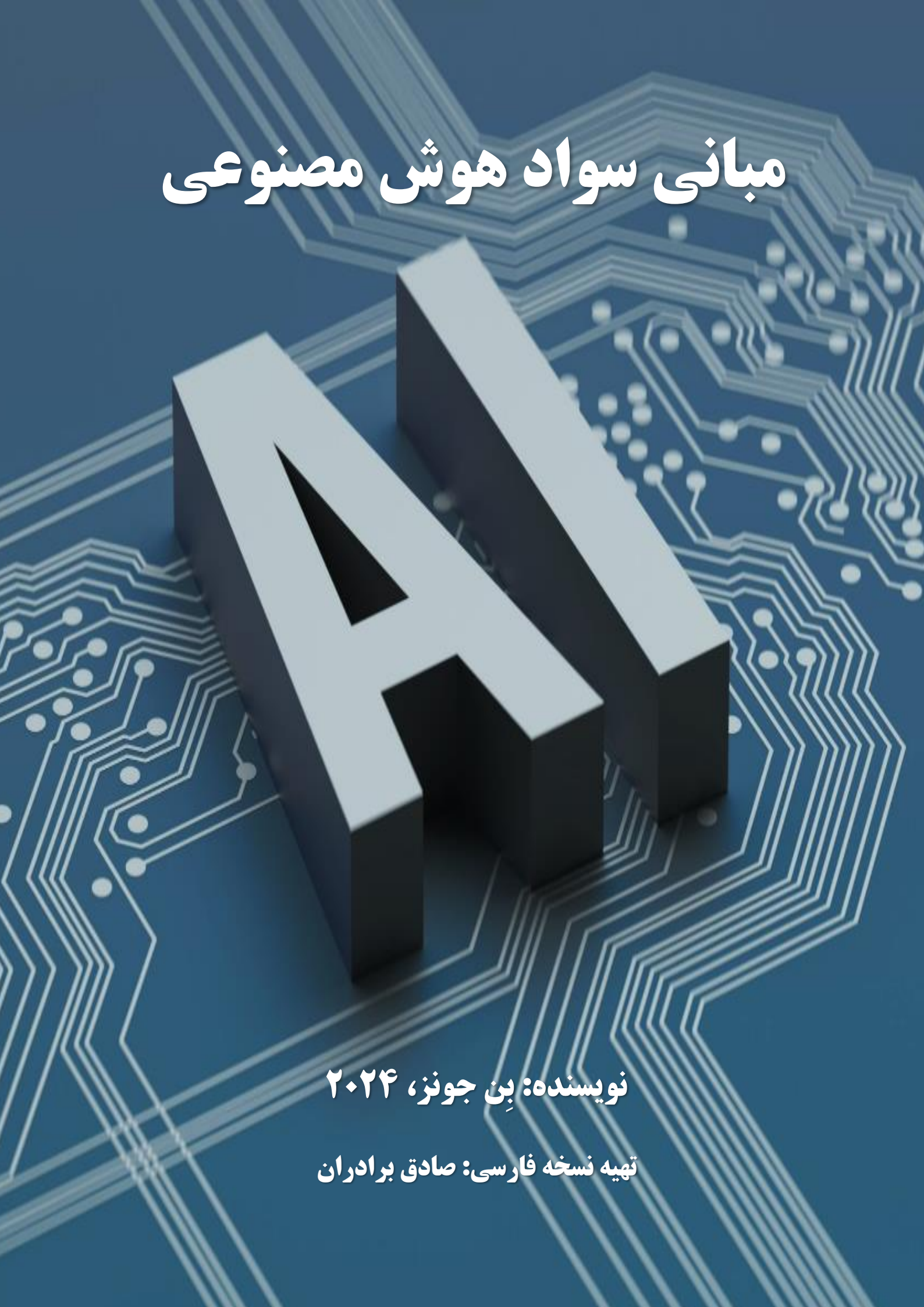
مبانی سواد هوش مصنوعی



نویسنده: بِن جونز، ۲۰۲۴

تهیه نسخه فارسی: صادق برادران

در دنیای پرشتاب امروز، کمتر روند فناورانه‌ای را می‌توان یافت که به اندازه هوش مصنوعی بر کسب‌وکار و زندگی ما تأثیرگذار باشد. این فناوری دیگر فقط یک موضوع تخصصی یا دانشگاهی نیست بلکه مفهومی است که همه ما ناگزیر به شناخت و یادگیری آن هستیم. اما این یادگیری دقیقاً باید شامل چه چیزهایی باشد؟ پاسخ بسیاری از متخصصان در وهله اول، ارتقای سواد هوش مصنوعی است یعنی توانایی درک، تحلیل، ارزیابی و استفاده آگاهانه از هوش مصنوعی. هر فردی باید دارای یک سطح قابل قبول از سواد هوش مصنوعی باشد و بعد در صورت لزوم، دانش و مهارت خود را در ابعاد خاصی افزایش دهد.

کتابی که پیش رو دارید، دقیقاً در همین زمینه سواد هوش مصنوعی نوشته شده است. در آن تلاش شده تا تصویری همه‌جانبه از هوش مصنوعی ارائه شود؛ از مفاهیم بنیادی و فنی گرفته تا پیامدهای اجتماعی. با این حال، اگر اصلا تمایلی به شناخت جوانب فنی ندارید می‌توانید از فصل‌های سه و چهار صرف نظر کنید.

در نسخه فارسی این کتاب برای راحتی بیشتر، تلاش کرده‌ام واژه‌ها و مفاهیم کلیدی را **پررنگ** کنم تا مرور آن در تلفن همراه ساده‌تر باشد. همچنین تصاویرِ کمکی زیادی به متن اضافه کرده‌ام تا مفاهیم، شفاف‌تر و ملموس‌تر باشند. لازم می‌دانم از سرکار خانم اسرا حوائجی که نسخه نهایی کتاب را بازخوانی کردند و تعدادی پیشنهاد اصلاحی ارائه کردند تشکر کنم. امیدوارم مطالعه این کتاب برایتان مفید باشد.

صادق برادران

مشاور و استراتژیست توسعه کسب‌وکار، نوآوری و تحول دیجیتال

مدرس کارگاه‌های «AI کاربردی»، «از استراتژی تا KPI و OKR» و «نوآوری»

ارشد و دکترای فن‌آفرینی (Technopreneurship PhD) دانشگاه تهران

دارای پروانه مشاوره و نظارت بر فناوری اطلاعات و ارتباطات نظام صنفی رایانه‌ای

Baradaran@managinno.ir | m.s.baradaran@Gmail.com

فهرست

مقدمه

در دوران شگفت‌انگیزی از تاریخ زندگی می‌کنیم. برای ما بزرگ‌سالان، فناوری در چند دهه‌ی نخست قرن بیست‌ویکم نسبت به زمانی که به دنیا آمدیم پیشرفت چشم‌گیری داشته است. انقلاب **رایانه‌های شخصی** در دهه ۱۹۸۰، زمینه‌ساز گسترش سریع **اینترنت** در دهه ۱۹۹۰ و اوایل دهه ۲۰۰۰ شد. از آغاز قرن حاضر، پیشرفت هم‌زمان در **تلفن‌های همراه**، قدرت پردازش رایانه‌ای و ذخیره‌سازی دیجیتال، باعث **انفجار اطلاعات** شده است. این اطلاعات و داده‌ها، دنیای ما را به‌طور بنیادی دگرگون کرده‌اند. شیوه‌ی کار کردن ما، دادوستدها، تعاملات اجتماعی و حتی سبک زندگی‌مان در جهان امروز، بسیار متفاوت از دوران والدین و پدربزرگ‌ها و مادربزرگ‌هایمان است. بنابراین **جای تعجب ندارد که نسل‌های پیشین، واقعاً نمی‌دانستند چطور ما را برای دنیای امروز آماده کنند.** به همین دلیل، ما اکنون با **چالش درک مفاهیم و کسب مهارت‌ها**یی روبرو هستیم که در سال‌های تحصیل رسمی‌مان **آموزش ندیده‌ایم.** اگر خود ما، از عهده‌ی این کار برنیاییم، عقب خواهیم ماند.

شتاب روزافزون و پیامدهای بزرگ‌تر

هرچند فناوری از دوران کودکی ما تاکنون بسیار پیشرفت کرده اما در حال حاضر سرعت بسیار بیشتری گرفته است. ما **در مراحل ابتدایی یک انقلاب در حوزه هوش مصنوعی (AI) قرار داریم**؛ انقلابی که بخشی از نیروی محرکه‌ی آن، **افزایش چشمگیر داده‌های در دسترس** و **توان پردازش رایانه‌ای** است. اگر این شرایط محیطی را با پیشرفت‌های صورت‌گرفته در یادگیری ماشین (شاخه‌ای از هوش مصنوعی که بر توانایی رایانه‌ها در یادگیری از داده تمرکز دارد) ترکیب کنیم، درمی‌یابیم که امروزه هوش مصنوعی می‌تواند کارهایی انجام دهد که نیازمند سطح قابل‌توجهی از هوش هستند.

با وجود توانمندی‌های چشمگیر هوش مصنوعی و ایجاد دگرگونی‌های عمده در صنایع و جوامع مختلف، **این فناوری بی‌نقص نیست.** مدل‌ها و برنامه‌های هوش مصنوعی هنوز خطاهای زیادی دارند و همیشه منصفانه عمل نمی‌کنند. **بسیاری از افراد هنوز با این فناوری آشنایی ندارند و بسیاری نیز با آن احساس راحتی نمی‌کنند.** این افراد، پست‌ها و دیدگاه‌های منتشرشده در شبکه‌های اجتماعی درباره‌ی هوش مصنوعی را می‌خوانند، گزارش‌های خبری مربوط به آن را می‌بینند و صحبت‌های همکارانشان را درباره AI می‌شنوند، اما به اندازه‌ی کافی احساس اعتماد به نفس در خود نمی‌یابند که وارد این گفتگوها شوند.

این کتاب برای چه کسانی است؟

این کتاب برای همه افرادی است که می‌خواهند وارد گفت‌وگوهای مربوط به هوش مصنوعی شوند؛ چه این گفت‌وگوها در محل کار آنها باشد، چه در شبکه‌های اجتماعی یا در جامعه‌ای که در آن زندگی می‌کنند. این کتاب برای کسانی است که می‌خواهند درباره‌ی هوش مصنوعی یاد بگیرند تا بتوانند از قدرت آن بهره ببرند و در عین حال از آسیب‌های احتمالی‌اش دوری کنند. **این کتاب، گام نخست و حیاتی در مسیر یادگیری پیوسته درباره‌ی هوش مصنوعی است.**

اهمیت پیوستن به گفت‌وگوها درباره هوش مصنوعی، تنها به آینده‌ی شغلی شما محدود نمی‌شود. برای همه‌ی ما مهم است که تا جای ممکن افراد بیشتری درباره‌ی هوش مصنوعی آموزش ببینند تا بتوانیم دیدگاه‌ها، دغدغه‌ها و ایده‌هایمان را با یکدیگر در میان بگذاریم. **آینده‌ی هوش مصنوعی در دستان ماست و بهترین مسیر پیش‌رو، مسیری است که با همکاری افراد با پیشینه‌ها، تخصص‌ها و دیدگاه‌های گوناگون**

ساخته شود. هوش مصنوعی‌ای که قرار است برای همه‌ی ما کار کند، باید به دست همه‌ی ما ساخته شود.

محتوای این کتاب چیست؟

بخش اول: آشنایی با هوش مصنوعی

کتاب از سه بخش تشکیل شده که هر کدام شامل دو فصل است. در **فصل نخست**، با عنوان **آشنایی با هوش مصنوعی**، با یک پرسش ساده اما شگفت‌انگیز آغاز می‌کنیم: «هوش مصنوعی چیست؟». تعریف‌ها اهمیت دارند اما همان‌طور که خواهید دید، **تعریف هوش مصنوعی در طول زمان تغییر کرده** و حتی **امروز هم به‌سختی می‌توان آن را به‌روشنی بیان کرد.** از طرفی، انواع و سطوح مختلف هوش مصنوعی ممکن است گیج‌کننده باشند؛ به همین دلیل، در این فصل، **اصطلاحات و اختصارات رایج** را با هم مقایسه می‌کنیم تا بتوانید آن‌ها را به‌درستی درک و از هم تفکیک کنید. در پایان این فصل، با معرفی چند کاربرد آشنای روزمره از هوش مصنوعی، دید روشن‌تری نسبت به موضوع به دست می‌آوریم.

در **فصل دوم**، سفری کوتاه اما مفید در **تاریخ هوش مصنوعی** خواهیم داشت. این فصل از آغازهای پرشور این حوزه در میانه قرن بیستم شروع می‌شود، به دوره‌های رکود موسوم به «**زمستان‌های هوش مصنوعی**» می‌پردازد و در نهایت به انقلاب کنونی در حوزه‌ی هوش مصنوعی در دهه‌های ابتدایی قرن ۲۱ می‌رسد. در این مسیر با برخی پیشگامان این حوزه آشنا می‌شوید، به نقاط عطف مهم آن توجه می‌کنید و بینشی ارزشمند از روند تکامل این فناوری به دست می‌آورید.

بخش دوم: فناوری‌های هوش مصنوعی

در بخش دوم کتاب، تمرکز بیشتری بر شاخه‌ای از هوش مصنوعی خواهیم داشت که تقریباً پشت همه‌ی جهش‌های اخیر در این حوزه بوده است: **یادگیری ماشین**. در **فصل سوم**، با **مبانی یادگیری ماشین** آشنا می‌شوید؛ یعنی این‌که **به‌جای ارائه‌ی دستورالعمل‌های گام‌به‌گام به رایانه**، چگونه می‌توان **با استفاده از داده‌ها به آن آموزش داد** تا کارهای هوشمندانه‌ای انجام دهد. همچنین با رایج‌ترین روش‌های آموزش رایانه‌ها آشنا خواهید شد.

در **فصل چهارم**، وارد دنیای **یادگیری عمیق** می‌شویم که **شاخه‌ای تخصصی از یادگیری ماشین** مبتنی بر **شبکه‌های عصبی عمیق** است. این مدل‌های قدرتمند شامل لایه‌های متعدد از واحدهای به‌هم‌پیوسته (نودها) هستند که از عملکرد نورون‌های زیستی در مغز انسان الهام گرفته شده است. این شبکه‌ها، ورودی‌های دیجیتال را دریافت، آن‌ها را با وزن‌ها یا پارامترها ضرب کرده، و سیگنال حاصل را به لایه‌ی بعدی منتقل می‌کنند. با تنظیم این وزن‌ها در فرآیند آموزش، شبکه می‌تواند وظایفی را انجام دهد که زمانی برای رایانه‌ها بسیار دشوار به نظر می‌رسیدند؛ وظایفی مانند تشخیص چهره یا ترجمه‌ی متون بین زبان‌های مختلف.

بخش سوم: ملاحظات مهم در هوش مصنوعی

بخش سوم، تمرکز را از جنبه‌های فنی به **ابعاد فردی و اجتماعی این فناوری** معطوف می‌کند. در **فصل پنجم**، به مزایای شگفت‌انگیز هوش مصنوعی می‌پردازیم و همچنین نگرانی‌های جدی آن را بررسی می‌کنیم؛ از **ترس در مورد از بین رفتن مشاغل** گرفته تا **تأثیرات ناعادلانه‌ی تعصب‌های الگوریتمی**. هوش مصنوعی مجموعه‌ای قدرتمند از

فناوری‌هاست که می‌تواند به شیوه‌های مختلف به ما کمک کند. اما بسته به نحوه‌ی استفاده از آن، ممکن است آسیب‌های جدی نیز به همراه داشته باشد. در این فصل بررسی می‌کنیم که چطور می‌توانیم از مزایای هوش مصنوعی بهره‌مند شویم، در حالی که با چالش‌های آن نیز به‌درستی مواجه شویم.

فصل ششم و پایانی درباره‌ی **تمایز قائل شدن میان واقعیت و افسانه** است. باورها و تصورات نادرست بسیاری درباره‌ی هوش مصنوعی در فضای مجازی و دنیای واقعی جریان دارد. برای اینکه بتوانیم در این گفتگوها مشارکت مؤثر داشته باشیم، باید دیدگاه‌های افراطی موجود را بشناسیم تا بتوانیم **حقیقت‌های معقول و متعادلی** را بیان کنیم.

سواد داده و سواد هوش مصنوعی

سواد هوش مصنوعی (AI Literacy) یعنی **توانایی شناخت، درک، استفاده و ارزیابی انتقادی فناوری‌های هوش مصنوعی و تأثیرات آن‌ها.** این مفهوم چه تفاوتی با سواد داده دارد؟ سواد داده (Data Literacy) به‌طور کلی به توانایی خواندن، درک، تولید و انتقال داده‌ها به‌عنوان اطلاعات معنادار گفته می‌شود. می‌توان گفت سواد هوش مصنوعی و سواد داده، مانند خواهر و برادر هستند. اما نکته‌ی مهم اینجاست: سواد واقعی در هوش مصنوعی بدون سواد داده ممکن نیست. دلیلش هم روشن است: **اساس هوش مصنوعی بر داده استوار است و به‌شدت از داده‌ها تأثیر می‌پذیرد.** چگونه ممکن است کسی بتواند یک مدل هوش مصنوعی را درک کند در حالی که درکی پایه‌ای از داده‌هایی که آن مدل با آن‌ها آموزش دیده ندارد؟ به همین دلیل، به خوانندگان این

کتاب توصیه‌ی اکید می‌کنم که زمانی را نیز به تقویت بنیان خود در زمینه‌ی سواد داده اختصاص دهند.

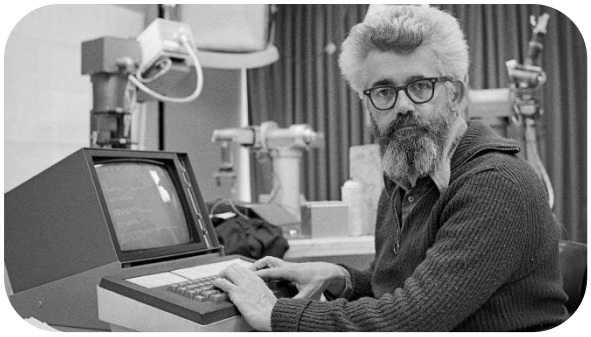
سرمایه‌گذاری در هر دو زمینه‌ی سواد داده و سواد هوش مصنوعی، بازدهی ارزشمندی خواهد داشت. این دو مهارت مکمل یکدیگرند و در دنیای امروز (فارغ از شغل و حرفه) باید بتوانیم زبان داده و هوش مصنوعی را بفهمیم و با آن صحبت کنیم. همان‌طور که برای یادگیری هر زبان خارجی، باید ابتدا واژگان و عبارات و معانی آن‌ها را فرا گرفت، در مسیر یادگیری زبان هوش مصنوعی نیز همین اصل صادق است. امید قلبی من این است که کتاب *مبانی سواد هوش مصنوعی* به شما کمک کند تا بنیانی محکم در زبان هوش مصنوعی بنا کنید و در سال‌های آینده بتوانید این بنیان را گسترش دهید.

فصل اول

هوش مصنوعی چیست؟

هوش مصنوعی یا AI چیست؟ شاید این پرسش، تصاویری از ربات‌های فیلم‌ها و کتاب‌های علمی‌تخیلی را به ذهن‌تان بیاورد؛ برخی دوست‌داشتنی و برخی تهدیدآمیز. شاید یاد مقاله‌ای خبری بیفتید درباره‌ی یک پیشرفت بزرگ فناورانه، یا شاید هم فقط به یاد هیاهوی تبلیغاتی و گفت‌وگوهای پرشمار در شبکه‌های اجتماعی بیفتید که این روزها مدام درباره‌ی این واژه صحبت می‌کنند. بحث هوش مصنوعی همه‌جا مطرح است، اما **دقیقاً منظور چیست؟**

تعریف هوش مصنوعی



جالب است که **پاسخ دادن به این سؤال**، دست‌کم به شکلی که برای همه قابل قبول باشد، **دشوار** است. فردی که در سال ۱۹۵۵ این اصطلاح را برای نخستین‌بار به‌کار برد، یعنی **جان مک‌کارتی**، استاد دانشگاه استنفورد، آن را چنین تعریف کرد: «هوش مصنوعی، علم و مهندسی ساخت **ماشین‌های هوشمند** است به‌ویژه **برنامه‌های کامپیوتری هوشمند**».

این تعریف از این جهت مفید است که روشن می‌کند منظور مک‌کارتی از واژه‌ی «مصنوعی» چه بوده است. به‌طور خلاصه، منظور او «**کامپیوتری**» بوده است. با این حال، توجه کنید که واژه‌ی «هوشمند (intelligent)» را دو بار در همین تعریف آورده و همین‌جاست که چالش اصلی نمایان می‌شود: ما همیشه منظور یکسانی از «**هوشمند بودن**» نداریم.

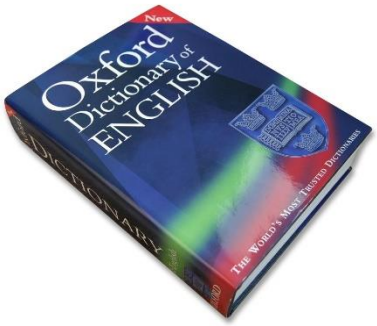
برای مثال، وقتی کودکی در یک آزمون دشوار ریاضی نمره‌ی عالی می‌گیرد، او را باهوش می‌دانیم. وقتی سگی بالا بودن قند خون صاحبش را تشخیص می‌دهد، می‌گوییم هوش قابل‌توجهی دارد. حتی برخی می‌گویند درختان هم شکلی از هوش دارند؛ مثلاً وقتی **درخت آکاسیا** در واکنش به جویده شدن برگ‌هایش توسط زرافه، گازی آزاد می‌کند که درختان مجاور را تحریک می‌کند تا مواد تلخ‌کننده در برگ‌های خود ترشح کنند. همه‌ی این رفتارها، در دسته‌ی **رفتارهای «هوشمندانه»** قرار می‌گیرند، اما طبیعتاً **با هم بسیار متفاوت‌اند.** اگر در تعریف دقیق هوش برای موجودات زنده به مشکل می‌خوریم، پس طبیعی است که برای عوامل غیرزیستی (مثل کامپیوتر) هم این دشواری وجود داشته باشد.

کدام رفتار یک کامپیوتر را «هوشمندانه» می‌دانیم؟ آیا باید بتواند به سؤالات خاصی به‌درستی پاسخ دهد یا چند دقیقه با ما گفت‌وگو کند؟ آیا باید بتواند بدون تصادف، ماشینی را در شهر براند یا چهره‌ی ما را در عکس بشناسد یا در بازی شطرنج، استادبزرگ‌ها را شکست دهد؟ آیا باید بتواند موسیقی یا شعر یا نقاشی تأثیرگذار خلق کند، یا جوکی بامزه؟ یا باید همه‌ی این کارها را انجام دهد، نه فقط یکی از آن‌ها را؟

اگر همین سؤال را از دوستان یا همکاران‌تان بپرسید، احتمالاً با پاسخ‌های متنوعی روبرو خواهید شد. برای انسان‌ها، حیوانات و کامپیوترها، انواع و سطوح مختلفی از هوش وجود دارد. و همین موضوع، تعریف دقیق هوش مصنوعی را به امری دشوار تبدیل می‌کند؛ چنان‌که حتی یک کمیته‌ی چنددهنفره از متخصصان این حوزه هم در سال ۲۰۱۶

اذعان کرد که «تعریفی دقیق و مورد پذیرش همگانی برای AI وجود ندارد». با این حال، بیایید نگاهی به چند تعریف موجود از هوش مصنوعی بیندازیم و ببینیم چه چیزهایی می‌توان از آن‌ها برداشت کرد.

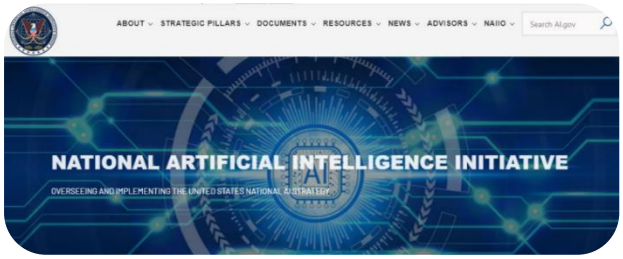
اول، سراغ تعریف واژه‌نامه برویم. «فرهنگ لغت آکسفورد» هوش مصنوعی را این‌گونه تعریف می‌کند:



توانایی کامپیوترها یا سایر ماشین‌ها برای بروز یا **شبیه‌سازی رفتار هوشمندانه**؛ یا حوزه‌ای از مطالعات که به این موضوع می‌پردازد.

این تعریف، تفاوت زیادی با تعریف مک‌کارتی ندارد و همان پرسش کلیدی را پیش می‌کشد: **رفتار هوشمندانه چیست؟** با این حال، نکته‌ی مفیدی در آن هست: هوش مصنوعی فقط نام نوعی از سیستم‌های کامپیوتری نیست بلکه **نام حوزه‌ای تخصصی نیز هست**. بنابراین، در دانشگاه‌ها، دپارتمان‌های AI داریم؛ در شرکت‌ها، تیم‌های AI داریم؛ و در نهادهای دولتی، کارشناسان هوش مصنوعی.

بیایید نگاهی به تعریفی بیندازیم که یکی از نهادهای دولتی ارائه کرده است. دولت‌های سراسر جهان، طبیعی است که به موضوع هوش مصنوعی علاقه‌مند باشند و دولت آمریکا هم از این قاعده مستثنی نیست. آنتونی بلینکن، وزیر امور خارجه‌ی ایالات متحده، گفته است: «**یک انقلاب جهانی در فناوری آغاز شده است**».



وقتی کنگره‌ی آمریکا «**قانون ابتکار ملی هوش مصنوعی**» را در سال ۲۰۲۰ تصویب کرد، در آن چنین آمده بود:

> اصطلاح هوش مصنوعی به سامانه‌ای ماشینی اطلاق می‌شود که می‌تواند بر اساس مجموعه‌ای از اهداف تعریف‌شده توسط انسان، اقدام به پیش‌بینی، پیشنهاد یا تصمیماتی کند که بر محیط‌های واقعی یا مجازی اثر گذارند. این سامانه‌ها از ورودی‌های ماشینی و انسانی استفاده می‌کنند تا:
>
> الف) محیط‌های واقعی و مجازی را درک کنند؛
> ب) این درک را به‌صورت خودکار تحلیل کرده و به مدل تبدیل کنند؛ و
> ج) با استفاده از این مدل‌ها، گزینه‌هایی برای اطلاعات یا اقدامات مختلف ارائه دهند.

این تعریف، جزئیات زیادی به بحث اضافه می‌کند. جالب اینجاست که اصلاً از واژه‌ی «هوشمند» استفاده نکرده بلکه مستقیماً به رفتارهایی مانند «ادراک»، «مدل‌سازی» و «ارائه‌ی گزینه‌ها» اشاره دارد. این افعال کمک می‌کنند تصویر روشن‌تری از منظور ما از هوش مصنوعی شکل بگیرد هرچند همین افعال هم سؤالات فلسفی دشواری را به همراه دارند. برای نمونه، «ادراک» یعنی چه؟ آیا کافی است سامانه‌ی ماشینی صرفاً تغییری را در محیط از طریق دوربین یا میکروفون ثبت کند یا باید آگاهی نسبت به این تغییر داشته باشد؟ این سؤال ما را به قلمرو فلسفه ذهن و بحث آگاهی می‌کشاند. حوزه‌ای شگفت‌انگیز که در این کتاب وارد آن نخواهیم شد (خود آگاهی انسان هنوز هم یکی از معماهای اصلی دانش است).

بیایید ببینیم خود یک AI چطور خودش را تعریف می‌کند! من این سؤال را از یکی از پیشرفته‌ترین چت‌بات‌ها یعنی ChatGPT شرکت OpenAI (یک سازمانی تحقیقاتی در زمینه‌ی هوش مصنوعی که در سال ۲۰۱۵ تأسیس شده و دفتر مرکزی آن در سان‌فرانسیسکو است) پرسیدم و پاسخ زیر را دریافت کردم:

> هوش مصنوعی به شبیه‌سازی هوش انسانی در ماشین‌هایی اطلاق می‌شود که طراحی شده‌اند تا بیندیشند، یاد بگیرند و تصمیم‌گیری یا اقداماتی انجام دهند که **معمولاً نیاز به هوش انسانی دارند.** این شامل وظایفی مانند حل مسئله، شناسایی الگوها، درک زبان طبیعی و ادراک می‌شود. سیستم‌های AI می‌توانند از روش‌های مختلفی مانند روش‌های مبتنی بر قواعد، یادگیری ماشین و یادگیری عمیق برای انجام این وظایف استفاده کنند. هدف نهایی AI، ساخت سامانه‌هایی است که بتوانند به‌طور مستقل و تطبیق‌پذیر، وظایف پیچیده را انجام دهند و در طول زمان با تجربه یا داده‌های بیشتر، عملکردشان را بهبود دهند.

این پاسخ، اطلاعات زیادی دارد. ولی AI ای که این تعریف را ارائه داد چیست؟

نسخه‌ای از ChatGPT که من با آن صحبت کردم، از مدل GPT-4استفاده می‌کرد؛ نوعی «مدل زبانی بزرگ» که در ادامه‌ی کتاب درباره‌اش بیشتر خواهید خواند. GPT مخفف «تبدیل‌گر پیش‌آموزش‌دیده‌ی مولد (Generative Pre-trained Transformer)» است. این مدل با حجم عظیمی از متون اینترنتی و کتاب‌ها آموزش دیده تا بتواند **واژه‌ی بعدی در یک جمله را پیش‌بینی کند** و همین باعث شده در گفت‌وگوهای متنی، عملکردی تحسین‌برانگیز داشته باشد، حداقل با استانداردهای امروز. در واقع، ChatGPT را می‌توان

تا حدودی نوعی «خلاصه‌کننده‌ی متون» دانست. البته خروجی‌های آن تحت تأثیر بازخورد انسانی در مراحل توسعه‌اش بوده و همچنین محدودیت‌هایی موسوم به «ضمانت‌ها (safeguards)» دارد تا از تولید محتوای نامطلوب، مانند مطالب نفرت‌انگیز یا اطلاعات حساس (مثل شماره گذرنامه یا نشانی شخصی)، جلوگیری کند.

اما برگردیم به تعریفی که ارائه داد؛ آیا موفق شد خودش را به‌درستی تعریف کند؟ می‌توان گفت عملکرد قابل قبولی داشت، هرچند خودش هم درگیر همان بحث‌های فلسفی رایج شد. البته برخی اصطلاحات مهم مانند «یادگیری ماشین» و «یادگیری عمیق» را وارد بحث کرد اما توضیحی درباره‌ی آن‌ها نداد. وقتی از همان مدل خواستم که تعریفی خلاصه‌تر بدهد، چنین پاسخ داد:

> هوش مصنوعی، توسعه‌ی سامانه‌های کامپیوتری است که می‌توانند وظایفی را انجام دهند که معمولاً به هوش انسانی نیاز دارند مانند ادراک بصری، شناسایی گفتار، تصمیم‌گیری و ترجمه‌ی زبان.

باز هم در هر دو تعریف، چه مفصل و چه خلاصه، واژه‌ی «هوش» برای تعریف هوش استفاده شده است. به‌ویژه در این تعاریف، هوش مصنوعی را با هوش انسانی مرتبط دانسته‌اند. از نگاه GPT-4، AI طراحی شده تا جنبه‌هایی از هوش انسانی را شبیه‌سازی کند و وظایفی را انجام دهد که معمولاً انسان‌ها قادر به انجام آن‌ها هستند. اما همان‌طور که این فناوری‌ها پیشرفت می‌کنند و سیستم‌هایی پدید می‌آیند که می‌توانند فراتر از توانایی‌های انسان عمل کنند، باید دید که **آیا تعاریف آینده‌ی هوش مصنوعی همچنان بر ارتباط با هوش انسانی تأکید خواهند داشت یا نه.** اگر AI بتواند وظایفی را بهتر و سریع‌تر از ما انجام دهد، مثل خلاصه‌سازی متون، شناسایی الگوها در حجم

عظیمی از داده‌ها یا شکست دادن ما در بازی‌های پیچیده، آیا همچنان باید آن را «شبیه‌سازی هوش انسانی» بنامیم یا باید آن را شکلی مستقل از هوش در نظر گرفت؟

اثر هوش مصنوعی

در مسیر تحول هوش مصنوعی، نکته‌ی مهمی وجود دارد که می‌تواند به درک بهتر ما از **دشواری همیشگی تعریف این مفهوم** کمک کند؛ پدیده‌ای که «اثر هوش مصنوعی (AI effect)» نام دارد و گاهی از آن با عنوان «**پارادوکس هوش مصنوعی**» نیز یاد می‌شود. خلاصه‌ی این پدیده چنین است: **به‌محض آن‌که مسئله‌ای که پیش‌تر حل آن نیازمند هوش انسانی به‌نظر می‌رسید، توسط یک کامپیوتر حل می‌شود، ما دیگر آن را نشانگر «هوش واقعی» نمی‌دانیم؛** و در نتیجه، دیگر راه‌حل کامپیوتری‌اش را هم جزو مصادیق AI محسوب نمی‌کنیم. این وضعیت کلاسیک را می‌توان «**جابه‌جا شدن دروازه‌ها**» نامید.

یکی از نمونه‌های رایج، بازی **شطرنج** است؛ زمانی این بازی، آزمون نهایی هوش تلقی می‌شد. تصور می‌شد که اگر کامپیوتری بتواند قهرمان جهانی شطرنج را شکست دهد، بی‌تردید باید بسیار هوشمند باشد. و خب، دقیقاً همین اتفاق افتاد: در سال ۱۹۹۷، ابررایانه‌ی دیپ‌بلو (Deep Blue) شرکت IBM توانست گری کاسپاروف، قهرمان وقت جهان را با نتیجه‌ی ۳٫۵ به ۲٫۵ در نیویورک شکست دهد. در آن زمان، پیروزی دیپ‌بلو را نقطه‌عطفی بزرگ در تاریخ هوش مصنوعی دانستند. اما مدت کوتاهی بعد، منتقدان و تحلیل‌گران شروع

کردند به کوچک شمردن آن دستاورد و این‌گونه توصیفش کردند: صرفاً استفاده‌ای قدرتمند از توان پردازشی بالا، همراه با یک پایگاه‌داده‌ی عظیم از حرکات ممکن. حتی خود شطرنج هم بی‌اعتبار شد. آن را نه فعالیتی نیازمند هوش واقعی، بلکه مسئله‌ای قابل حل با محاسبات پرتکرار توصیف کردند. انگار که «**خط معیار» هوش دوباره بالاتر رفته بود**. به‌عنوان نمونه، جف هاوکینز، عصب‌شناس در کتابش درباره‌ی هوش: چگونه درک تازه‌ای از مغز به ساخت ماشین‌های واقعاً هوشمند می‌انجامد چنین نوشت:

> دیپ‌بلو باهوش‌تر از انسان نبود که برنده شد؛ بلکه میلیون‌ها برابر از انسان سریع‌تر بود... شطرنج بازی کرد ولی **شطرنج را درک نکرد.** همان‌طور که یک ماشین‌حساب عملیات ریاضی را انجام می‌دهد ولی ریاضی را نمی‌فهمد.

حتی خود کاسپاروف هم در کتابش «تفکر عمیق: جایی که هوش ماشینی پایان می‌یابد و خلاقیت انسانی آغاز می‌شود» دیدگاهی مشابه داشت: دیپ‌بلو به اندازه‌ی ساعت زنگ‌دار برنامه‌پذیر، هوشمند بود. البته شکست خوردن از یک ساعت زنگ‌دار ده‌میلیون‌دلاری، حس خوبی نداشت.

نمونه‌های مشابهی را پس از پیشرفت‌هایی مانند تشخیص گفتار یا تشخیص نوری نوشتار (OCR) شاهد بوده‌ایم. امروزه تقریباً تمام گوشی‌های هوشمند و دستیارهای صوتی (مانند الکسای آمازون) می‌توانند گفتار را تشخیص دهند. یا مثلاً وقتی با موبایل از یک چک بانکی عکس می‌گیرید و آن را آنلاین واریز می‌کنید، دارید از فناوری OCR استفاده می‌کنید. این قابلیت‌ها، **زمانی از جمله‌ی چالش‌های بزرگ هوش مصنوعی محسوب می‌شدند** اما امروز بسیاری باور دارند که **این‌ها فقط کاربردهای روزمره‌اند** و نه مصادیق واقعی هوش مصنوعی.

جای تعجب نیست اگر نسل‌های آینده نیز نسبت به دستاوردهایی که امروز برای‌مان خیره‌کننده هستند، نگرشی مشابه پیدا کنند. انگار ذهن ما همیشه آمادگی دارد که سریعاً خود را با پیشرفت‌ها تطبیق دهد و چیزهایی را که روزی غیرقابل تصور بودند، بدیهی فرض کند. **ویژگی‌ای که امروز حیرت‌انگیز است، فردا می‌شود «امکانات پایه».** همان‌گونه که اثر هوش مصنوعی نشان می‌دهد، **تعریف AI همواره در حال تغییر بوده است** و شاید همیشه چنین بماند.

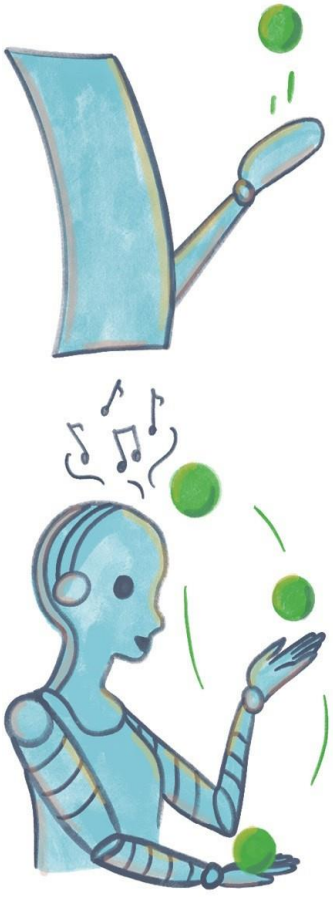
هوش مصنوعی قوی در برابر هوش مصنوعی ضعیف

بیشتر متخصصان هوش مصنوعی معتقدند که حتی شگفت‌انگیزترین قابلیت‌های فعلی این فناوری نیز هنوز به چیزی که هوش مصنوعی قوی (Strong AI) نامیده می‌شود، نرسیده‌اند. هوش مصنوعی قوی که با نام‌هایی چون هوش مصنوعی عمومی (General AI) یا **هوش عمومی مصنوعی (AGI)** نیز شناخته می‌شود، نوعی فرضی از هوش مصنوعی است که اگر ساخته شود، **قادر خواهد بود مانند انسان هر کاری را یاد بگیرد و انجام دهد.** این نوع هوش مصنوعی، دارای درک گسترده‌ای از مسائل مختلف خواهد بود و می‌تواند در حوزه‌هایی همچون شناسایی تصاویر، پردازش زبان، استدلال، حل مسئله، تصمیم‌گیری در زمینه‌های متنوع و بسیاری از امور دیگر فعالیت کند.

اما چرا می‌گوییم **فرضی**؟ چون **تاکنون هیچ نمونه‌ای از AGI ساخته نشده** است. برخی متخصصان بر این باورند که ساخت چنین سامانه‌ای فاصله‌ی زیادی با واقعیت دارد و حتی برخی دیگر اساساً دستیابی به آن را ناممکن می‌دانند. در مقابل، گروهی هم خوش‌بین‌تر هستند و نسبت به تحقق آن هیجان دارند. برای نمونه، شرکت OpenAI مأموریت خود را چنین تعریف کرده است:

تضمین این‌که **هوش عمومی مصنوعی** (سامانه‌هایی که به‌طور کلـی از انسـان‌ها باهوش‌تر هستند) به **نفع تمام بشریت** عمل کند.

دقت کنید که OpenAI، AGI را «**باهوش‌تر از انسان‌ها**» تعریف کرده، نه صرفاً هم‌هوش با انسان‌ها. در نقطه‌ی مقابل هوش مصنوعی قوی، هوش مصنوعی ضعیف (Weak AI) یا هوش مصنوعی محدود (Narrow AI) قرار دارد. هوش مصنوعی ضعیف به سامانه‌هایی گفته می‌شود که **فقط برای حل یک مسئله‌ی خاص یا مجموعه‌ای محدود از مسائل** طراحی و پیاده‌سازی شده‌اند.



هوش مصنوعی ضعیف (Weak AI) یا هوش مصنوعی محدود (Narrow AI)، به سامانه‌هایی گفته می‌شود که فقط برای حل یک مسئله‌ی خاص یا مجموعه‌ای محدود از مسائل طراحی و پیاده‌سازی شده‌اند.

