



محمّدجواد استادی  
Info@ostadiofficial.ir

پژوهشگر مطالعات فرهنگی



**وقتی از «هوش مصنوعی مسئولانه» سخن می‌گوییم، در نهایت چه کسی مسئول است: طراحان، شرکت‌ها،**

**دولت‌ها یا خود سامانه‌ها؟**

مسئولیت در این زمینه میان همه ذی‌نفعان توزیع شده است. طراحان در قبال نحوه پیاده‌سازی و ساخت سامانه مسئول‌اند؛ شرکت‌ها در برابر چگونگی به‌کارگیری و استقرار آن مسئولیت دارند؛ دولت‌ها مسئول تدوین چارچوب‌های حکمرانی و تنظیم‌گری هستند؛ و کاربران نیز در قبال انتخاب خود برای استفاده از این فناوری مسئول‌اند. اما خود سامانه‌ها کنشگران اخلاقی نیستند. هوش مصنوعی هرگز موجود یا نهادی نیست که بتوان مسئولیت اخلاقی را به آن نسبت داد.

**آیا «مسئولیت» در حوزه هوش**

**مصنوعی، در درجه نخست مفهومی**

**اخلاقی است یا باید آن را در قالب**

**اصطلاحات و سازوکارهای عینی**

**مهندسی عملیاتی کرد؟**

هر دو. مسئولیت باید از طریق رویه‌ها و اقدامات عینی و مشخص عملیاتی شود؛ اقداماتی که شامل استانداردهای فنی، فرایندهای سازمانی و سازوکارهای پاسخ‌گویی است. با این حال، این عملیاتی‌سازی باید بر اصول اخلاقی استوار باشد. اخلاق یک نتیجه نهایی یا محصول آماده نیست، بلکه یک فرایند است؛ فرایندی برای فهم انتخاب‌هایی که انجام می‌دهیم و پیامدهایی که این انتخاب‌ها به همراه دارند.

**آیا تلاش برای تحقق هوش مصنوعی**

**مسئولانه، می‌تواند در نهایت نوآوری**

**را محدود کند؟**

خیر. هوش مصنوعی مسئولانه، برخلاف این تصور، امکان نوآوری پایدار را فراهم می‌کند؛ زیرا از آسیب‌هایی جلوگیری می‌کند که در غیر این صورت می‌توانند به واکنش‌های محدود کننده، مقررات سخت‌گیرانه‌تر یا فرسایش اعتماد عمومی منجر شوند. هیچ حوزه فناوری‌انه‌ای بدون نوعی تنظیم‌گری دوام نمی‌آورد؛ از ترافیک گرفته تا پزشکی و برق و هر نمونه دیگری که در نظر بگیریم. کسانی که خلاف این را ادعا می‌کنند، در واقع غالباً امیدوارند خودشان کسانی باشند که قواعد را تعیین می‌کنند؛ قواعدی که به سودشان تمام می‌شود. تنظیم‌گری مانع نوآوری نیست، بلکه سکوی آغاز و شرط لازم برای نوآوری پایدار است.

**آیا باور دارید که سامانه‌های هوش**

**مصنوعی ممکن است نوعی «اخلاق**

**درونی» پیدا کنند؟**

خیر. سامانه‌های هوش مصنوعی بر اساس بهینه‌سازی عمل می‌کنند، نه بر پایه استدلال اخلاقی. آنچه ما از آن با عنوان «اخلاق» در هوش مصنوعی یاد می‌کنیم، باید از طریق انتخاب‌های طراحی که انسان‌ها انجام می‌دهند در این سامانه‌ها گنجانده شود؛ نه این که خود سامانه‌ها مستقلاً واجد نوعی اخلاق درونی باشند.

**برخی منتقدان می‌گویند «هوش**

**مصنوعی مسئولانه» مفهومی بیش**

**از حد مبهم یا شعاری است. چگونه**

هوش مصنوعی دیگر صرفاً یک فناوری آینده‌نگرانه یا موضوعی محدود به آزمایشگاه‌های علوم رایانه نیست؛ امروز در متن زندگی اجتماعی، اقتصاد، سیاست‌گذاری عمومی، آموزش، پزشکی، رسانه و حتی داورهای اخلاقی ما حضور دارد. همین حضور گسترده، پرسشی بنیادین را پیش‌روی ما می‌گذارد: اگر هوش مصنوعی خطا کند، تبعیض بسازد، به اعتماد عمومی آسیب بزند یا تصمیم‌هایی بگیرد که بر زندگی انسان‌ها اثر بگذارد، چه کسی باید پاسخ‌گو باشد؟

