ترجمه‌اش، نظرِ خطاپوشِ خواننده را چشم دارد.

همچنین، در ترجمهٔ این کتاب، از ابزارهای هوش مصنوعی نیز بهره برده‌ام؛ به‌ویژه از دو مدل زبانی بزرگِ (LLM) شناخته‌شده، ChatGPT و Claude. هرچند این ابزارها و ده‌ها و صدها ابزار مشابه امروزه در شرکت‌های بزرگ و کوچک به‌ویژه در حوزهٔ کاری من یعنی مهندسی نرم‌افزار به مؤلفه‌ای عادی ولی ضروری و جایگزین‌ناپذیر تبدیل شده‌اند، هدف من محک زدن عملکرد آن‌ها در حوزه‌های غیرفنیِ علوم انسانی بود که به طور عمده چنین تصور می‌شود که ماشین‌ها فاقد توانایی‌های لازم برای سودمندی در این حوزه‌ها هستند. ترجمهٔ متن، واکاوی برخی جمله‌ها و عبارت‌های پیچیده، توضیح و تفصیل مفاهیم و یافتن مشابهت‌ها در آثار نویسندگان دیگر و تهیهٔ پانویس برای عبارات فنی و فلسفی جزو مواردی بودند که سعی کردم کارایی این ابزارهای هوش مصنوعی را بیازمایم. همان‌طور که در متن کتاب هم از نظرتان خواهد گذشت، عملکرد هوش مصنوعی در این زمینه هنوز به طور کامل قابل اعتماد نیست، به‌خصوص در ترجمه به زبان فارسی که بیشتر ناشی از فقدان سرمایه‌گذاری‌های لازم در این عرصه و همچنین کمبود منابع زبانی فارسی برای آموزش مدل‌های زبانی بزرگ است. با این حال، نتیجه همچنان چشمگیر و در برخی موارد شگفت‌انگیز است، که خودِ این امر همان بیم و امیدهایی را

برمی‌انگیزد که نویسنده در کتاب به آن‌ها اشاره کرده است. شاید بتوان گفت این ابزارها در حال حاضر نقش یک دستیار جوان و باهوش را ایفا می‌کنند که به‌سرعت و با تفصیل بسیار خوب مطالبی را برای بررسی و نتیجه‌گیری نهاییِ کاربر فراهم می‌آورد، و خود نیز در این فرایند بیشتر یاد می‌گیرد. بی‌شک در سال‌های آتی پیشرفت‌های خیره‌کنندهٔ این دستیار جوان دیدنی خواهد بود!

۳۰ اوت ۲۰۲۳، ویگو، اسپانیا

پیشگفتار مجموعه

«دانش ضروری» نام مجموعه‌ای از کتاب‌های جیبی خوش‌طرح انتشارات دانشگاه ام‌آی‌تی است که به صورت فشرده و همه‌فهم به موضوعات روز می‌پردازد. این کتاب‌ها، که به قلم اندیشمندان پیشرو نگاشته می‌شوند، دیدگاه‌هایی کارشناسانه را دربارهٔ حوزه‌های مختلف، از فرهنگ و تاریخ گرفته تا علوم و فناوری، در اختیار خوانندگان و علاقه‌مندان قرار می‌دهند.

در دوران ما که عصر دسترسی فوری به اطلاعات و رفع سریع نیازهاست، به‌راحتی می‌توان به عقاید ظنّی، دلیل‌تراشی‌ها و توصیف‌های سطحی دست یافت، اما دستیابی به دانشی بنیادین که بتواند درکی اصولی و قاعده‌مند از جهان پدید آورد دشوار است. کتاب‌های مجموعهٔ دانش ضروری این خلأ را پر می‌کنند. هر یک از این کتاب‌های فشرده با در هم آمیختن دانش تخصصیِ قابل فهم برای غیرمتخصصان و پرداختن به مباحث حیاتی از راه توجه به مبادی مدخلی هستند که ورود به ایده‌های پیچیده را برای مخاطب عام ممکن می‌سازند.