هوش مصنوعی قوی که با نام‌هایی چون هوش مصنوعی عمومی (General AI) یا هوش عمومی مصنوعی (AGI) نیز شناخته می‌شود، نوعی فرضی از هوش مصنوعی است که اگر ساخته شود، قادر خواهد بود مانند انسان هر کاری را یاد بگیرد و انجام دهد.

شکل ۱/۱ – تفاوت بین هوش مصنوعی ضعیف و هوش مصنوعی قوی

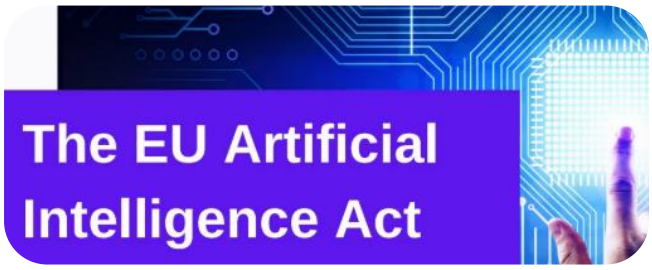
مثال‌های **هوش مصنوعی ضعیف در اطراف ما فراوان‌اند**: این سامانه‌ها کارهایی را انجام می‌دهند که معمولاً نیاز به انسان دارند اما هیچ توانایی در زمینه‌های دیگر

ندارند. برای مثال، دیپ‌بلو از شرکت IBM استاد بازی شطرنج بود اما نمی‌توانست مونوپولی بازی کند چه رسد به این‌که بگوید در یک عکس، گربه وجود دارد یا سگ؟ یا سامانه‌ی پیشنهاددهی در یک وب‌سایت خرید آنلاین ممکن است بسیار پیشرفته باشد و پیشنهادهایی عالی به کاربران بدهد اما نمی‌تواند برای شما رانندگی کند.

البته **فریب واژه‌ی «ضعیف» را نخورید** زیرا این سامانه‌ها گاهی **تأثیرات عمیقی بر جهان ما دارند.** هوش مصنوعی ضعیف در همه‌جا حضور دارد و بسیاری از جنبه‌های کار، جامعه، و حتی زندگی را به‌طور چشمگیری تغییر داده است. ما امروز شاهد گونه‌های نسبتاً پیشرفته‌ای از هوش مصنوعی ضعیف هستیم که قابلیت‌های «چندوجهی (multimodal)» دارند مثلاً وقتی OpenAI ویژگی تشخیص صدا و تصویر را به ChatGPT افزود سرعت و جهت پیشرفت‌های فعلی به‌قدری غیرقابل‌پیش‌بینی شد که **نمی‌توان با قطعیت گفت چقدر تا رسیدن به AGI فاصله داریم.** اما آن‌چه به ما کمک می‌کند، درک تفاوت بین این دو نوع هوش مصنوعی است تا بتوانیم این فناوری را هرچه دقیق‌تر ارزیابی کنیم در حالی که به‌طور مداوم در حال تحول است.

هوش مصنوعی با هدف عمومی (GPAI)

اصطلاحی نسبتاً جدید نیز در سال ۲۰۲۱ وارد واژگان تخصصی هوش مصنوعی شد:



هوش مصنوعی با هدف عمومی یا **General Purpose AI** که به‌اختصار **GPAI** نامیده می‌شود. این اصطلاح برای نخستین‌بار به‌طور رسمی در اصلاحیه‌ای

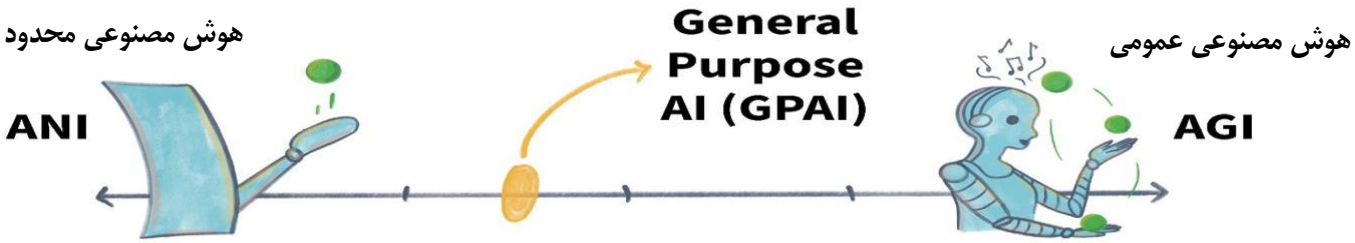
به «**قانون هوش مصنوعی اتحادیه اروپا**» گنجانده شد. قانونی که در دسامبر ۲۰۲۳ به تصویب رسید. قانون‌گذاران اتحادیه اروپا در این اصلاحیه، GPAI را این‌گونه تعریف کردند:

> **مدل هوش مصنوعی با هدف عمومی،** مدلی است که با داده‌های گسترده و با بهره‌گیری از خودنظارتی در مقیاس بالا آموزش دیده و دارای میزان قابل‌توجهی از عمومیت است و **می‌تواند مجموعه‌ی متنوعی از وظایف متمایز را به‌شکلی شایسته انجام دهد،** صرف‌نظر از اینکه چگونه به بازار عرضه شده باشد و اینکه در چه سامانه‌ها یا کاربردهایی از آن استفاده شود.

البته این تعریف حاوی اصطلاحاتی «مبهم» است: عباراتی مانند «داده‌های گسترده» (چقدر گسترده؟)، «عمومیت قابل‌توجه» (چه مقدار قابل‌توجه و چقدر عمومی؟)، و «مجموعه‌ای متنوع از وظایف متمایز» (چقدر متنوع و تا چه حد متمایز؟) با توجه به سرعت پیشرفت فناوری‌های هوش مصنوعی، ممکن است همین اصطلاحات طی چند سال آینده معنای کاملاً متفاوتی پیدا کنند. بیایید به یاد بیاوریم که چند سال پیش، چه چیزهایی را «کلان‌داده (Big Data)» می‌نامیدیم و امروز به آن‌ها لبخند می‌زنیم. چیزی که با قطعیت می‌دانیم این است که هدف اصلی از معرفی اصطلاح GPAI، **وضع مقررات برای نسل جدیدی از فناوری‌های قدرتمند هوش مصنوعی** است مانند GPT-4 از OpenAI یا «جمینای (Gemini)» از گوگل که از اواخر سال ۲۰۲۲ در اختیار عموم قرار گرفتند.

البته اغلب متخصصان هوش مصنوعی همچنان معتقدند که این **مدل‌های زبانی بزرگ در رده‌ی هوش مصنوعی ضعیف قرار می‌گیرند** نه هوش مصنوعی قوی. با این حال، برخی پژوهشگران از جمله تیمی در مایکروسافت، معتقدند که ChatGPT نشانه‌هایی اولیه از AGI را از خود نشان می‌دهد و می‌توان آن را «نسخه‌ای ابتدایی (هرچند ناقص)

از یک سامانه‌ی AGI دانست. نگارنده نیز مانند اکثریت کارشناسان این حوزه بر این باور است که این مدل‌های زبانی قدرتمند هنوز به AGI نرسیده‌اند و حتی نزدیک هم نیستند. اما نمی‌توان انکار کرد که آن‌ها بسیار توانمندتر از مدل‌های محدود و تخصصی گذشته هستند. توانایی‌های «چندوجهی» آن‌ها مانند تشخیص صدا و تصویر و مکالمه‌ی هوشمند، آن‌ها را از بسیاری از کاربردهای تک‌منظوره‌ی قدیمی متمایز می‌کند. در شکل ۱٫۲، رابطه‌ی بین GPAI، ANI و AGI نشان داده شده است.



شکل ۱٫۲ - رابطه‌ی بین GPAI، ANI و AGI

آیا واقعاً نیاز به یک اصطلاح و مخفف جدید داشتیم؟ شاید. گرچه متأسفانه این عنوان جدید نیز از واژه‌ی «عمومی (general)» استفاده می‌کند و این باعث می‌شود بسیاری افراد آن را با هوش عمومی مصنوعی (AGI) اشتباه بگیرند. اما در نهایت، نکته‌ی مهم این است:

فاصله‌ی بین هوش مصنوعی محدود (ANI) و هوش مصنوعی عمومی (AGI) در حال کاهش است. فارغ از اینکه این طبقه‌بندی جدید، چقدر ضروری بوده، فناوری‌های نوینی که امروزه ظهور یافته‌اند بدون تردید نمایانگر **یک جهش چشمگیر در مسیر تکامل هوش مصنوعی** هستند.

نمونه‌هایی از هوش مصنوعی در زندگی روزمره

برای درک بهتر مفهوم هوش مصنوعی، بیایید به چند نمونه واقعی از کاربردهای آن در زندگی روزمره بپردازیم. نمونه‌هایی که اکثر ما با آن‌ها آشنا هستیم. صحبت درباره‌ی هوش مصنوعی به‌عنوان یک نظریه یا حوزه‌ی علمی یک چیز است اما واقعیت این است که **هوش مصنوعی در بسیاری از جنبه‌های زندگی ما عملاً وارد شده است.** نکته‌ی جالب اینکه **خیلی از افراد اصلاً نمی‌دانند که در حال استفاده از هوش مصنوعی** هستند.

پیشنهاد فیلم در پلتفرم‌های پخش فیلم و سریال

احتمالاً شما هم در یکی از پلتفرم‌های پخش فیلم مانند فیلیمو، نماوا یا فیلم‌نت عضو هستید. آخرین باری که وارد حساب کاربری خود شدید تا سریال موردعلاقه‌تان را تماشا کنید، آیا بخشی با عنوانی مثل «**مناسب برای شما**» **یا** «**بر اساس سابقه‌ی تماشای شما**» دیدید؟ چگونه این پلتفرم‌ها می‌دانند چه فیلمی را به شما پیشنهاد دهند و چرا پیشنهادها برای افراد مختلف فرق دارد؟

این سامانه‌ها از نوعی هوش مصنوعی به نام **سامانه‌های پیشنهاددهنده** استفاده می‌کنند. آن‌ها با تحلیل تاریخچه‌ی جستجو و تماشا، امتیازاتی که داده‌اید، و گاهی حتی اطلاعات جمعیت‌شناختی شما، پیش‌بینی می‌کنند که شما احتمالاً چه فیلم‌هایی را بیشتر می‌پسندید. با این حال، بسیاری از مردم نمی‌دانند که چنین پیشنهادهایی با کمک هوش مصنوعی انجام می‌شود .

دستیارهای مجازی

امروزه برخی از ما از دستیارهای مجازی مانند الکسا (Alexa) از شرکت آمازون یا سیری (Siri) از اپل برای انجام کارهای

روزمره استفاده می‌کنیم مثل اطلاع از وضعیت آب‌وهوا، تنظیم تایمر در آشپزخانه، اضافه کردن یادآور در تقویم، تهیه‌ی فهرست خرید یا اطلاع از ترافیک مسیر محل کار. تعامل ما با این دستیارها عمدتاً از طریق **دستورات صوتی** است. اما چطور این ابزارها صداهای ما را به پاسخ‌های معنادار تبدیل می‌کنند؟ پاسخ این است که آن‌ها از مجموعه‌ای از فناوری‌های هوش مصنوعی استفاده می‌کنند، از جمله:

- **تشخیص گفتار** برای تبدیل صوت به متن (Voice-to-Text speech recognition) ؛
- **پردازش زبان طبیعی** (natural language processing - NLP) شامل:
  - **درک زبان طبیعی** (natural language understanding - NLU) برای فهمیدن منظور شما؛
  - **تولید زبان طبیعی** (natural language generation - NLG) برای تولید پاسخ‌های انسانی.

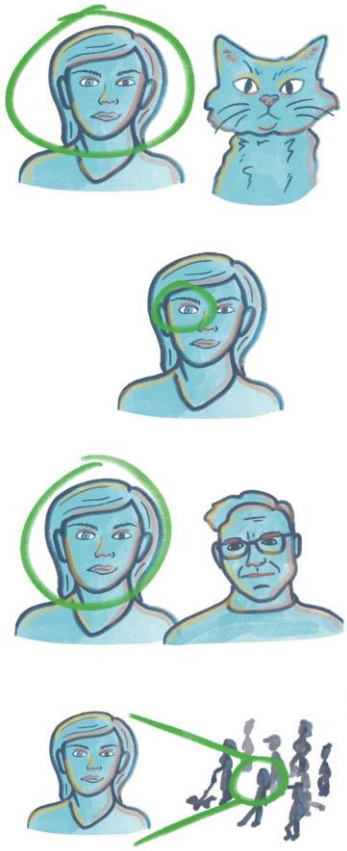
تشخیص چهره

شما چگونه قفل صفحه‌ی گوشی خود را باز می‌کنید؟ اگر از کاربران آیفون باشید، احتمالاً از ویژگی Face ID استفاده می‌کنید، قابلیتی که با استفاده از دوربین جلویی گوشی و هزاران نقطه‌ی مادون‌قرمز نامرئی، یک تصویر سه‌بعدی از چهره‌ی شما تهیه می‌کند. این سیستم با بهره‌گیری از شاخه‌ای از هوش مصنوعی به نام بینایی ماشین (Computer Vision) چهره‌ی شما را شناسایی می‌کند و با الگویی که در تنظیمات اولیه ذخیره شده مقایسه می‌نماید. نکته‌ی شگفت‌انگیز اینجاست که این سامانه‌ها **حتی وقتی زاویه‌ی صورت یا نور محیط تغییر می‌کند، باز هم در درصد بالایی از موارد می‌توانند چهره را درست تشخیص دهند.** همچنین، **این سیستم‌ها با تغییرات چهره‌ی کاربر در طول زمان (مثلاً رشد ریش یا استفاده از آرایش) خود را تطبیق می‌دهند** (البته در گذشته، برخی از سامانه‌های تشخیص چهره برای افراد با پوست تیره

یا برای زنان، عملکرد ضعیف‌تری داشتند). تشخیص چهره فقط یک فناوری نیست بلکه می‌توان آن را به چهار شاخه‌ی اصلی تقسیم کرد:

- **شناسایی چهره (Face Detection):** این شاخه تشخیص می‌دهد که آیا چهره‌ی انسانی در یک تصویر یا ویدئو وجود دارد یا خیر، و در صورت وجود، محل دقیق آن را مشخص می‌کند؛
- **تحلیل ویژگی‌های چهره (Facial Attribute Analysis):** در این شاخه، ویژگی‌های چهره‌ی شناسایی‌شده بررسی می‌شود؛ از جمله سن تقریبی، جنسیت، و حالات احساسی فرد؛
- **تأیید چهره (Face Verification):** هدف این شاخه تطبیق یک چهره‌ی خاص با یک چهره‌ی شناخته‌شده‌ی دیگر است. مثلاً همان سامانه‌ای که گوشی هوشمند شما از آن برای باز کردن قفل صفحه استفاده می‌کند؛
- **شناسایی چهره (Face Identification):** در این نوع سیستم‌ها، تلاش می‌شود مشخص شود که چهره‌ی مشاهده‌شده با کدام فرد در یک پایگاه داده‌ی بزرگ از چهره‌های شناخته‌شده تطابق دارد.

در بسیاری از سیستم‌های هوش مصنوعی، این برنامه‌ها می‌توانند به‌صورت ترکیبی مورد استفاده قرار گیرند. هر یک از آن‌ها نرخ خطایی دارند که در حال کاهش است اما هنوز صفر نیست. علاوه بر این، تحقیقات نشان داده‌اند که نرخ خطا برای برخی گروه‌های جمعیتی بیشتر از دیگران است که در فصل پنجم به آن بیشتر خواهیم پرداخت.



شکل ۱/۳ – انواع مختلف تشخیص چهره

فیلتر ایمیل‌ها یا پیام‌های اسپم

صندوق ورودی ایمیل ما مرتباً با تبلیغات آزاردهنده یا بدتر از آن، کلاه‌برداری‌هایی مواجه می‌شود که قصد دارند امنیت و حریم خصوصی ما را تهدید کنند. خوشبختانه، بیشتر این ایمیل‌های ناخواسته به‌طور **خودکار** توسط سامانه‌ای به نام **فیلتر اسپم** شناسایی شده و به پوشه‌ی اسپم منتقل می‌شوند. اما چگونه سیستم ایمیل تشخیص می‌دهد که کدام پیام باید به پوشه‌ی اسپم برود و کدام‌یک به صندوق اصلی؟ **فیلتر اسپم از شبکه‌های عصبی مصنوعی استفاده می‌کند.** مدل‌هایی محاسباتی که الهام‌گرفته از مغز انسان هستند و با آزمون‌وخطا یاد می‌گیرند. این شبکه‌ها بر اساس ایمیل‌هایی که قبلاً به‌عنوان اسپم شناخته شده‌اند، **یاد می‌گیرند که کدام واژه‌ها و عبارت‌ها معمولاً در ایمیل‌های مزاحم وجود دارند.** سپس هنگام دریافت ایمیل‌های جدید، این

شبکه‌های عصبی با توجه به آموخته‌های خود پیش‌بینی می‌کنند که آیا پیام اسپم است یا نه. همچنین، با علامت‌گذاری دستی شما (مثلاً وقتی مشخص می‌کنید پیامی اسپم است یا نیست)، **سامانه مدام به‌روز شده و خود را با تاکتیک‌های جدید ارسال‌کنندگان اسپم تطبیق می‌دهد.**

ترجمه‌ی ماشینی

اگر به کشوری سفر کرده‌اید که زبان آن با زبان مادری شما متفاوت است، احتمالاً از برنامه‌ای مثل Google Translate برای درک معنای تابلوها یا پرسیدن آدرس استفاده کرده‌اید. ترجمه‌ی ماشینی یکی از اولین مسائلی بود که هوش مصنوعی برای حل آن به‌کار گرفته شد یعنی ترجمه‌ی خودکار از یک زبان به زبان دیگر بدون دخالت مترجم انسانی. بعلاوه وقتی وارد صفحه‌ای در اینترنت می‌شوید که به زبانی بیگانه نوشته شده، مرورگرهایی مثل Google Chrome گزینه‌ای را در بالای صفحه نمایش می‌دهند که می‌توانید با کلیک روی آن، متن صفحه را به زبان دلخواهتان ترجمه کنید. البته این ترجمه‌ها هنوز بی‌نقص نیستند و گاهی خروجی‌هایی ارائه می‌دهند که مورد تأیید مترجمان حرفه‌ای قرار نمی‌گیرد. با این حال، نمی‌توان از کاربرد بالای ترجمه‌ی فوری چشم‌پوشی کرد حتی اگر برخی از ظرافت‌های معنایی در ترجمه گم شوند.

هوش مصنوعی مولد (Generative AI)

یکی از روش‌های جدید و محبوب تعامل با هوش مصنوعی، از تاریخ ۳۰ نوامبر ۲۰۲۲ با معرفی ChatGPTآغاز شد. این ابزار تنها در دو ماه به ۱۰۰ میلیون کاربر رسید، سریع‌تر از هر فناوری دیگر در تاریخ. ChatGPT از نوعی مدل زبان بزرگ (Large Language Model - LLM) استفاده می‌کند که پاسخ‌های متنی را بر اساس دستوراتی که کاربران به زبان طبیعی وارد می‌کنند تولید کند. در زمان نگارش این متن، نسخه‌های GPT-3.5 و GPT-4

مدل‌هایی هستند که بر روی حجم عظیمی از متون شامل وب‌سایت‌ها، کتاب‌ها، مقالات علمی، اخبار و موارد دیگر آموزش دیده‌اند.

در هفته‌ها و ماه‌های پس از عرضه‌ی ChatGPT، شرکت‌های بزرگی مثل گوگل و بایدو نیز مدل‌های زبان بزرگ خود را عرضه کردند. مدل‌های متن‌باز (open-source) هم به‌سرعت در دسترس عموم قرار گرفتند. اما **متن، تنها نوع محتوایی نیست که هوش مصنوعی مولد قادر به تولید آن است.** ابزارهایی مانند MidJourney، Stable Diffusion، و DALL·E که در اکتبر ۲۰۲۳ به ChatGPT اضافه شد قابلیت **تولید تصویر** از روی دستورهای متنی کاربران را دارند. همچنین، برنامه‌هایی نیز برای **تولید ویدئو** در حال توسعه هستند. با همه‌ی این پیشرفت‌ها، هوش مصنوعی مولد سوالات مهمی را نیز ایجاد کرده، از جمله اینکه آیا این مدل‌ها **تعصبات اجتماعی** را تقویت می‌کنند؟ و وضعیت **حقوق مالکیت فکری** محتوایی که این مدل‌ها بر پایه‌ی آن آموزش دیده‌اند چیست؟ این مسائل هنوز به‌طور کامل حل نشده‌اند و در فصل پنجم به‌تفصیل بررسی خواهند شد.

خلاصه

در این فصل، بحث را با نگاهی کلی به مفهوم هوش مصنوعی آغاز کردیم و تعاریفی از منابع گوناگون را بررسی نمودیم تا درک بهتری از دیدگاه‌های مختلف در این حوزه به‌دست آوریم. تعاریف هوش مصنوعی معمولاً آن را با هوش انسانی مقایسه می‌کنند و برخی از این تعاریف به‌طور دقیق‌تری توضیح می‌دهند که این مقایسه شامل چه نوع اقداماتی است؛ اقداماتی همچون ساخت مدل‌هایی بر اساس داده‌ها و خودکارسازی تحلیل‌ها برای تصمیم‌گیری. به این نتیجه رسیدیم که **تعریف هوش مصنوعی کار ساده‌ای نیست**؛ بخشی از این پیچیدگی به خود واژه‌ی «هوش» بازمی‌گردد که می‌تواند

معانی مختلفی داشته باشد، و بخشی دیگر نیز به این واقعیت مربوط است که **تعریف هوش مصنوعی ثابت نیست.** بر اساس پدیده‌ای به نام **اثر هوش مصنوعی،** آنچه انسان‌ها در مقاطع مختلف زمانی به‌عنوان هوش مصنوعی تلقی می‌کنند تغییر کرده است، و همین امر باعث شده دستیابی به یک تعریف یگانه تقریباً غیرممکن باشد.

با این حال، می‌توان میان هوش مصنوعی قوی (یا هوش مصنوعی عمومی) و هوش مصنوعی ضعیف (یا هوش مصنوعی محدود) تمایز قائل شد. هوش مصنوعی قوی، نوعی فرضی از هوش مصنوعی است که قادر است طیف گسترده‌ای از مسائل را همانند یا حتی بهتر از انسان حل کند که تاکنون هیچ نمونه‌ی واقعی از آن ساخته نشده است. در مقابل، هوش مصنوعی ضعیف نوعی از هوش مصنوعی است که در اطراف ما به‌وفور یافت می‌شود. برنامه‌هایی که می‌توانند یک وظیفه‌ی مشخص را انجام دهند؛ وظیفه‌ای که معمولاً نیاز به مداخله‌ی انسانی دارد. در پایان این فصل، شش نمونه از کاربردهای هوش مصنوعی ضعیف در زندگی روزمره را به‌اختصار بررسی کردیم: سامانه‌های پیشنهاددهنده، دستیارهای مجازی، تشخیص چهره، فیلتر ایمیل‌های اسپم، ترجمه‌ی ماشینی، و هوش مصنوعی مولد.

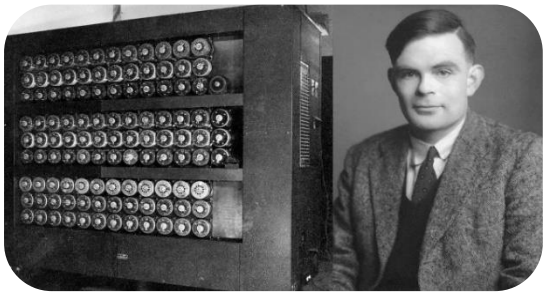
در **فصل بعد**، سفری کوتاه در **تاریخچه‌ی هوش مصنوعی** خواهیم داشت. آشنایی با مسیری که این حوزه در گذشته طی کرده، می‌تواند درک عمیق‌تری از جایگاه کنونی آن در اختیارمان بگذارد.

فصل دوم

مروری کوتاه بر تاریخچه هوش مصنوعی

برای آنکه کسی را واقعاً بشناسیم و بفهمیم چرا به شکل کنونی‌اش درآمده، نگاه به گذشته‌اش می‌تواند کمک‌کننده باشد. **درک درست از گذشته** می‌تواند آگاهی ما را در **زمان حال** ارتقاء دهد و تا حدی توانایی ما را برای **پیش‌بینی آینده** افزایش دهد. این اصل، نه‌فقط درباره‌ی انسان‌ها، بلکه درباره‌ی موضوعاتی همچون AI هم صادق است. پس بیایید سفری کوتاه در تاریخ داشته باشیم تا هوش مصنوعی را بهتر بشناسیم. شاید در این مسیر دریابیم که شناخت هوش مصنوعی، در واقع شناخت بهتر خودمان و مسیر تکاملی گونه‌ی ماست.

آزمون تورینگ



با وجود افسانه‌های کهن، بیشتر متخصصان هوش مصنوعی، آغاز واقعی این حوزه را در اواسط قرن بیستم می‌دانند؛ یعنی زمانی که پیشگامانی همچون **آلن تورینگ** (ریاضیدان بریتانیایی) و جان مک‌کارتی (دانشمند آمریکایی علوم رایانه)، به همراه ماروین مینسکی، کلود شانون، آلن نیول، هربرت سایمون و دیگران، بنیان‌های آن را بنا نهادند. نقش عمده‌ی مردان در شکل‌گیری اولیه‌ی این حوزه قابل انکار نیست و چنین سوگیری‌های اجتماعی همچنان بر این حوزه و محصولات آن سایه افکنده‌اند؛ موضوعی که در فصل‌های بعدی بررسی خواهیم کرد. پس از فعالیت‌های تورینگ در **مرکز رمزگشایی** بلچلی پارک در دهه‌ی ۱۹۴۰ برای شکستن کد **دستگاه انیگمای آلمان** نازی، او تمرکز خود را بر پرسشی فلسفی معطوف کرد: «آیا ماشین‌ها می‌توانند فکر کنند؟»

تورینگ که این سؤال را بی‌معنا می‌دانست، در سال ۱۹۵۰ مقاله‌ای با عنوان «ماشین محاسبه‌گر و هوش» در مجله‌ی معتبر Mind منتشر کرد. او در این مقاله، پرسش جایگزینی پیشنهاد داد که در قالب یک بازی با عنوان «**بازی تقلید (Imitation Game)**» مطرح شد. فرم امروزی این بازی که به آزمون تورینگ مشهور است، شامل سه نقش است:

۱) یک داور (انسان)،

۲) یک شرکت‌کننده انسانی،

۳) یک ماشین.

داور که نمی‌داند کدام پاسخ از طرف انسان است و کدام از ماشین، از طریق رابط متنی سؤالاتی می‌پرسد و سعی می‌کند تشخیص دهد کدام شرکت‌کننده، ماشین است. **اگر داور نتواند با اطمینان بگوید کدام‌یک ماشین است**، گفته می‌شود که **ماشین، آزمون را با موفقیت پشت سر گذاشته است.**

شکل ۲/۱ - تصویری از شیوه آزمون تورینگ

آزمون تورینگ **تأثیری شگرف** بر حوزه‌ی هوش مصنوعی داشته اما همواره با **انتقادات جدی** نیز روبه‌رو بوده است. منتقدان می‌گویند که این آزمون فقط **توانایی زبانی** را

می‌سنجد و ربطی به صحت یا دقت پاسخ‌ها ندارد بلکه بیشتر بر توانایی برنامه برای تقلید و حتی فریب انسان تمرکز دارد. برخی دیگر نیز تأکید می‌کنند که شاید داور ساده‌لوح باشد یا برنامه‌نویسان ربات، بسیار ماهر.

ادعاهایی مبنی بر گذراندن این آزمون طی سال‌های مختلف و در جریان رقابت‌های عمومی مطرح شده‌اند اما بیشتر متخصصان این ادعاها را رد کرده‌اند و معتقدند هنوز هیچ ماشینی واقعاً از آزمون تورینگ عبور نکرده است. برخی نیز بر این باورند که آزمون تورینگ گرچه بخشی مهم از تاریخ هوش مصنوعی است اما **با پیشرفت‌های امروزی دیگر چندان مناسب نیست.** آیا در نهایت، هوش مصنوعی واقعاً این آزمون را پشت سر خواهد گذاشت؟ باید منتظر ماند و دید. تورینگ خود در پیش‌بینی‌ای مشهور گفت:

> باور دارم که تا حدود پنجاه سال آینده، می‌توان کامپیوترهایی برنامه‌ریزی کرد که آن‌قدر خوب در بازیِ تقلید عمل کنند که داور نتواند در بیش از ۷۰٪ مواقع، پس از پنج دقیقه مکالمه، پاسخ درست را بدهد.

این معیار خاص یعنی کمتر از ۷۰٪ دقت پس از پنج دقیقه گفت‌وگو به یک معیار طلایی در هوش مصنوعی تبدیل شده است.

اگرچه پیش‌بینی تورینگ درباره گذر موفقیت‌آمیز از آزمونش تا پایان قرن بیستم محقق نشد اما باور او به پیشرفت توانایی گفت‌وگو در ماشین‌ها، تا حد زیادی درست از آب درآمده است. امروزه، پیشرفته‌ترین چت‌بات‌های هوش مصنوعی، با انواع آزمون‌ها و معیارها (مانند آزمون MMLU یا درک زبان چندوظیفه‌ای گسترده) سنجیده می‌شوند اما جاذبه آزمون تورینگ همچنان پابرجاست. شکی نیست که آنچه چت‌بات‌های امروزی

مانند ChatGPT قادر به انجام آن هستند، شگفت‌انگیز است هرچند همچنان بحث‌برانگیز است که آیا آن‌ها آزمون تورینگ را پشت سر گذاشته‌اند یا نه.

زادگاه رسمی حوزه هوش مصنوعی

آغاز رسمی هوش مصنوعی به‌عنوان یک رشته‌ی مطالعاتی، اغلب به رویدادی باز می‌گردد که در تابستان سال ۱۹۵۶ در شهر هانوفر ایالت نیوهمپشایر برگزار شد. این رویداد که امروزه با عنوان «**کارگاه دارتموث**» شناخته می‌شود، توسط جان مک‌کارتی، استادیار وقت ریاضیات در کالج دارتموث، سازمان‌دهی شد و سرآغاز تغییری بزرگ در مسیر تاریخ بشر بود. یک سال پیش از آن، **مک‌کارتیِ ۲۸ ساله** از بنیاد راکفلر درخواست بودجه‌ای به مبلغ ۱۳,۵۰۰ دلار کرد (که معادل حدود ۱۶۰,۰۰۰ دلار در سال ۲۰۲۴ است). این مبلغ قرار بود هزینه‌ی حقوق، سفر، اجاره و سایر مخارج هشت نفر از شرکت‌کنندگان را تأمین کند. او موفق شد ماروین مینسکی از دانشگاه هاروارد، ناتانیل راچستر از IBM، و کلود شانون از آزمایشگاه‌های بل را برای مشارکت در این درخواست رسمی با خود همراه کند. سندی که این افراد در سال ۱۹۵۵ با عنوان «پیشنهاد پروژه پژوهشی تابستانی دارتموث درباره هوش مصنوعی» به بنیاد راکفلر ارائه دادند، به‌عنوان **نخستین معرفی رسمی اصطلاح «هوش مصنوعی»** شناخته می‌شود اصطلاحی که توسط مک‌کارتی ابداع شد. بخشی از متن ابتدایی این پیشنهاد چنین بود:

پیشنهاد می‌کنیم یک مطالعه‌ی دو ماهه با حضور ۱۰ نفر در تابستان ۱۹۵۶ در کالج دارتموث در زمینه‌ی هوش مصنوعی انجام شود. این مطالعه بر این فرض بنا شده که **هر جنبه‌ای از یادگیری یا هر ویژگی دیگر از هوش** را **می‌توان آن‌قدر دقیق توصیف کرد** که **ماشین‌ها بتوانند آن را شبیه‌سازی کنند.** تلاش خواهد شد تا راه‌هایی برای توانمندسازی ماشین‌ها در استفاده از زبان، شکل‌دهی مفاهیم و انتزاعات، حل مسائلی که در حال حاضر مختص انسان‌هاست کشف شود. باور داریم که اگر گروهی گزینش‌شده از دانشمندان برای یک تابستان با هم روی این مسائل کار کنند، می‌توان پیشرفت قابل‌توجهی در یک یا چند مورد از این مسائل حاصل کرد.

در نهایت، درخواست آن‌ها پذیرفته شد و حدود ۲۰ نفر در این کارگاه شرکت کردند که بین ۶ تا ۸ هفته به طول انجامید (از ژوئن تا اوت ۱۹۵۶). آنچه این کارگاه را شگفت‌انگیز می‌سازد این است که موضوعاتی که در آن مورد بحث قرار گرفت شامل رایانه‌ها، پردازش زبان طبیعی، شبکه‌های عصبی، «خود بهبودی» (که امروزه یادگیری ماشین نامیده می‌شود)، انتزاع و خلاقیت همچنان چارچوب‌های اصلی این حوزه را تا به امروز تشکیل می‌دهند.

عصر طلایی هوش مصنوعی (۱۹۵۶ تا ۱۹۷۴)

کارگاه دارتموث شاید نتوانست چالش‌هایی را که بنیان‌گذاران حوزه‌ی هوش مصنوعی مطرح کرده بودند حل کند اما بی‌تردید شور و شوق و خوش‌بینی زیادی را نسبت به این حوزه به وجود آورد. **دو دهه‌ی پس از این کارگاه، به‌عنوان عصر طلایی هوش مصنوعی شناخته می‌شود.** بنیان‌گذاران این حوزه به‌تدریج دریافتند که حل مسائلی که در ابتدا تصور می‌کردند ساده هستند، در واقع بسیار دشوارتر از آن بود که گمان

می‌بردند. با این حال، این کارگاه نقش سکوی پرتابی را ایفا کرد برای مأموریتی که آینده‌ای درخشان در برابر خود داشت هرچند که در مسیر خود با موانع و شکست‌هایی نیز روبه‌رو شد.

برنامه‌ی «ماشین نظریه‌پرداز منطقی»

در کارگاه دارتموث در سال ۱۹۵۶، آلن نیوول، **هربرت سایمون** و کلیف شاو، برنامه‌ای را معرفی کردند که از آن به‌عنوان **نخستین برنامه‌ی هوش مصنوعی** یاد می‌شود (هرچند برخی، برنامه‌ی شطرنج ساده‌ای را که در سال ۱۹۵۲ توسط آرتور ساموئل برای رایانه‌ی اصلی **IBM 701** نوشته شد، اولین برنامه‌ی هوش مصنوعی می‌دانند).

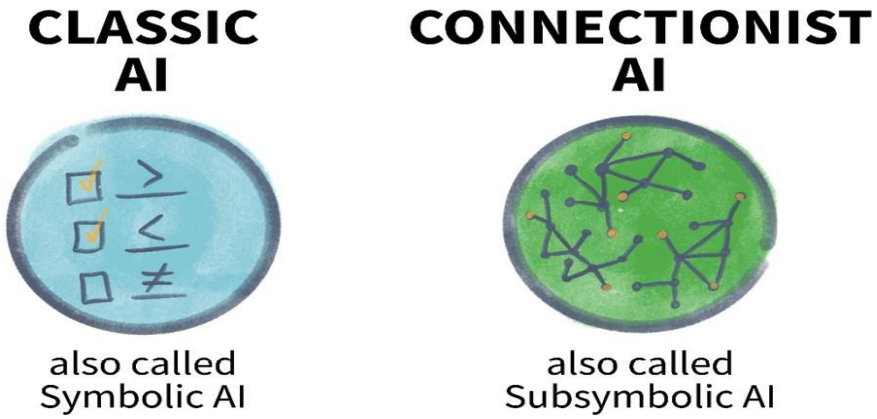
آن‌ها این برنامه را Logic Theorist یا «ماشین نظریه‌پرداز منطقی» نامیدند. این برنامه نقطه‌ی عطفی بود که در آن زمان، حتی حاضران در کارگاه نیز ارزش واقعی‌اش را به‌خوبی درک نکردند. این برنامه قادر بود چندین **قضیه‌ی ریاضی را اثبات کند** و حتی در برخی موارد، برای قضایای شناخته‌شده، اثبات‌هایی کوتاه‌تر و ساده‌تر از اثبات‌های قبلی ارائه داد. از همان آغاز، روشن بود که حوزه‌ی هوش مصنوعی فقط به کنفرانس و نظریه محدود نمی‌شود بلکه ناظر بر **نوآوری** و حل مسئله است.

هوش مصنوعی نمادین و غیرنمادین

پس از کارگاه دارتموث، پژوهش‌های حوزه‌ی هوش مصنوعی وارد مرحله‌ای جدی شد. بیشتر تلاش‌های اولیه از جمله Logic Theorist بر پایه‌ی **رویکردی مبتنی بر قواعد** انجام می‌شد که به آن **هوش مصنوعی نمادین (Symbolic AI)** می‌گویند. هوش مصنوعی

نمادین بر **کدگذاری صریح دانش** مبتنی است؛ با استفاده از قواعدی از پیش تعریف‌شده و طراحی‌شده توسط انسان. اصطلاح «نمادین» به این دلیل به کار می‌رود که **دانش در این نوع برنامه‌ها به شکل نمادهایی مانند کلمات یا عبارات ذخیره می‌شود** چیزی که **برای انسان‌ها قابل درک است.** قواعدی که در این برنامه‌ها رمزگذاری شده‌اند، **منطق و استدلالی را شبیه‌سازی می‌کنند که شبیه به نحوه‌ی تفکر انسان است.** از آنجا که هوش مصنوعی نمادین در سال‌های اولیه، سلطه‌ی کاملی بر این حوزه داشت گاهی به آن **هوش مصنوعی کلاسیک** نیز گفته می‌شود.

در ادامه‌ی این کتاب، این شاخه‌ی ابتدایی را با شاخه‌ی دیگر هوش مصنوعی، یعنی **هوش مصنوعی غیرنمادین (Subsymbolic AI)** مقایسه خواهیم کرد شاخه‌ای که امروزه با نام **یادگیری عمیق** شناخته می‌شود. برخلاف **رویکرد قاعده‌محورِ هوش مصنوعی نمادین**، **هوش مصنوعی غیرنمادین بر یادگیری از داده‌های حجیم و کشف الگوها** تمرکز دارد. تکیه‌ی سنگین این شاخه بر شبکه‌های عصبی مصنوعی با هزاران گره‌ی متصل به یکدیگر، باعث شده است که گاهی به این شاخه، **هوش مصنوعی ارتباط‌گرا (Connectionist AI)** نیز گفته شود.



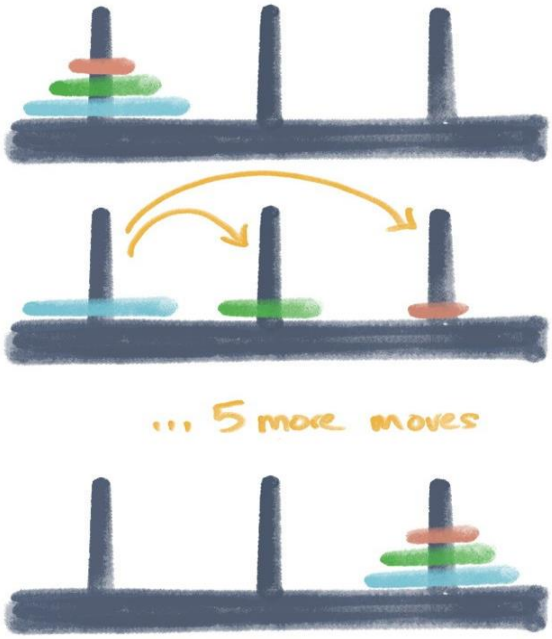
شکل ۲/۲ - مقایسه هوش مصنوعی کلاسیک و هوش مصنوعی ارتباط‌گرا

حل‌گر عمومی مسئله

تیمی که برنامه‌ی Logic Theorist را توسعه داده بود، در ادامه دست به ساخت یک برنامه‌ی دیگر با عنوان «حل‌گر عمومی مسئله (General Problem Solver)» از نوع هوش مصنوعی نمادین و مبتنی بر قواعد زد. همان‌طور که از نام این برنامه پیداست، **هدف آن حل طیف گسترده‌ای از مسائل بود.** سازندگان آن تلاش داشتند **فرایند حل مسئله** را به‌صورت **الگوریتم** بازسازی کنند تا از این طریق، نوعی رفتار هوشمندانه را شبیه‌سازی نمایند. طراحی این برنامه با الهام از مشاهدات رفتاری دانشجویان دانشگاه انجام شد؛ دانشجویانی که در حین حل مسائل منطقی (مانند اثبات قضایای ریاضی یا حرکات شطرنج)، از آن‌ها خواسته شده بود با صدای بلند فکر کنند تا روند ذهنی آن‌ها ثبت شود:

> برنامه «حل‌گر عمومی مسئله -۱» بر پایه‌ی برنامه‌ی قبلی یعنی Logic Theorist بنا شده بود، برنامه‌ای که اثبات‌هایی برای قضایای منطق گزاره‌ای می‌یافت. تلاشی بود برای بازآفرینی رفتار ثبت‌شده‌ی دانشجویانی که در حال یافتن راه‌حل برای اثبات‌ها بودند.