پروفسور ویرجینیا دیگنوم، یکی از چهره‌های برجسته جهانی در حوزه هوش مصنوعی مسئولانه، دقیقاً در مرکز همین بحث ایستاده است. او استاد هوش مصنوعی

**می‌توان آن را به معیارهای قابل**

**سنجش تبدیل کرد؟**

این کار از طریق ارزیابی‌های اثرگذاری، ممیزی‌های الگوریتمی، پروتکل‌های مشارکت ذی‌نفعان، الزامات مربوط به مستندسازی و سازوکارهای مشخص

جبران و رسیدگی

به خسارت

امکان‌پذیر است.

هوش مصنوعی

مسئولانه صرفاً

مجموعه‌ای

از اصول کلی و

انتزاعی نیست،

بلکه باید در عمل و در

قالب فرایندها، معیارها و

سازوکارهای اجرایی مشخص،

عملیاتی شود.

**آیا چارچوب‌هایی که**

**در کتاب خود پیشنهاد**

**می‌کنید، به‌طور واقع‌بینانه**

**از سوی شرکت‌های بزرگ**

**فناوری قابل اجرا هستند،**

**یا تا حدی آرمان‌گرایانه‌اند؟**

این چارچوب‌ها واقع‌بینانه‌اند،

اما اجرای آن‌ها نیازمند

اراده نهادی است. آنچه این

چارچوب‌ها مطالبه می‌کنند،

صرفاً چند اصلاح فنی یا راه‌حل

تکنیکی محدود نیست، بلکه

نوعی تغییر سازمانی است؛

تغییری که دشوار است،

اما ضرورت دارد. علاوه بر

این، اجرای واقعی آن‌ها به

اراده سیاسی و سازوکارهای

دموکراتیک برای اعمال و

نظارت نیز نیاز دارد، که تحقق

آن حتی دشوارتر است. به همین

دلیل است که من از توافق‌های

جهانی بر سر اصول بنیادین دفاع

می‌کنم. همان‌طور که در مورد

ترافیک، لازم نیست همه کشورها دقیقاً از

قواعد یکسان پیروی کنند یا همه در یک سمت

جاده رانندگی کنند؛ مهم این است که در همه

جا قواعدی وجود داشته باشد و ابزارهایی

برای اجرای آن قواعد نیز در دسترس باشد.

**آیا تمرکز بر اخلاق، این خطر را ندارد که**

**توجه ما را از مسائل ساختاری، مانند**

**تمرکز قدرت و شکل‌گیری انحصارها در**

**صنعت فناوری، منحرف کند؟**

فکر نمی‌کنم چنین باشد؛ هر چند بسیار آسان است که بحث را در سطح همین انحراف‌ها و حاشیه‌ها نگه داریم و به جای اقدام واقعی، صرفاً درباره آن‌ها سخن بگوییم. کار اخلاقی

مسئولانه در دانشگاه اومئوسوئد و مدیر آزمایشگاه سیاست‌گذاری هوش مصنوعی در این دانشگاه است. دیگنوم همچنین عضو نهادهای مهم بین‌المللی در حوزه حکمرانی هوش مصنوعی بوده و در گفت‌وگوهای جهانی درباره نسبت فناوری، اخلاق، حقوق بشر و سیاست‌گذاری عمومی نقشی اثرگذار دارد.

کتاب او با عنوان هوش مصنوعی مسئولانه، از جمله آثار مهمی است که تلاش می‌کند بحث اخلاق هوش مصنوعی را از سطح شعارهای کلی فراتر ببرد و آن را به مسئله‌ای عملی، نهاده‌ای و قابل پیگیری تبدیل کند. دیگنوم در این گفت‌وگو بر نکته‌ای کلیدی تأکید می‌کند: ماشین‌ها مسئول نیستند؛ آن‌ها فهم اخلاقی ندارند و تصمیم اخلاقی

تکیه صرف بر خودتنظیم‌گری شرکت‌ها و واگذاری کامل مسئولیت به خود صنعت کافی نیست و نمی‌تواند تضمین کند که اصول مسئولیت‌پذیری در عمل رعایت شوند.