بروس تیدور
استادتمام مهندسی زیستی و علوم رایانه
مؤسسهٔ فناوری ماساچوست (ام‌آی‌تی)

سپاسگزاری

این کتاب نه‌تنها محصول فعالیت خود من در حوزهٔ اخلاق ناظر به هوش مصنوعی است، بلکه منعکس‌کنندهٔ دانش و تجربهٔ فعلی موجود در این حوزه هم هست. نام بردن از همهٔ افرادی که طی سال‌های اخیر با آن‌ها گفتگو کرده‌ام و از آن‌ها آموخته‌ام غیرممکن است، اما لازم می‌دانم از برخی افراد و مجامع در حال رشدی که در این زمینه فعالیت می‌کنند نام ببرم و تشکر کنم، پژوهشگران هوش مصنوعی مانند جوآنا برایسون و لوسین استیلز؛ همکارانم در حوزهٔ فلسفهٔ فناوری مثل شانون والور و لوچیانو فلوریدی؛ اساتید دانشگاهی که در حوزهٔ نوآوری مسئولانه در هلند و انگلستان فعال هستند، از جمله بِرْنْد اشتال در دانشگاه دومونت‌فورت؛ افرادی که در وین با آن‌ها ملاقات کردم، مثل رابرت تراپل، سارا اشپیکرمن و ولفگانگ (بیل) پرایس؛ و اعضای همکار من در حوزهٔ مشاورهٔ سیاست‌گذاری گروه عالی کارشناسی (کمیسیون اروپا در امور هوش مصنوعی) و شورای رباتیک و هوش مصنوعی اتریش که فقط می‌توانم به نام برخی از آن‌ها اشاره کنم، مثل راجا چاتیلا، ویرجینیا دیگنام، خرون فن دن هوون، سابینه کوسگی و ماتیاس شوتس.

همچنین می‌خواهم از زاخاری استورمز به خاطر کمک به نمونه‌خوانی و قطع‌بندی و لنا اشتارکل و ایزابل والترر به خاطر حمایت در جستجوی منابع صمیمانه تشکر کنم.

۱

ای آینهٔ روی دیوار

هراس و هیاهوهای هوش مصنوعی: ای آینهٔ روی دیوار، چه کسی از همه باهوش‌تر است؟[۱]

ماه مارس ۲۰۱۶ وقتی نتایج را اعلام می‌کنند، چشمان لی سدول[۲] از اشک پر می‌شود. آلفاگو،[۳] هوش مصنوعی توسعه‌یافته توسط شرکت گوگل دیپ‌مایند،[۴] در بازی گُو[۵] با نتیجهٔ چهار بر یک لی را شکست می‌دهد. دو دههٔ قبل نیز گری کاسپاروف، استاد بزرگ شطرنج، از رایانه‌ای به نام دیپ بلو[۶] شکست خورده بود و حالا آلفاگو لی سدول، قهرمان هیجده دوره مسابقات جهانی، را در این بازی پیچیده که گمان می‌رفت فقط انسان‌ها به مدد تفکر راهبردی و شهودشان توانایی بازی کردن آن را دارند شکست داده بود. آلفاگو این پیروزی را با دنبال کردن دستورالعمل‌های برنامه‌نویسانش کسب نکرد، بلکه از طریق اِعمال یادگیری ماشینی بر روی داده‌های میلیون‌ها بازی قبلی گُو و همچنین بازی کردن در برابر خودش به دست آورد. در چنین موردی،

---

۱. این جمله تلمیحی دارد به جمله‌ای قالبی در داستان عامیانهٔ سفیدبرفی که از زبان ملکهٔ شیطان‌صفت آن داستان ادا می‌شود. ملکه روبه‌روی آینه‌ای جادویی می‌ایستد و از آن می‌پرسد: «ای آینهٔ روی دیوار، چه کسی از همه زیباتر است؟» (تمامی پانویس‌ها از مترجم است.)

2. Lee Sedol
3. AlphaGo
4. Google's DeepMind
5. Go
6. Deep Blue

برنامه‌نویسان داده‌ها را آماده می‌کنند و الگوریتم‌ها را می‌نویسند، اما نمی‌توانند حرکت بعدی ماشین را بدانند؛ هوش مصنوعی خودش یاد می‌گیرد. سرانجام لی پس از چند حرکت غیرمعمول و غافلگیرکنندهٔ ماشین مجبور به تسلیم شد (Borowiec 2016).