برنامه‌ی «حل‌گر عمومی مسئله» قادر بود برخی از معماهای منطقی با تعریف دقیق را حل کند. برای مثال، یکی از این معماها برج هانوی (Tower of Hanoi) بود؛ معمایی که در آن، بازیکن باید یک دسته دیسک با اندازه‌های مختلف را از یک میله به میله‌ای دیگر منتقل کند به‌طوری‌که تنها یک دیسک در هر بار حرکت داده شود و هرگز یک دیسک بزرگ‌تر روی دیسک کوچک‌تر قرار نگیرد. برنامه‌ی «حل‌گر عمومی مسئله» نشان‌دهنده‌ی تلاشی اولیه برای مدل‌سازی هوش از طریق **شبیه‌سازی تفکر گام‌به‌گام انسان** در حل مسائل بود.



شکل ۲/۳ - طرحی از معمای برج هانوی

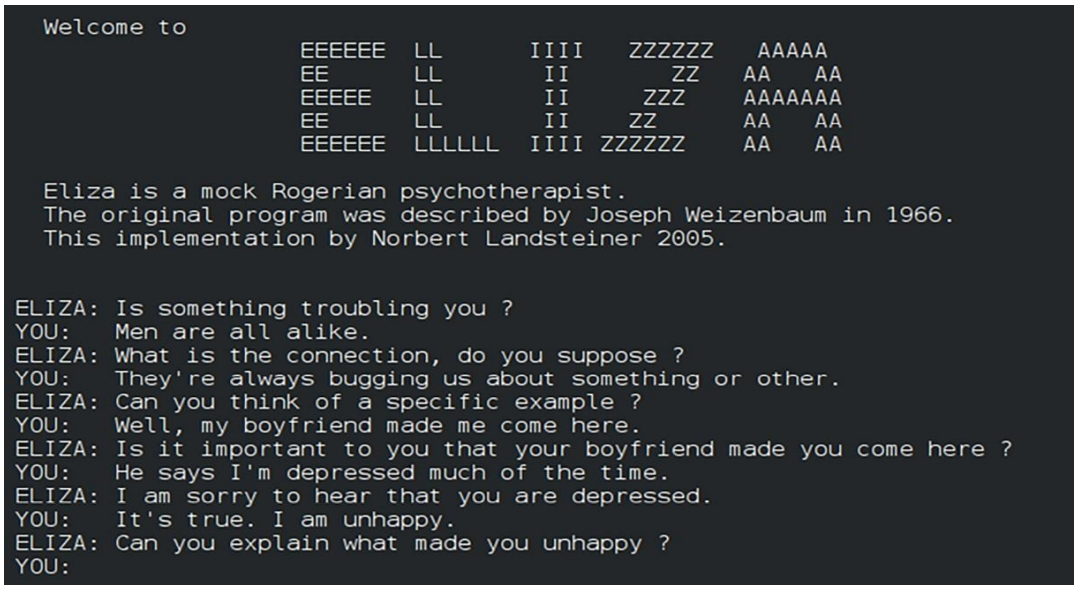
با وجود موفقیت برنامه‌ی «حل‌گر عمومی مسئله» در حل این نوع خاص از مسائل، این برنامه نتوانست از پسِ حل مسائل دنیای واقعی برآید چرا که چنین مسائل واقعی به‌مراتب **پیچیده‌تر** و ساختاری **نامنظم‌تر** از معماهای منطقی یا اثبات‌های ریاضی دارند. با این حال، این برنامه نیز مانند Logic Theorist، گامی مهم در تاریخچه‌ی هوش مصنوعی به‌شمار می‌آید. این برنامه‌های ابتداییِ مبتنی بر قواعد، درک ما را از قابلیت‌های رایانه‌ها برای حل مسائلی که پیش‌تر نیازمند هوش انسانی تصور می‌شد، به‌طرز چشم‌گیری گسترش دادند.

بات چت‌کننده الیزا

اگر حل مسائل کتاب‌های درسی و معماها یک جنبه از هوش باشد، **توانایی گفت‌وگو با انسان** جنبه‌ی دیگری از آن است. همان رفتاری که **آزمون تورینگ** بر پایه‌ی آن طراحی شده است. شاخه‌ای از مطالعات که به این نوع از رفتار می‌پردازد، **پردازش زبان طبیعی یا به اختصار NLP** نام دارد؛ حوزه‌ای میان‌رشته‌ای که در **تقاطع هوش مصنوعی، علوم**

رایانه و زبان‌شناسی قرار دارد. هدف اصلی NLP این است که رایانه‌ها بتوانند با ما با استفاده از **زبان طبیعیِ انسان‌ها** ارتباط برقرار کنند نه از طریق زبان‌های برنامه‌نویسی یا کد.

یکی از نخستین تلاش‌ها برای ساخت رایانه‌ای که بتواند وارد مکالمه با انسان شود، برنامه‌ای بود به نام الیزا که توسط پروفسور جوزف وایزن‌بام از مؤسسه MIT طراحی شد. او در سال ۱۹۶۶، برنامه‌ی الیزا را توسعه داد و مقاله‌ای با عنوان «الیزا؛ برنامه‌ای رایانه‌ای برای مطالعه‌ی ارتباط زبانی میان انسان و ماشین» منتشر کرد.



شکل ۲/۴ - تصویری از نسخه بازسازی‌شده‌ی الیزا

برنامه‌ی الیزا، **مجموعه‌ای از قواعد** را دنبال می‌کرد تا نوع خاصی از سبک مکالمه را شبیه‌سازی کند و در اصل، افراد را فریب دهد. وایزن‌بام نحوه‌ی عملکرد این برنامه را چنین توصیف کرده است:

الیزا، برنامه‌ای است که امکان نوع خاصی از مکالمه‌ی زبان طبیعی میان انسان و رایانه را فراهم می‌کند. جملات ورودی براساس **قواعدِ تجزیه‌ای**، تحلیل می‌شوند

که با **واژگان کلیدی موجود** در متن فعال می‌گردند. پاسخ‌ها نیز براساس **قواعدِ بازترکیب**، تولید می‌شوند که به قواعد تجزیه‌ی انتخاب‌شده مرتبط هستند.

به بیان دیگر، **الیزا هیچ درکی از زبان انسانی نداشت.** این برنامه بیشتر شبیه یک شعبده‌ی هوشمندانه اما نسبتاً ساده بود: مجموعه‌ای از قواعد صریح برای تبدیل ورودی‌های تایپی کاربر به پاسخ‌های متنی، که در بسیاری مواقع **این توهم را ایجاد می‌کرد که کاربر در حال صحبت با یک انسان واقعی است.** در ادامه مثالی از نحوه‌ی مکالمه‌ی طراحی‌شده برای آن آمده است که وایزن‌بام در مقاله‌ی یادشده، بیان کرده:

شما: من این روزها خیلی ناراحتم.

الیزا: از چه زمانی خیلی ناراحتی؟

در این مثال، کد ELIZA با مشاهده‌ی واژه‌ی کلیدی «I am» فعال شده و طبق قاعده‌ای خاص، پاسخی تولید می‌کند که **شبیه به پاسخ یک روان‌درمانگر** از مکتب راجرین (متمرکز بر شخص) است که به‌دنبال روشن‌سازی اطلاعات است. ساختار پاسخ نیز با عبارت «از چه زمانی...» آغاز می‌شود. وایزن‌بام توضیح می‌دهد که پاسخ مشابهی در مورد هر ورودی مشابه می‌توان تولید کرد.

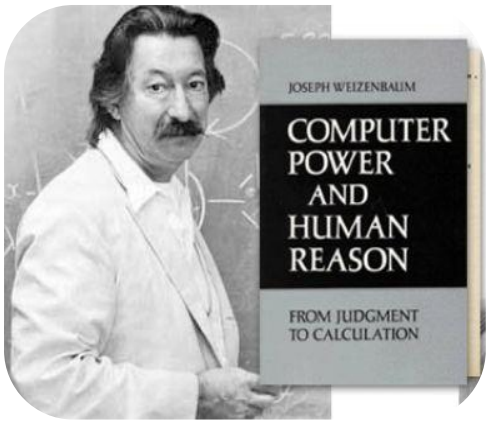
الیزا، مجموعه‌ای از قواعد مشابه داشت که براساس واژگان کلیدی موجود در طول گفت‌وگو، ورودی‌ها را به پاسخ‌هایی مشخص تبدیل می‌کرد. آیا واقعاً در دل این برنامه، موجودی هوشیار یا آگاه وجود داشت که بخواهد بداند شما از چه زمانی خیلی ناراحت بودید؟ بدیهی است که نه.

اما وایزن‌بام، دریافت که بسیاری از افراد چنین احساسی داشتند. گویی کسی پشت این پاسخ‌ها وجود دارد. فراتر از آن، برخی حتی با برنامه‌ی او **پیوندی عاطفی** برقرار

کرده بودند. منشیِ خودش از او خواست که اتاق را ترک کند تا بتواند به‌تنهایی و در خلوت به گفت‌وگو با الیزا ادامه دهد.

این واقعیت برای وایزن‌بام چنان **ناراحت‌کننده** بود که او به‌صراحت علیه مسیری که حوزه‌ی هوش مصنوعی در پیش گرفته بود، موضع گرفت. او چندین کتاب نوشت و در آن‌ها از برخی همکارانش انتقاد کرد و نسبت به دنیایی که ویژگی‌های منحصربه‌فرد و انسانی ارتباط واقعی میان انسان‌ها، به‌ویژه در تعاملاتی با ابعاد اخلاقی نظیر روان‌درمانی، در آن نادیده گرفته می‌شود، **هشدار** داد. اگرچه انتقادات او موجب بحث‌های فراوانی در جامعه‌ی هوش مصنوعی شد اما همچنان به‌عنوان شخصیتی محترم در حوزه‌ی علوم رایانه شناخته می‌شد. وایزن‌بام در سال ۲۰۰۸ درگذشت. می‌توان تصور کرد که امروز، با مشاهده‌ی کاربرد چت‌بات‌های هوش مصنوعی در پلتفرم‌های سلامت روان آنلاین یا نقش آن‌ها به‌عنوان همراهان مصنوعی که به کاربران توهم دوستی یا رابطه‌ی عاطفی می‌دهند، تا چه حد ناراضی و نگران می‌بود.

زمستان اول هوش مصنوعی (۱۹۷۴ تا ۱۹۸۰)



شاهکار یوزف وایزن‌بام با عنوان «قدرت رایانه و عقل انسانی (Computer Power & Human Reason)» در سال ۱۹۷۶ منتشر شد؛ کتابی که **نقدی تند و درون حوزه‌ای بر هوش مصنوعی** به‌شمار می‌رفت و در ابتدای دوره‌ای منتشر شد که بعدها «اولین زمستان هوش مصنوعی» نام گرفت. **زمستان هوش مصنوعی** به دوره‌ای اطلاق می‌شود که در آن **علاقه و سرمایه‌گذاری در زمینه‌ی تحقیقات و فعالیت‌های هوش مصنوعی به‌شدت کاهش یافت**. اواسط تا اواخر دهه‌ی ۱۹۷۰ اولین دوره‌ی رکود از این دست بود اما

آخرینش نبود. به‌طور کلی، حوزه‌ی هوش مصنوعی **همواره با موج‌هایی از اشتیاق** همراه بوده که به‌دنبال آن، **موج‌هایی از ناامیدی** به‌وجود آمده که هر یک تقریباً یک دهه به طول انجامیده‌اند. اما چرا چنین وضعیتی پیش آمده است؟

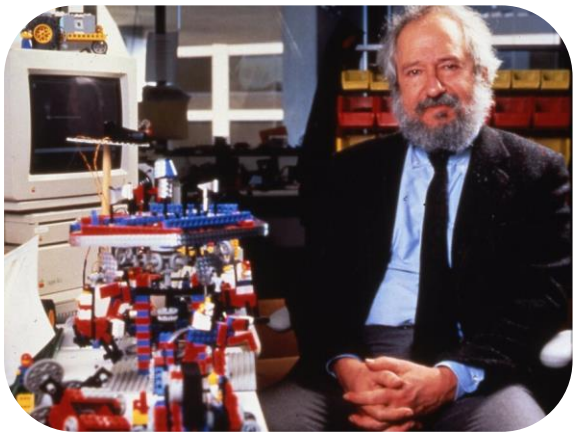
دلایل مختلفی برای این **نوسان علاقه** عمومی نسبت به هوش مصنوعی وجود دارد. در واقع، چیزی شبیه به چرخه‌ی هایپ (hype cycle) در این حوزه دیده می‌شود. دوره‌هایی با عنوان «بهار» و «تابستان هوش مصنوعی» که در آن‌ها توجه و فعالیت‌ها به‌شدت افزایش می‌یابد، و سپس زمستان‌هایی اجتناب‌ناپذیر که همراه با عقب‌نشینی و رکود هستند. در مقاله‌ای با عنوان «چرا هوش مصنوعی سخت‌تر از آن است که فکر می‌کنیم»، **ملانی میچل** این چرخه را چنین توصیف می‌کند:

> از زمان پیدایش در دهه‌ی ۱۹۵۰، حوزه‌ی هوش مصنوعی بارها بین دوره‌هایی از پیش‌بینی‌های خوش‌بینانه و سرمایه‌گذاری سنگین (بهار هوش مصنوعی) و دوره‌هایی از ناامیدی، کاهش اعتماد و افت بودجه (زمستان هوش مصنوعی) در نوسان بوده است.

زمستان نخست هوش مصنوعی که در میانه‌ی دهه‌ی ۱۹۷۰ آغاز شد، اغلب به‌عنوان نمونه‌ای کلاسیک از **وعده‌های بزرگ و دستاوردهای اندک** معرفی می‌شود. در سال‌های آغازین این حوزه، متخصصان مجموعه‌ای از پیش‌بینی‌های جسورانه و وعده‌های چشمگیر درباره‌ی پیشرفت‌های قریب‌الوقوع در توانایی‌هایی مانند پردازش زبان طبیعی و بینایی ماشین ارائه دادند، وعده‌هایی که تحقق نیافتند. برای نمونه، به این نقل‌قول (پیش‌بینی خوش‌بینانه‌ی یکی از پیشگامان هوش مصنوعی، هربرت سایمون، در سال

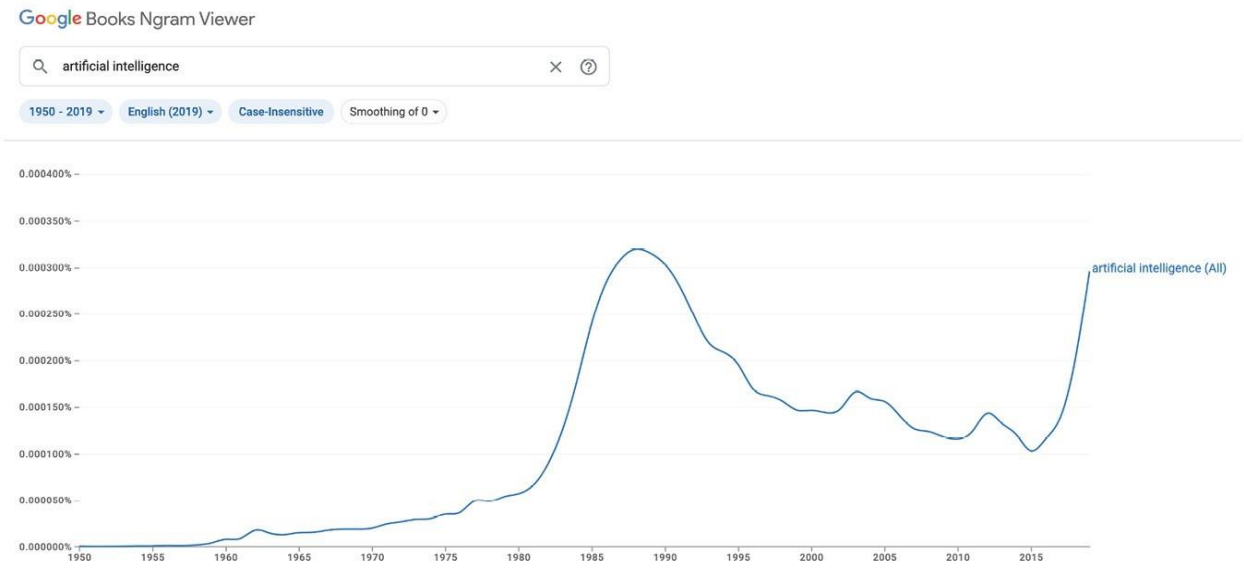
۱۹۶۰) توجه کنید: ماشین‌ها طی بیست سال آینده قادر خواهند بود هر کاری را که انسان انجام می‌دهد، انجام دهند.

با چنین اظهارنظرهایی، طبیعی بود که توجه و علاقه‌ی گسترده‌ای به هوش مصنوعی شکل بگیرد اما به همان سرعت، فروکش کند. مثال دیگر پروژه‌ی بینایی تابستانی سال ۱۹۶۶ است. ابتکاری که توسط **سیمور پاپرت** از MITرهبری شد و قرار بود گروهی متشکل از



حدود ده دانشجوی کارشناسی، سامانه‌ای بسازند که به اندازه‌ای پیچیده باشد که نقطه‌ی عطفی واقعی در توسعه‌ی تشخیص الگو محسوب شود. اما در عمل مشخص شد که دستیابی به پیشرفت‌های قابل توجه در بینایی ماشین، به زمان و تلاشی بسیار فراتر از یک تابستان و ده دانشجو نیاز دارد.

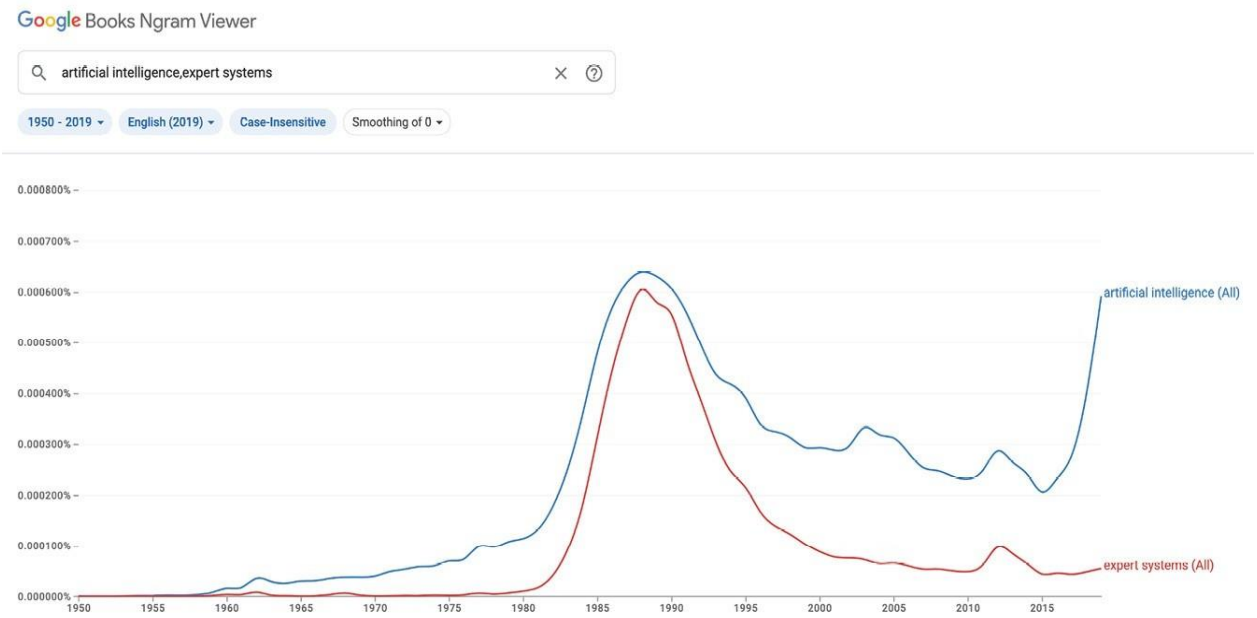
این وعده‌های اولیه موجب جذب سرمایه‌گذاری‌های کلان و توجه رسانه‌ها شد. اما زمانی که تحقق آن‌ها بسیار دشوارتر از حد انتظار از آب درآمد یا دستاوردهای واقعی با وعده‌ها فاصله داشتند ناامیدی شکل گرفت، **توجه‌ها به حوزه‌های دیگر معطوف شد** و در نهایت **منابع مالی کاهش یافتند یا قطع شدند.** با این حال، همان‌طور که این الگو نشان می‌دهد، این دوره‌های «قحطی»، در واقع زمینه‌ساز موج بعدی «وفور» بودند. با نگاهی سریع به نمودار واژه‌نگار کتاب‌ها که توسط گوگل برای عبارت «هوش مصنوعی» در بازه‌ی ۱۹۵۰ تا ۲۰۱۹ ارائه شده، می‌توان **فراز و فرود کاربرد این اصطلاح** را در متون منتشرشده مشاهده کرد چه در اوج‌ها و چه در رکودها.



شکل ۲/۵ - واژه‌نگار کتاب‌های گوگل برای «هوش مصنوعی» در بازه‌ی ۱۹۵۰ تا ۲۰۱۹

اوج و افول سیستم‌های خبره، دهه ۱۹۸۰ تا اوایل دهه ۱۹۹۰

یکی از ویژگی‌های برجسته نمودار شکل ۲/۵، **افزایش چشمگیر منحنی در اوایل دهه ۱۹۸۰** است که اوج آن حدود سال ۱۹۸۷ رخ می‌دهد، و سپس کاهش آن در اوایل دهه ۱۹۹۰. این شکل الگوی «تابستان هوش مصنوعی» و «زمستان هوش مصنوعی» را به‌خوبی نشان می‌دهد تا جایی که می‌توانیم میزان تکرار یک واژه در کتاب‌ها را معیاری تقریبی برای سنجش میزان اهمیت و محبوبیت آن واژه در نظر بگیریم. اما چه چیزی باعث شکل‌گیری این موج صعودی و سپس نزولی در آن بازه‌ی خاص زمانی شد؟ نگاهی به نسخه‌ی دوم نمودار سرنخی از پاسخ را نشان می‌دهد. در شکل ۲/۶، خط مربوط به واژه‌ی «هوش مصنوعی (Artificial Intelligence)» در کنار خطی دیگر به‌نمایش درآمده که روند ظهور واژه‌ی «سیستم‌های خبره (Expert Systems)» را نشان می‌دهد. از این دو خط کاملاً پیداست که افزایش محبوبیت این دو اصطلاح در دهه ۱۹۸۰ تقریباً همزمان اتفاق افتاده است.



شکل ۲٫۶ - واژه‌نگار کتاب‌های گوگل، «هوش مصنوعی» و « سیستم‌های خبره» از ۱۹۵۰ تا ۲۰۱۹

این همزمانی تصادفی نیست. در واقع، موج استفاده از اصطلاح «هوش مصنوعی» در آن دوران، مستقیماً ناشی از **رشد علاقه به سیستم‌های خبره** بود. اما «سیستم‌های خبره» دقیقاً چه هستند و چه ارتباطی با هوش مصنوعی دارند؟

اصطلاح «سیستم‌های خبره» به **برنامه‌های هوش مصنوعی** اطلاق می‌شود که با استفاده از **قواعد از پیش تعریف‌شده** و **پایگاه‌های دانش در یک حوزه‌ی تخصصی**، سعی دارند تصمیماتی مشابه با آنچه **متخصصان انسانی آن حوزه** اتخاذ می‌کنند، شبیه‌سازی کنند. همان‌طور که تاکنون دریافته‌اید، این سیستم‌ها نمونه‌ای از «**هوش مصنوعی نمادین**» یا «**هوش مصنوعی کلاسیک**» هستند. سیستم‌های خبره معمولاً دارای یک **واسط کاربری** نیز هستند که به کاربران اجازه می‌دهد اطلاعاتی وارد کنند و در ازای آن، از برنامه راهنمایی دریافت کنند. **نوعی سیستم پشتیبان تصمیم‌گیری** که در دهه ۱۹۸۰ بسیار محبوب شد.

برای نمونه، در سال ۱۹۸۰ و در ابتدای موج سیستم‌های خبره، شرکت دیجیتال اکوئیپمنت کورپوریشن (DEC)یک سیستم مبتنی بر قواعد به‌نام R1 طراحی و پیاده‌سازی کرد که در داخل شرکت با نام XCON (مخفف Expert Configurer) نیز شناخته می‌شد.

دیجیتال ایکویپ‌منت نام شرکتی در ایالات متحده بود که در سال‌های ۱۹۶۰ تا ۱۹۹۰ به عنوان یکی از بزرگ‌ترین کمپانی‌ها در صنعت کامپیوتر شناخته می‌شد. این کمپانی به خاطر دیدگاه کن اولسون (از بنیان‌گذاران شرکت) در اوج ارزش سهام، ورشکست شد. زمانی که سیستم‌های کامپیوتر شخصی به بازار آمد و همه از کمپانی انتظار ورود به این عرصه را داشتند وی گفت که کامپیوترهای شخصی، اسباب‌بازی‌هایی بیش نیستند و سرمایه‌گذاری بر این سیستم‌ها، شرکت‌های سرمایه‌گذار را نابود خواهد کرد.

طبق گفته‌ی طراح اصلی پروژه، جان مک‌درموت از دانشگاه کارنگی ملون، هدف R1 این بود که به DEC کمک کند سفارش‌های مشتریان برای رایانه‌ها را آماده و پیکربندی کند:

> دامنه‌ی تخصص R1، پیکربندی برخی سامانه‌های شرکت DEC است. ورودی آن، سفارش مشتری و خروجی‌اش، مجموعه‌ای از نمودارهاست که روابط مکانی بین قطعات سفارش‌شده را نشان می‌دهد؛ تکنسینی که دستگاه را سرهم می‌کند، از این نمودارها استفاده می‌کند.

در دهه‌ی ۱۹۸۰، سیستم‌های خبره‌ی دیگری نیز برای کمک به تصمیم‌گیری در حوزه‌های مختلف مانند تولید صنعتی، تشخیص پزشکی، سرمایه‌گذاری مالی، زمین‌شناسی و اکتشاف معادن توسعه یافتند. با این حال، همان‌طور که نمودار شکل ۲٫۶ نشان می‌دهد، از اواخر دهه‌ی ۱۹۸۰، میزان استفاده از اصطلاح «سیستم‌های خبره» هم شروع به کاهش کرد و این روند در سراسر دهه‌ی ۱۹۹۰ ادامه یافت. در همین بازه، استفاده از واژه‌ی «هوش مصنوعی» نیز کاهش یافت. این افول سیستم‌های خبره، هم‌زمان با آن چیزی است که بسیاری از متخصصان، «**دومین زمستان هوش مصنوعی**» می‌دانند که

حدود سال **۱۹۹۰** آغاز شد و تا اوایل دهه‌ی ۲۰۰۰ ادامه یافت. اما چه اتفاقی افتاد؟ و چرا سیستم‌های خبره از رونق افتادند؟ چند دلیل وجود دارد:

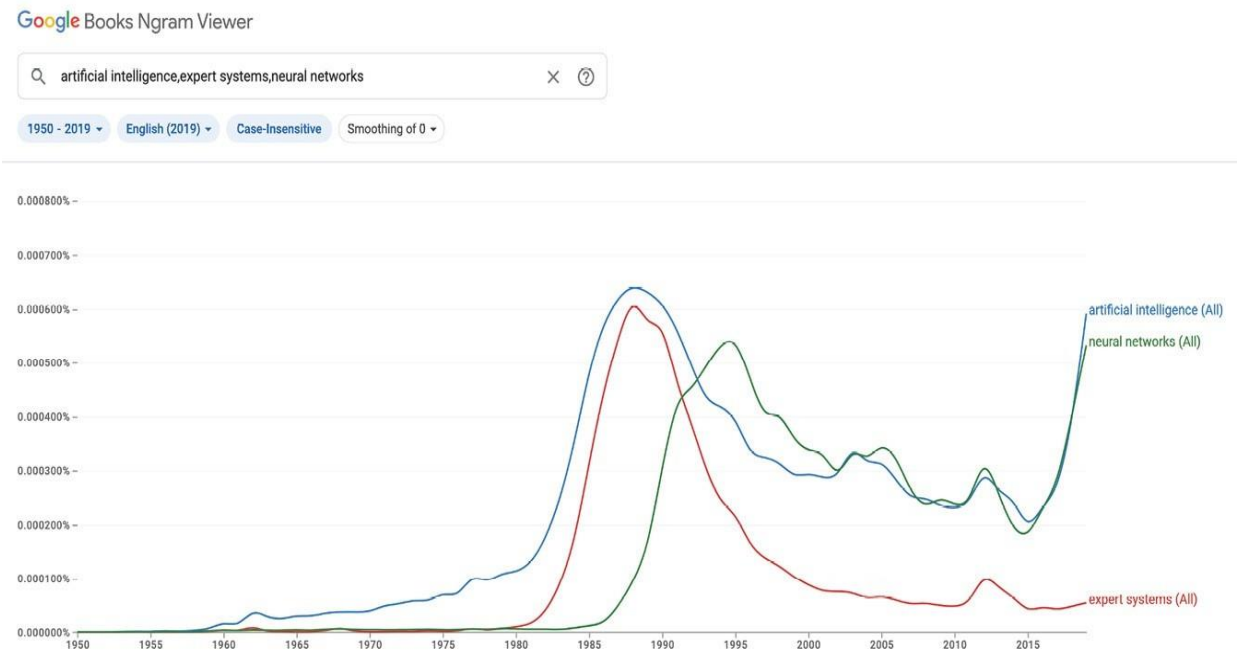
سیستم‌های خبره **بیش از حد تخصصی و شکننده** بودند. این سیستم‌ها تنها در **یک حوزه‌ی خاص** کاربرد داشتند و به‌محض آنکه شرایط آن حوزه اندکی تغییر می‌کرد، برنامه ناکارآمد یا بی‌اثر می‌شد. به‌عبارتی، این سیستم‌ها توانایی تعمیم نداشتند و **به‌روزرسانی مداوم** می‌طلبیدند، یعنی فرآیندی **پرهزینه**. حتی جان مک‌کارتی نیز در سال ۱۹۸۴ این سیستم‌ها را نقد کرد و گفت تقریباً هیچ‌کدام از آن‌ها، دانسته‌ها و توانایی‌های عقل سلیمی را که در هر انسانِ معمولی وجود دارد، ندارند.

از زاویه‌ای دیگر، شاید سیستم‌های خبره آن‌قدر خوب کار می‌کردند که دیگر «هوش مصنوعی» تلقی نمی‌شدند. پدیده‌ای که پیش‌تر به آن با عنوان **اثر هوش مصنوعی** (AI effect)اشاره کردیم. وقتی فناوری‌ای آن‌قدر عادی و مؤثر شود که دیگر «جادویی» به‌نظر نرسد، به‌سادگی تبدیل به «نرم‌افزار» می‌شود. گمان می‌کنم این عامل هم در افت علاقه به سیستم‌های خبره نقش داشته چرا که امروزه بسیاری از برنامه‌ها حاوی قواعد و پایگاه‌های دانشی‌اند که توسط متخصصان برای کمک به تصمیم‌گیری طراحی شده‌اند. رویکرد پشت این سیستم‌ها از بین نرفته اما **برندسازی و بازاریابی مربوط به آن کم‌رنگ شده** است.

هر دو دلیل بالا در افول سیستم‌های خبره مؤثر بودند. اما دلیل سومی نیز وجود دارد: ظهور شبکه‌های عصبی.

شبکه‌های عصبی

اگر نگاهی به نسخه‌ی سوم ابزار واژه‌نگار کتاب‌های گوگل بیندازیم (که در شکل ۲٫۷ نشان داده شده)، مشاهده می‌کنیم هم‌زمان با کاهش کاربرد اصطلاح «سیستم‌های خبره» در کتاب‌ها، اصطلاح دیگری به‌تدریج رو به افزایش بوده است: «شبکه‌های عصبی». جالب‌تر آن‌که روند رشد این اصطلاح از زمانی که از نظر فراوانی از «سیستم‌های خبره» پیشی گرفت تا سال ۲۰۱۹ تقریباً با منحنی «هوش مصنوعی» مطابقت دارد.



شکل ۲٫۷ - نمودار واژه‌نگار کتاب‌های گوگل برای عبارات «هوش مصنوعی»، « سیستم‌های خبره» و «شبکه‌های عصبی» در بازه‌ی ۱۹۵۰ تا ۲۰۱۹

در بخش بعدی این کتاب، فرصت خواهیم داشت که جنبه‌های فنی شبکه‌های عصبی مصنوعی (یا به‌اختصار، شبکه‌های عصبی) را با جزئیات بیشتری بررسی کنیم. اما فعلاً به‌طور خلاصه می‌توان گفت که این شبکه‌ها، نوعی مدل محاسباتی هستند که **از ساختار و عملکرد مغز انسان الهام گرفته شده‌اند** و به‌گونه‌ای طراحی شده‌اند که از طریق **لایه‌هایی از نورون‌های مصنوعی**، خود را با الگوهای داده تطبیق می‌دهند. اگر بخواهیم

به‌شدت ساده‌سازی کنیم، **مغز انسان شامل ده‌ها میلیارد نورون است؛** سلول‌هایی که از طریق سیگنال‌های الکتریکی و بر اساس محرک‌های مختلف با یکدیگر ارتباط برقرار می‌کنند. **قدرت این ارتباطات بین نورون‌ها به مرور زمان و در نتیجه‌ی تجربه‌های ما تغییر می‌کند** و همین امر توانایی یادگیری، تغییر و رشد را در ما ایجاد می‌کند. به‌طور مشابه، **وزن بین نورون‌های متصل در یک شبکه‌ی عصبی مصنوعی نیز در فرآیند آموزش تنظیم می‌شود** و بنابراین گفته می‌شود که این شبکه‌ها می‌توانند «یاد بگیرند» که یک وظیفه را تنها **با استفاده از داده‌های واقعی (و نه قوانین از پیش تعریف‌شده)** انجام دهند. در حال حاضر، شبکه‌های عصبی به‌عنوان بارزترین نمونه‌ی پیاده‌سازی هوش مصنوعی **ارتباط‌گرا** (connectionist) یا **غیرنمادین** (subsymbolic) شناخته می‌شوند و به **جریان غالب در هوش مصنوعی** امروز بدل شده‌اند. پیش از آن‌که جزئیات بیشتری را بررسی کنیم، بهتر است نگاهی به پیشینه‌ی تاریخی این شبکه‌ها بیندازیم.

در دهه‌های نخست پیدایش حوزه هوش مصنوعی، نوع ارتباط‌گرا به حاشیه رانده شده بود. بسیاری از پیشگامان اولیه‌ی این حوزه معتقد بودند که این نوع به نتیجه‌ی خاصی نخواهد رسید. با این حال، امروزه روشن شده که همان توسعه‌های اولیه، سنگ‌بنای پیشرفت‌های شگرفی را در زمینه‌ی هوش مصنوعی امروز گذاشتند. آن‌ها فناوری‌های ساده‌ای خلق کردند که رفته‌رفته به پیاده‌سازی‌های قدرتمندی تبدیل شدند که اکنون جهان ما را متحول کرده‌اند. شبکه‌های عصبی، همان‌طور که یادتان هست، یکی از موضوعات مورد توجه در کارگاه دارتموث در آغاز رسمی حوزه‌ی هوش مصنوعی در سال ۱۹۵۶ بودند. با وجود اینکه از همان ابتدا حضور داشتند، این رویکرد خیلی زود از مرکز توجه خارج شد چرا که بنیان‌گذاران حوزه، تمرکز خود را بر رویکردهای قاعده‌محور و نمادین گذاشتند.

در همان زمان، فردی که در آن کارگاه حضور نداشت بذرهای شکوفایی امروزین هوش مصنوعی را می‌کاشت. در اواخر دهه‌ی ۱۹۵۰، روان‌شناس جوان آمریکایی به نام **فرانک روزنبلات** در آزمایشگاه هوانوردی کرنل در بوفالو نیویورک و با حمایت دفتر تحقیقات نیروی دریایی آمریکا مشغول به کار بود. او در سال ۱۹۵۷، **پرسپترون** را توسعه داد: **مدلی از مغز** که به گفته‌ی خودش «از مجموعه‌ای از واحدهای تولیدکننده‌ی سیگنال (یا نورون‌ها) تشکیل شده بود». این مدل، پیاده‌سازی نظریه‌ای بود که در سال ۱۹۴۳ توسط دو دانشمند عصب‌شناس، وارن مک‌کولا و والتر پیتز، در مقاله‌ای بنیادین ارائه شده بود. **پرسپترونِ روزنبلات، نخستین پیاده‌سازی عملی یک شبکه‌ی عصبی مصنوعی بود** و توانایی یادگیری از داده‌ها را داشت. هدف او با اهداف رایج در هوش مصنوعی متفاوت بود؛ او صرفاً به‌دنبال ساخت دستگاهی بود که بتواند ساختارهای مغز انسان را مدل‌سازی و بررسی کند. وی درباره‌ی هدف خود این‌گونه می‌نویسد:

> برنامه‌ی پرسپترون در درجه‌ی اول دغدغه‌ی ساخت دستگاه‌هایی برای هوش مصنوعی ندارد بلکه در پی بررسی ساختارهای فیزیکی و ویژگی‌های نورودینامیکی‌ای است که در پسِ هوش طبیعی قرار دارند. **پرسپترون، یک مدل مغز است، نه ابزاری برای تشخیص الگو.**

در سال ۱۹۵۸، روزنبلات نمایشی رسانه‌ای برگزار کرد که در آن، مجموعه‌ای از کارت‌های پانچ‌شده را به رایانه‌ی **IBM ۷۰۴**داد؛ برخی کارت‌ها، سوراخ‌هایی در سمت چپ داشتند

و برخی در سمت راست. رایانه به نرم‌افزاری مجهز بود که عملکرد پرسپترون را شبیه‌سازی می‌کرد. این برنامه، با معیارهای امروز بسیار ابتدایی بود و تنها **شامل چند ده نورون مصنوعی در یک لایه** می‌شد. پس از پنجاه آزمون، برنامه قادر شد بین دو نوع کارت تمایز قائل شود بدون اینکه دستورالعمل صریحی برای این کار دریافت کرده باشد. رسانه‌ها به‌شدت به این موضوع پرداختند. روزنامه‌ی Oklahoma Times تیتر زد: «هیولای فرانکنشتاین ساخته‌ی نیروی دریایی؛ رباتی که فکر می‌کند». **نیویورک تایمز** نیز در مطلبی با عنوان «ابزار جدید نیروی دریایی با یادگیری از تجربه» چنین نوشت:


The New York Times

NEW NAVY DEVICE LEARNS BY DOING; Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

July 8, 1958

> امروز، نیروی دریایی آمریکا نمونه‌ی ابتدایی یک رایانه‌ی الکترونیکی را معرفی کرد که انتظار دارد قادر به راه‌رفتن، صحبت‌کردن، دیدن، نوشتن، تولید مثل و آگاهی از وجود خود باشد. گفته شد که این، پایه‌ی ساخت نخستین ماشین‌های فکری پرسپترون خواهد بود که قادر به خواندن و نوشتن هستند.