**آیا باید برای**

**هوش مصنوعی**

**استانداردهای**

**جهانی وجود**

**داشته باشد، یا**

**هر جامعه‌ای باید**

**چارچوب اخلاقی**

**خاص خود را**

**تعریف کند؟**

هر دو. از یک

سو، باید اصول

جهان‌شمول حقوق

بشر به‌عنوان مبنای

مشترک مورد توجه

قرار گیرد؛ اصولی که فراتر از

مرزهای ملی و تفاوت‌های فرهنگی

اعتبار دارند. از سوی دیگر، نحوه

اجرای این اصول باید متناسب با

زمینه‌های خاص هر جامعه باشد و

ارزش‌های محلی، شرایط فرهنگی

و چارچوب‌های حقوقی هر کشور را

بازتاب دهد.

**نقش نهادهایی مانند**

**اتحادیه اروپا را در حکمرانی**

**هوش مصنوعی چگونه**

**ارزیابی می‌کنید؟**

اتحادیه اروپا با پیشگامی در تدوین

قانون هوش مصنوعی و قرار دادن

حقوق بشر در مرکز بحث، چارچوب

درستی را برگزیده است. با این

حال، مسئله اصلی این است که

اجرای این چارچوب‌ها در عمل

مؤثر باشد و صرفاً به نوعی نمایش

تبعیت از مقررات یا اجرای صورتی

الزامات قانونی تبدیل نشود.

**آیا کشورهای در حال توسعه**

**در این گفت‌وگوهای جهانی**

**به حاشیه رانده می‌شوند؟**

نه لزوماً. روشن است که بخش بزرگی از

نمی‌گیرند. مسئولیت همچنان بر عهده انسان‌ها، نهاده‌ها، شرکت‌ها، دولت‌ها و سازوکارهای حکمرانی است. اهمیت دیدگاه او در این است که هوش مصنوعی مسئولانه را نه مانعی در برابر نوآوری، بلکه شرط نوآوری یادار می‌داند. از نظر او، تنظیم‌گری، شفافیت، ارزیابی اثرات، نظارت مستقل و پاسخ‌گویی دموکراتیک، دشمن فناوری نیستند؛ بلکه همان چیزهایی‌اند که می‌توانند از تبدیل هوش مصنوعی به ابزاری برای تمرکز قدرت، بی‌عدالتی و سلب مسئولیت انسانی جلوگیری کنند. این مصاحبه تلاشی است برای فهم این پرسش مهم که در عصر هوش مصنوعی، چگونه می‌توان هم از ظرفیت‌های فناوری بهره برد و هم کرامت، عدالت و مسئولیت انسانی را از یاد نبرد.

**پاسخ‌گویی چگونه ممکن است تحول پیدا کند؟**

پاسخ‌گویی باید از سطح فردی صرف فراتر برود و به امری سیستمی و ساختاری تبدیل شود. ما باید پاسخ‌گویی را در سراسر زنجیره‌های توسعه، طراحی، استقرار و به‌کارگیری سامانه‌ها بفهمیم؛ نه این که آن را فقط به کاربران نهایی یا یک کنشگر منفرد محدود کنیم.

**اگر یک سامانه هوش مصنوعی**

**تصمیمی «اخلاقی» بگیرد که با**

**ارزش‌های یک جامعه خاص در تعارض**

**باشد، چه کسی باید حرف نهایی را**

**بزند؟**

ماشین‌ها تصمیم‌های اخلاقی نمی‌گیرند. آن‌ها صرفاً بر اساس پارامترهایی که ما تعیین می‌کنیم، فرایند بهینه‌سازی را انجام می‌دهند. در نهایت، آنچه به آن نیاز داریم فرایند‌های حکمرانی دموکراتیک است؛ نه این که تصمیم‌گیری درباره مسائل ارزشی و اخلاقی را به نوعی تطبیق الگوهای فنی و محاسباتی فروبکاهیم.

**آیا هوش مصنوعی مسئولانه،**

**می‌تواند به ابزاری برای تحمیل**

**ارزش‌های فرهنگی خاص، برای مثال**

**هنجارهای غربی، در مقیاس جهانی**

**تبدیل شود؟**

ما پیشتر با چنین وضعیتی در شبکه‌های اجتماعی و اینترنت روبه‌رو بوده‌ایم. یعنی خطر تحمیل یا گسترش ارزش‌های فرهنگی خاص از طریق فناوری، پدیده‌ای تازه نیست و همین حالا نیز در زیرساخت‌های دیجیتال جهانی تجربه شده است.