این اتفاق موفقیتی چشمگیر برای هوش مصنوعی بود. اما از طرفی نگرانی‌هایی نیز برمی‌انگیزد. حرکت‌های زیبای هوش مصنوعی تحسین‌برانگیز، و از طرفی ناراحت‌کننده، و حتی ترسناک، است. امیدهایی وجود دارد که هوش مصنوعی‌هایی حتی باهوش‌تر بتوانند در ایجاد تحولات بزرگ در عرصهٔ بهداشت و درمان یا یافتن راه‌حل‌هایی برای همهٔ انواع مشکلات اجتماعی ما را یاری کنند، اما در عین حال بیم آن می‌رود که ماشین‌ها بر ما مسلط شوند. آیا ماشین‌ها می‌توانند بر ما استیلا یابند و کنترلمان کنند؟ آیا هوش مصنوعی در حال حاضر فقط یک ابزار است، یا به‌آرامی ولی به طور حتم ارباب ما خواهد شد؟ این ترس‌ها یادآور حرف‌های هال،[1] همان رایانهٔ مبتنی بر هوش مصنوعی، است که در فیلم علمی‌ـتخیلیِ **۲۰۰۱: اودیسهٔ فضایی**،[2] ساختهٔ استنلی کوبریک، در جواب فرمان «درهای سفینه رو باز کن!»، به فضانورد گفت: «متأسفم دیوید! نمی‌تونم این کار رو بکنم!» و اگر ترسناک هم نباشد، شاید احساس ناراحتی یا ناامیدی بکنیم. داروین و فروید باورهایمان در مورد استثنایی بودنمان، احساساتمان در مورد برتری‌مان، و خیالاتمان در مورد کنترلگری را به زیر کشیدند؛ و حالا به نظر می‌رسد هوش مصنوعی بر تصور انسان از خودش ضربهٔ دیگری می‌زند. اگر ماشینی بتواند چنان کارهایی کند، چه چیزی [به عنوان ویژگی ممتاز] برایمان باقی می‌ماند؟ ما چه هستیم؟ آیا فقط ماشینیم، ماشین‌هایی پَست با بی‌شمار ایراد؟[3] چه چیزی قرار است بر

1. HAL 2. *2001: A Space Odyssey* 3. bugs

سرمان بیاید؟ آیا بردهٔ ماشین‌ها خواهیم شد، یا حتی بدتر، صرفاً منابع انرژی برای آنان، مثل فیلم ماتریکس؟!

تأثیرات واقعی و فراگیر هوش مصنوعی

اما دستاوردهای هوش مصنوعی به دنیای بازی‌ها یا فیلم‌های علمی‌ـتخیلی محدود نیست. در حال حاضر هوش مصنوعی در زندگی ما حضوری گسترده دارد، اغلب در دل ابزارهای روزمرهٔ ما و به عنوان بخشی از سیستم فناورانهٔ پیچیدهٔ آن‌ها (Bodding-ton 2017). به لطف رشد نُماییِ قدرت رایانه‌ها، در دسترس بودن داده‌ها(ی کلان) که معلول ظهور شبکه‌های اجتماعی است و استفادهٔ گسترده از میلیاردها گوشی هوشمند و شبکه‌های موبایلیِ پرسرعت، هوش مصنوعی، مخصوصاً یادگیری ماشینی، پیشرفت‌های مهمی داشته است. این موضوع الگوریتم‌ها را قادر ساخته بسیاری از فعالیت‌های ما را ــ نظیر برنامه‌ریزی، سخن گفتن، تشخیص چهره و تصمیم‌گیری ــ بر عهده بگیرند. هوش مصنوعی در بسیاری از حوزه‌ها کاربرد دارد، از جمله حمل‌ونقل، بازاریابی، بهداشت و درمان، امور مالی و بیمه، امنیت و امور نظامی، علوم، آموزش، امور اداری، دستیار شخصی (مثل گوگل داپلکس[1])،[1] سرگرمی، هنر (مثل آهنگسازی و بازیابی اطلاعات موسیقی)، کشاورزی و همچنین تولید.