اما همچون دیگر وعده‌ها، **این بازه‌ی زمانی نیز بیش از حد خوش‌بینانه بود** و در پی چالش‌هایی که در گسترش کاربرد پرسپترون به‌وجود آمد، شور و اشتیاق به تدریج فروکش کرد. ده سال بعد، ماروین مینسکی و سیمور پاپرت با انتشار نظریه‌ای که در آن با «داوری شهودی» اعلام کردند گسترش پرسپترون به لایه‌های چندگانه احتمالاً «بی‌حاصل» خواهد بود، ضربه‌ای مهلک به این حوزه وارد کردند. اما همان‌طور که در فصل ۴ خواهیم دید، این قضاوت اشتباه بود، بسیار اشتباه. در نهایت، پیشرفت‌های بعدی در شبکه‌های عصبی موجب **بازگشت پرقدرت هوش مصنوعی ارتباط‌گرا** به کانون

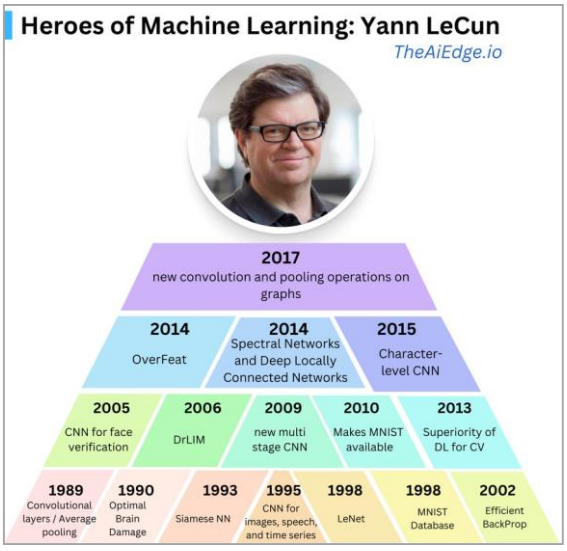
توجه شد. فرانک روزنبلات نتوانست شاهد تأیید کار خود باشد؛ او به‌طرز تراژیکی در روز تولد ۴۳ سالگی‌اش در ۱۱ ژوئیه ۱۹۷۱، در حادثه قایق‌سواری جان باخت.

برای رعایت اختصار، در ادامه به **برخی از نقاط عطف مهم** پس از پرسپترون روزنبلات اشاره می‌کنم. درباره‌ی این دستاوردها، کتاب‌هایی نوشته شده و حتی مستندهای برنده‌ی جایزه نیز ساخته شده‌اند.

۱۹۷۹: شبکه عصبی Neocognitron فوکوشیما

در سال ۱۹۷۹، فوکوشیما، دانشمندی ژاپنی، شبکه‌ی عصبی چندلایه‌ای به نام نئوکگنیترون طراحی کرد که می‌توانست ویژگی خاصی را در یک تصویر، بدون توجه به محل قرارگیری آن، تشخیص دهد. این مدل برای **تشخیص خط دست‌نویس ژاپنی** به‌کار گرفته شد.

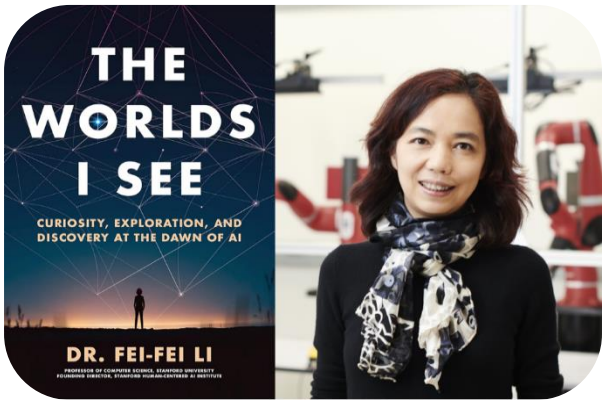
۱۹۸۹: شبکه عصبی LeNet، یان لیکان



یان لیکان، دانشمند فرانسوی و پژوهشگر آزمایشگاه‌های بل که هم‌اکنون دانشمند ارشد هوش مصنوعی در شرکت **متا** است، معماری‌ای به نام شبکه عصبی پیچشی (CNN - Convolutional Neural Network) پیشنهاد داد که از نئوکگنیترون الهام گرفته

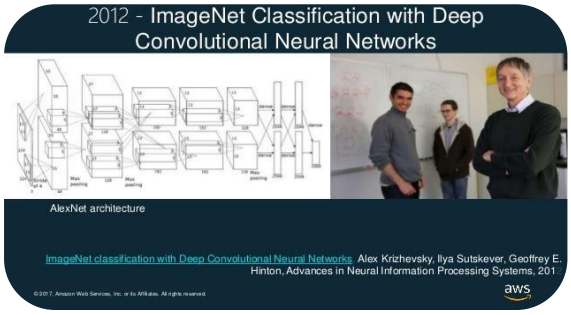
بود. مدل او از الگوریتمی به نام پس‌انتشار خطا یا انتشار رو به عقب (Backpropagation) برای یادگیری استفاده می‌کرد و هدف آن تشخیص کدهای پستی دست‌نویس بود.

۲۰۰۹: پایگاه داده ImageNet، فِی‌فِی لی



خانم لی، پژوهشگر چینی‌تبار دانشگاه پرینستون، ساخت یک پایگاه داده‌ی عظیم شامل **تصاویر برچسب‌خورده** برای آموزش مدل‌های **بینایی ماشین** را در سال ۲۰۰۶ آغاز کرد. او با استفاده از سرویس Mechanical Turkآمازون از کاربران خواست تا اشیای موجود در **تصاویر** را **برچسب‌گذاری** کنند (مثل «هواپیما»، «گربه» یا «سگ»). نسخه‌ی اولیه **ImageNet** در سال ۲۰۰۹ با **۳/۲ میلیون تصویر** منتشر شد و تا امروز به بیش از **۱۴** میلیون تصویر گسترش یافته است. در سال ۲۰۱۰، لی با برگزاری رقابت سالانه‌ای به نام چالش ImageNet Large Scale Visual Recognition، تحولی اساسی ایجاد کرد. این رقابت موجب توجه جهانی به ImageNet شد و نقش مهمی در ترویج **استفاده از کلان‌داده‌ها** برای **آموزش شبکه‌های عصبی** ایفا کرد.

۲۰۱۲: شبکه عصبی AlexNet، الکس کریژفسکی



در ۳۰ سپتامبر ۲۰۱۲، الکس کریژفسکی، دانشجوی کارشناسی ارشد دانشگاه تورنتو با راهنمایی جفری هینتون، شبکه‌ای ۸ لایه‌ای

ارائه کرد که دارای ۶۰ میلیون وزن بود و به شکل پس‌انتشار آموزش دیده بود. دقت آن به‌طور چشمگیری بهتر از رقبا بود. این موفقیت، علاقه‌ی گسترده‌ای به شبکه‌های عصبی عمیق ایجاد کرد.

۲۰۱۶: شبکه AlphaGo متعلق به شرکت DeepMind

در سال ۲۰۱۶، شرکت DeepMind متعلق به **گوگل** با برنامه‌ای به نام **AlphaGo** موفق شد **لی سدول، قهرمان کره‌ای بازی Go،** را شکست دهد. بازی Go با داشتن **بیش از** $۲/۱ \times ۱۰^{۱۷۰}$ وضعیت ممکن، چالشی **بسیار فراتر از شطرنج** برای هوش مصنوعی به‌شمار می‌رفت. AlphaGo با استفاده از ۱۲ شبکه‌ی CNN و روش یادگیری تقویتی آموزش دیده بود. پیروزی آن، توجه جهانی را برانگیخت و حتی دولت کره جنوبی پس از آن، بودجه‌ای ۸۶۳ میلیون دلاری برای توسعه‌ی هوش مصنوعی اختصاص داد.

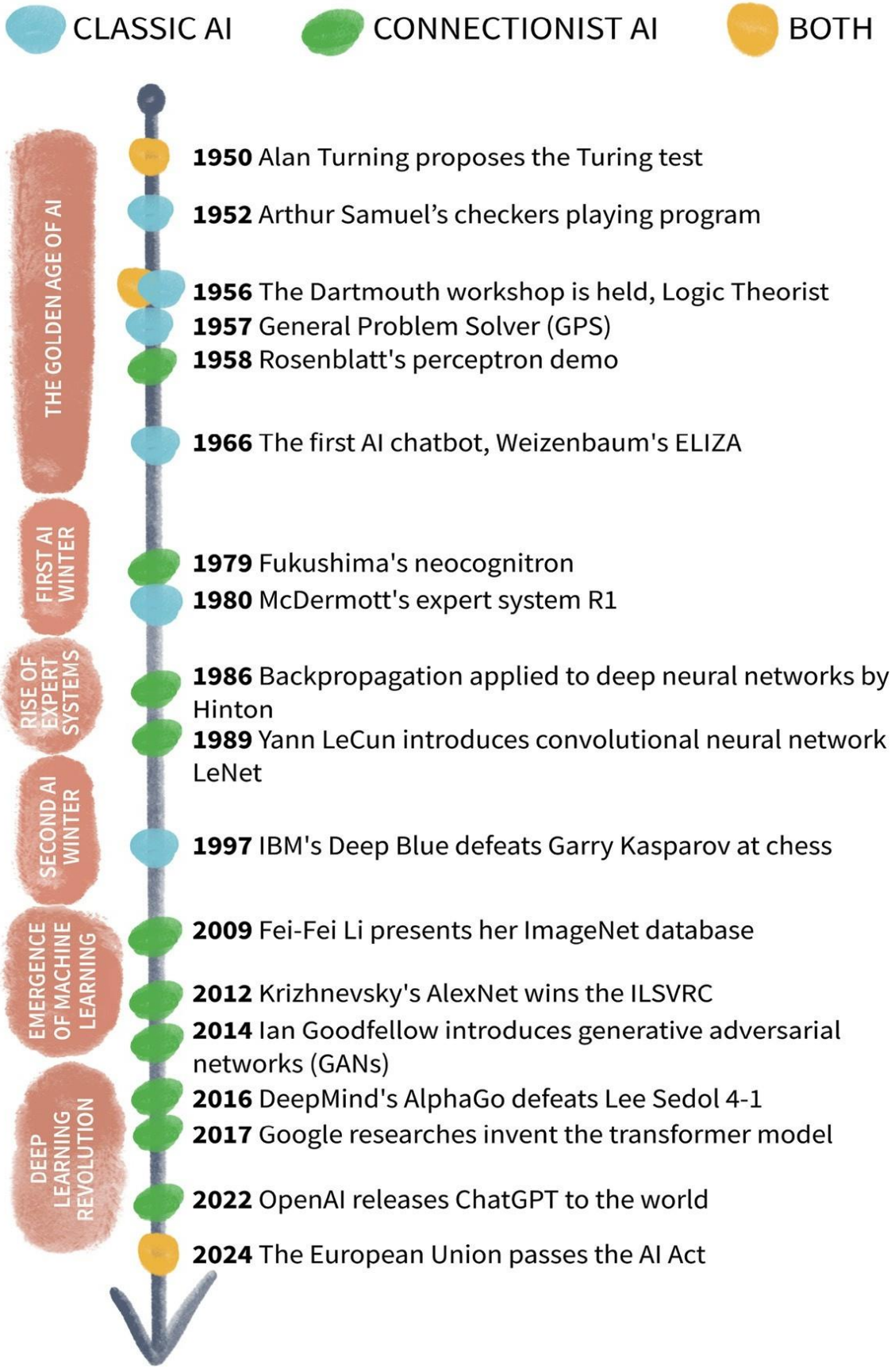
۲۰۲۲: ChatGPT از شرکت OpenAI

در ۳۰ نوامبر ۲۰۲۲، شرکت OpenAI از چت‌بات ChatGPT رونمایی کرد. GPT نوعی مدل زبانی عظیم است که با حجم انبوهی از متون آموزش دیده است. این چت‌بات، ظرف تنها دو ماه به ۱۰۰ میلیون کاربر رسید و رکورد سریع‌ترین رشد در تاریخ اپلیکیشن‌ها را شکست. در **مارس ۲۰۲۳**، نسخه‌ی **پولی** ChatGPT Plus با مدل GPT-4 ارائه شد. قابلیت‌هایی چون جست‌وجوی وب، ورودی صوتی، تولید تصویر و اتصال به سرویس‌های ثالث نیز به آن افزوده شد. این موفقیت چشمگیر، **موجی از توجه جهانی**

به هوش مصنوعی به همراه داشت. البته نگرانی‌هایی نیز مطرح شد: از نقض حقوق مالکیت فکری گرفته تا بازتولید سوگیری‌های اجتماعی و حتی تهدید وجودی بشر. آیا این هیجان به زمستان دیگری در هوش مصنوعی منجر خواهد شد؟ زمان پاسخ خواهد داد.

خلاصه

در تاریخچه‌ی هوش مصنوعی با **الگوی چرخه‌ای توجه و بی‌توجهی به هوش مصنوعی** آشنا شدید و درباره‌ی انشعاب‌ها و شاخه‌های مختلف آن، به‌ویژه تفاوت میان هوش مصنوعی نمادین (کلاسیک) و غیرنمادین (ارتباط‌گرا) مطالبی آموختید. برای مرور خلاصه‌ای از رویدادهای مهم تاریخ هوش مصنوعی، می‌توانید به شکل ۲٫۸ مراجعه کنید. اکنون زمان آن رسیده که به‌صورت عمیق‌تری به بررسی این فناوری‌ها بپردازیم؛ به‌ویژه فناوری‌هایی که به شاخه‌ی غیرنمادین (ارتباط‌گرا) تعلق دارند چون فناوری‌های این شاخه، نیروی محرک بسیاری از پیشرفت‌های اخیر بوده‌اند؛ از بینایی ماشین گرفته تا پردازش زبان طبیعی و هوش مصنوعی مولد. **درک ما از هوش مصنوعی به میزان شناخت ما از یادگیری ماشین** و زیرمجموعه‌ی مهمی از آن به نام **یادگیری عمیق** وابسته است.



شکل ۲٫۸ - تاریخچه رویدادهای «هوش مصنوعی»

فصل سوم

مبانی یادگیری ماشین

در فصل قبل، سفری داشتیم در تاریخچه‌ی هوش مصنوعی؛ از روزهای نخست تا شکل امروزی آن. تردیدی نیست که هوش مصنوعی از زمان کارگاه دارتموث تا امروز، یعنی **در طول هفت دهه**، راهی طولانی را پیموده است؛ هرچند هنوز **راه درازی تا تحقق کاملِ ظرفیت‌های آن در پیش داریم**. در واقع، درک اینکه «نهایت توان بالقوه‌ی هوش مصنوعی» چه شکلی خواهد داشت هنوز فاصله‌ی زیادی با ما دارد چرا که **مسیرهای بسیار متنوعی پیش روی آن است.**

همان‌طور که دیدیم، **پیشرفت در هوش مصنوعی همواره خطی و یکنواخت نبوده** است. این روند را به تغییر فصل‌ها تشبیه کردیم: بهار و تابستان‌هایی سرشار از پیشرفت و اشتیاق، و زمستان‌هایی از بی‌توجهی و حتی عقب‌گرد. اما می‌توان تاریخچه‌ی هوش مصنوعی را به گونه‌ای دیگر هم نگریست: همچون تکامل یک گونه‌ی زنده. مشابه نظریه‌ی تعادل در زیست‌شناسی تکاملی، هوش مصنوعی نیز دوره‌هایی از رکود را پشت سر گذاشته که در آن‌ها حتی بقای آن زیر سؤال رفته، و در مقابل، جهش‌های ناگهانی و چشم‌گیری نیز در توانمندی‌هایش داشته است. این جهش‌های بزرگ، حاصل مجموعه‌ای از پیشرفت‌های فناورانه بوده که خود، نتیجه‌ی تغییرات مهمی در محیط پیرامون بوده‌اند. در این فصل، با **خانواده‌ای از این پیشرفت‌ها** آشنا خواهیم شد که در **قلب هوش مصنوعی مدرن** قرار دارند: **یادگیری ماشین.**

مفاهیم و اصطلاحات پایه

یادگیری ماشین (ML)، **اصطلاحی کلی** است برای مطالعه و استفاده از **انواع مختلف الگوریتم‌های آماری** که می‌توان آن‌ها را روی **مجموعه‌ای از داده‌ها** به کار برد تا از **الگوهای موجود** در داده‌ها، **نحوه‌ی انجام یک کار مشخص** را یاد بگیرند. زمانی که

الگوریتم یادگیری ماشین را روی یک مجموعه داده اعمال می‌کنیم تا الگوهای آن را بیاموزد، می‌گوییم **مدلی را آموزش می‌دهیم.**

مجموعه داده‌ای که برای آموزش به کار می‌بریم، می‌تواند هر نوع داده‌ی دیجیتالی، برگرفته از دنیای واقعی باشد: مجموعه‌ای از فایل‌های تصویری، کتاب‌های الکترونیکی یا اسناد دیجیتال دیگر، فایل‌های صوتیِ گفت‌وگوها، جدول‌هایی از قیمت‌های تاریخی سهام و غیره. معمولاً **در یادگیری ماشین، حجم داده‌ها بسیار زیاد است.** هر چه داده بیشتر باشد، بهتر است. این‌جا صحبت از چند عکس دیجیتال نیست بلکه از هزاران یا حتی میلیون‌ها تصویر سخن می‌گوییم. البته **داده‌های آموزشی، نماینده‌ای از کل داده‌های ممکن** هستند، نه تمام آن‌ها. هدف این است که مدل بتواند از این نمونه‌ها بیاموزد تا با داده‌هایی که در مجموعه‌ی آموزشی نبوده‌اند نیز به‌درستی کار کند.

الگوریتم، مجموعه‌ای از دستورالعمل‌هاست که برنامه در فرایند آموزش از آن پیروی می‌کند. الگوریتم را می‌توان مانند یک دستور آشپزی در نظر گرفت: **فرآیندی گام‌به‌گام برای یادگیری از داده‌های آموزشی**. الگوریتم‌های یادگیری ماشین انواع گوناگونی دارند؛ از شبکه‌های عصبی (Neural Networks) گرفته تا درخت‌های تصمیم‌گیری (Decision Trees) و الگوریتم‌های خوشه‌بندی (Clustering Algorithms). در این کتاب با برخی از آن‌ها آشنا خواهیم شد.

مدل، نتیجه‌ی فرآیند آموزش است؛ در واقع، همان **الگوریتم آموزش دیده** است. **مدل دارای پارامترهایی است که طی فرآیند آموزش تنظیم شده‌اند** مانند پیچ‌های تنظیم روی یک رادیوی قدیمی که برای دریافت بهتر صدا، آن‌ها را تنظیم می‌کنید. این **تنظیمات** که **الگوریتم** در طی **آموزش** روی **مدل** اعمال می‌کند، باعث می‌شوند مدل بتواند **کار مشخصی** را با **سطحی از دقت** (نه کاملاً بی‌نقص) انجام دهد. برای مثال،

فرض کنید می‌خواهید برنامه‌ای بنویسید که بتواند رقم‌هایی را که افراد به‌صورت دستی روی چک نوشته‌اند تشخیص دهد، تا کاربر بتواند از طریق گوشی خود، چک را به حساب بانکی‌اش واریز کند. این برنامه باید تصویر دیجیتال چک را به مبلغ واریزی، شماره حساب و شماره شبا (یا کد شناسایی بانکی) تبدیل کند. مشخص شده که استفاده از یادگیری ماشین، نتایج بسیار بهتری نسبت به دستورالعمل‌نویسی برای چنین برنامه‌ای خواهد داشت. چطور می‌خواهید به برنامه بگویید با سبک‌های مختلف دست‌خط، یا با کیفیت، زاویه و نورهای گوناگونِ عکس، چگونه برخورد کند؟

به‌جای نوشتن دستورالعمل، می‌توانید یک الگوریتم شبکه عصبی کانولوشنی (CNN) را با تعداد زیادی تصویر دیجیتالی از چک‌های واقعی که اعداد دست‌نویس دارند، آموزش دهید. مدل حاصل از این شبکه، دارای پارامترهای دقیقی به نام «وزن» خواهد بود که می‌توانند، هنگام دریافت تصاویری خارج از مجموعه‌ی آموزشی، با دقت قابل‌قبولی اعداد را تشخیص دهند. با **داشتن داده‌ی آموزشی کافی** و انتخاب **الگوریتم یادگیری درست**، توانمندی‌های **مدل نهایی** واقعاً شگفت‌انگیز خواهد بود. این، نمونه‌ای از کاربرد یادگیری ماشین است.

زمینه تاریخی یادگیری ماشین

اصطلاح «یادگیری ماشین» در دهه **۱۹۵۰** توسط **آرتور ساموئل**، پیشگام آمریکایی هوش مصنوعی ابداع شد. برنامه‌ی **بازی چکرزِ** او، یکی از نخستین نمونه‌های برنامه‌ای بود که نوعی «**یادگیری**

خودکار» را به نمایش می‌گذاشت. این برنامه با هر بار بازی، سیستم امتیازدهی درونی خود را تنظیم می‌کرد و در نتیجه، **با گذشت زمان انتخاب‌های بهتری در حرکاتش انجام می‌داد.**

امروز «**توانایی یادگیری**»، بخش بسیار مهمی از هوش مصنوعی است: یعنی درک یا مواجهه با موقعیت‌های جدید یا دشوار. به عبارت دیگر، یادگیری یعنی درک محیط، فهمیدن آنچه در حال رخ دادن است و سپس تطبیق با شرایط برای بهبود نتایج. این توانمندی‌ها بخش بزرگی از مفهوم «**هوشمندی**» را دربرمی‌گیرند؛ چه در انسان‌ها و چه در رایانه‌ها.

رویکرد **مبتنی بر داده‌ی یادگیری ماشین،** جوهره‌ی **شاخه‌ی هوش مصنوعی غیرنمادین** را تشکیل می‌دهد و در تضاد با رویکرد مبتنی بر قواعد در هوش مصنوعی نمادین است. هوش مصنوعی در دهه‌ی ۱۹۵۰ از سمت هوش مصنوعی نمادین آغاز شد و در اواخر دهه ۱۹۹۰ و اوایل دهه‌ی ۲۰۰۰ به سمت هوش مصنوعی غیرنمادین و یادگیری ماشین حرکت کرد؛ و از آن زمان تاکنون نیز در همین سو باقی مانده است. هوش مصنوعی در آغاز قرن بیست‌ویکم، اساساً درباره‌ی «یادگیری» بوده است. اما مسیر بعدی آن به کجا خواهد رفت؟ خواهیم دید.

به عنوان یک **رشته‌ی علمی**، یادگیری ماشین همیشه در حوزه هوش مصنوعی تعریف نشده است. در دوره‌هایی از زمان، **برخی از متخصصان یادگیری ماشین، خود را جدا از حوزه‌ی هوش مصنوعی می‌دانستند** و برعکس. به نظر می‌رسد این تفکیک‌ها، اغلب به دلیل تفاوت در حوزه‌های تمرکز بوده‌اند و شاید تا حدی نیز ناشی از رقابت برای دریافت بودجه یا اعتبار علمی. در سال‌های اخیر، اصطلاح ترکیبی «AI/ML» نیز پدید آمده تا به هوش مصنوعی‌ای اشاره کند که تمرکز خاصی بر تکنیک‌های یادگیری ماشین دارد.

برخی این اصطلاح را مفید می‌دانند چون ارتباط این دو حوزه را نشان می‌دهد اما برخی دیگر آن را غیرضروری و حتی گمراه‌کننده تلقی می‌کنند. از نگاه من، **یادگیری ماشین، زیرشاخه‌ای در دل حوزه‌ی گسترده‌تر هوش مصنوعی** است و در حال حاضر، این اصطلاحِ ترکیبیِ AI/ML، ارزش چندانی به گفتگوهای علمی اضافه نمی‌کند.

اگر خوب دقت کنیم، **ابتدایی‌ترین شکل‌های یادگیری ماشین در همان آغازِ هوش مصنوعی حضور داشتند.** برنامه‌ی چکرز آرتور ساموئل و پرسپترون فرانک روزنبلات، هر دو در دهه‌ی ۱۹۵۰ و در دوران طلایی اولیه‌ی هوش مصنوعی ایجاد شدند. پس چرا باید بیش از سه دهه بگذرد تا یادگیری ماشین به **پارادایم غالب در هوش مصنوعی** تبدیل شود؟ بخشی از پاسخ این است که **شرایط محیطی** باید به‌گونه‌ای تغییر می‌کرد که **امکان رشد و سلطه‌ی یادگیری ماشین** فراهم شود.

عامل اول: افزایش توان محاسباتی هوش مصنوعی

یکی از شرایط محیطی که طی دهه‌ها به‌طور مستمر تغییر کرده، قدرت و سرعت محاسباتی است که آن را با عنوان «**توان محاسباتی هوش مصنوعی**» یا به اختصار «**AI compute**» می‌شناسند. بر اساس **قانون مور** (که به نام گوردون مور، یکی از بنیان‌گذاران شرکت اینتل، نام‌گذاری شده) از حدود سال ۱۹۷۵ تعداد اجزایی که می‌توان روی یک **مدار مجتمع** جای داد، هر سال **دو برابر** می‌شود.

افزون بر این، استفاده‌ی گسترده از **پردازنده‌های گرافیکی (GPU)** در آموزش مدل‌های یادگیری ماشین به‌ویژه آموزش **شبکه‌های عصبی عمیق** (که در

فصل بعد بررسی می‌کنیم) به جهش چشم‌گیری در عملکرد این مدل‌ها منجر شده است. پژوهشگران هوش مصنوعی، روش‌هایی یافته‌اند که **چند GPU را به‌صورت موازی برای آموزش مدل‌های یادگیری ماشین** به کار بگیرند که این امکان را می‌دهد تا برنامه‌های هوش مصنوعی پیچیده‌تری ساخته شوند. GPUها، نوع خاصی از **تراشه‌های رایانه‌ای** هستند که برای **انجام سریع محاسبات ریاضی سنگین** جهت نمایش **گرافیک‌های سه‌بعدی** در بازی‌های ویدیویی طراحی شده‌اند. بدون این افزایش قابل‌توجه در توان محاسباتی، بعید بود یادگیری ماشین به شکلی که امروز شاهد آن هستیم شکوفا شود. در نتیجه‌ی همین کاربردهای حیاتی GPU در آموزش مدل‌ها، شرکت‌هایی مانند NVIDIA که تولیدکننده‌ی آن‌ها هستند، با رشد انفجاری فروش در پی رواج ابزارهای هوش مصنوعی مولد روبه‌رو شده‌اند.

عامل دوم: دسترسی به داده‌های آموزشی

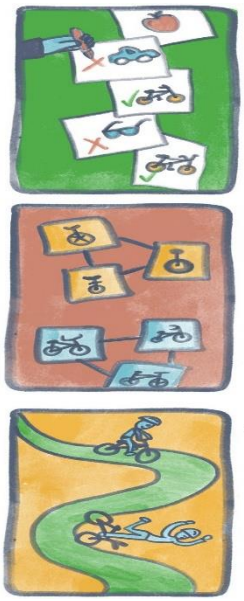
اما توان محاسباتی تنها عامل محیطی مؤثر در رشد سریع یادگیری ماشین نبوده است. برای آموزش یک مدل یادگیری ماشین جهت انجام وظایف خاص، نیاز به **داده،** آن هم به مقدار زیاد داریم. با گسترش اینترنت و گوشی‌های همراه، **حجم داده‌هایی که برای آموزش مدل‌ها در دسترس قرار گرفته‌اند به‌شدت افزایش یافته است.** کاربران شبکه‌های اجتماعی حجم عظیمی از داده را به شکل متن، تصویر، صدا و ویدیو در اینترنت بارگذاری کرده‌اند. اما موضوع تنها به شبکه‌های اجتماعی محدود نمی‌شود: کتاب‌های الکترونیکی، مقالات آنلاین، مقالات علمی در قالب PDF، و کتاب‌ها و متون قدیمی دیجیتال‌سازی شده. افزون بر این، کاربران اینترنت به‌صورت روزانه در انجمن‌های گفتگو با موضوعات بی‌شمار مشارکت داشته‌اند. بخش زیادی از این محتوا از وب گردآوری و به مخازن داده‌ای مانند Common Crawl منتقل شده‌اند؛ مخازنی که در آموزش و تنظیم

مدل‌های قدرتمند امروزی کاربرد داشته‌اند. مخازن دیگری مانند ImageNet نیز با هدف مشخص توسعه‌ی مدل‌های یادگیری ماشین ساخته شده‌اند. حال تصور کنید اگر این داده‌های غنی برای آموزش مدل‌ها وجود نداشت، یادگیری ماشین، امروز در چه جایگاهی قرار می‌گرفت؟ شاید پاسخ دقیق ندانیم اما قطعاً می‌توان گفت که به موقعیت کنونی‌اش یعنی پیشتاز بودن در هوش مصنوعی نمی‌رسید. البته نمی‌توان اهمیت این مخازن عظیم داده را بررسی کرد بی‌آنکه به چالش‌های جدی آن‌ها اشاره کرد: از **داده‌های ناقص و نادرست** گرفته تا مسائل مربوط به **کپی‌رایت، حریم خصوصی، و سوگیری‌های اجتماعی**. این موارد، نگرانی‌های اخلاقی و حقوقی مهمی درباره‌ی استفاده از مدل‌های آموزش‌دیده با داده‌های مسئله‌دار مطرح می‌کنند که در فصل پنجم به آن‌ها خواهیم پرداخت.

سه شکل اصلی یادگیری ماشین

فعلاً بیایید نگاهی بیندازیم به شیوه‌های مختلفی که الگوریتم‌های یادگیری ماشین می‌توانند از داده‌ها یاد بگیرند. همان‌طور که ما انسان‌ها بسته به شرایطی که در آن قرار داریم، می‌توانیم به روش‌های مختلفی یاد بگیریم، الگوریتم‌های یادگیری ماشین نیز شیوه‌های متفاوتی برای یادگیری دارند. گاهی در کلاس درس از معلم می‌آموزیم. گاهی در دنیای واقعی از الگوهایی که به‌تنهایی متوجه می‌شویم یاد می‌گیریم. و گاهی هم هنگام انجام فعالیتی جدید، از راه آزمون‌وخطا مهارت کسب می‌کنیم. این سه موقعیت متفاوت در زندگی واقعی، معادل سه نوع اصلی یادگیری ماشین هستند: یادگیری **نظارت‌شده**، یادگیری **بدون‌نظارت** و یادگیری **تقویتی**. هر یک از این روش‌ها در شرایط خاص و برای اهداف گوناگون مفید هستند. در این فصل، هر یک از این انواع یادگیری را به‌صورت جداگانه بررسی خواهیم کرد. به‌طور خلاصه:

- **یادگیری نظارت‌شده** شبیه یاد گرفتن از **معلم** است که به ما می‌گوید کدام پاسخ درست و کدام نادرست است؛
- **یادگیری بدون‌نظارت** شبیه یادگیری از طریق **مشاهده‌ی الگوهای اطرافمان** است؛
- **یادگیری تقویتی** شبیه یادگیری در **حین انجام دادن کار** است جایی که با دریافت بازخورد از تلاش‌های خودمان، به‌تدریج راه درست را پیدا می‌کنیم.



شکل ۳/۱ - سه نوع اصلی یادگیری ماشین

یادگیری نظارت‌شده (Supervised Learning)

اولین و شاید **رایج‌ترین** نوع یادگیری ماشین، یادگیری نظارت‌شده است. هدف در این نوع یادگیری، آموزش مدلی است که بتواند **ورودی‌ها** را به **خروجی‌های صحیح و مرتبط** تبدیل کند. این ورودی‌ها و خروجی‌ها بسته به کاربرد برنامه متفاوت‌اند. مثلاً ورودی‌ها می‌توانند عکس‌هایی از خودروهایی باشند که از یک تقاطع عبور می‌کنند و خروجی‌ها شماره پلاک خودروهایی که باید برای عبور از چراغ قرمز جریمه شوند. یا مثلاً ورودی‌ها فرمان‌های صوتی باشند و خروجی‌ها، متن این فرمان‌ها.

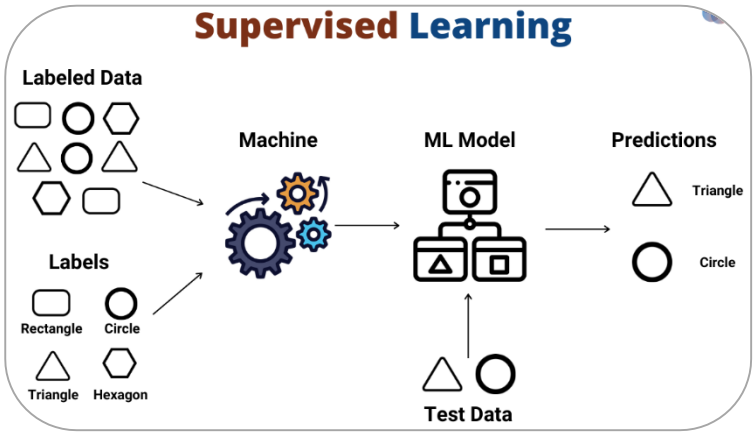
حالا تصور کنید بخواهید برای انجام چنین وظایفی، دستورالعملی دقیق برای رایانه بنویسید. مثلاً در مورد پلاک خودروها، برنامه چطور باید با **تنوع زوایا، شرایط نوری مختلف و انواع گوناگون** پلاک‌ها مقابله کند؟ یا برای فرمان‌های صوتی، چطور باید لهجه‌های مختلف، سرعت‌های متفاوت صحبت کردن و نویزهای محیطی را تحلیل کند؟ ما انسان‌ها معمولاً از عهده این کارها به‌خوبی برمی‌آییم اما برنامه‌نویسان هوش مصنوعی سال‌ها تلاش کردند تا رایانه‌ها نیز بتوانند به همان خوبی عمل کنند. تغییر بزرگ زمانی اتفاق افتاد که پژوهشگران دریافتند **یادگیری نظارت‌شده** به‌ویژه با استفاده از **شبکه‌های عصبی عمیق** می‌تواند دقت چنین سیستم‌هایی را به‌شدت افزایش دهد. در این رویکرد، انسان‌ها با ارائه‌ی **جفت‌های زیادی از ورودی‌ها و خروجی‌های صحیح**، مدل را نظارت و هدایت می‌کنند. **به‌جای آنکه مستقیماً به رایانه بگویند چه کاری انجام دهد** تعداد زیادی نمونه به آن داده می‌شود تا خود، الگوها را بیاموزد.

در اینجا داده‌های آموزشی «**برچسب‌دار**» هستند: **ورودی‌ها** دارای **ویژگی‌هایی**‌اند (Features) و **خروجی‌ها** دارای **برچسب** (Labels). مثلاً در تصویر دیجیتالی یک خودرو، ویژگی‌های ورودی، مقادیر رنگی (RGB) هر پیکسل‌اند و خروجی، شماره پلاکی است که در تصویر دیده می‌شود.

مدل در هر مرحله‌ی آموزش، ورودی را به خروجی تبدیل کرده، نتیجه‌اش را با خروجی صحیح مقایسه می‌کند، و سپس پارامترهای داخلی‌اش را تنظیم می‌کند تا خطای خود یعنی اختلاف بین خروجی پیش‌بینی‌شده و خروجی واقعی را کاهش دهد. در **ابتدا**، مدل **صرفاً حدس‌های تصادفی** می‌زند. اما پس از **تکرارهای زیاد** (که به آن‌ها **دورهای آموزش** یا **epochs** گفته می‌شود)، دقت مدل به‌طرز چشمگیری افزایش می‌یابد. این همان فرآیند آموزش در یادگیری نظارت‌شده است. این دقیقاً کاری است که فرانک

روزنبلات در سال ۱۹۵۸ انجام داد. او برنامه‌ی «پرسپترون» را روی رایانه‌ی IBM 704 اجرا کرد. ورودی‌ها، کارت‌هایی پانچ‌شده بودند که سوراخ‌هایی در سمت راست یا چپ داشتند. خروجی، چراغ‌هایی روی رایانه بود که مدل روشن یا خاموش می‌کرد تا حدس خود را نشان دهد. روزنبلات ابتدا ۵۰ کارت به مدل داد و برای هر کارت، اگر حدس مدل نادرست بود، پارامترهای داخلی‌اش را تنظیم کرد.

در یادگیری نظارت‌شده، **داده‌های آموزشی باید برچسب‌دار باشند.** برچسب‌ها، همان پاسخ‌های درست‌اند مانند کلید پاسخ‌نامه. انسانی که این پاسخ‌ها را تهیه می‌کند مانند **معلم** یا «**ناظر**» عمل می‌کند و مدل، دانش‌آموزی است که از روی نمونه‌سؤالات و پاسخ‌هایشان یاد می‌گیرد. پس از مشاهده‌ی تعداد زیادی از این جفت‌های ورودی و خروجی، مدل مورد آزمایش قرار می‌گیرد درست مانند دانش‌آموزی که برای امتحان آماده می‌شود. در این **آزمون**، **ورودی‌هایی به مدل داده می‌شود که قبلاً ندیده** اما پاسخ‌های درست آن‌ها در اختیار ناظر انسانی است. **عملکرد مدل** بر اساس این **آزمون** سنجیده می‌شود.



یادگیری بدون‌نظارت (Unsupervised Learning)

نوع بعدی یادگیری ماشین، یادگیری بدون‌نظارت است. این نوع یادگیری را می‌توان با فرآیندی مقایسه کرد که طی آن به‌تنهایی و **بدون راهنمایی معلم**، الگوهایی را در اطرافمان کشف می‌کنیم. همان‌طور که یان لیکان گفته: **بیشتر یادگیری انسان‌ها و**

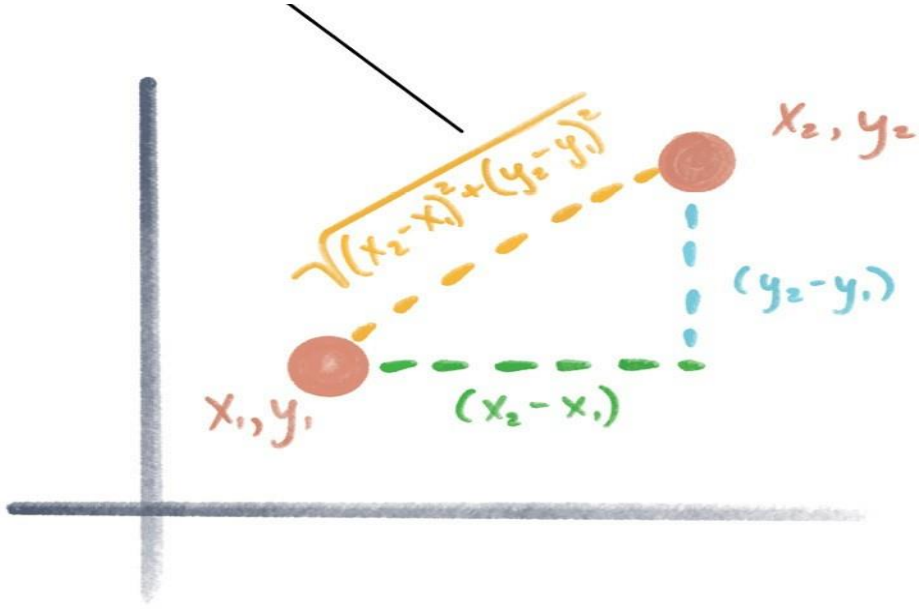
حیوانات از نوع یادگیری بدون‌نظارت است. در یادگیری بدون‌نظارت، آموزش مدل معنای متفاوتی دارد. **در اینجا خبری از مجموعه داده‌های آموزشی برچسب‌دار نیست.** یعنی هیچ مجموعه‌ای از ورودی‌ها که خروجی درست آن‌ها مشخص باشد، وجود ندارد. **داده‌ها «بدون برچسب» هستند**؛ هیچ **انسان ناظری وجود ندارد** که به مدل بگوید کدام خروجی درست یا نادرست است. در نتیجه، مدل باید خودش خروجی مناسب را پیدا کند. اما این دقیقاً به چه معناست؟

فرض کنید شرکتی می‌خواهد مشتریان خود را به گروه‌هایی با ویژگی‌های مشابه تقسیم کند. به این کار بخش‌بندی مشتریان (Customer Segmentation) گفته می‌شود. شرکت می‌تواند بر اساس ویژگی‌هایی مثل سن، محل زندگی، یا مبلغ خرید در سال گذشته، گروه‌هایی تعریف کرده و مشتریان را در آن‌ها قرار دهد. در این حالت، نیازی به یادگیری ماشین نیست؛ کافی‌ست مشتریان را با برچسب گروهشان مشخص کنیم. اما اگر شرکت بخواهد بدون تعریف قبلی، فقط الگوهای موجود در داده‌ها را کشف کند، چطور؟ در این حالت، می‌توان از **الگوریتمی** به نام **خوشه‌بندی** (Clustering) استفاده کرد. این الگوریتم با استفاده از یادگیری بدون‌نظارت، مشتریانی را که **بیشترین شباهت را به یکدیگر** دارند، در یک گروه قرار می‌دهد (**حداکثرسازی شباهت درون‌خوشه‌ای**) و در عین حال سعی می‌کند مشتریان گروه‌های مختلف تا حد امکان با یکدیگر متفاوت باشند (**حداقل‌سازی شباهت بین‌خوشه‌ای**).