**اگر در طراحی هوش مصنوعی ناچار**

**باشید یکی را در اولویت قرار دهید،**

**کارآمدی یا اخلاق را انتخاب می‌کنید؟**

**چرا؟**

این یک دوگانه نادرست است. سامانه‌هایی که اخلاق را نادیده می‌گیرند، در نهایت شکست می‌خورند؛ چه از طریق آسیب‌هایی که ایجاد می‌کنند، چه از مسیر دعاوی حقوقی و چه به دلیل از دست رفتن اعتماد عمومی. اخلاق در برابر کارآمدی قرار ندارد، بلکه امکان شکل‌گیری کارآمدی پایدار و قابل تکرار فراهم می‌کند.

**برخی خوانندگان معتقدند کتاب شما**

**چارچوب مفهومی نیرومندی ارائه**

**می‌کند، اما کمتر با چالش‌های فنی**

**اجرای آن درگیر می‌شود. آیا این را**

**محدودیتی برای کتاب می‌دانید؟**

این کتاب برای مخاطب عمومی نوشته شده و تمرکز اصلی آن بر چارچوب‌های حکمرانی هوش مصنوعی است. بنابراین، هدف کتاب ارائه یک راهنمای فنی یا دستورالعمل تخصصی برای اجرای جزئیات مهندسی نیست، بلکه می‌خواهد تصویری روشن‌تر از مسئولیت، حکمرانی و الزامات اجتماعی و نهادی هوش مصنوعی مسئولانه ارائه دهد.

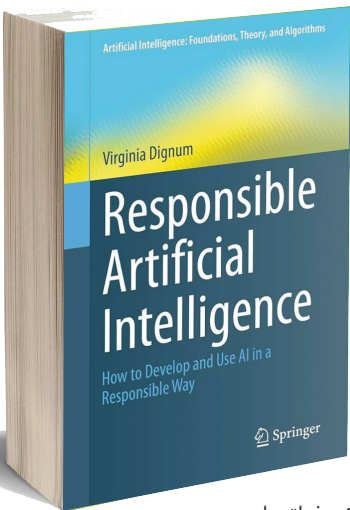
**اگر قرار بود امروز، با توجه به**

**پیشرفت‌های سریع هوش مصنوعی،**

**کتاب را دوباره بنویسید، چه چیزی را**

**اصلاح یا به آن اضافه می‌کردید؟**

هیچ چیز. ادعاهایی که من در کتاب مطرح کرده‌ام، بر موج‌های زودگذر، هیجان‌های مقطعی یا تبلیغات مربوط به فناوری‌های خاص بنا نشده‌اند؛ بنابراین با تغییر شتابان مصادیق و پیشرفت‌های جدید هوش مصنوعی نیز اصل آن استدلال‌ها نیازمند بازنویسی یا اصلاح اساسی نیست.



نمی‌گیرند.

**دیگنوم در**

**گفت‌وگوهای**

**جهانی درباره**

**نسبت فناوری،**

**اخلاق،**

**حقوق بشر و**

**سیاست‌گذاری**

**عمومی نقشی**

**اثرگذار دارد.**

**کتاب او با عنوان**

**هوش مصنوعی**

**مسئولانه، از آثار**

**مهمی است که**

**تلاش می‌کند**

**بحث اخلاق**

**هوش مصنوعی**

**را از سطح**

**شعارهای کلی**

**فراتر ببرد و آن**

**را به مسئله‌ای**

**عملی، نهادی**

**و قابل پیگیری**

**تبدیل کند**

مطرح کنند و مسئولیت را به ماشین‌ها نسبت

دهند. بزرگ‌ترین تهدید، تسلط ماشین‌ها بر

انسان نیست، بلکه کنار گذاشتن، فراموش

کردن یا انکار مسئولیت خود ماست. همچنین

خطر مهم دیگر این است که انسان‌ها از

ماشین‌ها برای وارد کردن آسیب و انجام

کارهای زیان‌بار استفاده کنند. خودکارسازی

تصمیم‌ها نباید به معنای خودکارسازی

پاسخ‌گویی باشد. انسان‌ها همچنان در

قبال طراحی سامانه‌ها و نحوه به‌کارگیری و

استقرار آن‌ها مسئول‌اند.

**در جهانی که با هوش مصنوعی**

**پیشرفته شکل می‌گیرد، مفهوم**