شرکت‌های حوزهٔ فناوری اطلاعات و اینترنت هوش مصنوعی را ایجاد و از آن استفاده کرده‌اند. مثلاً، گوگل همواره از هوش مصنوعی در موتور جستجوی خود استفاده کرده است. فیسبوک هوش مصنوعی را در تبلیغات هدفمند و برچسب‌گذاری تصاویر به کار می‌برد. مایکروسافت و اپل برنامه‌های دستیار دیجیتال[2] را به کمک هوش

1. Google Duplex 2. digital assistant

مصنوعی به کار انداخته‌اند. اما کاربرد هوش مصنوعی وسیع‌تر از آن چیزی است که حوزهٔ فناوری اطلاعات به معنایی محدود تعریف کرده است. مثلاً، طرح‌های عملیاتی بسیاری برای خودروهای خودران وجود دارد و آزمایش‌های بسیاری در مورد آن‌ها انجام گرفته است، که اساس همهٔ آن‌ها هوش مصنوعی است. پهپادها از هوش مصنوعی بهره می‌برند و نیز سلاح‌های خودکار که می‌توانند بدون دخالت انسان آتش کنند. هوش مصنوعی همچنین در فرایند تصمیم‌گیری دادگاه‌ها استفاده شده است. مثلاً، در آمریکا از سیستم کُمپاس[1] برای پیش‌بینی این‌که احتمال می‌رود چه کسی دوباره مرتکب جرم شود استفاده کرده‌اند. هوش مصنوعی وارد حوزه‌هایی نیز شده است که به طور معمول بسیار شخصی یا خصوصی تلقی می‌کنیم. مثلاً، امروزه ماشین‌ها می‌توانند چهرهٔ ما را بخوانند: نه‌فقط برای تشخیص هویتمان، بلکه برای تشخیص احساساتمان و بیرون کشیدن همه نوع اطلاعات.

لزوم بحث در باب مسائل اخلاقی و اجتماعی

ممکن است هوش مصنوعی فواید زیادی داشته باشد. ممکن است به بهبود خدمات عمومی و تجاری منجر شود. مثلاً، تشخیص تصویر[2] به منظور استفاده در کاربردهای پزشکی چشم‌اندازی نویدبخش دارد: می‌تواند در تشخیص بیماری‌هایی نظیر سرطان و آلزایمر یاری‌بخش باشد. اما همین کاربردهای روزمرهٔ هوش مصنوعی همچنین نشان می‌دهد که چگونه این فناوری‌های جدید نگرانی‌هایی اخلاقی برمی‌انگیزند. اجازه دهید به برخی پرسش‌های مربوط به اخلاق هوش مصنوعی اشاره کنم.

---

1. COMPAS

۲. image recognition: یک فرایند ویدئویی یا تصویری دیجیتال برای شناسایی یک شیء یا ویژگی است و هوش مصنوعی به طور فزاینده‌ای در استفاده از فناوری مؤثر بوده است.

آیا ضروری است که خودروهای خودران قیود اخلاقی درونی داشته باشند؟ و اگر پاسخ مثبت است، چه نوع قیودی باید داشته باشند، و آن قیود چگونه باید مشخص شوند؟ مثلاً موقعیتی را در نظر بگیرید که خودرویِ خودرانی یا باید کودکی را زیر بگیرد، یا به سمت دیواری براند تا جان کودک حفظ شود، ولی با این کار سرنشینانش در معرض مرگ قرار می‌گیرند. خودرو کدام را باید انتخاب کند؟ آیا واقعاً وجود سلاح‌های مرگبار خودکار باید مجاز باشد؟ چه تعداد و چه گستره‌ای از چنین تصمیم‌گیری‌هایی را می‌خواهیم به هوش مصنوعی تفویض کنیم؟ و در مواقعی که مشکلی پیش می‌آید، چه کسی پاسخگو خواهد بود؟ در یک پرونده، قضات به نتایج الگوریتم کُمپاس اعتماد بیشتری داشتند تا مستنداتی که از دفاعیات و پیگیری‌های قضایی به دست آمده بود.(۲) آیا با گذشت زمان ما بیش از اندازه به هوش مصنوعی متکی خواهیم بود؟ الگوریتم کُمپاس به این دلیل هم به‌شدت مناقشه‌برانگیز است که تحقیقات نشان داده است مصادیق نتایج مثبت کاذبِ[1] الگوریتم (افرادی که الگوریتمْ ارتکاب مجدد جرم از سوی آن‌ها را پیش‌بینی کرده بود، ولی جرمی مرتکب نشدند) به طور نامتناسبی سیاهپوست بودند (Fry 2018). بنابراین، ممکن است هوش مصنوعی سوگیری[2] و تبعیض ناعادلانه را تشدید کند. در الگوریتم‌هایی که در فرایندهای تصمیم‌گیری دربارهٔ درخواست تسهیلات مالی و درخواست استخدام کاری به کار می‌روند ممکن است مشکلات مشابهی رخ دهد. به عنوان مثالی دیگر، کنترل پیشگیرانه[3] را در نظر بگیرید: از الگوریتم‌ها برای پیش‌بینی محدوده‌های وقوع جرم (مثلاً، کدام محله‌های یک شهر) و افرادی که احتمال دارد آن جرم‌ها را مرتکب شوند استفاده می‌کنند، اما خروجی