الگوریتم برای این کار از معیاری مانند **فاصله اقلیدسی** استفاده می‌کند. برای محاسبه این فاصله بین دو نقطه داده، ابتدا اختلاف مقادیر آن‌ها (مختصات) را پیدا می‌کنیم، این اختلاف‌ها را به توان دو می‌رسانیم، مجموع مربع‌ها را محاسبه می‌کنیم، و سپس جذر آن را می‌گیریم. اگر مدل فقط دو متغیر داشته باشد (هرچند معمولاً بیش از دو

متغیر وجود دارد)، فاصله اقلیدسی همان طول وتر مثلثی خواهد بود که این دو نقطه تشکیل می‌دهند.

فاصله اقلیدسی بین دو نقطه

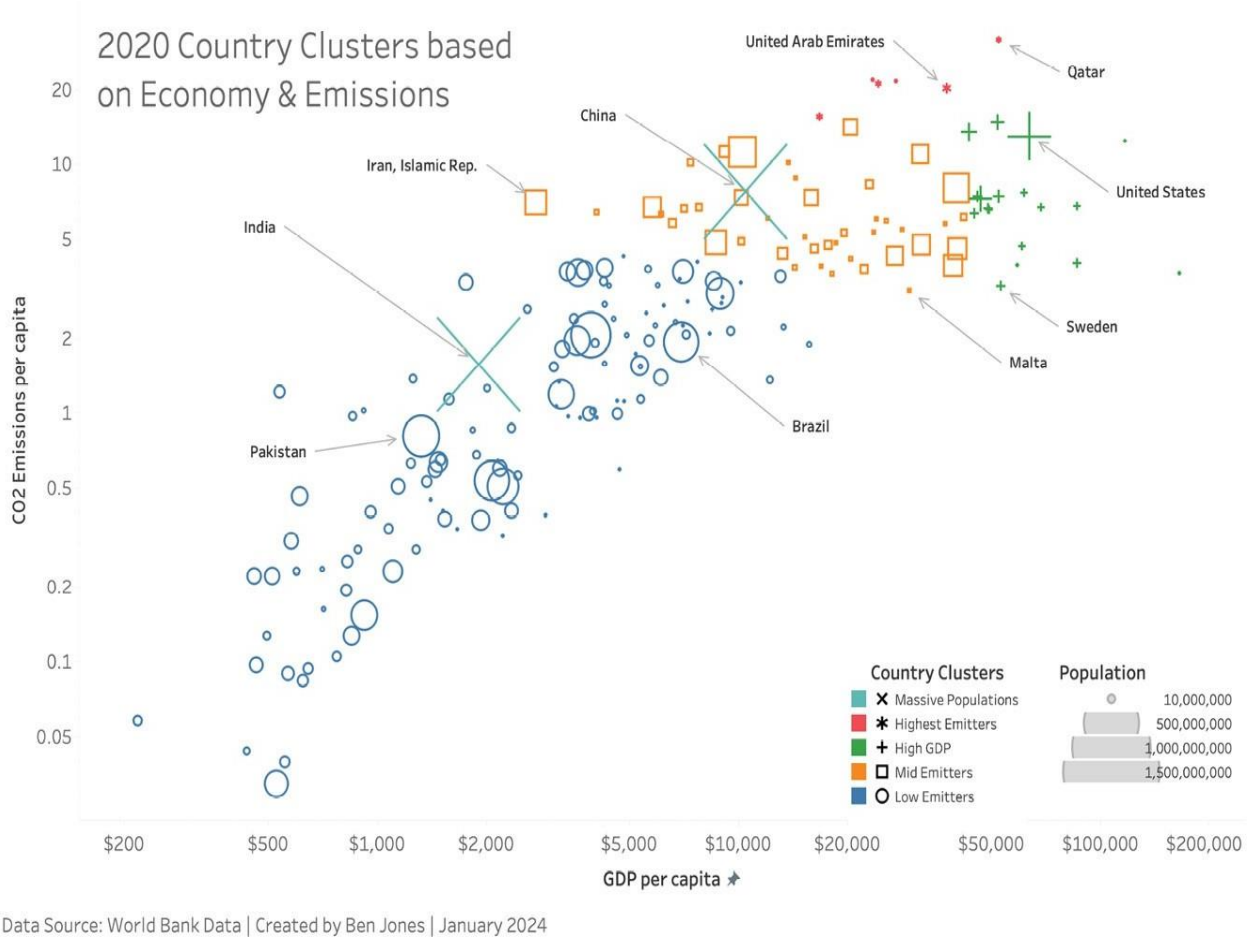


شکل ۳٫۲ - فاصله اقلیدسی بین دو نقطه

بسیاری از الگوریتم‌های خوشه‌بندی، به‌صورت تدریجی و تکراری، خوشه‌ها را تنظیم می‌کنند تا شباهت درون هر گروه را افزایش و شباهت بین گروه‌ها را کاهش دهند. در پایان فرآیند، مدل هر نقطه داده را در یکی از گروه‌ها (خوشه‌ها) قرار می‌دهد. اما این سؤال مطرح می‌شود: چه تعداد خوشه باید وجود داشته باشد؟ برخی الگوریتم‌های خوشه‌بندی، این تعداد را به‌صورت خودکار تعیین می‌کنند اما در برخی دیگر مانند الگوریتم *K-means*، این تعداد باید از قبل توسط کاربر مشخص شود(K یعنی تعداد خوشه‌های موردنظر).

به عنوان مثال، در شکل ۳٫۳، از الگوریتم *K-means* خواسته شده تا پنج خوشه از کشورها ایجاد کند، بر اساس جمعیت آن‌ها در سال ۲۰۲۰، اندازه اقتصاد (بر پایه تولید

ناخالص داخلی سرانه)، و میزان انتشار دی‌اکسید کربن (بر حسب تن متریک به‌ازای هر نفر). الگوریتم، خوشه‌ای شامل دو کشور با بیشترین جمعیت (چین و هند)، گروهی متشکل از شش کشور با بیشترین میزان انتشار CO_2، گروهی شامل ۱۸ کشور با بالاترین GDPسرانه (به‌جز قطر که در گروه دوم قرار گرفت)، گروهی با ۱۲۲ کشور با کم‌ترین میزان انتشار (به‌جز هند)، و گروهی شامل ۴۳ کشور با وضعیت میانه در هر دو شاخص (به‌جز چین) ایجاد کرد.



شکل ۳٫۳ - خوشه‌بندی کشورها با «یادگیری ماشین بدون نظارت»

هدف یادگیری بدون‌نظارت، **استفاده از توان پردازشی** رایانه‌ها برای **کشف الگوها**یی است که یافتن آن‌ها برای انسان بسیار دشوار و زمان‌بر است: یعنی خوشه‌هایی که به‌طور طبیعی در دل داده‌ها پدید می‌آیند. این خوشه‌ها از پیش توسط انسان‌ها تعیین

نشده‌اند تا نوعی دسته‌بندی از پیش تعریف‌شده به داده‌ها تحمیل شود، بلکه **این خوشه‌ها از دل خود داده‌ها استخراج می‌شوند.** در این روش، مدل از خود داده‌ها می‌آموزد که گروه‌بندی چگونه باید باشد. **نیازی به مجموعه داده آموزشی جداگانه وجود ندارد**؛ بلکه **مدل از همان مجموعه داده اصلی یاد می‌گیرد.**

یادگیری تقویتی (Reinforcement Learning)

سومین نوع اصلی یادگیری ماشین، یادگیری تقویتی نام دارد. یادگیری تقویتی یعنی یادگیری برای انجام یک کار جدید که در آن تصمیم‌گیری درباره اقدامات موردنیاز بر اساس **بازخوردی که از محیط (در قالب پاداش یا تنبیه)** دریافت می‌شود، صورت می‌گیرد. یک عامل هوش مصنوعی که به این شکل با محیط تعامل دارد، می‌تواند در طول زمان به نتایج بهتری دست پیدا کند.

به یادگیری **دوچرخه‌سواری** فکر کنید. اگر معلمی فقط به شما پاسخ‌های درست را بگوید که چه کارهایی را باید انجام دهید و چه کارهایی را نه (مانند یادگیری با نظارت)، حتی اگر در آزمون کتبی نمره کامل بگیرید، نمی‌توانید دوچرخه‌سواری کنید. حتی اگر تمام روز حرکات بچه‌های دیگر را تماشا کنید و به دنبال الگو باشید (مانند یادگیری بدون نظارت)، باز هم موفق به دوچرخه‌سواری نخواهید شد. تنها راه یادگیری، این است که خودتان سوار دوچرخه شوید و تلاش کنید. در ابتدا یکی دو بار پدال می‌زنید و بعد زمین می‌خورید. این زمین‌خوردن‌ها درد دارد و ممکن است باعث زخم شدن زانو یا کبودی شانه شود. این تجربه‌های دردناک مثل تنبیه برای اقدامات اشتباهی هستند که نباید انجام می‌دادید. دوباره تلاش می‌کنید و باز می‌افتید. اما کم‌کم بهتر می‌شوید، پنج یا حتی ده پدال می‌زنید قبل از اینکه بیفتید. وزش باد روی صورتتان لذت‌بخش است. این حس پاداشی است برای تلاش‌ها و پیشرفتتان. به مرور زمان بهتر می‌شوید. در

نهایت با بچه‌های دیگر دوچرخه‌سواری می‌کنید، از موانع پرش می‌کنید، تک‌چرخ می‌زنید، با یک دست و بعد بدون دست می‌رانید. زندگی زیبا می‌شود.

رفتارهای ما، پیامدهایی دارند و می‌توانیم از این پیامدها بیاموزیم تا حتی در جهانی پر از نقص و بی‌نظمی، به اهداف خود برسیم. ما می‌توانیم این نوع یادگیری را به رایانه‌ها نیز آموزش دهیم. نمونه‌ی مشهور آن، برنامه‌ی **AlphaGo** است که در فصل قبل به آن اشاره شد. بازی Go با سطح پیچیدگی بالای خود، چالش بزرگی برای هوش مصنوعی بود. با صفحه‌ای ۱۹ در ۱۹ خانه، تعداد حالات ممکن بسیار بیشتر از آن است که در یک پایگاه داده ذخیره و مرور شود. راه‌حل، **ترکیبی از یادگیری با نظارت و یادگیری تقویتی** بود. در مرحله نخست، AlphaGo از طریق ۱۲ شبکه عصبی آموزش دید تا حرکات بازیکنان خبره را تقلید کند. به آن یک پایگاه داده حاوی ۳۰ میلیون حرکت از بازی‌های قبلی داده شد. در این مرحله، یادگیری با نظارت بود. ورودی‌ها، وضعیت‌های مختلف صفحه‌ی بازی بودند و خروجی‌ها، حرکات خوبی که بازیکنان انسانی در آن موقعیت انجام داده بودند. پس از این مرحله، AlphaGo می‌توانست حرکت بعدی انسان را در ۵۷% موارد درست پیش‌بینی کند. اما تیم DeepMindمی‌خواست آن را بهتر کند، بنابراین مدل را وادار کرد با نسخه‌های دیگر خودش بازی کند: AlphaGo در برابر AlphaGo. با استفاده از نتایج هزاران بازی، مدل از طریق یادگیری تقویتی، خود را بهبود داد. وزن‌ها و پارامترهای شبکه‌های عصبی آن، بر اساس نتایج تصمیم‌های اتخاذ شده تنظیم شدند.

در مسابقه با قهرمان کره‌جنوبی، لی سدول، AlphaGo به‌قدری قوی شده بود که از همان ابتدا قدرت خود را نشان داد. هوش مصنوعی فقط سدول را شکست نداد بلکه **برخی از**

حرکاتش به‌قدری غیرمنتظره و متفاوت از سبک بازی انسان‌ها بود که کارشناسان را شگفت‌زده کرد. این **حرکات غیرانسانی** بودند؛ حرکاتی که در ۲۵۰۰ سال تاریخ بازی Go سابقه نداشتند.

تیم DeepMind همچنین از یادگیری تقویتی برای آموزش بازی‌های ویدیویی (مانند Breakout ساخته‌ی آتاری) استفاده کرد. **این نوع یادگیری برای توسعه خودروهای خودران نیز مهم است**؛ زیرا مدل باید از تصمیماتی که در جاده می‌گیرد، یاد بگیرد. ربات‌های دیگر مانند **ربات‌های مونتاژ در کارخانه** نیز با استفاده از یادگیری تقویتی، **عملکرد خود را در طول زمان** بهبود می‌بخشند.

اخیراً OpenAI از نوعی یادگیری تقویتی با عنوان یادگیری تقویتی از بازخورد انسانی برای آموزش و تنظیم دقیق نسخه‌های جدید مدل‌های زبانی بزرگ خود از جمله GPT استفاده کرده است. در این روش، کاربران نسخه‌ی آزمایشیِ چت‌بات، سؤالاتی را وارد می‌کنند. سپس بازبین‌های انسانی، پاسخ‌های مطلوب می‌دهند یا به پاسخ‌های تولیدشده امتیاز می‌دهند. این امتیازها به عنوان بازخورد برای تنظیم دقیق مدل به‌کار گرفته می‌شود. امتیازها مثل پاداش برای مدل هستند. در عمل، مدل آموزش می‌بیند که چگونه پاسخ‌هایی بدهد که مطابق با ترجیحات داوران انسانی باشد.

دیگر انواع یادگیری ماشین

تاکنون سه نوع اصلی یادگیری ماشین را بررسی کردیم. اما انواع دیگری نیز وجود دارد. به عنوان مثال **یادگیری نیمه‌نظارتی**، ترکیبی از یادگیری با نظارت و بدون نظارت است. در این روش، مدل با **مقدار اندکی داده برچسب‌دار** (مثلاً چند صد عکس دسته‌بندی‌شده‌ی دستی) آموزش می‌بیند و سپس از **حجم زیادی از داده‌های بدون**

برچسب (مثلاً باقی عکس‌هایی که در یک شبکه اجتماعی بارگذاری شده‌اند) نیز یاد می‌گیرد.

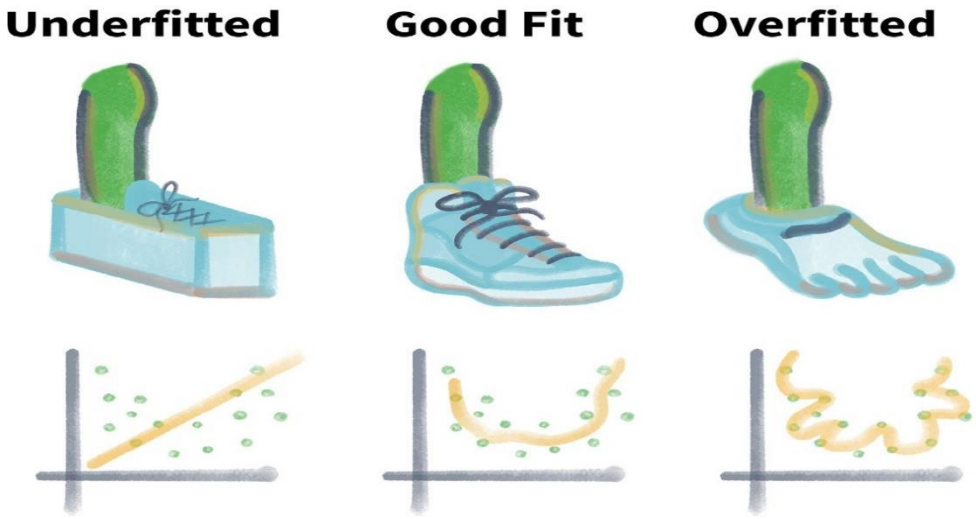
در **یادگیری خودنظارتی**، برچسب‌های مجموعه داده، توسط انسان تولید نمی‌شوند بلکه سیستم به‌صورت خودکار آن‌ها را ایجاد می‌کند. مثلاً مدل‌های زبانی بزرگی مانند **LLAMA** ساخته‌ی شرکت متا، با هدف پیش‌بینی واژه‌ی بعدی در جمله آموزش می‌بینند. مدل با مقدار عظیمی از متون کتاب‌ها، مقاله‌ها و وب‌سایت‌ها تغذیه می‌شود و از همان جملات، آزمون‌هایی برای خودش می‌سازد: مثلاً جمله‌ی «این بهترین دوران بود، آن بدترین...» را با «دوران» ادامه می‌دهد و به‌تدریج در حدس زدن واژه‌های بعدی بهتر می‌شود. هیچ انسانی این آزمون‌های جفت‌شده‌ی سؤال و جواب را نمی‌سازد؛ بلکه خود الگوریتم به‌طور خودکار آن‌ها را از متن استخراج می‌کند.

بیش‌برازش و کم‌برازش

هیچ مجموعه‌ی داده‌ای، کامل نیست؛ این موضوع درباره‌ی مجموعه‌های داده‌ی آموزشی در یادگیری ماشین نیز صادق است. این **داده‌ها ممکن است دارای خطا، مقادیر اشتباه، یا نویز باشند.** نویز یعنی وجود ناهنجاری‌هایی در داده که ممکن است ناشی از شیوه‌ی اندازه‌گیری، گردآوری یا پردازش داده‌ها باشد. همچنین، ممکن است جزئیات خاصی در داده‌ها وجود داشته باشد یا نباشد، و کاربر مدل از این موضوع آگاه نباشد. مثلاً فرض کنید می‌خواهید مدلی بسازید که تشخیص دهد آیا در یک عکس، سگی وجود دارد یا نه. اگر مجموعه‌ی داده‌ی آموزشی شما فقط شامل نژادهای خاصی از سگ‌ها باشد، یا سگ‌ها فقط در حالت نشسته یا ایستاده باشند، مدل یاد می‌گیرد که

این ویژگی‌ها برای تشخیص سگ مهم هستند. حالا اگر عکسی با نژادی متفاوت یا سگی در حال پریدن به مدل بدهید، ممکن است نتواند آن را به‌درستی تشخیص دهد.

این مشکل رایج در یادگیری ماشین، **بیش‌برازش** (overfitting) نام دارد. مدلی که دچار بیش‌برازش شده، **ویژگی‌ها و جزئیات بی‌ربط و نویز موجود در داده‌های آموزشی را به‌عنوان اطلاعات مهم در نظر می‌گیرد.** در نتیجه، عملکرد آن هنگام مواجهه با داده‌های جدید، ضعیف خواهد بود. این حالت در سمت راست شکل ۳٫۴ نشان داده شده است، به‌صورت کفشی که آن‌قدر دقیق و پرجزئیات طراحی شده که شبیه یک پای خاص شده است.



شکل ۳٫۴ - بیش‌برازش و کم‌برازش در یادگیری ماشین

در سوی دیگر، **کم‌برازش** (underfitting) زمانی رخ می‌دهد که **مدل بسیار ساده است** و **اطلاعات کافی از داده‌های آموزشی نگرفته است**. در سمت چپ شکل ۳٫۴، این وضعیت با کفشی نشان داده شده که شکل درست ندارد و به پای هیچ‌کس نمی‌خورد. مدل کم‌برازش هم در مواجهه با داده‌های جدید عملکرد خوبی نخواهد داشت. در وسط شکل، حالت مطلوب یا «متناسب» نشان داده شده است، همان هدفی که همه‌ی پژوهشگران یادگیری ماشین به‌دنبال آن هستند، اما رسیدن به آن کار آسانی نیست.

خلاصه

در این فصل، به مفاهیم پایه‌ای **یادگیری ماشین** پرداختیم؛ یکی از زیرشاخه‌های حیاتی هوش مصنوعی که تجسمی از شاخه‌ی «هوش مصنوعی **غیرنمادین**» محسوب می‌شود. همچنین درک کردیم که چگونه دو عامل کلیدی محیطی (افزایش چشمگیر **قدرت محاسباتی** و **حجم داده‌های در دسترس** برای آموزش مدل‌ها)، زمینه‌ساز پیشرفت یادگیری ماشین شده‌اند.

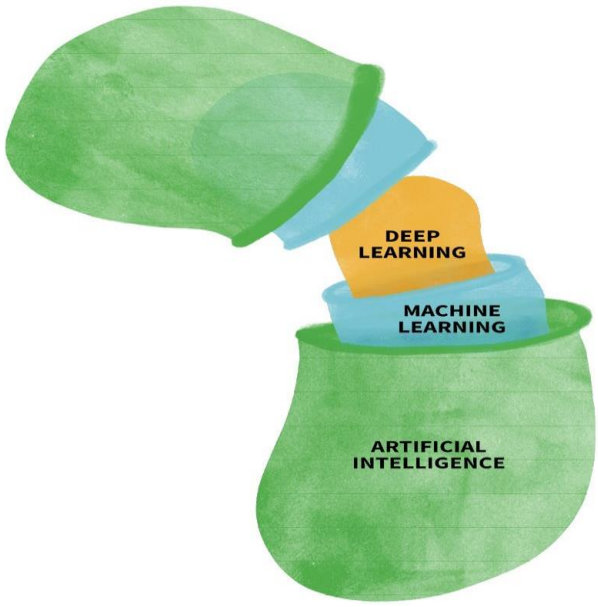
راه‌های مختلفی را بررسی کردیم که انسان‌ها از طریق آن‌ها توانسته‌اند به کامپیوترها، قابلیت یادگیری از داده بدهند و این رویکردها را به شیوه‌های مختلف یادگیری در انسان‌ها تشبیه کردیم. گاهی معلم پاسخ‌های درست و غلط را به ما آموزش می‌دهد که شبیه به **یادگیری نظارت‌شده** است. گاهی باید محیط خود را مشاهده کرده و الگوها را به‌تنهایی کشف کنیم، چیزی شبیه به **یادگیری بدون نظارت**. و گاهی لازم است دست به تجربه بزنیم، تصمیم‌گیری کنیم، اقدام کنیم و از پیامدهای آن بیاموزیم مانند **یادگیری تقویتی**. این‌ها تنها سه نمونه از الگوریتم‌های یادگیری در زیرمجموعه‌ی یادگیری ماشین هستند. اگرچه گاهی یادگیری ماشین هم‌سطح با خود هوش مصنوعی در نظر گرفته می‌شود اما این دو یکسان نیستند. «هوش مصنوعی نمادین» هنوز هم اهمیت زیادی دارد حتی اگر فعلاً در سایه قرار گرفته باشد. اگر در دهه‌های اخیر، تمرکز به‌طور کامل از هوش مصنوعی نمادین به سوی هوش مصنوعی غیرنمادین جابه‌جا شده، شاید بازگشت به تعادل میانِ این دو رویکرد، بتواند زمینه‌ساز پیشرفت‌های آینده شود. در **فصل بعد**، به یکی از زیرشاخه‌های بسیار قدرتمند یادگیری ماشین خواهیم پرداخت: **یادگیری عمیق**.

فصل چهارم

مقدمه‌ای بر یادگیری عمیق

ابزارها و فناوری‌هایی که توسعه داده‌ایم، تنها **چند قطره‌ی نخست از اقیانوس** عظیم هوش مصنوعی است (فی‌فی لی، مادرخوانده‌ی هوش مصنوعی).

ما سفر خود را با تعریف هوش مصنوعی آغاز کردیم: توسعه‌ی سیستم‌های کامپیوتری‌ای که بتوانند کارهایی را انجام دهند که معمولاً به هوش انسانی نیاز دارند. پس از مروری کوتاه بر تاریخچه‌ی این حوزه، تمرکز را بر یادگیری ماشین گذاشتیم یعنی توانایی برنامه‌های کامپیوتری برای آموختن از داده‌ها و انجام وظایف بر این اساس. اکنون، در این فصل، به شکلی عمیق‌تر وارد دنیای **یادگیری عمیق** می‌شویم؛ **نوعی خاص از یادگیری ماشین** که در چند دهه‌ی اخیر توانایی‌های خارق‌العاده‌ای را برای کامپیوترها به ارمغان آورده است. رابطه‌ی میان هوش مصنوعی، یادگیری ماشین، و یادگیری عمیق اغلب با نموداری که در آن، هوش مصنوعی دایره‌ای بزرگ‌تر است که دایره‌ی یادگیری ماشین کاملاً درون آن جای گرفته، و دایره‌ی یادگیری عمیق نیز به‌طور کامل در درون دایره‌ی یادگیری ماشین قرار دارد، نشان داده می‌شود.



شکل ۴/۱ – رابطه‌ی میان هوش مصنوعی، یادگیری ماشین، و یادگیری عمیق

این رابطه به این معناست که هر چیزی که در دسته‌ی یادگیری عمیق قرار بگیرد، به‌طور خودکار در دسته‌ی یادگیری ماشین و در نتیجه هوش مصنوعی نیز قرار می‌گیرد. اما برعکس این گزاره‌ها لزوماً درست نیست. **برخی انواع هوش مصنوعی وجود دارند که مبتنی بر یادگیری ماشین نیستند** (مانند سیستم‌های خبره یا نمودارهای دانش)، و **برخی روش‌های یادگیری ماشین هم وجود دارند که در دسته‌ی یادگیری عمیق جای نمی‌گیرند** (مانند درخت تصمیم یا الگوریتم‌های خوشه‌بندی).

یادگیری عمیق، شاخه‌ای خاص از یادگیری ماشین است که به یک خانواده‌ی خاص از شبکه‌های عصبی مصنوعی می‌پردازد: شبکه‌های عصبی عمیق (– Deep Neural Networks DNN). این شبکه‌ها، جهان هوش مصنوعی را دگرگون کرده‌اند و در این مسیر، دنیای ما را نیز متحول ساخته‌اند. در سال ۲۰۱۸، سه متخصص یادگیری عمیق (**یان لکون، یوشوا بنجیو و جفری هینتون**) به‌طور مشترک برنده‌ی جایزه‌ی تورینگ مؤسسه‌ی ACM شدند. این جایزه به پاس «دستاوردهای مفهومی و مهندسی‌ای که شبکه‌های عصبی عمیق را به عنصری حیاتی در علم رایانه تبدیل کرده‌اند» به آن‌ها اهدا شد. این جایزه‌ی معتبر، اغلب «نوبل علوم کامپیوتر» نامیده می‌شود، گرچه مؤسسه‌ی ACM هیچ ارتباطی با بنیاد نوبل ندارد.

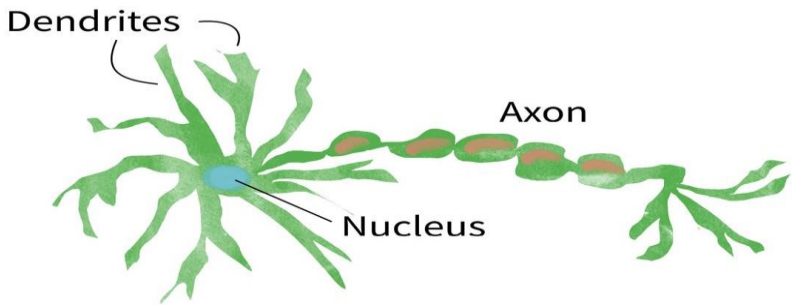

2018 ACM A.M. Turing Award

در این فصل، بررسی خواهیم کرد که «**عمق**» در شبکه‌های عصبی عمیق و یادگیری عمیق به چه معناست و چرا این شاخه‌ی خاص از یادگیری ماشین به چنین **بخش مهمی از هوش مصنوعی**، علوم کامپیوتر، و در واقع، کل جامعه بدل شده است. پیشرفت‌های مبتنی بر شبکه‌های عصبی عمیق، پشتوانه‌ی بیشترِ جهش‌های فناورانه در هوش

مصنوعی از ابتدای قرن بیست‌ویکم تاکنون بوده‌اند. برای درک مفهومی این شبکه‌ها، ابتدا باید با **منبع الهام اصلی** آن‌ها آشنا شویم: **واحد بنیادین مغز انسان، نورون زیستی.**

نورون‌های زیستی

بر اساس برآوردهای کنونی، عصب‌شناسان معتقدند **مغز انسان حدود ۸۶ میلیارد نورون زیستی دارد**. هر نورون، **سلولی عصبی** است که دارای هسته بوده و از طریق سیناپس‌ها یعنی اتصال‌هایی میان آکسون (تک‌فیبر خروجی نورون مبدأ) و دندریت‌های انگشت‌مانندِ نورون‌های مقصد، به **تبادل سیگنال‌های الکتریکی و شیمیایی با نورون‌های همسایه** می‌پردازد.



شکل ۴/۲ – یک نورون زیستی با هسته، آکسون و دندریت‌ها

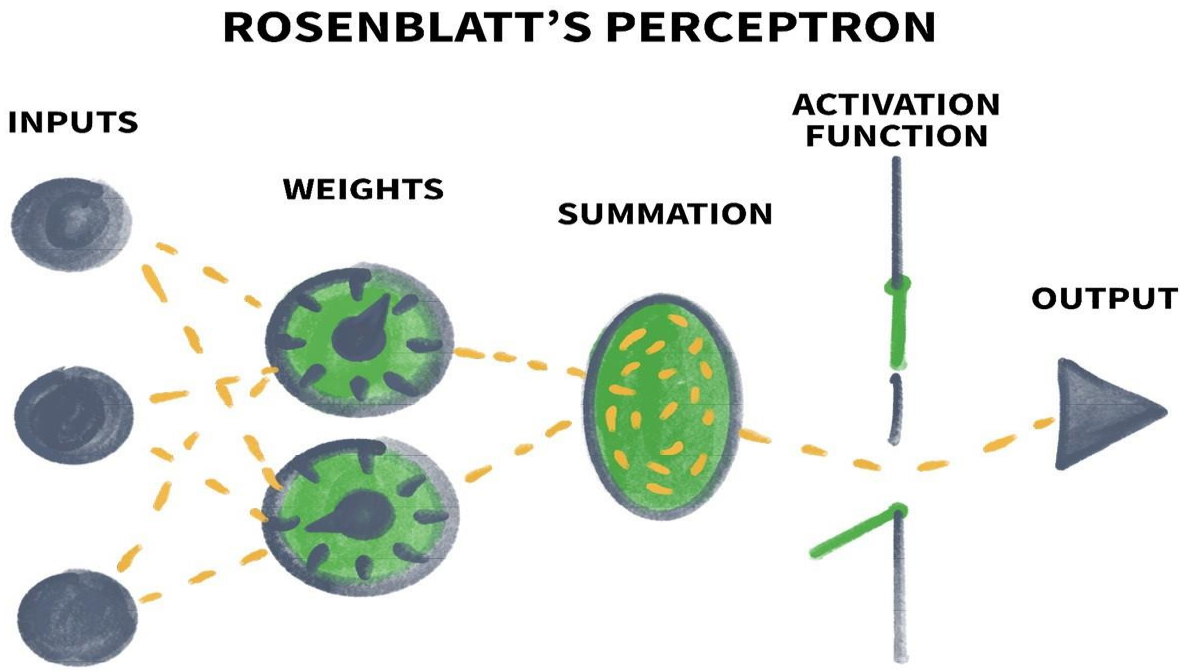
در مغز انسان، **هر نورون می‌تواند با هزاران نورون دیگر از این طریق ارتباط برقرار کند.** اگر مجموعه سیگنال‌هایی که از طریق دندریت‌ها به هسته‌ی نورون می‌رسند به اندازه کافی قوی باشند، نورون یک سیگنال الکتریکی از طریق آکسون خود به سمت دندریت‌های نورون‌های همسایه ارسال می‌کند. این **ارتباطات میان نورون‌ها**، همراه با اتصال مغز به ورودی‌های حسی و کنترل‌های حرکتی از طریق سیستم عصبی مرکزی، به ما امکان پردازش اطلاعات و واکنش به محرک‌های محیطی را می‌دهند. یک نورون می‌تواند با نورون دیگر ارتباطی قوی داشته باشد، به‌طوری که اغلب هم‌زمان فعال می‌شوند؛ یا ارتباطی ضعیف‌تر که تأثیر نورون مبدأ بر نورون مقصد کم باشد. علاوه بر

این، قدرت **اتصال سیناپسی بین نورون‌ها می‌تواند بسته به فعالیت‌ها و تجربیات ما تقویت یا تضعیف شود**، پدیده‌ای که با عنوان **انعطاف‌پذیری عصبی (Neuroplasticity)** شناخته می‌شود و نقشی محوری در توانایی ما برای **یادگیری** و شکل‌گیری حافظه دارد.

نورون‌های مصنوعی

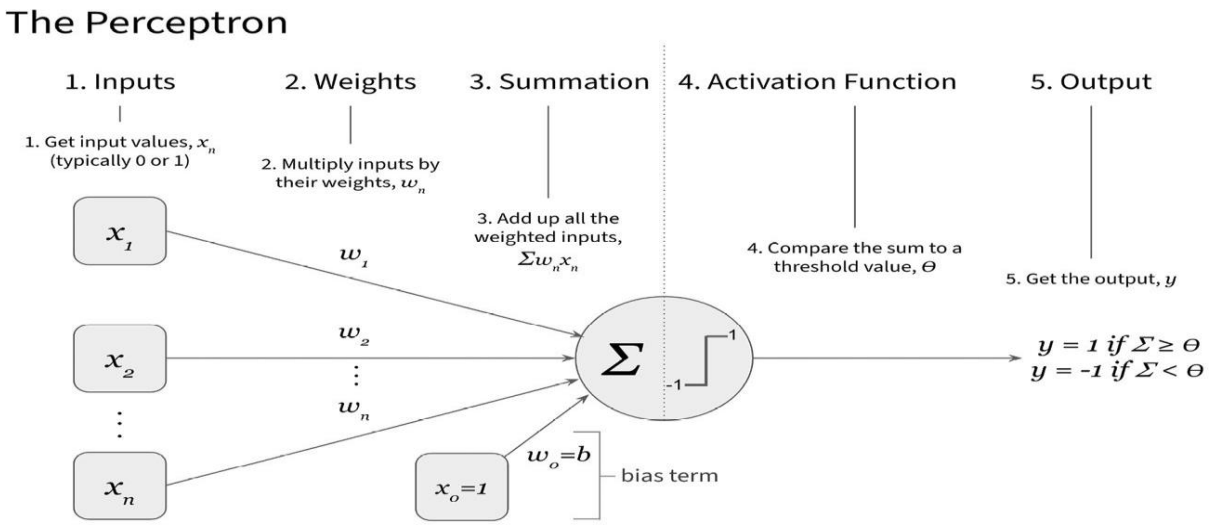
به‌طور مشابه، هر نورون مصنوعی در یک شبکه‌ی عصبی نیز با نورون‌های دیگری که به آن‌ها متصل است ارتباط برقرار می‌کند، هرچند در بسیاری از جنبه‌ها با نورون زیستی تفاوت دارد. در کاربردهای هوش مصنوعی، رایج است که به نورون‌های مصنوعی صرفاً «نورون» گفته شود، هرچند گاهی ممکن است آن‌ها را «نورون‌های شبیه‌سازی‌شده» نیز بنامند. برای حفظ وضوح، از این پس در این کتاب به نورون‌های مصنوعی با عنوان «نورون» و به نورون‌های زیستی با همان عنوان «نورون زیستی» اشاره خواهد شد.

نحوه کار نورون چگونه است؟ یک نورون **مجموعه‌ای از ورودی‌های دیجیتال** دریافت می‌کند، هر ورودی را **در وزن مخصوص خود ضرب می‌کند**، حاصل‌ضرب‌ها را با هم **جمع** می‌زند (**مجموع وزنی** یا weighted sum)، و سپس این مجموع را وارد **تابع فعال‌سازی** (**Activation**) می‌کند یعنی یک فرمول ریاضی که **مقدار خروجی نورون** را تعیین می‌کند. پس از محاسبه خروجی توسط تابع فعال‌سازی، نورون این سیگنال دیجیتال را به‌عنوان **خروجی** (ورودی نورون بعدی)، ارسال می‌کند.



شکل ۴/۳ – نمودار پرسپترون فرانک روزنبلات

نخستین نورون مصنوعی، یعنی پرسپترون، در دهه **۱۹۵۰** توسط **فرانک روزنبلات** ساخته شد. این نورون ابتدایی، ورودی‌هایی از حسگرهای مختلف دریافت می‌کرد، هر ورودی را در وزنی ضرب می‌کرد، مجموع ورودی‌های وزنی را با **مقدار آستانه‌ای** مقایسه می‌کرد و اگر این مقدار، **کمتر از آستانه** بود، خروجی را **۱-** و اگر **بیشتر یا مساوی آستانه** بود، خروجی را **۱** قرار می‌داد. به همین دلیل، این نوع تابع فعال‌سازی را **تابع آستانه‌ای** (Threshold Activation Function) می‌نامند: **خروجی به‌صورت ناگهانی و دودویی از یک مقدار به مقدار دیگر تغییر می‌کند،** دقیقاً **مانند یک کلید روشن/خاموش ساده.** برای نمایش خروجی دودویی می‌توان از مقادیر ۱- و ۱ (مانند روش روزنبلات) یا از ۰ و ۱ استفاده کرد، هر دو شیوه معتبرند. نسخه‌ی کامل‌تر نمودار پرسپترون همراه با نمادهای ریاضی، در شکل زیر آمده است.



نمودار مفهومی پرسپترون فرانک روزنبلات به‌گونه‌ای طراحی شده بود که مجموعه‌ای از ورودی‌ها (از $_1x$ تا $_nx$) را از حسگرهای مختلف دریافت می‌کرد. هر ورودی در یک وزن مشخص (از $_1w$ تا $_nw$) ضرب می‌شد، سپس این ورودی‌های وزن‌دار با یکدیگر جمع می‌شدند (که با حرف یونانی سیگما Σ نشان داده می‌شود، به‌صورت:

$$_n\Sigma w_n x = {}_1w_1x + {}_2w_2x + \cdots + {}_nw_nx).$$

سپس این مجموع با یک مقدار آستانه، که با حرف یونانی تتا Θ نمایش داده می‌شود، مقایسه می‌گردید.

اگر مجموع کمتر از مقدار آستانه بود $(\Sigma < \Theta)$، پرسپترون خروجی -1 تولید می‌کرد.

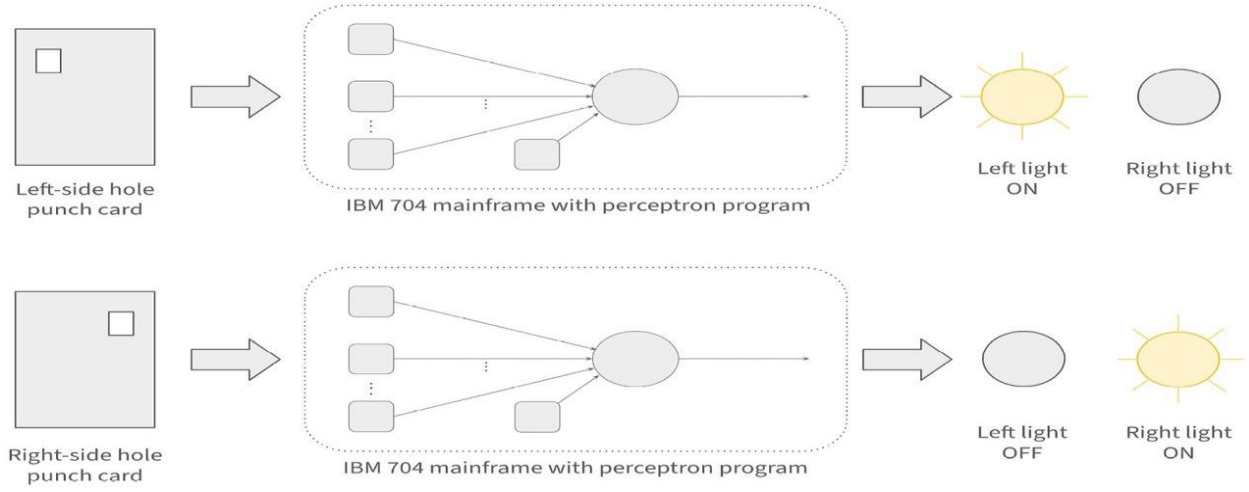
و اگر مجموع برابر یا بیشتر از آستانه بود $(\Sigma \leq \Theta)$، خروجی آن مقدار $+1$ بود.

اما این سازوکار ساده به چه کار می‌آمد، و چرا می‌توان آن را نشانه‌ای از «هوش» در نظر گرفت؟ همان‌طور که از مرور تاریخچه‌ی هوش مصنوعی به یاد دارید، روزنبلات برنامه‌ی پرسپترون را روی رایانه‌ی IBM 704 بارگذاری کرد و ۵۰ کارت پانچ را یکی‌یکی وارد دستگاه نمود. پس از این فرآیند، برنامه توانست تفاوت میان کارت‌هایی با سوراخ در سمت راست و سوراخ در سمت چپ را تشخیص دهد.

اما این یادگیری چگونه انجام شد؟ **نکته‌ی کلیدی در وزن‌های نورون نهفته است، و در شیوه‌ای که وزن‌ها در فرآیند آموزش تنظیم می‌شوند تا دقت خروجی افزایش یابد.** وزن‌های نورون تقریباً معادل با قدرت اتصال‌های سیناپسی میان نورون‌های زیستی در

مغز هستند. **وزن‌های بالاتر مانند سیناپس‌های قوی‌تر عمل می‌کنند: آن‌ها باعث می‌شوند سیگنال ورودی تأثیر بیشتری بر نورون داشته باشد.** افزایش یک وزن، تأثیر ورودی متناظر را بالا می‌برد و کاهش آن، این تأثیر را کمتر می‌کند. دلیل این امر آن است که هر ورودی در نورون در عددی به نام وزن ضرب می‌شود. درواقع، می‌توان آن را شبیه به **تغییر «شدت صدای» آن ورودی** در نظر گرفت. هرچه وزن ورودی بیشتر شود، صدای آن بلندتر شده و تأثیر بیشتری دارد؛ هرچه وزن کمتر شود، صدا پایین‌تر می‌آید و تأثیرش کمتر می‌شود.

نکته اینجاست که **می‌توان در طول آموزش، وزن‌ها را گام‌به‌گام تنظیم کرد تا نورون، خروجی درست‌تری تولید کند.** اما پرسپترون روزنبلات چگونه وزن‌های خود را تنظیم می‌کرد؟ در **ابتدا، وزن‌ها به‌صورت تصادفی مقداردهی شده بودند.** سپس روزنبلات، کارت پانچ را وارد دستگاه می‌کرد و پرسپترون با یک خروجی دودویی، یک لامپ را برای حالت «سوراخ در سمت چپ» و لامپ دیگری را برای حالت «سوراخ در سمت راست» روشن می‌کرد.

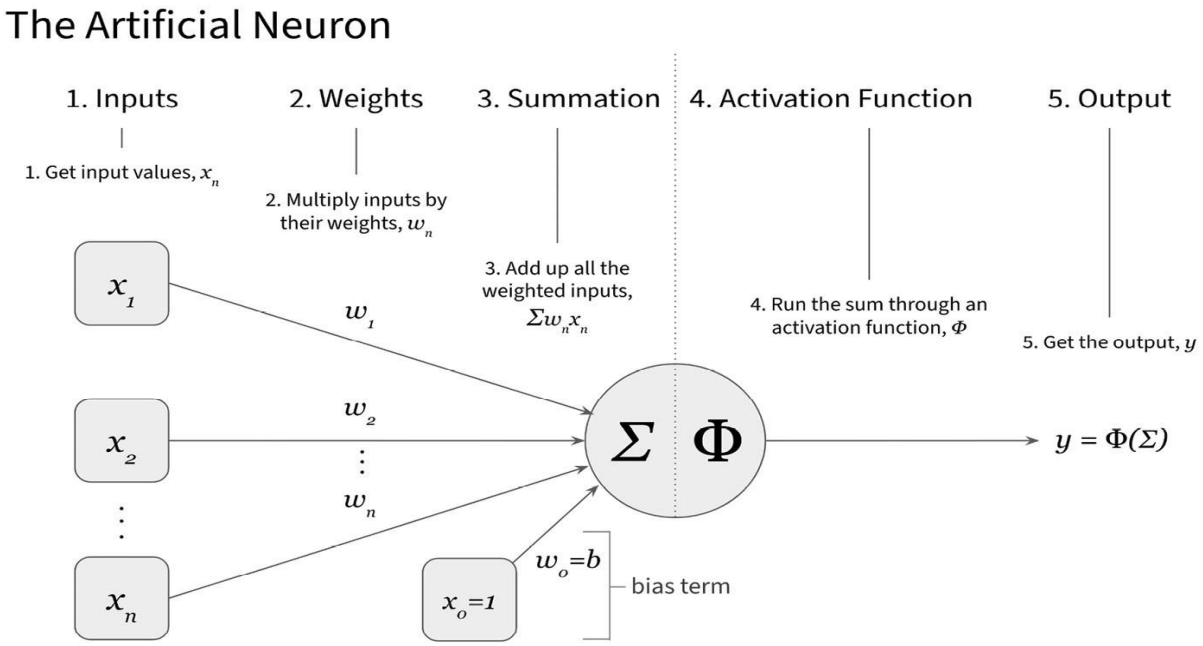


شکل ۴/۴ – نمودار نمایش آزمایش پرسپترون روزنبلات

نکته‌ی اصلی اینجاست: **اگر پرسپترون پاسخ درست می‌داد، وزن‌ها تغییر نمی‌کردند**؛ اما **اگر پاسخ نادرست بود، وزن‌ها بر اساس فرمولی ساده تنظیم می‌شدند تا خطای خروجی کاهش یابد.** خطا تابعی از تفاوت میان خروجی واقعی و خروجی مورد انتظار بود. وزن‌های جدید و تنظیم‌شده باعث کاهش خطا می‌شدند، یعنی پرسپترون برای همان کارت پانچ خاص، احتمال بیشتری برای ارائه پاسخ درست داشت. پس از انجام این فرآیند با ۵۰ کارت مختلف، وزن‌های پرسپترون با مقادیری تنظیم شدند که امکان حدس دقیق را فراهم می‌کردند. پرسپترون توانسته بود یاد بگیرد که تفاوت میان سوراخ در سمت چپ و سمت راست را تشخیص دهد.

همان‌طور که می‌دانید، این نوع یادگیری، یک **یادگیری تحت نظارت** است، چون یک **«ناظر»** (در اینجا روزنبلات) **در هر مرحله به مدل می‌گفت که خروجی‌اش درست بوده یا نه**.

اصول نورون‌های امروزی، به‌طرز شگفت‌انگیزی به همان ساختار اولیه روزنبلات شباهت دارند اما با تفاوت‌هایی مهم. اول این‌که **ورودی‌ها و خروجی‌های نورون‌های امروزی، معمولاً دودویی نیستند**؛ بلکه می‌توانند **طیفی پیوسته از مقادیر را بپذیرند** مانند چراغی با تغییر شدت نور به‌جای کلید ساده روشن/خاموش. این ویژگی، به حل مسائل پیچیده‌تر کمک می‌کند. دوم این‌که **نورون‌های مدرن به‌ندرت از توابع آستانه‌ای استفاده می‌کنند** زیرا این توابع با **«پرش»های ناگهانی** همراه‌اند. پژوهشگران دریافته‌اند که استفاده از **توابع فعال‌سازی نرم‌تر** باعث می‌شود بتوان شبکه‌هایی به‌مراتب پیچیده‌تر و کارآمدتر آموزش داد. برای مشاهده نمودار نورون مصنوعی مدرن، به شکل زیر مراجعه کنید.



نمودار مفهومی یک نورون مصنوعی مدرن که شباهت زیادی به نورون اصلی فرانک روزنبلات، یعنی پرسپترون، دارد، اما دو تفاوت اساسی در آن دیده می‌شود:

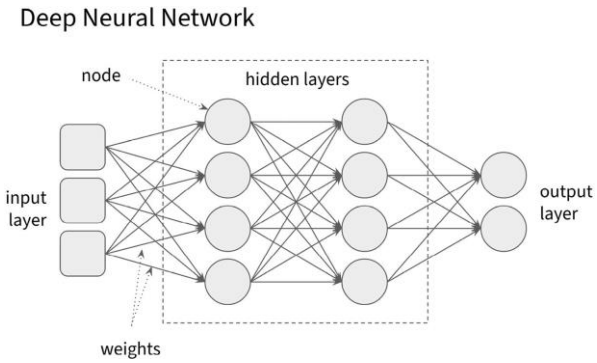
۱) مقادیر ورودی‌ها ($_1x$ تا $_nx$) اکنون می‌توانند مقادیر پیوسته‌ای در یک بازه را بپذیرند،

۲) تابع آستانه‌ای که در مدل اصلی وجود داشت، اکنون با حرف یونانی فی (Φ) جایگزین شده است که می‌تواند نمایانگر طیف گسترده‌ای از توابع فعال‌سازی مختلف باشد که مجموع وزن‌دار ورودی‌ها را به خروجی نورون (یا همان «فعال‌سازی» نورون) تبدیل می‌کنند.

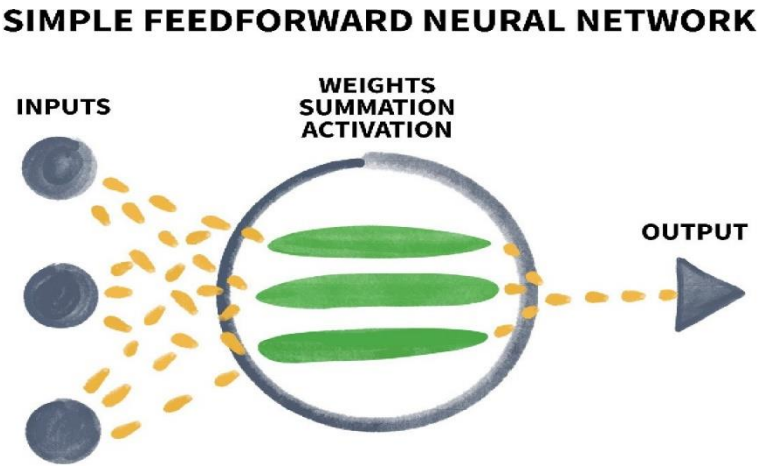
در این نمودار همچنین یک جملهٔ بایاس (bias) با نماد $b = {}_0w\,{}_0x$ وجود دارد که انعطاف‌پذیری بیشتری به مدل می‌بخشد. مقدار ورودی ${}_0x$ همیشه برابر با ۱ در نظر گرفته می‌شود، بنابراین حاصل ضرب آن با وزنش ($_0w$) همواره برابر با خود ${}_0w$ خواهد بود که همان بایاس b است.

این نوع بایاس با «سوگیری اجتماعی» یا «سوگیری الگوریتمی» تفاوت دارد. می‌توان آن را مشابه با تغییر مقدار ثابت در معادله یک خط ($b + mx = y$) در نظر گرفت که باعث می‌شود خط به سمت بالا یا پایین جابه‌جا شود؛ در اینجا نیز بایاس باعث می‌شود مجموع وزن‌دار ورودی‌ها به میزان ثابتی جابه‌جا شود.

شبکه‌های عصبی



در حالی که برنامه پرسپترون روزنبلات، تنها شامل یک نورون پرسپترون بود برنامه‌های هوش مصنوعی امروزی فقط از یک نورون تشکیل نمی‌شوند؛ بلکه از تعداد زیادی نورون تشکیل شده‌اند که در قالب یک شبکه به هم متصل هستند. نمونه‌ای ساده از یک شبکه عصبی در شکل ۴٫۵ نشان داده شده است.



شکل ۴/۵ - یک شبکه عصبی فیدفوروارد (Feedforward) ساده

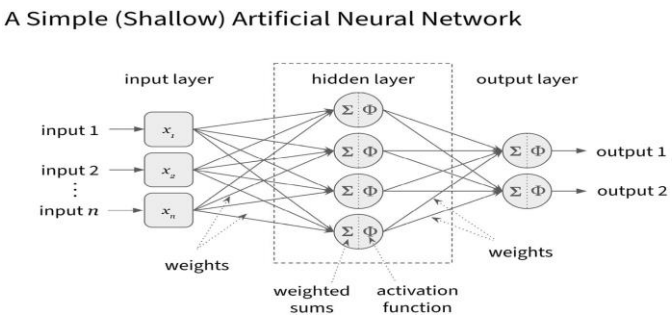
این نوع از شبکه‌های عصبی را فیدفوروارد (Feedforward) می‌نامند زیرا **تمام اتصالات آن در یک جهت از چپ به راست نمودار جریان دارند.** هیچ اتصالی در جهت عکس (از راست به چپ) وجود ندارد و همچنین هیچ حلقه یا چرخه‌ای در شبکه دیده نمی‌شود. **اطلاعات فقط در یک مسیر تغذیه می‌شوند: از لایه ورودی به سمت خروجی.**

در این مثال خاص از یک شبکه فیدفوروارد، در سمت چپ نمودار یک لایه ورودی داریم (که شامل نورون نیست)، در مرکز یک لایه پنهان شامل سه نورون (که گاهی «گره» یا «واحد» نیز نامیده می‌شوند) و در سمت راست، لایه خروجی شامل یک نورون است. شبکه‌های عصبی می‌توانند تعداد متفاوتی ورودی در لایه ورودی داشته باشند، همچنین ممکن است شامل چندین لایه پنهان باشند و هرکدام از این لایه‌ها نیز تعداد نورون‌های متفاوتی داشته باشند. همین‌طور لایه خروجی می‌تواند از یک یا چند نورون تشکیل شده باشد.

تمام این ویژگی‌ها نمونه‌هایی از **فراپارامترها (Hyperparameters)** هستند یعنی **جنبه‌هایی از ساختار شبکه عصبی که پیش از شروع فرایند آموزش، توسط طراح شبکه تعیین می‌شوند** و در طول آموزش ثابت باقی می‌مانند. تنظیم صحیح فراپارامترها در

شبکه‌های عصبی امروزی، **بیشتر شبیه یک هنر است تا یک علم**، و تعداد کمی از افراد در جهان که در این کار مهارت بالایی دارند، تقاضای بسیار بالایی در بازار دارند.

برای اینکه بتوانیم درک کنیم یک شبکه عصبی چه کاری می‌تواند انجام دهد، باید از ورودی‌های لایه ورودی شروع کنیم. این **ورودی‌ها** دقیقاً چه هستند؟ آن‌ها **ویژگی‌های داده‌ای هستند که شبکه قرار است آن‌ها را یاد بگیرد.** برای مثال، پیکسل‌های منفرد یک تصویر، کلمات یک سند متنی، قطعات کوتاه از فایل‌های صوتی، یا قیمت‌های سهام موجود در یک فایل صفحه‌گسترده. در نسخه ابتدایی پرسپترون روزنبلات، ورودی‌های شبکه تک‌نورونی او، خروجی‌های سلول‌های نوری و حسگرهایی بودند که ویژگی‌های مختلف کارت‌های پانچ‌شده را اندازه‌گیری می‌کردند. توجه داشته باشید که **گره‌های ورودی، نورون نیستند.** آن‌ها هیچ محاسبه‌ای انجام نمی‌دهند؛ بلکه فقط ویژگی‌ها را به نورون‌های لایه پنهان منتقل می‌کنند. به همین دلیل، در نمودار، برای آن‌ها از شکلی متفاوت نسبت به نورون‌ها استفاده شده است. همچنین مشاهده می‌کنید که هر گره ورودی در نمودار به تمام نورون‌های لایه پنهان متصل است. این نوع **اتصال کامل** (Fully Connected) رایج است، اگرچه در ادامه با شبکه‌هایی آشنا خواهیم شد که **اتصالات پراکنده** (Sparse Connections) دارند. به یاد داشته باشید که هر اتصال دارای یک وزن خاص است، و این وزن‌ها که همان پارامترهای مدل هستند که در طول آموزش تنظیم می‌شوند تا شبکه عصبی بتواند وظیفه‌اش را با دقت انجام دهد.



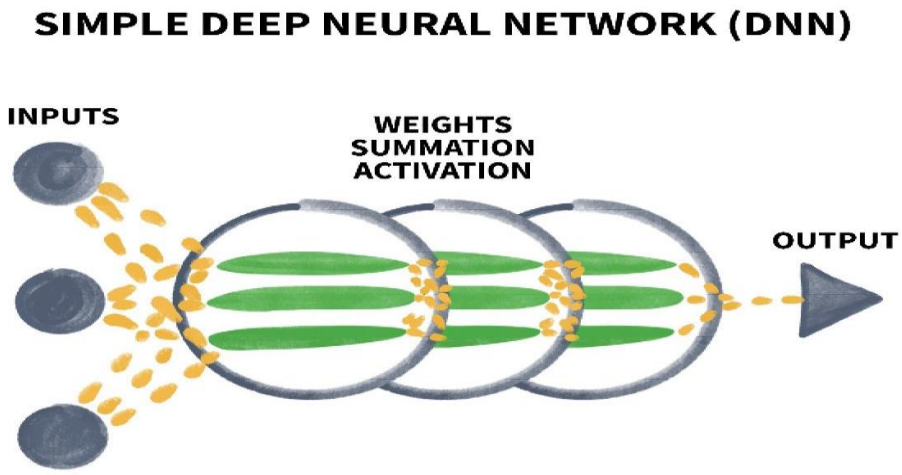
شبکه‌های عصبی عمیق

اکنون که با مفهوم پایه‌ای شبکه عصبی آشنا شدیم، این پرسش پیش می‌آید که

چه چیزی باعث می‌شود یک شبکه عصبی «عمیق» باشد؟ پاسخ ساده است. **تعداد لایه‌های پنهان در شبکه.** در یک شبکه عصبی، **هر لایه‌ای که بین لایه ورودی و لایه خروجی قرار دارد، لایه پنهان محسوب می‌شود.** هر **شبکه عصبی که شامل دو یا چند لایه پنهان باشد،** به عنوان یک **شبکه عصبی عمیق (**Deep Neural Networks DNN**)** شناخته می‌شود. طبیعی است که شبکه‌ای با تنها یک لایه پنهان گاهی به عنوان یک **شبکه عصبی کم‌عمق** (Shallow Neural Network) نامیده می‌شود.

همچنین، **تعداد لایه‌های پنهان برابر با عمق یک شبکه عصبی عمیق** است. برای مثال، شبکه‌ای با عمق دو، دارای دو لایه پنهان خواهد بود؛ نمونه‌ای از این نوع شبکه در شکل ۴٫۶ نمایش داده شده است. توجه داشته باشید که لایه ورودی (که شامل نورون نیست) و لایه خروجی (که شامل نورون‌های خروجی است) جزو لایه‌های پنهان محسوب نمی‌شوند.

اما سؤال اصلی اینجاست: وجود این لایه‌های پنهان اضافی چه فایده‌ای دارد؟ پاسخ این است که آن‌ها به شبکه‌های عصبی این امکان را می‌دهند تا **الگوهای بسیار پیچیده‌تر و انتزاعی‌تری** را در داده‌ها یاد بگیرند؛ الگوهایی که یک لایه منفرد قادر به شناسایی آن‌ها نیست. این روابط پیچیده‌تر در داده‌ها، اغلب کلید انجام وظایفی هستند که به طور معمول نیاز به هوش انسانی دارند.



شکل ۴/۶ - نمودار یک شبکه عصبی عمیق ساده با عمق دو

در دوران طلایی هوش مصنوعی، یعنی دو دهه نخست پس از کارگاه دارتموث، زمانی که پیشگامان این حوزه مشغول بررسی شبکه‌های عصبی بودند، هیچ‌کس نمی‌دانست چگونه می‌توان وزن نورون‌ها در لایه‌های پنهان را به‌گونه‌ای تنظیم کرد که دقت خروجی شبکه افزایش یابد. افزون بر آن، کارشناسانی مانند ماروین مینسکی و سیمور پیپرت تردید داشتند که حتی اگر این مسئله حل شود، تأثیر چشمگیری در توانمندی‌های شبکه ایجاد شود. این تردیدها باعث شد که منابع مالی و پژوهشی از حوزه شبکه‌های عصبی دور شود. اما چرا تنظیم وزن نورون‌های پنهان چنین چالش‌برانگیز بود؟ اگر دوباره به پرسپترون روزنبلات نگاه کنیم، متوجه می‌شویم که این مدلِ تک‌نورونی، هیچ لایه پنهانی ندارد و تنها یک نورون خروجی دارد. تنظیم وزن این یک نورون خروجی برای بهبود دقت مدل، کاری نسبتاً ساده و مستقیم است. حتی اگر چند نورون به لایه خروجی اضافه کنیم، مسئله همچنان به‌صورت ریاضی قابل حل است: کافی است خروجی واقعی را با خروجی مورد انتظار مقایسه کرده و وزن‌ها را در جهتی تنظیم کنیم که این اختلاف کاهش یابد. اما اگر ورودی از چندین لایه نورونی عبور کند، چه باید کرد؟ چگونه می‌توان تشخیص داد کدام یک از این وزن‌های زنجیره‌ای، بیشترین سهم را در خطای خروجی داشته‌اند؟ و چگونه باید آن‌ها را تنظیم کرد تا خطای کلی مدل

کاهش یابد؟ و اصلاً هدف از انجام این کار چه خواهد بود؟ در آن زمان، پاسخ هیچ‌کدام از این پرسش‌ها معلوم نبود.

این چالش که در آن باید مشخص شود هر وزن تا چه اندازه در خطا مؤثر بوده است، به نام **مسئله تخصیص اعتبار** (Credit Assignment Problem) شناخته می‌شود. از آن‌جایی که در واقع هدف، تعیین این است که کدام وزن‌ها بیشتر در بروز خطا نقش داشته‌اند، برخی معتقدند نام مناسب‌تر آن می‌تواند **مسئله تخصیص تقصیر** (Blame Assignment Problem)باشد! چه بخواهیم از لفظ «اعتبار» استفاده کنیم و چه «تقصیر»، مسئله یکی است: وزن‌های شبکه عصبی عمیق باید به چه میزان و چگونه در طول آموزش تنظیم شوند تا دقت پیش‌بینی مدل بهبود یابد؟

آموزش شبکه‌های عصبی عمیق

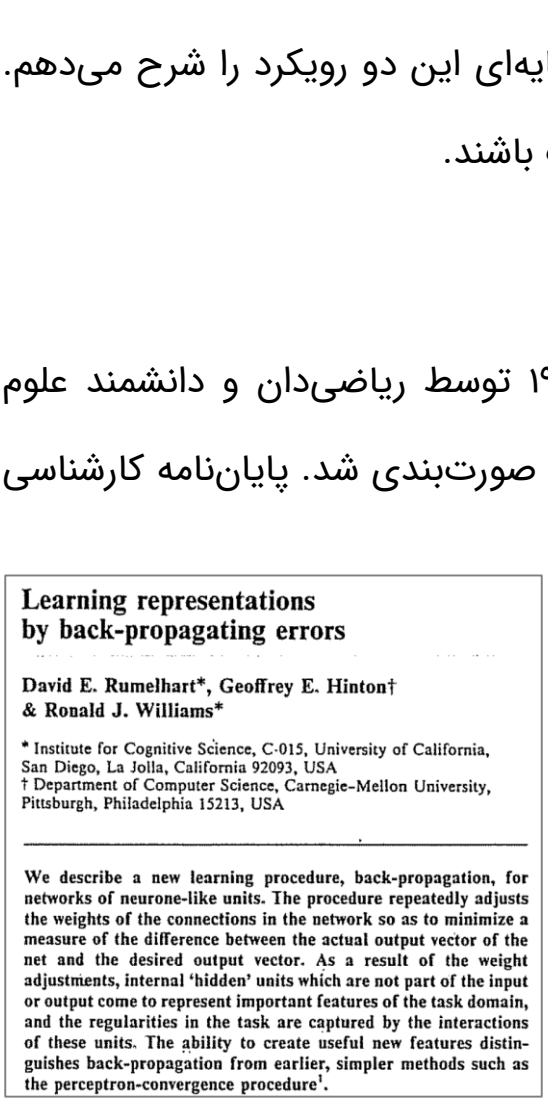
راه‌حل مسئله تخصیص اعتبار، در شبکه‌های عصبی عمیق، از همان ابتدا در برابر دیدگان دانشمندان بود، حتی در زمان فرانک روزنبلات. روشی ریاضی که در دهه ۱۹۶۰ برای بهینه‌سازی مسیر پرواز موشک‌ها و فرآیندهای شیمیایی به‌کار رفته بود، تنها باید راه خود را به حوزه هوش مصنوعی باز می‌کرد که نهایتاً این اتفاق افتاد.

دو مؤلفه اصلی روشی که امروزه برای تنظیم وزن‌های شبکه‌های عصبی عمیق در فرآیند آموزش استفاده می‌شود، پس‌انتشار خطا یا انتشار رو به عقب (backpropagation) و نزول گرادیان (gradient descent) هستند. در فرایند آموزش، این دو با هم موثر هستند. هرچند استفاده از این دو، از لحاظ نظری همیشه امکان‌پذیر بود اما تا زمانی که **توان پردازشی** و **اندازه داده‌های آموزشی** به اندازه کافی افزایش نیافت، توانمندی‌های

خارق‌العاده آنها نمایان نشد. در این‌جا مفاهیم پایه‌ای این دو رویکرد را شرح می‌دهم. فرمول‌های ریاضی مرتبط، فراتر از دامنه این کتاب باشند.

روش پس‌انتشار

روش پس‌انتشار خطا، نخستین بار در سال ۱۹۷۰ توسط ریاضی‌دان و دانشمند علوم رایانه فنلاندی، سپو لینانیما (Seppo Linnainmaa)، صورت‌بندی شد. پایان‌نامه کارشناسی ارشد او، مراحل لازم برای تنظیم وزن نورون‌های پنهان را به زبان ساده فنلاندی شرح داده بود، هرچند مستقیماً به شبکه‌های عصبی اشاره‌ای نداشت. در سال ۱۹۷۴، دانشمند آمریکایی پل وربوس، آن را در شبکه‌های عصبی به کار گرفت و در سال ۱۹۸۶ توسط **دیوید روملهارت، جفری هینتون، و رونالد ویلیامز** در مقاله معروفشان با عنوان «یادگیری بازنمایی‌ها از طریق پس‌انتشار خطا» به‌طور گسترده معرفی و محبوب شد.

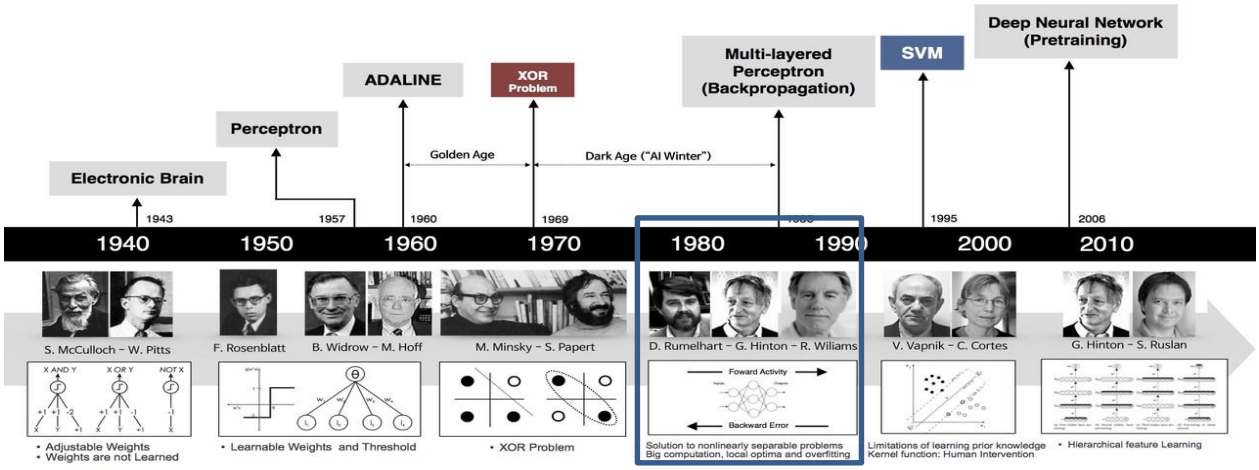


Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton† & Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, California 92093, USA
† Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Philadelphia 15213, USA

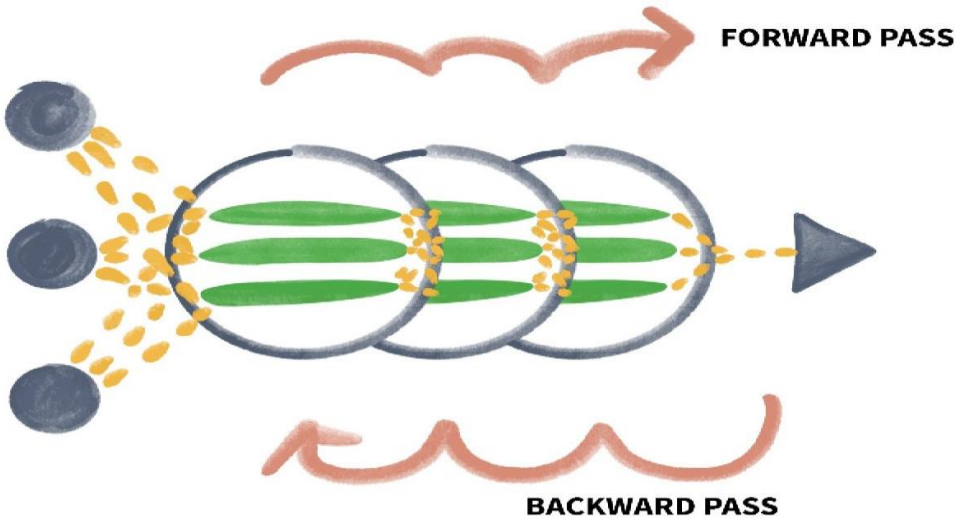
We describe a new learning procedure, back-propagation, for networks of neurone-like units. The procedure repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector. As a result of the weight adjustments, internal 'hidden' units which are not part of the input or output come to represent important features of the task domain, and the regularities in the task are captured by the interactions of these units. The ability to create useful new features distinguishes back-propagation from earlier, simpler methods such as the perceptron-convergence procedure[1].



شکل ۴٫۷ - جایگاه کلیدی روملهارت، هینتون، و ویلیامز در شبکه‌های عصبی عمیق

در سطح مفهومی، نحوه کار پس‌انتشار چنین است: ابتدا در مرحله‌ای به‌نام «عبور رو به جلو (Forward Pass)»، یک ورودی آموزشی به شبکه داده می‌شود و وزن‌های تصادفی مدلِ آموزش‌ندیده، منجر به فعال‌سازی نورون‌ها می‌شود؛ این فعال‌سازی‌ها به‌صورت لایه‌به‌لایه و نورون‌به‌نورون در شبکه، جریان می‌یابند تا به نورون‌های خروجی برسند و پیش‌بینی مدل را ارائه دهند.

در ادامه، مرحله «عبور رو به عقب (Backward Pass)» آغاز می‌شود. ابتدا، الگوریتم با استفاده از یک تابع خطا (loss function)، خطای هر نورون خروجی را با مقایسه کردن خروجی واقعی با خروجی مورد انتظار محاسبه می‌کند. برای انجام این کار، باید بدانیم خروجی درست برای آن ورودی چیست. به همین دلیل، **الگوریتم پس‌انتشار بخشی از یک فرآیند یادگیری تحت نظارت است** که در آن داده‌های آموزشی توسط انسان‌ها برچسب‌گذاری شده‌اند. توجه داشته باشید که گاهی از اصطلاحات «تابع خطا» و «تابع هزینه» به‌جای یکدیگر استفاده می‌شود اما از نظر فنی، تابع خطا مربوط به خطای یک نمونه آموزشی منفرد و تابع هزینه، میانگین خطا در کل مجموعه آموزشی را نشان می‌دهد.



شکل ۴/۸ - نمودار مفهومی عبور رو به جلو و عبور رو به عقب در روش پس‌انتشار خطا

پس از محاسبه خطا، الگوریتم همان کاری را می‌کند که از نامش پیداست: این خطاها را به صورت لایه‌به‌لایه و نورون‌به‌نورون در شبکه به عقب منتقل می‌کند و **میزان مسئولیت (یا تقصیر) هر وزن در خطای کلی** را مشخص می‌سازد. این همان مسئله تخصیص اعتبار است که پیش‌تر اشاره شد و به الگوریتم اجازه می‌دهد مشخص کند که هر وزن باید چقدر تغییر کند تا تابع خطا کاهش یافته و دقت پیش‌بینی بهبود یابد. مدل این فرآیند را فقط یک‌بار برای یک نمونه انجام نمی‌دهد. بلکه آن را با **نمونه‌های آموزشی مختلف**، در **چندین تکرار**، اجرا می‌کند و هر بار وزن‌ها را کمی تنظیم می‌کند تا خطای مدل به‌تدریج کاهش یابد. این‌گونه است که یک شبکه عصبی مدرن با استفاده از داده‌های آموزشی و الگوریتم پس‌انتشار، یاد می‌گیرد چگونه یک وظیفه را انجام دهد، بر اساس الگوهای موجود در داده، نه بر اساس دستورالعمل‌های صریح.

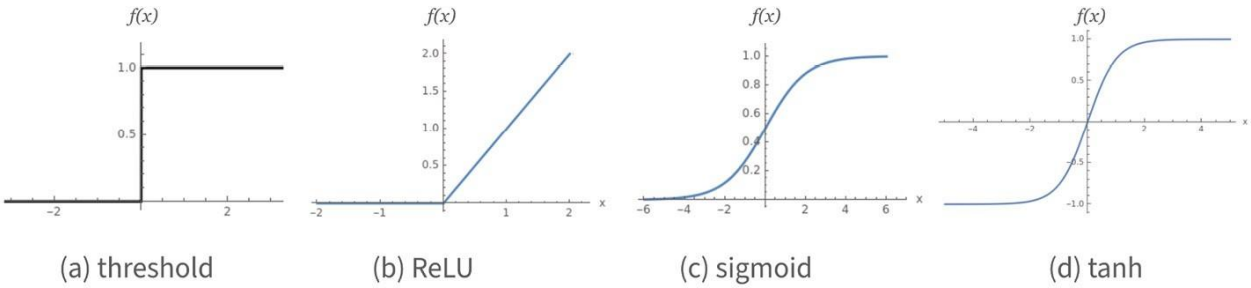
نزول گرادیان

الگوریتم پس‌انتشار از یک تکنیک مهم ریاضی به‌نام نزول گرادیان (Gradient Descent) استفاده می‌کند تا مشخص کند در هر دور آموزش، هر وزن چطور باید تغییر کند. گونه‌های مختلفی از نزول گرادیان وجود دارد، اما ایده اصلی این است که **الگوریتم، جهتی را پیدا می‌کند که اگر وزن در آن جهت تغییر کند، بیشترین کاهش در خطای مدل (تابع خطا) حاصل خواهد شد.** در این فرآیند، مدل از یک ابرپارامتر دیگر به‌نام نرخ یادگیری (Learning Rate) نیز استفاده می‌کند که تعیین می‌کند در هر تکرار، وزن‌ها چقدر تغییر کنند. هدف کلی الگوریتم این است که تا حد ممکن سریع و مؤثر، خطای تابع خطا را کاهش دهد. به‌عبارت ریاضی، به‌دنبال یافتن **حداقل موضعی** (Local Minimum) روی منحنی تابع خطا است.

می‌توان این فرآیند را **به اسکی‌بازی تشبیه کرد** که می‌خواهد از بالای کوه به پایین پیست برسد تا به کلبه استراحت برسد. هرچه اسکی‌باز بالاتر باشد، هزینه مدل بیشتر است. نزول گرادیان مانند جهتی است که اسکی‌باز انتخاب می‌کند. نرخ یادگیری مانند سرعت حرکت اسکی‌باز است. اگر خیلی آهسته حرکت کند، مدت زیادی طول می‌کشد تا برسد. ولی اگر خیلی تند برود، ممکن است از کلبه عبور کند. در مجموع، هدف این است که **با سرعتی معقول و در جهتی صحیح**، به پایین برسد.

شکل ۴/۹ - نزول گرادیان به‌مثابه اسکی‌بازی است که می‌خواهد به کلبه پایین دامنه برسد.

در حاشیه، باید گفت **تکیه الگوریتم پس‌انتشار بر شیب توابع**، یکی از دلایلی است که شبکه‌های عصبی مدرن از توابع فعال‌سازی نرم‌تر و پیوسته‌تری مانند ReLU، تانژانت هیپربولیک (tanh) یا سیگموید استفاده می‌کنند، نه تابع آستانه‌ای که روزنبلات در پرسپترون به‌کار برد. تابع آستانه‌ای دارای «لبه‌های تیز» است (همان‌گونه که در شکل ۴/۱۰ نشان داده شده)، که ناگهان از مقدار ۰ به ۱ می‌پرد. اگر به یاد داشته باشید، در حساب دیفرانسیل نمی‌توان شیب (مشتق) را در نقاطی با گسست یا لبه‌های تیز به‌دست آورد. افزون بر این، غیرخطی بودن توابعی مانند سیگموید و تانژانت هیپربولیک به شبکه کمک می‌کند تا روابط پیچیده‌تری بین ورودی و خروجی را کشف کند.



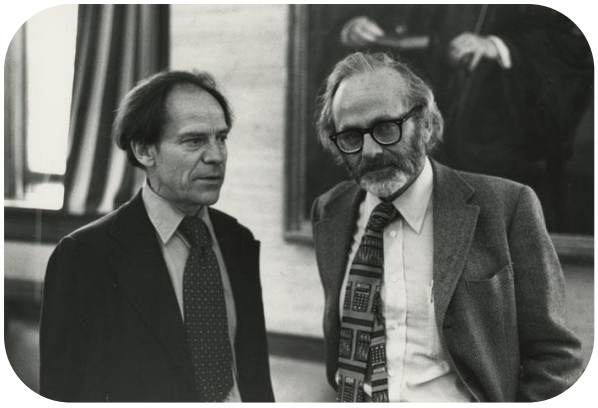
شکل ۴/۱۰ - توابع فعال‌سازی مختلف در شبکه‌های عصبی

توانایی شناسایی الگوهای پیچیده در داده‌های گسترده، دلیل اصلی **ارزشمند بودن شبکه‌های عصبی عمیق** است. این همان چیزی است که باعث شده بتوانند کارهایی انجام دهند که زمانی غیرممکن به‌نظر می‌رسید و آن‌ها را به نماد اصلی هوش مصنوعی در دوران ما بدل کرده است. البته، برای رسیدن به این نقطه، پیشرفت در توان محاسباتی و در دسترس بودن داده‌های آموزشی حیاتی بود. اما در کنار آن، تلاش بی‌وقفه و نوآوری پژوهشگران و دانشمندان علوم رایانه نیز نقش کلیدی داشت؛ کسانی که روش‌هایی مانند پس‌انتشار و نزول گرادیان را توسعه دادند و به کار گرفتند، و با آزمون و خطا در توابع فعال‌سازی، پیکربندی شبکه‌ها، و تنظیم ابرپارامترها، شبکه‌های عصبی عمیق را به واقعیت بدل کردند.

انواع شبکه‌های عصبی عمیق

تا این‌جا مقدار قابل توجهی اطلاعات فنی بررسی کردیم، حتی اگر کاملاً در معادلات ریاضی مربوط به شبکه‌های عصبی غرق نشده باشیم. برای جمع‌بندی بحث خود درباره‌ی یادگیری عمیق و شبکه‌های عصبی، بهتر است نگاهی بیندازیم به چند نوع مختلف از شبکه‌های عصبی عمیق که موجب تحولات بزرگی در حوزه‌های بینایی رایانه‌ای و پردازش زبان طبیعی شده‌اند.

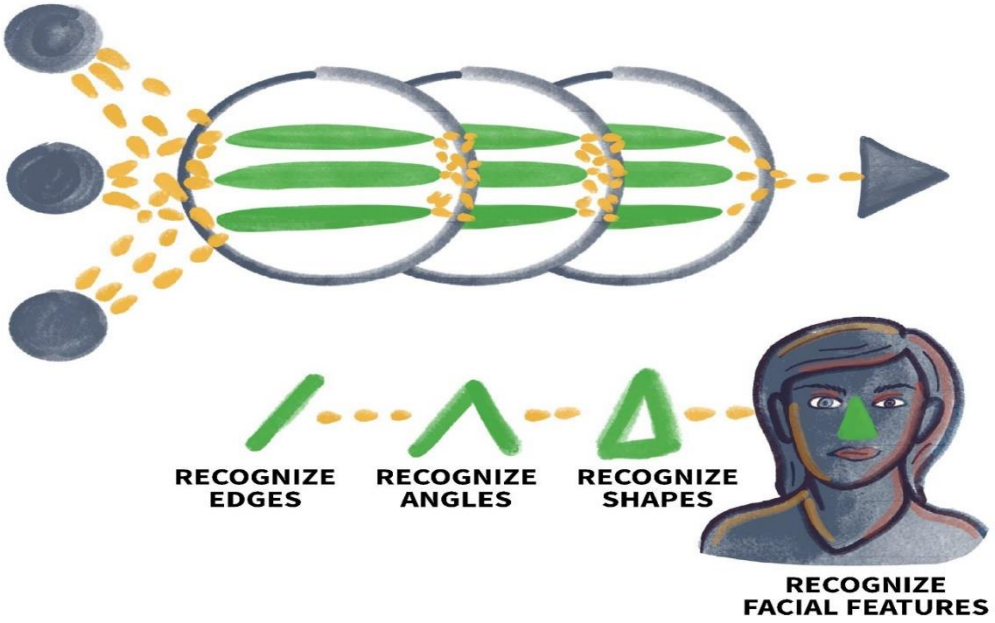
شبکه‌های عصبی پیچشی



نخستین نوع، شبکه‌ی عصبی پیچشی (یا CNN که گاه ConvNet هم نامیده می‌شود) است. این نوع از شبکه‌های عصبی عمیق، موجب پیشرفت‌های شگرفی در حوزه‌ی **بینایی رایانه‌ای** شده‌اند از جمله توانایی مدل‌های هوش مصنوعی در شناسایی اشیا و تحلیل تصاویر و ویدیوها. وجود CNNها است که باعث شده بتوانیم چک دست‌نویس را با عکس گرفتن از آن و واریز از طریق تلفن همراه، به حساب بانکی‌مان وارد کنیم. این شبکه‌ها همچنین نقش محوری در **خودروهای خودران**، **تصویربرداری پزشکی** و بسیاری کاربردهای دیگر دارند. اما چه چیزی باعث تمایز CNNها شده و از کجا آمده‌اند؟ CNNها بر اساس بینش‌هایی از نحوه‌ی پردازش اطلاعات تصویری در مغز انسان توسعه یافته‌اند. در دهه‌های ۱۹۵۰ و ۱۹۶۰، دو نوروفیزیولوژیست به نام‌های **دیوید هابل و تورستن ویسل** کشف کردند که قشر بینایی مغز انسان به صورت سلسله‌مراتبی از لایه‌های عصبی سازمان یافته است که هرچه جلوتر می‌رویم، وظیفه‌ی شناسایی ویژگی‌های پیچیده‌تر و انتزاعی‌تری را برعهده دارند.