1. false positive 2. bias 3. predictive policing

این الگوریتم‌ها ممکن است گروه‌های اجتماعی-اقتصادی یا نژادی خاصی را شامل شود که به‌خطا تحت نظارت پلیس قرار بگیرند. در حال حاضر در آمریکا از الگوریتم‌های کنترل پیشگیرانه استفاده می‌شود، و طبق گزارش اخیر نهاد الگوریتم‌واچ[1] (2019) در اروپا نیز به کار برده شده است.[۳] فناوریِ تشخیصِ چهرهٔ مبتنی بر هوش مصنوعی نیز که اغلب برای مقاصد نظارتی استفاده می‌شود می‌تواند حریم خصوصی افراد را نقض کند. این فناوری همچنین می‌تواند گرایشات جنسی را نیز کمابیش پیش‌بینی کند. هیچ اطلاعاتی از تلفن شما و هیچ نوع دادهٔ زیست‌سنجی مورد نیاز نیست. دستگاه کارش را از دور انجام می‌دهد. با دوربین‌هایی که در خیابان‌ها و مکان‌های عمومی نصب شده، ما و حس و حالمان شناسایی و «خوانده» می‌شود. با تحلیل داده‌هایمان، بدون این‌که خودمان خبردار شویم، وضعیت سلامت روانی و جسمی‌مان پیش‌بینی می‌شود. کارفرماها می‌توانند از این فناوری برای نظارت بر عملکرد ما استفاده کنند. و الگوریتم‌هایی که در شبکه‌های اجتماعی فعال‌اند می‌توانند نفرت‌پراکنی کنند و اخبار نادرست پخش کنند؛ مثلاً، بات‌های سیاسی[2] می‌توانند در نقش کاربران واقعی ظاهر شوند و محتوای سیاسی تولید کنند. از موارد مشهور آن در سال ۲۰۱۶ چت‌باتِ[3] مایکروسافت به نام تای[4] است که قرار بود گفتگوهای سرگرم‌کننده در توئیتر داشته باشد، اما وقتی باهوش‌تر شد، شروع به توئیت مطالب نژادپرستانه کرد. برخی الگوریتم‌های هوش مصنوعی می‌توانند حتی سخنرانی‌های ویدئویی جعلی تولید کنند، مانند ویدئویی که به شکل گمراه‌کننده‌ای یک سخنرانی از باراک اوباما را جعل کرده بود.[۴]

1. AlgorithmWatch
2. political bots
3. chatbot
4. Tay

اهداف و نیات فناوری اغلب خیرخواهانه است. اما این مشکلات اخلاقی معمولاً پیامدهای ناخواستهٔ فناوری است. بسیاری از این موارد، مثل سوگیری یا نفرت‌پراکنی، از ابتدا مد نظر توسعه‌دهندگان یا کاربران آن فناوری‌ها نبودند. علاه بر این، پرسشی حیاتی که باید همواره مطرح شود این است که بهبود و ارتقای برآمده از هوش مصنوعی بهبود و ارتقای چه کسی است؟ حکومت یا شهروندان؟ پلیس یا کسانی که پلیس به‌شان مظنون است؟ فروشندگان یا مصرف‌کنندگان؟ قضات یا متهم؟ از سوی دیگر، پرسش‌های ناظر بر قدرت اثرگذاری هم مطرح می‌شوند، مثلاً زمانی که سمت‌وسوی کلی فناوری را فقط چند ابرشرکت تعیین می‌کنند (Nemitz 2018). چه کسی آیندهٔ هوش مصنوعی را رقم می‌زند؟