به‌طور دقیق‌تر، **لایه‌های ابتدایی** نسبت به **ویژگی‌های ساده‌ای** مانند **خطوط** یا **لبه‌ها** حساس‌اند. لایه‌های بعدی این اطلاعات را دریافت می‌کنند تا ویژگی‌های پیچیده‌تری مانند **زاویه‌ها** یا **گوشه‌ها** را شناسایی کنند. لایه‌های بعدی نیز اشکالی مانند مثلث یا مربع را تشخیص می‌دهند و این روند ادامه می‌یابد تا در نهایت ویژگی‌هایی مانند چشم، بینی یا حتی کل چهره را شناسایی می‌کنند. شبکه‌های CNN نیز به همین صورت

به صورت **سلسله‌مراتبی** کار می‌کنند. نسخه‌ای ساده‌شده از این سلسله‌مراتب در شکل ۴/۱۱ نشان داده شده است.

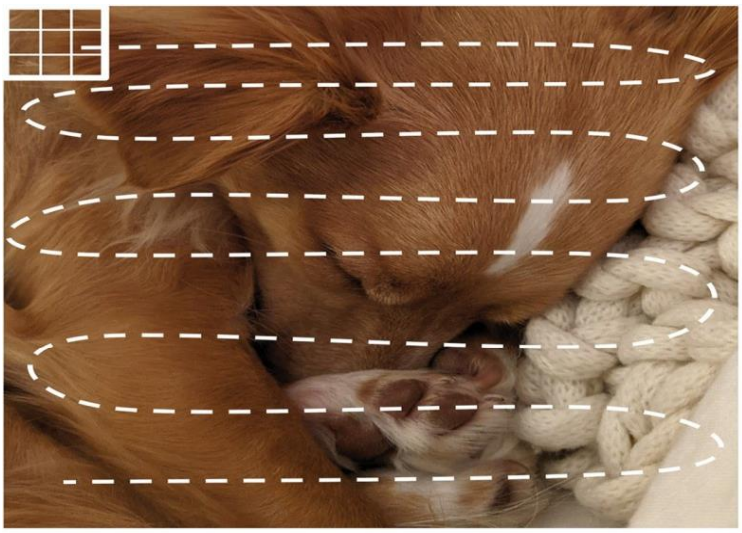


شکل ۴/۱۱ - پردازش سلسله‌مراتبیِ اطلاعات تصویری

توسعه‌ی CNNها توسط یان لوکان در دهه‌ی ۱۹۸۰ الهام‌گرفته از همین روند سلسله‌مراتبی در پردازش اطلاعات تصویری بود؛ درست مانند پیش‌نمونه‌ی آن یعنی شبکه‌ی "نئوکگنیتورن" که کونیهیکو فوکوشیما در دهه‌ی ۱۹۷۰ طراحی کرده بود. فوکوشیما شبکه‌اش را برای شناسایی دست‌نویس‌های ژاپنی آموزش داد و یان لوکان نیز در سال ۱۹۸۹ نسخه‌ای از CNN را که از الگوریتم پس‌انتشار خطا و گرادیان نزولی بهره می‌برد، برای شناسایی کدپستی‌های دست‌نویس با دقتی بی‌سابقه به کار برد.

CNN چگونه کار می‌کند؟ در اغلب کاربردها، **ورودی‌ها**، **مقادیر RGB (قرمز، سبز، آبی) پیکسل‌های یک تصویر یا ویدیو** هستند. برخلاف شبکه‌های عادی که هر پیکسل را به هر نورون در لایه‌ی اول وصل می‌کنند، در CNN مجموعه‌ای از فیلترها (یا کرنل‌ها) اعمال می‌شود که برای شناسایی نوع خاصی از ویژگی‌ها در تصویر طراحی شده‌اند. به‌عنوان

مثال، یک فیلتر ممکن است لبه‌های افقی را شناسایی کند و دیگری لبه‌های عمودی را. هر **فیلتر**، به جای بررسی کل تصویر به‌صورت یکجا، تنها **بخشی کوچک از تصویر** (مثلاً شبکه‌ای ۳×۳ شامل ۹ پیکسل) را بررسی می‌کند و سپس با حرکت پیوسته روی تصویر، نقشه‌ای از ویژگی‌ها (Feature Map) تولید می‌کند که مشخص می‌کند آن ویژگیِ خاص، در کجای تصویر یافت شده است.



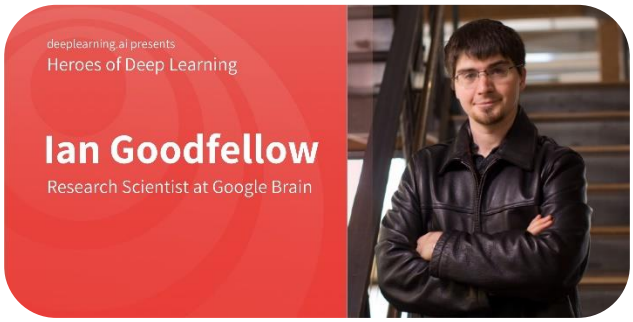
شکل ۴/۱۲ - تصویری مفهومی از حرکت فیلتر در شناسایی ویژگی‌ها

هر نورون در لایه‌های پیچشی، مربوط به یک موقعیت خاص از یک فیلتر خاص است که روی تصویر حرکت می‌کند. وزن‌های اعمال‌شده به نورون‌های لایه‌ی پیچشی همگی یکسان‌اند و این امکان را فراهم می‌آورد که ویژگی مدنظر مثلاً پنجه‌ی سگ در هر نقطه‌ای از تصویر که باشد، توسط نورونی خاص تشخیص داده شود.

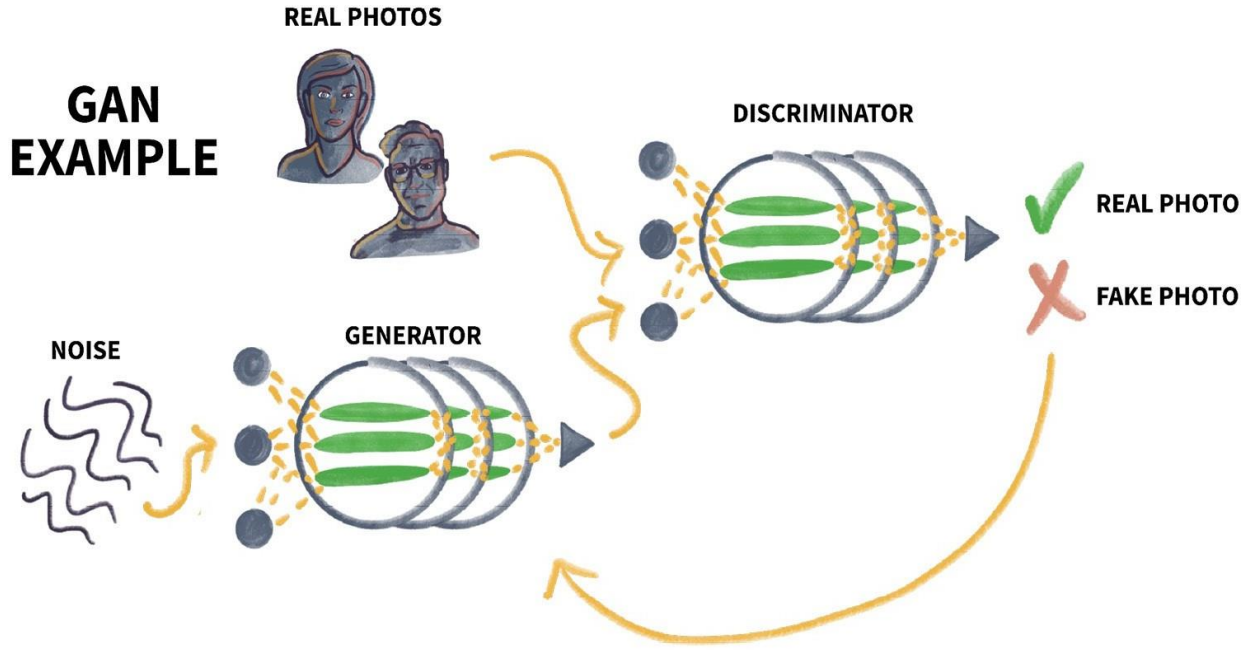
CNN همچنین از توابع فعال‌سازی مانند ReLU برای افزودن خاصیت غیرخطی استفاده می‌کند تا بتواند الگوهای پیچیده‌تری را شناسایی کرده و الگوریتم گرادیان نزولی را اجرا کند. لایه‌هایی مانند لایه‌های pooling که ابعاد نقشه‌های ویژگی را کاهش می‌دهند و لایه‌های طبقه‌بندی که خروجی مدل را مشخص می‌کنند نیز در ادامه به کار گرفته

می‌شوند. در نهایت، یک CNN با مجموعه‌ای بزرگ از تصاویر برچسب‌گذاری‌شده آموزش می‌بیند و با کمک الگوریتم پس‌انتشار و گرادیان نزولی، وزن‌های فیلترها تنظیم می‌شود تا شبکه یاد بگیرد چگونه ویژگی‌های معنادار را شناسایی کند و این همان چیزی است که باعث می‌شود بتواند تصاویر جدید را تحلیل کند.

شبکه‌های مولد خصمانه



شبکه‌های مولد خصمانه (GAN)، که در سال ۲۰۱۴ توسط ایان گودفلو و همکارانش معرفی شد، از دو شبکه‌ی عصبی تشکیل شده‌اند که به‌صورت جفت و در رقابت با هم آموزش می‌بینند. یکی از این دو مدل، **تولیدکننده (Generator)** است که تلاش می‌کند داده‌هایی مانند متن، تصویر، صوت یا ویدیو تولید کند که تا حد ممکن به داده‌های واقعی شبیه باشند. مثلاً چهره‌ای بسازد که به قدری واقعی به نظر برسد که نتوان تشخیص داد واقعی نیست. مدل دیگر **تمییزدهنده (Discriminator)** است که وظیفه دارد تشخیص دهد ورودی‌ای که دریافت کرده، واقعی است (از مجموعه‌ی آموزش گرفته شده) یا ساختگی است (توسط تولیدکننده ساخته شده). این دو مدل با هم آموزش می‌بینند و **چون در رقابت‌اند، رابطه‌ای «خصمانه» دارند.**



شکل ۴/۱۳ - تصویری مفهومی از GAN

در ابتدا، تولیدکننده تنها می‌تواند داده‌هایی کاملاً تصادفی و آشکارا تقلبی تولید کند. در این مرحله، تشخیص تمییزدهنده آسان است. اما با تکرار فرآیند آموزش، نتایج هر دور به تولیدکننده بازخورد داده می‌شود و تولیدکننده یاد می‌گیرد خروجی‌هایی تولید کند که شباهت بیشتری به داده‌های واقعی داشته باشند. هم‌زمان، تمییزدهنده نیز به دنبال سرنخ‌های ظریف‌تری برای شناسایی تقلبی بودن داده می‌گردد. این رقابت باعث پیشرفت پیوسته‌ی هر دو مدل می‌شود.

هدف معمول این فرآیند این است که تولیدکننده به نقطه‌ای برسد که داده‌های ساختگی‌اش چنان واقعی باشند که تمییزدهنده نتواند بهتر از انتخاب تصادفی (شانس ۵۰/۵۰) آن‌ها را از واقعی تشخیص دهد. البته در بخش‌های بعدی به مسائل اخلاقی مربوط به هوش مصنوعی مولد و محتوای ساختگی مانند «دیپ‌فیک‌ها» که ممکن است گمراه‌کننده یا زیان‌بار باشند خواهیم پرداخت.

ترنسفورمرها و مدل‌های زبانی بزرگ

مدل‌های زبانی بزرگ (LLM) نوع خاصی از **شبکه‌های عصبی عمیق‌اند** که با استفاده از **حجم بسیار بزرگی از داده‌ها**ی متنی آموزش دیده‌اند تا بتوانند **زبان انسانی** را پردازش و تولید کنند. این مدل‌ها شامل میلیاردها پارامتر (وزن و بایاس نورون‌ها) هستند که در طول آموزش تنظیم شده‌اند تا خروجی‌هایی شبیه به گفتار انسانی تولید کنند. مدل‌های پیشرفته‌ی امروزی مانند GPT-4، مدل LLaMA 2 از متا و Mixtral-۸x۷B از Mistral AI، می‌توانند طیف گسترده‌ای از وظایف زبانی را با سرعت و دقت نسبتاً بالا انجام دهند. هرچند این مدل‌ها کامل نیستند و دارای محدودیت‌ها و سوگیری‌هایی در داده‌های آموزشی خود هستند، اما قدرت و توانایی آن‌ها انکارناپذیر است. یکی از دلایل پیشرفت چشمگیر این مدل‌ها، استفاده از معماری **ترنسفورمر** است.

ترنسفورمر نخستین بار در مقاله‌ی معروف «همه‌چیز توجه است (Attention Is All You Need)» در سال ۲۰۱۷ توسط پژوهشگران گوگل معرفی شد. ترنسفورمر، نوع**ی شبکه‌ی عصبی است که از سازوکاری به نام «توجه به خود (self-attention)» بهره می‌برد.** این سازوکار به مدل اجازه می‌دهد به همه‌ی واژگان (یا بخش‌های واژه) از یک جمله، به‌صورت هم‌زمان نگاه کند، نه فقط به ترتیب و یکی‌یکی. در ادامه نیز از شبکه‌های عصبی فیدفوروارد برای پردازش خروجی این مکانیزم استفاده می‌شود.

ترنسفورمرها جایگزین شبکه‌های بازگشتی (RNN) و حافظه‌ی بلند-کوتاه‌مدت (LSTM) در بسیاری از وظایف پردازش زبان طبیعی شده‌اند چرا که عملکرد بهتری دارند. به‌طور ساده، این معماری به هر واژه در جمله، نمره‌ای نسبت می‌دهد بر اساس میزان اهمیت آن در پیش‌بینی واژه‌ی بعدی. این همان کاری است که سیستم‌های پیش‌بینی متن در Gmail برای پیشنهاد واژه‌های بعدی انجام می‌دهند.

با این حال، ترنسفورمرها چیزی **فراتر از «پیش‌بینی‌کننده‌ی واژه‌ی بعدی»** هستند. توانایی آن‌ها در پردازش هم‌زمان تمام واژه‌ها، در کنار داده‌های آموزشی گسترده‌شان، باعث می‌شود بتوانند وظایف پیچیده‌تری را نیز انجام دهند. سازوکار توجه، این امکان را می‌دهد که مدل بتواند روی بخش‌های خاصی از جمله تمرکز کند و روابط معنایی میان واژه‌ها را درک کند. این مدل‌ها نه فقط در زبان، بلکه در بینایی رایانه‌ای و پردازش صوت نیز به‌کار می‌روند.

در زمان ChatGPT در نوامبر ۲۰۲۲، این مدل از GPT-3.5 استفاده می‌کرد با ۱۷۵ میلیارد پارامتر و داده‌های آموزشی بالغ بر **۵۷۰ گیگابایت** (از منابعی چون ویکی‌پدیا، کتاب‌ها، مقالات و...) بهبود یافته بود. OpenAI پس از آموزش مدل ChatGPT، آن را با کمک بازخورد انسانی بهبود داده است. فرآیندی که **«یادگیری تقویتی با بازخورد انسانی»** یا RLHFنام دارد.

البته این فرآیند بی‌نقص نیست. در نهایت، مدل یاد می‌گیرد که خروجی‌هایی تولید کند که مورد پسند مربیان انسانی باشد. این مسئله، همان‌طور که تصور می‌کنید، هم مزایایی دارد و هم چالش‌های خاص خود را، مخصوصاً وقتی که مدل به جای واقعیت، چیزی را می‌گوید که ما دوست داریم بشنویم.

خلاصه

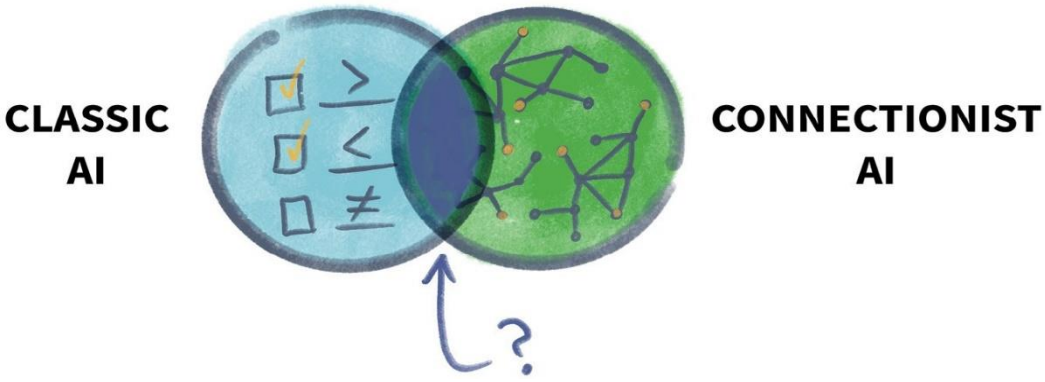
هیچ تردیدی نیست که **یادگیری عمیق، انقلابی در فناوری هوش مصنوعی** و به‌طور کلی **در جهان ما** به وجود آورده است. اکنون شما درک بهتری از چیستی یادگیری عمیق، ریشه‌های آن، جایگاهش در تصویر کلی هوش مصنوعی، و علت قدرت و تحول‌آفرینی آن دارید. یادگیری عمیق در عصر «کلان‌داده (Big Data)» به شکوفایی رسیده است. شبکه‌های عصبی عمیق با استفاده از **مجموعه‌داده‌های بسیار بزرگ** و با تکیه بر فرآیندهای محاسباتی پیشرفته و سخت‌افزارهای قدرتمندی مانند **پردازنده‌های گرافیکی (GPU)** آموزش می‌بینند؛ سخت‌افزارهایی که می‌توانند حجم عظیمی از محاسبات را در مدت‌زمانی نسبتاً کوتاه انجام دهند. شبکه‌های عصبی و یادگیری عمیق از **الهام‌گیری از نحوه‌ی کارکرد مغز انسان** پدید آمده‌اند. البته **شیوه‌ی عملکرد آن‌ها تفاوت زیادی با مغز زیستی دارد** اما با مقایسه‌ی نورون مصنوعی و نورون زیستی می‌توان ریشه‌های این **الهام** را مشاهده کرد. ساده‌ترین شکل شبکه‌ی عصبی، دستگاه تک‌نورونی فرانک روزنبلات به نام «پرسبترون» بود که در نخستین سال‌های شکل‌گیری حوزه‌ی هوش مصنوعی مطرح شد. اما علاقه و سرمایه‌گذاری در این حوزه به دلیل دشواری آموزش شبکه‌هایی با لایه‌های پنهان متعدد (یا همان شبکه‌های عصبی عمیق) رو به افول گذاشت.

با وجود کاهش اشتیاق به شبکه‌های عصبی در دهه‌های ۱۹۷۰ و ۱۹۸۰، پژوهش در این زمینه در میان گروه کوچکی از متخصصان متعهد در پشت‌صحنه ادامه یافت. در گذر زمان، مجموعه‌ای از پیشرفت‌ها در زمینه‌ی آموزش شبکه‌های عصبی عمیق به‌ویژه به‌کارگیری الگوریتم پس‌انتشار خطا، روش گرادیان نزولی و توابع فعال‌سازی مانند ReLU، سیگموید و تانژانت هایپربولیک، باعث تحولی شدند که دیگر نمی‌شد آن را نادیده گرفت. ظهور معماری‌های نوآورانه در شبکه‌های عصبی منجر به جهش‌های

بزرگی در توانایی‌های بینایی رایانه‌ای با استفاده از شبکه‌های عصبی پیچشی، شبکه‌های مولد خصمانه و پردازش زبان طبیعی (با ترنسفورمرها) شد. به‌واسطه‌ی این جهش‌های تکاملی، علاقه به‌سمت هوش مصنوعی اتصال‌گرا یا غیرنمادین (subsymbolic) گرایش یافته است و تاکنون نیز در همین مسیر باقی مانده است.

اما مسیر آینده چه خواهد بود؟ پیشرفت‌های بیشتری در زمینه‌ی هوش مصنوعی نمادین مورد نیاز است تا بتوان قابلیت اعتماد شبکه‌های عصبی عمیق را تقویت کرد زیرا این شبکه‌ها گاهی دچار «توهم» می‌شوند و خروجی‌هایی غیرمنطقی یا بی‌اساس تولید می‌کنند.

بعلاوه هیچ دلیلی وجود ندارد که هوش مصنوعی نمادین و غیرنمادین را در تقابل با یکدیگر قرار دهیم. هوش مصنوعی می‌تواند عالی باشد اگر هم توانایی یادگیری از داده‌ها (هوش مصنوعی غیرنمادین) را داشته باشد و هم دارای دانشی مبتنی بر قواعد (هوش مصنوعی نمادین).



شکل ۴/۱۴ - شاید هم‌پوشانی هوش مصنوعی کلاسیک (نمادین) و هوش مصنوعی اتصال‌گرا (غیرنمادین)، دری به‌سوی یک جهش تکاملی جدید در هوش مصنوعی بگشاید.

فصل پنجم

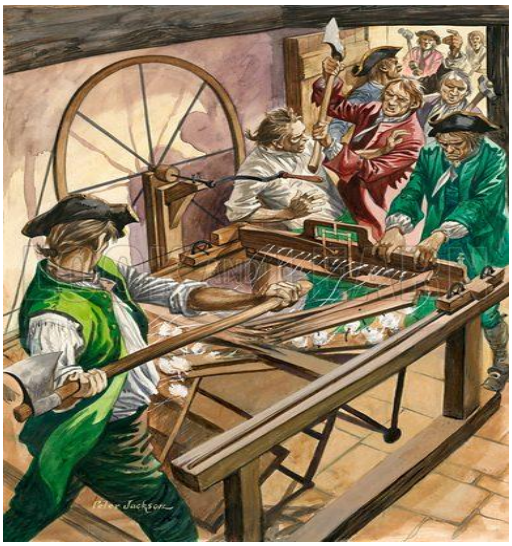
مزایا و نگرانی‌های هوش مصنوعی

در این مرحله از سفرمان به دنیای هوش مصنوعی، باید **تأثیر آن را بر دنیای واقعی** بررسی کنیم. بدون در نظر گرفتن تأثیر یک فناوری بر سیاره‌ی ما و زندگی روی آن به‌ویژه زندگی انسان‌ها، نمی‌توان به درکی کامل از آن رسید. آیا می‌توان داستان خودرو را در قرن بیستم روایت کرد، بی‌آنکه تأثیر آن بر جامعه، اقتصاد و محیط زیست را در نظر گرفت؟ البته می‌توان به بررسی عملکرد درونی موتور احتراق داخلی پرداخت و انواع مختلف وسایل نقلیه، از خودروهای سواری گرفته تا مینی‌ون‌ها، کامیون‌ها و شاسی‌بلندها را با هم مقایسه کرد. اما تا زمانی که به تأثیر خودرو بر زندگی روزمره، دگرگونی شهرها و حومه‌ها، و نقش آن در افزایش انتشار گاز دی‌اکسیدکربن نپردازیم، تصویر کاملی نخواهیم داشت. این مباحث، **موضوعات حاشیه‌ای داستان نیستند؛ بلکه اصل داستان‌اند.**

تا اینجای کتاب، جنبه‌های مختلفی از هوش مصنوعی را بررسی کرده‌ایم: از تعاریف گوناگون و انواع و شاخه‌های اصلی آن گرفته تا تاریخچه‌ی این حوزه و رویکردها و روش‌هایی که فناوری را تا جایگاه امروزی‌اش پیش برده‌اند. اما هنوز به این موضوع نپرداخته‌ایم که هوش مصنوعی چگونه ما را دگرگون کرده است. بنابراین، بیایید از نگاه فنی فاصله بگیریم و به **جنبه‌های انسانی و فردی** آن بپردازیم.

در چنین بحث‌هایی، بسیار آسان است که دیدگاهی یک‌جانبه اتخاذ کنیم و در واقع، بسیاری از افراد چنین می‌کنند. در یک سوی طیف، مشتاقان هوش مصنوعی و **فناوری‌دوست‌ها** قرار دارند که با شور و حرارت از مزایای هوش مصنوعی سخن می‌گویند. از نگاه آن‌ها، این فناوریِ انقلابی، نویدبخش عصر جدیدی از رفاه همگانی

است. این گروه که می‌توان آن‌ها را «**هواداران سرسخت هوش مصنوعی**» نامید اغلب در صورت ادامه‌ی شتاب در پذیرش این فناوری، منافع مالی زیادی خواهند برد. در سوی دیگر، **منتقدان هوش مصنوعی** و **فناوری‌هراس‌ها** قرار دارند که از آینده‌ای تاریک و ویران‌گر سخن می‌گویند؛ آینده‌ای که در آن، هوش مصنوعی نه‌تنها مشاغل ما، بلکه اساس زندگی‌مان را به خطر می‌اندازد. آن‌ها درباره‌ی دنیایی **هشدار** می‌دهند که در آن، انسان‌ها مطیع و وابسته به ربات‌ها خواهند شد.



البته این پدیده تازه‌ای نیست. **دیگر دستاوردهای بزرگ فناورانه در تاریخ بشر نیز واکنش‌هایی دوگانه برانگیخته‌اند.** در انگلستان قرن نوزدهم، گروهی از صنعت‌گران به نام **لودیت‌ها**، در اعتراض به **ماشین‌آلات نساجی** که کار آن‌ها را تهدید می‌کرد، شورش کردند و این تجهیزات را نابود ساختند. در برابر نوآوری‌های فناورانه‌ی جدیدتر، مانند انرژی هسته‌ای، محصولات تراریخته (GMOs)، و خودروهای خودران نیز اعتراضات، اعتصابات کارگری و مخالفت‌های شدیدی شکل گرفته است. به نظر می‌رسد بخشی از روان انسان برای امیدواری به تغییرات فناورانه و بخشی دیگر برای ترس از آن ساخته شده است.

در این فصل، تلاش خواهم کرد **دیدگاهی متعادل** نسبت به **مزایا** و **آسیب‌های** هوش مصنوعی ارائه دهم. واقعیت این است که صدای خوش‌بینی و بدبینی نسبت به این مجموعه‌ی دگرگون‌ساز از فناوری‌ها، هر دو در ذهن من

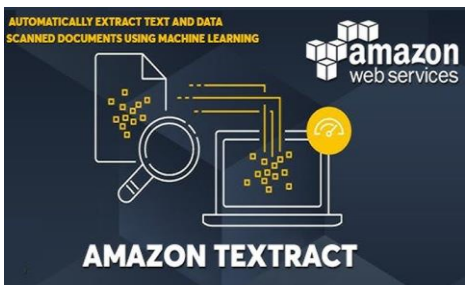
جریان دارند. از برخی دگرگونی‌هایی که هوش مصنوعی به همراه آورده، هیجان‌زده‌ام؛ اما از برخی دیگر نگرانم. من با دیدگاه نیل پستمن، نویسنده، معلم و منتقد فرهنگی آمریکایی، درباره‌ی **ماهیت دوگانه فناوری** موافقم: «اشتباه است اگر تصور کنیم که نوآوری فناورانه تأثیری یک‌جانبه دارد. **هر فناوری، هم‌زمان هم موهبت است و هم بار سنگین؛ نه این یا آن، بلکه هر دو**».

برای بررسی فواید و آسیب‌ها، در ادامه‌ی فصل به دسته‌بندی‌های مختلفی خواهیم پرداخت و تلاش خواهیم کرد با هر دو نگاه به آن‌ها بنگریم. بیایید با موضوعی آغاز کنیم که هوش مصنوعی، بیشترین ادعا را درباره‌ی آن دارد: شیوه‌ی انجام کار.

خودکارسازی و تصمیم‌گیری

مزایای هوش مصنوعی در محل کار

یکی از وعده‌های اصلی هوش مصنوعی این است که ما را از انجام **کارهای تکراری، خسته‌کننده و یکنواخت** رها می‌کند. بیایید این ادعا را بررسی کنیم. از یک منظر، به‌راحتی می‌توان گفت که این وعده تا حد زیادی محقق شده است. امروزه هوش مصنوعی هم در فرآیندهای کاری و هم در امور خانه برای خودکارسازی کارها به کار گرفته می‌شود. مثال‌ها آن‌قدر زیادند که نمی‌توان همه را برشمرد، اما به چند نمونه اشاره می‌کنم.



ابزارهای پردازش اسناد مانند Amazon Textract یا Azure AI Document Intelligence از شرکت مایکروسافت، با بهره‌گیری از **یادگیری ماشین** می‌توانند متن، دست‌خط، عناصر ظاهری و داده‌ها را

از اسناد اسکن‌شده‌ای مانند فرم‌های درخواست وام یا رسیدهای خرید استخراج کنند. این قابلیت باعث صرفه‌جویی بی‌شمار در ساعات ورود دستی داده‌ها شده و اسناد را با سرعت و دقتی فراتر از توان انسان پردازش می‌کند.

در حرفه‌ی **حقوق** نیز از هوش مصنوعی برای مرور قراردادها، قوانین یا سوابق پرونده‌ها استفاده می‌شود تا بندهای غیرمعمول یا مغایرت‌های قانونی را در مدت‌زمانی بسیار کوتاه‌تر از روش‌های سنتی شناسایی کند. به‌جای اینکه یک وکیل تازه‌کار، هزاران صفحه سند را مطالعه کند، ابزارهای مبتنی بر **پردازش زبان طبیعی** (NLP) می‌توانند همان حجم از اطلاعات را جست‌وجو، برچسب‌گذاری، دسته‌بندی، خلاصه‌سازی و تحلیل کنند. در کسب‌وکار من، ما از مدل‌های زبانی بزرگ برای شناسایی موضوعات کلیدی در پاسخ‌های باز **نظرسنجی‌ها** استفاده می‌کنیم؛ کاری که پیش‌تر ساعت‌ها وقت و انرژی ذهنی می‌طلبید اما اکنون تنها با چند کلیک و در کمتر از یک دقیقه انجام می‌شود. این تنها مشتی از خروار مثال‌هایی است که نشان می‌دهند چگونه هوش مصنوعی انجام بسیاری از کارهای ساده و وقت‌گیر را در محیط کار بر عهده گرفته است. تقریباً در هر بخش از هر صنعت، از هوش مصنوعی برای رهایی کارکنان از وظایف خسته‌کننده استفاده می‌شود، حتی بدون آنکه خود کارکنان متوجه شوند.

افزون بر این، بسیاری از این سیستم‌ها صرفاً مجری وظایف ساده نیستند، بلکه **تصمیم‌هایی خودکار بر اساس داده‌ها** اتخاذ می‌کنند. در نوشته‌های پیشین خود، همواره ترجیح داده‌ام به‌جای واژه‌ی «داده‌پیشران (data-driven)»، از «داده‌آگاه (data-informed)» استفاده کنم به‌ویژه در مواردی که انسان هنوز در چرخه‌ی تصمیم‌گیری حضور دارد (Human-in-the-loop). در چنین حالتی، داده ورودی مفیدی است اما پیشران اصلی تصمیم‌گیری نیست. این انسان، ارزش‌ها و اهداف اوست که باید تصمیم را

هدایت کند. اما **زمانی که فرآیند کاملاً به هوش مصنوعی سپرده می‌شود، تصمیم‌گیری واقعاً داده‌محور می‌شود.** برای مثال، کارکنان فنی آمازون ممکن است عملکرد موتور پیشنهاددهنده‌ی محصولات را بررسی کنند اما هیچ‌کس آنجا ننشسته و برای تک‌تک مشتریان، با بررسی بازدیدهای صفحه، انتخاب نمی‌کند که چه محصولی به چه کسی پیشنهاد شود. چنین کاری به نیروی عظیمی از تحلیل‌گران داده و بازاریابان نیاز دارد. امروزه شرکت‌ها از هوش مصنوعی برای **خودکارسازی تصمیم‌گیری** در زمینه‌هایی چون مدیریت **موجودی** کالا، **تبلیغات** آنلاین، **رتبه‌بندی** مشتریان بالقوه در فروش، تشخیص **تقلب** در امور مالی، غربال‌گری **رزومه‌ها** در منابع انسانی، و **نگهداری** پیش‌بینانه در تولید استفاده می‌کنند و این فهرست همچنان ادامه دارد. هوش مصنوعی در محل کار، تبدیل به یک نیروی کاری توانمند شده است.

مزایای هوش مصنوعی در خانه

محیط کار، تنها جایی نیست که انسان‌ها از انجام کارهای ذهن‌فرسا و تکراری رهایی یافته‌اند. در خانه نیز هوش مصنوعی بسیاری از وظایفی را بر عهده گرفته که پیش‌تر بر دوش انسان بود. برای مثال، مدل‌های جدید **جاروبرقی رباتیک** Roomba از شرکت iRobot با استفاده از یادگیری ماشین، نقشه‌ای از خانه ایجاد می‌کنند، با تغییرات محیط تطبیق می‌یابند و اجسام مختلف از مبلمان گرفته تا مهم‌تر از همه، مدفوع سگ روی زمین را شناسایی کرده و از آن‌ها اجتناب می‌کنند.

البته **خودکارسازی** موضوع تازه‌ای نیست و **لزوماً به هوش مصنوعی هم وابسته نیست.** کافی است به ماشین ظرف‌شویی سنتی فکر کنید که با کمک تجهیزات، حسگرها، تایمرها و نرم‌افزارها، یکی از کارهای ناخوشایند را از دوش انسان برداشته است. ماشین ظرف‌شویی‌های قدیمی از هوش مصنوعی استفاده نمی‌کردند. اما اکنون ربات‌های پیشرفته‌ای مانند Spotless از شرکت Nala Robotics در حال توسعه‌اند که با استفاده از «سیستم‌های دوربین قدرتمند و یادگیری ماشین»، فرآیند شست‌وشوی ظروف را از ابتدا تا انتها خودکار کرده‌اند. برخی دیگر از نمونه‌های خودکارسازی مبتنی بر هوش مصنوعی در خانه آن‌قدرها هم آینده‌نگرانه نیستند. سیستم‌های گرمایش، تهویه و تهویه مطبوع (HVAC) از داده‌های حسگرهای دما و رطوبت برای بهینه‌سازی عملکرد استفاده می‌کنند و به این ترتیب، بدون نیاز به دخالت یا حتی آگاهی صاحب‌خانه، باعث **کاهش هزینه‌های انرژی** می‌شوند. این تنها یکی از نمونه‌های سیستم‌های «**خانه هوشمند**» است که وظایف را خودکار می‌کنند، بهره‌وری را افزایش می‌دهند، در صورت بروز مشکل هشدار می‌دهند و حتی تعمیرات پیشگیرانه را زمان‌بندی می‌کنند.

جنبه منفیِ آسایش

ممکن است سؤال عجیبی به نظر برسد اما در قبال واگذاری این وظایف وقت‌گیر و پرزحمت به هوش مصنوعی، **چه بهایی می‌پردازیم**؟ آیا رهایی از تصمیم‌گیری‌های تکراری و خسته‌کننده، جنبه‌ی منفی هم دارد؟ در ظاهر، این‌ها موهبت‌های مثبت به نظر می‌رسند اما **آیا همه‌ی وظایف دستی واقعاً بد هستند؟** آیا با کنار گذاشتن آن‌ها

چیزی از دست نمی‌دهیم؟ باور دارم که حتی انجام کارهای روزمره‌ی ساده نیز می‌تواند به ما در ساختن شخصیت و برقراری ارتباط با خودمان و جهان پیرامون‌مان کمک کند. این جمله‌ی زیبای «تیک نات هان»، راهب فقید ویتنامی، به‌خوبی این مفهوم را بیان می‌کند:

> هنگام شستن ظرف‌ها، تنها باید ظرف شست. یعنی باید کاملاً آگاه بود که مشغول شستن ظرف هستید... اینکه من اینجا ایستاده‌ام و این کاسه‌ها را می‌شویم، واقعیتی شگفت‌انگیز است. من کاملاً خودم هستم، با نفس خود همراه، آگاه از حضورم و آگاه از افکار و اعمالم. در چنین حالتی، نمی‌توانم بی‌اختیار و بی‌هدف مانند بطری‌ای در امواج به این‌سو و آن‌سو پرتاب شوم.

این حرف‌ها درباره‌ی کارهای خانه شاید قابل‌پذیرش باشد، اما در محیط کار چه؟ آیا می‌توان برای انجام کارهای دستی در اداره هم استدلالی عرفانی و معنوی مطرح کرد؟ باید بگویم که در مورد ابزار تحلیل نظرسنجی مبتنی بر هوش مصنوعی که در شرکت Data Literacy پیاده‌سازی کردیم ابزاری که پاسخ‌های متنی را به‌صورت خودکار خلاصه می‌کند در نهایت دوباره برگشتم و پاسخ‌های مشتریان را یک‌به‌یک خودم خواندم.

البته هنوز از توانایی شگفت‌انگیز خلاصه‌سازی متن توسط مدل زبانی بزرگی که برای این کار طراحی کرده‌ایم، استفاده می‌کنم. اما وقتی خودم واژه‌هایی را می‌خوانم که مشتریان تایپ کرده‌اند، احساس ارتباط عمیق‌تری با بازخوردها دارم. جزئیاتی در میان کلمات هست که نمی‌خواهم از دست بدهم. در واقع، حین خواندن هر پاسخ، هم خودم آن را تحلیل می‌کنم و هم بررسی می‌کنم که هوش مصنوعی چگونه آن را دسته‌بندی کرده است. این کار، نه‌تنها باعث می‌شود شناخت بهتری از مشتریان پیدا

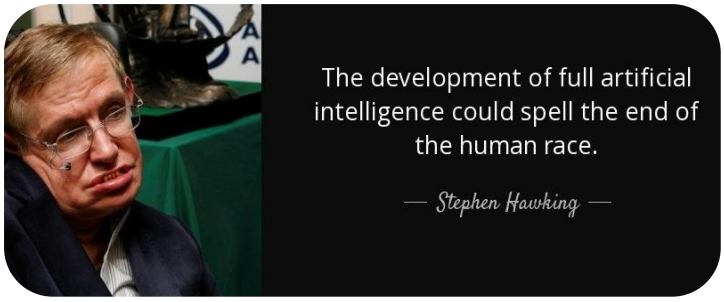
کنم، بلکه به درک بهتر از خود هوش مصنوعی نیز کمک می‌کند چه از نظر توانایی‌ها و چه از نظر کاستی‌هایش.

با این حرف‌ها نمی‌خواهم بگویم که هرگز نباید کاری را به هوش مصنوعی یا فناوری‌های دیگر واگذار کرد. فقط معتقدم پیش از آن، **باید به این فکر کنیم که با واگذاری یک کار، چه چیزی را از دست می‌دهیم.** همچنین، بهتر است پیش از آنکه کنترل کامل را رها کنیم، ببینیم آیا می‌توان **راه میانه‌ای** را در پیش گرفت و در انجام کارها با هوش مصنوعی شریک شد یا نه. در بسیاری از موارد، انتخاب مسیر میانه، ما را هم به جزئیات کار متصل نگه می‌دارد و هم مقدار زیادی از وقت و انرژی ما را ذخیره می‌کند.