این سؤال اهمیت سیاسی و اجتماعی هوش مصنوعی را نشان می‌دهد. اخلاق هوش مصنوعی دربارهٔ تغییر فناورانه و تأثیرات آن در زندگی فردی انسان‌هاست، اما در عین حال به دگرگونی‌ها در اجتماع و اقتصاد نیز ناظر است. مسائل مربوط به سوگیری و تبعیض به‌روشنی نشان‌دهندهٔ جنبهٔ اجتماعی هوش مصنوعی است. اما هوش مصنوعی دارد اقتصاد را هم دگرگون می‌کند، که این امر خود ممکن است دگرگونی ساختارهای اجتماعی جامعه را نیز در پی داشته باشد. به عقیدهٔ اریک برینجلفسن[1] و اندرو مک‌آفی[2] (2014) ما وارد عصر ماشینی دوم شده‌ایم که در آن ماشین‌ها نه‌فقط مکمل انسان ــ مانند دوران انقلاب صنعتی ــ که جایگزین او هستند. از آن‌جایی که همهٔ حرفه‌ها و کارها تحت تأثیر هوش مصنوعی قرار خواهند گرفت، پیش‌بینی می‌شود جامعهٔ ما تغییر چشمگیری کند، زیرا فناوری‌هایی که زمانی در داستان‌های علمی‌ـتخیلی توصیف می‌شدند وارد دنیای

1. Erik Brynjolfsson
2. Andrew McAfee

واقعی می‌شوند (McAfee and Brynjolfsson 2017). آیندهٔ بازار کار چگونه است؟ وقتی هوش مصنوعی شغل‌ها را تصاحب کند، زندگی «ما» به چه شکل خواهد شد؟ و «ما» یعنی چه کسانی؟ چه کسی از این دگرگونی سود خواهد برد، و چه کسی متضرر خواهد شد؟

این کتاب

به دلیل دستاوردهای چشمگیر هوش مصنوعی، هیاهوی پردامنه‌ای پیرامون آن شکل گرفته است. در حال حاضر هوش مصنوعی در طیف گسترده‌ای از حوزه‌های دانش و مسائل انسانی عملی استفاده می‌شود. در حوزهٔ دانش، گمانه‌زنی‌هایی غیرعادی در خصوص آینده‌ای فناورمحور و بحث‌های فلسفی جالبی در مورد معنای انسان بودن شکل گرفته است. کاربرد هوش مصنوعی در مسائل انسانی عملی نیز نوعی احساس ضرورت در کنشگران اخلاق و سیاست‌گذاران ایجاد کرده است، تا اطمینان حاصل کنند که این فناوری نه عامل ایجاد مشکلات غیرقابل حل برای افراد و جوامع، که به سود ما خواهد بود. این نگرانی‌های اخیر عملی‌تر و فوری‌تر هستند.

این کتاب، که اثر فیلسوفی دانشگاهی و صاحب تجربه در امر مشاوره‌دهی به سیاست‌گذاران است، به هر دو جنبه می‌پردازد، و اخلاق را با تمامی سؤالات فوق مرتبط در نظر می‌گیرد. هدف کتاب فراهم کردن دیدی مناسب از مسائل اخلاقیِ هوش مصنوعی به معنای موسّع آن برای خواننده است، از روایت‌های اثرگذار در باب آیندهٔ هوش مصنوعی و سؤالات فلسفی در مورد ماهیت و آیندهٔ انسان، تا نگرانی‌های اخلاقی در مورد مسئولیت‌پذیری و سوگیری و ترسیم خط‌مشی‌هایی برای مواجهه با مسائل عملی دنیای واقعی که فناوری به بار آورده ــ با این ترجیح که قبل از آن‌که کار از کار بگذرد فکری شود.