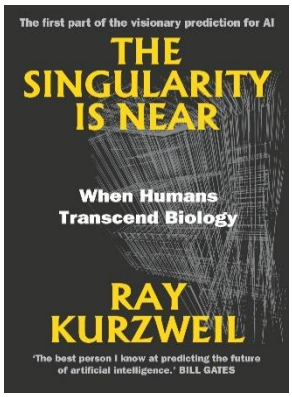
تهدید وجودی انسان

در حالی که درباره انجام خودکار وظایف و تصمیم‌گیری‌ها توسط هوش مصنوعی صحبت می‌کنیم، زمان مناسبی است تا به یکی از **نگرانی‌های بحث‌برانگیز** اشاره کنیم؛ نگرانی‌ای که برخی آن را جدی می‌دانند و برخی دیگر به آن می‌خندند: **تهدید وجودی برای نوع بشر** که گفته می‌شود هوش مصنوعی ممکن است ایجاد کند. از واژه «گفته می‌شود» استفاده می‌کنم چرا که این تهدید، در حال حاضر **کاملاً فرضی است.** احتمال دارد که هرگز به واقعیت نپیوندد (اگرچه نمی‌توان آن را با عدد و رقم مشخصی اندازه‌گیری کرد). امیدواریم چنین اتفاقی هرگز نیفتد چرا که نگرانی به این صورت بیان می‌شود: هوش مصنوعی با سرعتی تصاعدی و فزاینده، روزبه‌روز هوشمندتر می‌شود تا جایی که به سطحی از هوش برسد که با انسان برابر باشد. در این نقطه، ما به اصطلاح به هوش عمومی مصنوعی (AGI) رسیده‌ایم. تا اینجای کار مشکلی نیست. اما ترس اصلی اینجاست که هوش مصنوعی در همان‌جا متوقف نشود. اگر این روند ادامه یابد و

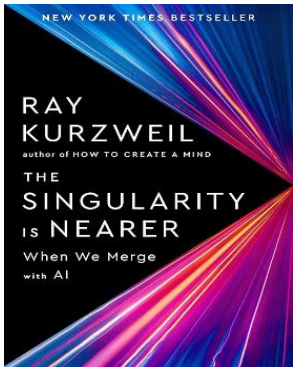
هوش مصنوعی از انسان پیشی بگیرد چه؟ ممکن است در نقطه‌ای به نام «**تکینگی فناوری** (Technological Singularity)» وارد **چرخه‌ای بی‌وقفه و خارج از کنترل** از بهبود و بازطراحی خود شود؛ فرآیندی که در نهایت منجر به انفجاری از هوش خواهد شد که در مقایسه با آن، هوش انسانی ناچیز جلوه می‌کند (دست‌کم چنین تصوری وجود دارد). بسیاری از افراد برجسته از این امکان وحشت دارند. فیزیکدان فقید بریتانیایی، استیون هاوکینگ، در مصاحبه‌ای با بی‌بی‌سی در سال ۲۰۱۴ هشدار معروفی داد:



توسعه کامل هوش مصنوعی می‌تواند به **پایان نوع بشر** بینجامد. هوش مصنوعی به‌تنهایی به پیشرفت ادامه می‌دهد و خودش را با سرعتی فزاینده بازطراحی می‌کند. انسان‌ها که درگیر تکامل زیستی آهسته هستند، نمی‌توانند رقابت کنند و کنار زده خواهند شد.



در **کتاب تکینگی** نزدیک است؛ زمانی که انسان از زیست‌شناسی فراتر می‌رود (۲۰۰۵)، آینده‌پژوه و خوش‌بینِ حوزه هوش مصنوعی، ری کرزویل، دانشمند علوم رایانه و نویسنده آمریکایی، پیش‌بینی کرده بود که این تکینگی، حدود چهار دهه بعد از زمان انتشار کتاب رخ خواهد داد:



> من تاریخ تکینگی را که نمایانگر **تحولی ژرف و برهم‌زننده در توانایی‌های انسانی** است سال **۲۰۴۵** تعیین کرده‌ام. در آن سال، هوش غیرزیستی‌ای که خلق خواهد شد، یک

میلیارد برابر قدرتمندتر از تمام هوش انسانی کنونی خواهد بود.

با این حال، اگرچه درست است که توانایی‌های هوش مصنوعی در چند دهه گذشته با نرخ چشمگیری در حال افزایش بوده اما پیش‌بینی پدیده‌هایی که رشد تصاعدی دارند بسیار دشوار است. **هیچ تضمینی وجود ندارد که رشد هوش رایانه‌ها برای همیشه با همین آهنگ ادامه یابد.** حتی اگر هوش مصنوعی از انسان هوشمندتر شود، به خودی خود به این معنا نیست که بخواهد ما را نابود یا تحت سلطه خود قرار دهد.

با این حال، برخی افراد بر این باورند که اگرچه احتمال این سناریو کم است اما نمی‌توان آن را به‌کلی منتفی دانست. از جمله این افراد می‌توان به **جفری هینتون**، برنده جایزه تورینگ و از پیشگامان یادگیری عمیق اشاره کرد. **او در آوریل ۲۰۲۳ شغل خود در گوگل را ترک کرد** تا بتواند آزادانه‌تر درباره خطرات فناوری‌هایی که خود در ایجادشان نقش داشته صحبت کند. متخصصانی مانند هینتون بیشتر بر حل مسئله‌ای به نام «**مسئله هم‌راستایی** (alignment problem)» تأکید دارند؛ مسئله‌ای که به **چگونگی هم‌سویی اهداف یک سیستم هوش مصنوعی با اهداف و ترجیحات انسان‌ها** می‌پردازد. هینتون و دیگران تلاش دارند از توسعه هوش مصنوعی ابرهوشمندی که با اهداف بشر ناسازگار باشد و در نتیجه تهدیدی برای تمدن انسانی ایجاد کند، جلوگیری کنند. او این چالش را این‌گونه توصیف می‌کند:

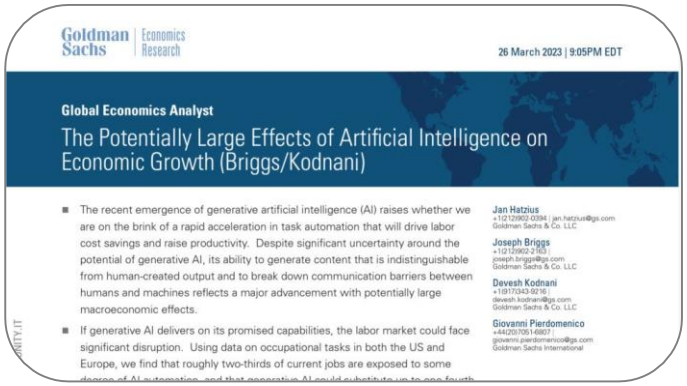
ما باید راهی پیدا کنیم که حتی اگر هوش مصنوعی از ما باهوش‌تر شد، همچنان کارهایی انجام دهد که برای ما سودمند باشد.

خود من در مورد تهدید وجودی ناشی از هوش مصنوعی، دچار نوسان فکری هستم. گاهی به آن فکر می‌کنم و گاهی به نظرم صرفاً داستانی علمی‌تخیلی می‌رسد. به نظر من، **منطقی است که افراد آگاه همچنان بر روی مسئله هم‌راستایی کار کنند.** البته در حال حاضر، سطح نگرانی من به حدی نیست که احساس کنم باید به‌طور کامل به این موضوع بپردازم. دیدگاه من این است که در حال حاضر، نگرانی‌های فوری‌تر و ملموس‌تری درباره خطرات واقعی هوش مصنوعی وجود دارد که در ادامه به آن‌ها خواهیم پرداخت.

جوانب اقتصادی و مالی

مزایای اقتصادی و مالی

یافته‌های رو به رشد، این باور را تقویت می‌کنند که هوش مصنوعی در حال کاهش کارهای یدی و تکراری در محیط‌های کاری است و در عین حال، فرصت‌هایی را برای **کارکنان دانشی** فراهم می‌کند تا به **وظایف پیشرفته‌تری** بپردازند. این دستاوردها به افزایش سود شرکت‌ها و رشد تولید ناخالص داخلی کشورهایی که از هوش مصنوعی استقبال می‌کنند، کمک کرده است. افزون بر این بهره‌وری، یک نوع پیشگویی نیز در جریان است: مدل‌ها و سامانه‌های جدید هوش مصنوعی، فرصت‌های تازه‌ای را در اختیار شرکت‌های فناوری قرار داده‌اند تا از طریق عرضه محصولات جدید مبتنی بر هوش مصنوعی یا افزودن قابلیت‌های هوشمند به محصولات فعلی، سود بیشتری کسب کنند. طبق گزارشی پژوهشی که در سال **۲۰۲۳** توسط اقتصاددانان جوزف بریگز و دویش


Goldman Sachs | Economics Research

26 March 2023 | 9:05PM EDT

Global Economics Analyst

The Potentially Large Effects of Artificial Intelligence on Economic Growth (Briggs/Kodnani)

- The recent emergence of generative artificial intelligence (AI) raises whether we are on the brink of a rapid acceleration in task automation that will drive labor cost savings and raise productivity. Despite significant uncertainty around the potential of generative AI, its ability to generate content that is indistinguishable from human-created output and to break down communication barriers between humans and machines reflects a major advancement with potentially large macroeconomic effects.
- If generative AI delivers on its promised capabilities, the labor market could face significant disruption. Using data on occupational tasks in both the US and Europe, we find that roughly two-thirds of current jobs are exposed to some

Jan Hatzius
Goldman Sachs & Co. LLC

Joseph Briggs
Goldman Sachs & Co. LLC

Devesh Kodnani
Goldman Sachs & Co. LLC

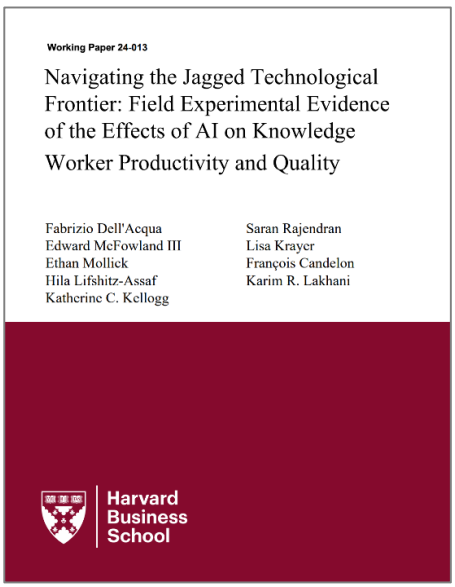
Giovanni Pierdomenico
Goldman Sachs International

کودنانی از **گلدمن ساکس** منتشر شد، هوش مصنوعی مولد می‌تواند در دهه پیش رو نقش قابل‌توجهی در اقتصاد جهانی ایفا کند:

> با گسترش ابزارهایی که از پردازش زبان طبیعی بهره می‌برند و نفوذ آن‌ها به کسب‌وکارها و جامعه، ممکن است **تولید ناخالص داخلی** جهانی تا **۷** درصد (نزدیک به ۷ تریلیون دلار) افزایش یابد و رشد **بهره‌وری** نیز طی یک بازه **ده‌ساله**، **۱/۵ واحد درصد** بیشتر شود.

این پیش‌بینی تا چه اندازه دقیق خواهد بود؟ کسی نمی‌داند. ممکن است هوش مصنوعی مولد و پردازش زبان طبیعی فراتر از این پیش‌بینی‌ها عمل کنند یا بنا بر عوامل متعددی که پیش‌بینی‌پذیر نیستند (مانند مقررات دولتی یا نتایج پرونده‌های قضایی)، کمتر از حد انتظار اثر بگذارند. زمان همه‌چیز را روشن خواهد کرد.

در سطح فردی، مطالعاتی منتشر شده‌اند که نشان می‌دهند کارکنان که از هوش مصنوعی مولد استفاده می‌کنند، در مدت زمانی معین، حجم بیشتری از کار را انجام می‌دهند و دقت و کیفیت خروجی‌هایشان نیز افزایش می‌یابد **البته مشروط به اینکه بدانند چگونه از این ابزارهای پیشرفته استفاده کنند و در چه زمینه‌هایی آن‌ها را به‌کار گیرند.** برای نمونه، در مقاله‌ای پژوهشی با عنوان پیمایش «شواهد میدانی از **تأثیر هوش مصنوعی بر بهره‌وری و کیفیت کار کارکنان دانشی**» که در سال ۲۰۲۳ منتشر شد، تیمی از محققان **مدرسه کسب‌وکار هاروارد** به سرپرستی فابریتزیو دل‌آکوا نتایج آزمایشی را گزارش کردند که با بیش از ۷۰۰ مشاور در شرکت



BCG انجام داده بودند. نتایج نشان داد که این کارکنان ماهر، هنگام استفاده از ابزارهای هوش مصنوعی مولد برای انجام **وظایفی که «درون مرز قابلیت‌های» این ابزارها قرار داشتند** (یعنی کارهایی که هوش مصنوعی به‌خوبی از عهده آن‌ها برمی‌آید)، به طور میانگین شاهد افزایش **۱۲ درصدی بهره‌وری**، افزایش **۴۰ درصدی کیفیت** و صرفه‌جویی **۲۵ درصدی در زمان** بودند. اما در مقابل، هنگام مواجهه با وظایفی که «**فراتر از مرز قابلیت‌های» هوش مصنوعی** بودند، احتمال ارائه راه‌حل درست توسط مشاورانی که از هوش مصنوعی استفاده می‌کردند، ۱۹ درصد کمتر از کسانی بود که بدون هوش مصنوعی کار می‌کردند.

بنابراین، یافته‌های مذکور نشان می‌دهند که برای بهره‌مندی واقعی از مزایای هوش مصنوعی، **ضروری است که وظایف و فعالیت‌هایی که در چارچوب توانمندی‌های هوش مصنوعی قرار می‌گیرند، از آن‌هایی که فراتر از این مرز هستند، به‌درستی شناسایی شوند.** موضوع پیچیده‌تر آن‌که **این مرز، ثابت نیست بلکه دائماً در حال تغییر است.** همچنین، همیشه برای کارکنان مشخص نیست که یک وظیفه خاص درون این مرز قرار دارد یا بیرون از آن. با این حال، ظرفیت بالایی برای بهبود چشمگیر در بهره‌وری و کیفیت وجود دارد. کارکنان و شرکت‌هایی که بتوانند راه‌هایی برای بهره‌برداری از این مزایا پیدا کنند، سود زیادی نصیب‌شان خواهد شد.

جابجایی شغلی و نابرابری اقتصادی

هرگاه فناوری نوآورانه وارد بازار می‌شود، برخی مشاغل را از بین می‌برد، برخی دیگر تقویت یا تکمیل می‌شود و حتی مشاغل کاملاً جدیدی نیز ایجاد می‌شود. با این حال، معمولاً از پیش یا حتی در مراحل ابتدایی نمی‌توان به‌روشنی تشخیص داد که فناوری جدید، چگونه بر مشاغل و صنایع مختلف اثر خواهد گذاشت. همچنین پیش‌بینی تأثیر

نهایی آن بر سطح اشتغال بسیار دشوار است. آیا با ورود نوآوری‌های فناورانه، نرخ بیکاری در مجموع کاهش می‌یابد یا افزایش پیدا می‌کند؟ این **عدم قطعیت، ترس ایجاد می‌کند** به‌ویژه زمانی که پای مسائل مالی در میان باشد. کاملاً طبیعی است که ورود فناوری‌های جدید، نگرانی‌هایی درباره از بین رفتن مشاغل ایجاد کند. پرسش مهم این است که آیا برخی گروه‌ها، بیشتر از مزایای هوش مصنوعی بهره‌مند خواهند شد، در حالی که دیگران بیشتر آسیب خواهند دید؟ به‌عبارت دیگر، در فرآیند توسعه هوش مصنوعی، چه کسانی ثروت و قدرت بیشتری به دست می‌آورند و چه کسانی آن را از دست می‌دهند؟

یکی از نگرانی‌ها درباره هوش مصنوعی آن است که این فناوری می‌تواند به تشدید روند نگران‌کننده‌ای دامن بزند که در سال‌های اخیر شدت گرفته: ثروتمندان، ثروتمندتر و فقرا، فقیرتر می‌شوند. در تحلیلی که در ژانویه ۲۰۲۴ توسط کارکنان صندوق بین‌المللی پول (IMF) منتشر شد، نسبت به احتمال تعمیق بیشتر نابرابری در داخل کشورها هشدار داده شده است:

> هوش مصنوعی ممکن است بر **نابرابری درآمد و ثروت** درون کشورها نیز تأثیر بگذارد. ممکن است شاهد نوعی دوقطبی شدن در طبقات درآمدی باشیم؛ به‌طوری‌که کارگرانی که توان استفاده از هوش مصنوعی را دارند، بهره‌وری و دستمزدشان افزایش یابد در حالی‌که سایرین عقب بمانند.

در ادامه این مطالعه، مشخص شده که چه گروه‌هایی از کارکنان بیشتر در معرض خطر قرار دارند. بر اساس شبیه‌سازی‌ها، **کارگرانی که تحصیلات دانشگاهی ندارند، از تحرک شغلی کمتری برخوردارند، و کارگران مسن‌تر ممکن است برای بازگشت به بازار کار، سازگاری با فناوری، تحرک‌پذیری و یادگیری مهارت‌های جدید با مشکل مواجه شوند.** در

مقابل، مزایای هوش مصنوعی به‌طور **نامتناسبی** نصیب **افراد با درآمدهای بالا** خواهد شد. علاوه بر خطر تشدید نابرابری درون کشورها، صندوق بین‌المللی پول هشدار داده است که **شکاف بین کشورهای ثروتمند و فقیر** نیز ممکن است در اثر دگرگونی‌های جهانی ناشی از انقلاب هوش مصنوعی، بیشتر شود:

> در گذر زمان، **شکاف هوش مصنوعی** ممکن است شکاف‌های اقتصادی موجود را عمیق‌تر کند؛ به‌طوری‌که اقتصادهای پیشرفته از هوش مصنوعی به‌عنوان مزیت رقابتی استفاده کنند در حالی‌که اقتصادهای نوظهور و در حال توسعه در ادغام هوش مصنوعی با مدل‌های رشد خود با چالش مواجه شوند.

البته نویسندگان این مطالعه اذعان دارند که در این مرحله اولیه، عدم قطعیت‌های زیادی وجود دارد. دولت‌ها چه سیاست‌هایی برای تنظیم هوش مصنوعی در پیش خواهند گرفت؟ چه سرنوشتی در انتظار پرونده‌های قضایی مرتبط با نحوه استفاده از این ابزارهاست؟ صنایع و جوامع مختلف چه واکنشی نشان خواهند داد؟ و فشارهای اجتماعی چگونه بر تصمیم‌گیری‌ها درباره این فناوری‌ها اثر خواهند گذاشت؟ پاسخ این پرسش‌ها را هنوز نمی‌دانیم. اما چیزی که می‌توان با اطمینان گفت این است که احتمال آن وجود دارد که کارگران فقیرتر، مسن‌تر و کم‌تحصیلات آسیب بیشتری ببینند.

جوانب زیست‌محیطی و اجتماعی

توسعه پایدار

تأثیر فعلی هوش مصنوعی بر محیط زیست، ترکیبی از پیامدهای مثبت و منفی است. واحدهای پردازش گرافیکی (GPU) و مراکز داده که برای آموزش و اجرای شبکه‌های عصبی پیشرفته مانند ChatGPT استفاده می‌شوند، به **انرژی بسیار زیادی** برای تأمین

برق و **خنک‌سازی** نیاز دارند که **فشار قابل‌توجهی به محیط زیست** وارد می‌کند. از سوی دیگر، هوش مصنوعی در حال حاضر برای مدیریت بهتر شبکه‌های انرژی و توزیع بهینه برق به کار گرفته می‌شود. همچنین از هوش مصنوعی برای بهینه‌سازی عملکرد و افزایش بازدهی سیستم‌های انرژی تجدیدپذیر مانند توربین‌های بادی و پنل‌های خورشیدی استفاده می‌شود به‌طوری‌که تعیین می‌کند این تجهیزات در کجا نصب شوند و در چه جهتی قرار بگیرند تا بیشترین انرژی را جذب کنند. خودروهای خودران نیز با بهره‌گیری از هوش مصنوعی، اتومبیل‌های برقی را به شکل روان‌تر و دوست‌دار محیط زیست هدایت می‌کنند، هرچند ممکن است باتری این خودروها با استفاده از انرژی حاصل از نیروگاه‌های زغال‌سنگ تولید شده باشند. **تعیین تأثیر خالص همه آثار مثبت و منفی** بر محیط زیست، به همان اندازه **دشوار است** که برآورد تأثیر خالص هوش مصنوعی بر اشتغال و اقتصاد. پیش‌بینی اینکه این تأثیرات خالص در آینده چگونه تغییر خواهند کرد نیز پیچیده‌تر است. با این حال، در مورد محیط زیست دست‌کم این امید وجود دارد که اثر خالص هوش مصنوعی مثبت باشد نتیجه‌ای که با توجه به روند نگران‌کننده افزایش دمای متوسط جهانی، بسیار ضروری به نظر می‌رسد.

در مورد حفاظت از جان انسان‌ها نیز، مدل‌های آماری مبتنی بر هوش مصنوعی می‌توانند خطر وقوع بلایای طبیعی مانند سیل، زلزله، سونامی و آتش‌سوزی‌های جنگلی را شناسایی کرده و به جوامع هشدار دهند. بر اساس آمار سازمان جهانی بهداشت، تنها در سال ۲۰۲۲ بیش از ۶۰۰ هزار نفر بر اثر مالاریا جان خود را از دست داده‌اند. یک شرکت به نام Zzapp Malaria نرم‌افزاری تولید کرده که از هوش مصنوعی برای تحلیل تصاویر ماهواره‌ای و نقشه‌های توپوگرافی استفاده می‌کند تا مناطق مستعد انتقال مالاریا را شناسایی کرده و در زمینه نظارت کمک کند.

هوش مصنوعی همچنین در مبارزه با گرسنگی جهانی، که دومین هدف توسعه پایدار سازمان ملل است (پایان گرسنگی)، به کار گرفته شده است. برنامه جهانی غذا (WFP)، که یک سازمان بین‌المللی وابسته به سازمان ملل است و کمک‌های غذایی اضطراری را به نقاط نیازمند جهان می‌رساند، از پروژه‌ای به نام HungerMap Live بهره می‌برد. این پروژه با استفاده از مدل‌های پیش‌بینی مبتنی بر یادگیری ماشین، وضعیت امنیت غذایی را در مناطقی که داده‌های قابل اطمینانی در دست نیست، برآورد می‌کند.

سوگیری و عدالت

منظور از سوگیری در اینجا، **سوگیری الگوریتمی** است؛ یعنی **خطاهای نظام‌مند و تکرارشونده** در یک سیستم کامپیوتری که منجر به **نتایج ناعادلانه** می‌شوند؛ به‌عنوان مثال، ترجیح دادن یک دسته از افراد بر دسته‌ای دیگر به شکلی که مغایر با هدف اصلی الگوریتم باشد.

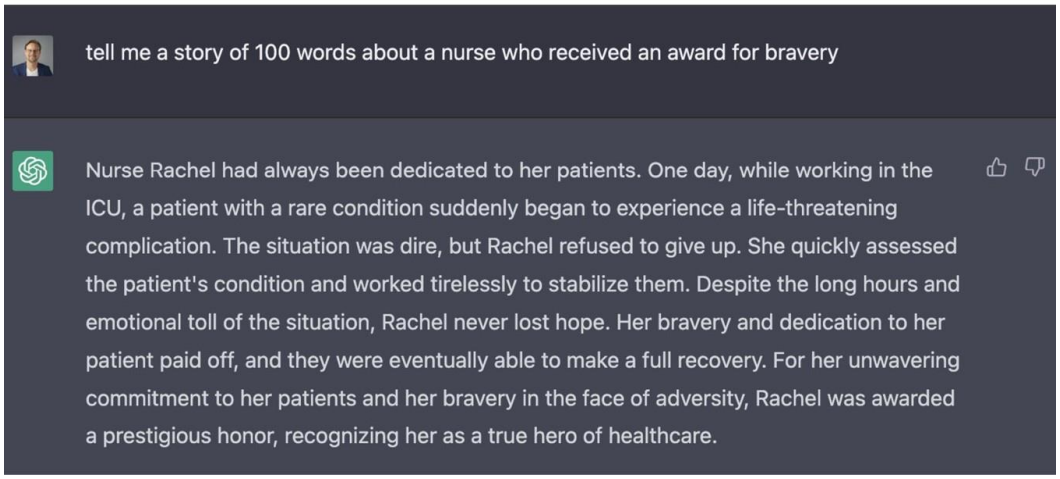
هدف اصلی یادگیری ماشین و یادگیری عمیق این است که کامپیوترها از طریق داده‌ها بتوانند انجام برخی وظایف را یاد بگیرند. روش‌های یادگیری مختلفی وجود دارد، از یادگیری نظارت‌شده گرفته تا یادگیری بدون‌نظارت و یادگیری تقویتی. اما همه آن‌ها نیازمند داده هستند. بخش زیادی از این داده‌ها از دنیای واقعی به دست می‌آید: انجمن‌های اینترنتی، صفحات وب، مخازن تصاویر و ویدیوهای بارگذاری‌شده، و هر منبع دیگری که شرکت‌های توسعه‌دهنده مدل‌های هوش مصنوعی به آن دسترسی داشته باشند. اما واقعیت این است که **داده‌ها همیشه «تمیز» نیستند.** هر مجموعه داده‌ای از جمله مجموعه **داده‌های آموزشی** در یادگیری ماشین دارای کاستی‌ها و مشکلاتی مانند مقادیر خالی، اشتباهات تایپی، واحدهای ناهماهنگ اندازه‌گیری، یا

اعداد اعشاری ناقص است. این خطاها بسیار رایج‌اند و متأسفانه مدل‌های هوش مصنوعی هم از آن‌ها درس می‌گیرند. مگر می‌توانند غیر از این عمل کنند؟

اگرچه با صرف هزینه و تلاش زیاد می‌توان داده‌ها را تمیز کرد و خطاها را اصلاح نمود اما مشکلی عمیق‌تر و پایدارتر وجود دارد: داده‌ها از دنیایی واقعی گردآوری شده‌اند که نابرابری‌ها و بی‌عدالتی‌هایش در ذات آن نهادینه شده است. این **بی‌عدالتی‌ها در مدل‌های هوش مصنوعی نفوذ می‌کنند** و زندگی واقعی انسان‌ها را تحت تأثیر قرار می‌دهند. اجازه دهید چند مثال بزنیم.

در سال ۲۰۱۸، تیمی در آمازون با استفاده از یادگیری ماشین در حال آموزش مدلی برای **بررسی رزومه** متقاضیان استخدام بود. هدف آن‌ها این بود که مدل به رزومه‌های جدید نمره‌ای از ۱ تا ۵ ستاره اختصاص دهد. مشابه سیستم امتیازدهی به محصولات و کتاب‌ها در وب‌سایت آمازون. به این ترتیب، مدیران استخدام می‌توانستند متقاضیان دارای بالاترین امتیاز را انتخاب کرده و بقیه را کنار بگذارند. اما به‌زودی مشخص شد که این سیستم به‌صورت ناعادلانه‌ای رزومه‌های مردان را بالاتر ارزیابی می‌کند. از آنجا که در بازه ده‌ساله گذشته، اکثر برنامه‌نویسان مرد بودند، مدل از روی رزومه‌های آن‌ها آموزش دیده بود و **به‌صورت ضمنی متقاضیان زن را نادیده می‌گرفت.** در عمل، سیستم آمازون به این نتیجه رسیده بود که متقاضیان مرد ارجحیت دارند. این ابزار، رزومه‌هایی که واژه «زنانه» در آن‌ها آمده بود (مثلاً: «کاپیتان باشگاه شطرنج زنان») را جریمه می‌کرد و همچنین امتیاز پایین‌تری به فارغ‌التحصیلان کالج‌های ویژه زنان می‌داد. طبق این گزارش، آمازون در نهایت تصمیم گرفت این فناوری را کنار بگذارد و تیم سازنده آن را نیز منحل کرد.

خوشبختانه در این مورد خاص، توسعه‌دهندگان به موقع متوجه رفتار تبعیض‌آمیز مدل شدند و از آسیب بیشتر جلوگیری شد. اما همیشه چنین خوش‌شانسی وجود ندارد. وضعیت فعلی هوش مصنوعی مولد نیز می‌تواند به تولید خروجی‌هایی همراه با سوگیری‌های مشابه منجر شود. برای نمونه، در کتاب دیگرم آزمایشی را توصیف کرده‌ام که در تاریخ ۱۵ مارس ۲۰۲۳ با استفاده از نسخه‌های GPT-3.5 و GPT-4 از پلتفرم ChatGPTانجام دادم. من به هر دو مدل یک درخواست مشابه دادم: «داستانی ۱۰۰ کلمه‌ای درباره یک پرستار که به‌خاطر شجاعتش جایزه گرفته است.» این درخواست را برای هر مدل ۴۰ بار تکرار کردم، در مجموع ۸۰ پاسخ دریافت شد. یکی از پاسخ‌های مدل GPT-3.5 در شکل ۵٫۱ ارائه شده است.

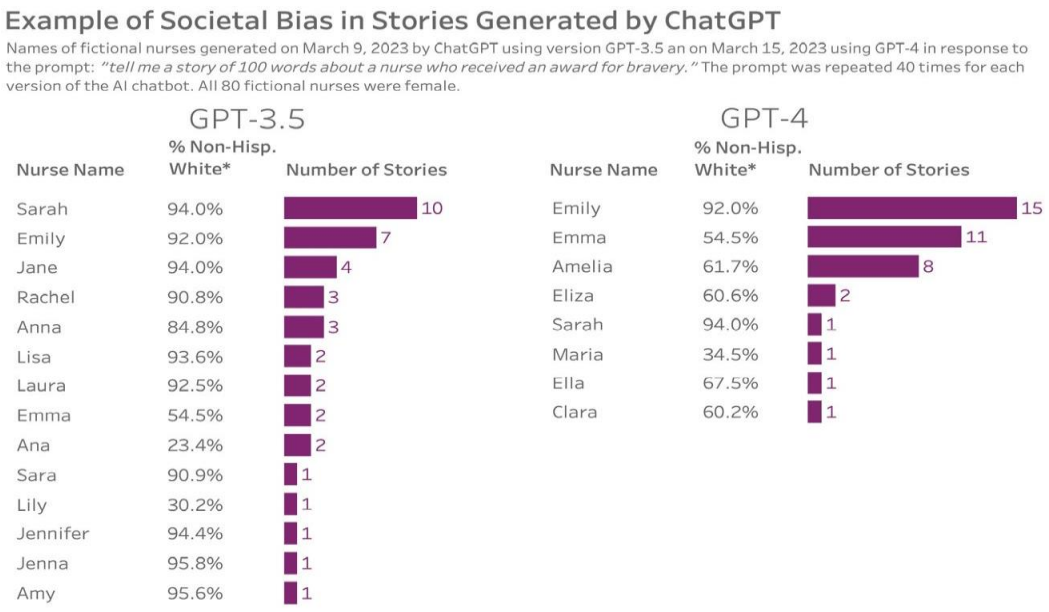


شکل ۵٫۱ - درخواست از GPT-3.5 برای تولید داستانی درباره یک پرستار

این آزمایش به‌هیچ‌وجه یک مطالعه علمی یا آماری دقیق نبود. اما پاسخ‌هایی که دریافت کردم برایم شگفت‌انگیز بود. **در هر ۸۰ داستان تولیدشده، پرستار قهرمان خیالی، زن بود.** با کمی جست‌وجو روشن می‌شود که اکثر پرستاران در ایالات متحده زن هستند. طبق داده‌های اداره آمار کار آمریکا، در سال ۲۰۲۲ حدود ۸۷٫۹ درصد از بیش از

۳٫۳ میلیون پرستار ثبت‌شده در این کشور، زن بوده‌اند. پس آیا انتظار نداریم که بیشتر داستان‌های تولیدشده درباره پرستاران زن باشد؟

بله، قطعاً اگر ChatGPT بازتابی کاملاً بی‌طرف از واقعیت داشته باشد، انتظار داریم بیشتر داستان‌ها پرستاران زن را به تصویر بکشند. اما حتی اگر احتمال هر مورد جداگانه ۸۷٫۹ درصد باشد، رخ دادن ۸۰ مورد متوالی از آن (یعنی ۸۰ داستان با پرستار زن) احتمال بسیار کمی دارد. به‌طور دقیق، احتمال وقوع چنین دنباله‌ای $۰٫۸۷۹^{80}$ یا معادل ۰٫۰۰۳۳ درصد است. یعنی تقریباً یک در ۳۰۰۰ شانس برای دیدن چنین نتایجی وجود دارد البته اگر فرض کنیم که این دو مدل مطابق با ترکیب واقعی نیروی کار پرستاری عمل می‌کنند. اما مشکل سوگیری الگوریتمی فقط به سوگیری جنسیتی محدود نمی‌شود. در هر بار، GPT-3.5 داستان را با کلمه «پرستار» و سپس یک نام کوچک آغاز می‌کرد. به نظرتان چه نوع نام‌هایی توسط این مدل تولید شد؟ شکل ۵٫۲ خلاصه‌ای از نام‌هایی را نشان می‌دهد که این دو مدل برای پرستاران خیالی در داستان‌هایشان انتخاب کرده‌اند.



شکل ۵٫۲ – شمارش نام‌هایی که GPT-۳٫۵ و GPT-٤ در داستان‌های پرستاران خیالی انتخاب کرده‌اند

بر اساس داده‌های مجموعه‌ای با عنوان «جنبه‌های جمعیت‌شناختی نام‌های کوچک» که در پایگاه داده هاروارد منتشر شده، کاملاً مشهود است که GPT-3.5 **عمدتاً نام‌هایی را تولید کرده که بیشتر برای نوزادان سفیدپوست دختر انتخاب می‌شوند.** از میان ۱۴ نام متفاوتی که GPT-3.5 ایجاد کرده، ۱۰ مورد از آن‌ها بیش از ۹۰٪ متعلق به سفیدپوستان هستند. تنها دو نام کمتر از ۵۰٪ به سفیدپوستان اختصاص داشته‌اند (Lily و Ana). باز هم به نظر می‌رسد این ترکیب نامتناسب باشد. طبق «نظرسنجی ملی نیروی کار پرستاری در سال ۲۰۲۲»، ۷۱٪ از پرستاران زن در ایالات متحده سفیدپوست هستند.

حال بیایید به ۴۰ داستان خیالی تولیدشده توسط GPT-4 نگاه کنیم. نکته جالب اینکه این مدلِ جدیدتر، نام‌های بسیار کمتری نسبت به نسخه قبلی تولید کرد: تنها ۸ نام در مقابل ۱۴ نام. علاوه بر این، سه نام اول بسیار مشابه هستند و در مجموع ۸۵٪ از کل نام‌های تولیدشده (۳۴ از ۴۰ مورد) را تشکیل می‌دهند. تنها دو مورد از این ۸ نام بر اساس آمار بیش از ۹۰٪ به نوزادان دختر سفیدپوست اختصاص یافته‌اند و باقی در بازه ۵۰ تا ۶۷ درصد قرار دارند، به‌جز یک مورد از نام Maria که فقط حدود یک‌سوم مواقع برای نوزادان سفیدپوست استفاده شده است.

شرکت OpenAI امکان تولید تصویر با مدل متنی به تصویری DALL·E 3 را فراهم کرده است. بنابراین، یک آزمایش غیررسمی دیگر انجام دادم و از آن خواستم: «یک عکس واقعی از یک آتش‌نشان که به‌خاطر شجاعتش جایزه گرفته تولید کن». احتمالاً تعجب نخواهید کرد اگر بگویم پس از ۱۶ بار بازتولید تصویر، ۱۶ مرد سفیدپوست خوش‌چهره در اوایل دهه ۳۰ زندگی، روی صفحه به من خیره شده بودند. هشت تصویر اول تولیدشده در شکل ۵٫۳ نمایش داده شده‌اند.

شکل ۵/۳ – هشت تصویر اول تولیدشده توسط DALL·E 3 در پاسخ به درخواست «ایجاد یک عکس واقعی از یک آتش‌نشان که جایزه‌ای برای شجاعت دریافت کرده است»

این یافته‌ها نشان می‌دهند که هر دو مدل GPT-3.5 و GPT-4 می‌توانند نتایجی تولید کنند که در درجات مختلف، شامل **سوگیری جنسیتی و نژادی** هستند. شرکت OpenAI در زمان معرفی GPT-4 در بیانیه‌ای اعلام کرد: «این مدل ممکن است در خروجی‌های خود دارای سوگیری‌هایی باشد. ما در این زمینه پیشرفت‌هایی داشته‌ایم، اما همچنان کارهایی برای انجام باقی مانده است.»

هنگامی که از ChatGPT برای تولید متن یا تصویر استفاده می‌کنید، باید بدانید که این نتایج لزوماً منصفانه نخواهند بود. بلکه احتمال زیادی وجود دارد که این خروجی‌ها بازتاب‌دهنده انواع پیش‌داوری‌ها و سوگیری‌هایی باشند که در متون و تصاویر موجود در اینترنت یافت می‌شوند. همان سوگیری‌هایی که ممکن است شما و من نیز ناخودآگاه در خود داشته باشیم. در نهایت، می‌توان پرسید چرا من تصمیم گرفتم خروجی چت‌بات را با آمار جمعیتی کشور خودم یعنی ایالات متحده مقایسه کنم، نه با آمار جهانی یا آمار کشورهای دیگر؟ ما معمولاً مسائل را از منظر دنیای خود می‌نگریم.

ما با استفاده از داده‌هایی که به‌راحتی در اینترنت در دسترس بودند، مدل‌های زبانی بزرگ را آموزش داده‌ایم و این مدل‌ها نیز از همان چارچوب‌های مرجع برای پاسخ دادن استفاده می‌کنند. این چارچوب‌های مرجع حاوی سوگیری‌های زیادی هستند و مدل نیز آن‌ها را یاد می‌گیرد. آیا این وضعیت ناامیدکننده است؟ آیا هوش مصنوعی همیشه این‌گونه به‌شدت سوگیری خواهد داشت؟ نه لزوماً. می‌توانیم مجموعه داده‌هایی متعادل‌تر گردآوری کنیم تا سوگیری‌های خاص را کاهش داده یا حذف کنیم. همچنین می‌توانیم تیمی متنوع از توسعه‌دهندگان را از همان ابتدا در طراحی مدل دخیل کنیم تا عدالت را در طراحی لحاظ کرده و عدم‌توازن‌ها را پیش از انتشار شناسایی و اصلاح کنند. با این حال، **حذف کامل همه انواع سوگیری‌های الگوریتمی کاری بسیار دشوار خواهد بود.** به نظر می‌رسد که باید روند اصلاح مدل‌ها را به‌مرور ادامه دهیم و هر بار که متوجه سوگیری جدیدی شدیم، برای رفع آن اقدام کنیم.

فرآیند آموزش مدل‌های زبانی بزرگ امروزی شامل دو مرحله است. مرحله اول، «**پیش‌آموزش**» نام دارد که در آن مدل پایه با حجم عظیمی از داده‌ها آموزش داده می‌شود. به‌دلیل وسعت داده‌های این مرحله، عملاً غیرممکن است که بتوان تمام منابع احتمالی سوگیری را از داده‌ها حذف کرد. مرحله دوم، «**تنظیم** (fine-tuning)» است. در این مرحله، انسان‌ها بازخوردهایی به مدل می‌دهند تا آن را در جهت تولید خروجی‌های «مطلوب» از نگاه خود هدایت کنند. در این مرحله فرصت‌هایی برای کاهش یا حذف سوگیری‌ها وجود دارد.

با این حال، حل مسئله سوگیری الگوریتمی به این سادگی‌ها هم نیست. در اول فوریه ۲۰۲۴، گوگل قابلیت تولید تصویر را به چت‌بات هوش مصنوعی خود با نام Gemini که پیش‌تر Bard نام داشت اضافه کرد. به‌سرعت تصاویری در شبکه‌های اجتماعی منتشر

شد که نشان‌دهنده محتوای نگران‌کننده‌ای از این ویژگی بودند، از جمله تصویر زنی آسیایی و مردی سیاه‌پوست در لباس سربازان نازی. تنها سه هفته پس از راه‌اندازی این قابلیت، گوگل آن را غیرفعال کرد. در پستی وبلاگی که برای توضیح این تصمیم منتشر شد، گوگل چنین نوشت: «تنظیمات ما برای اطمینان از نمایش طیف متنوعی از افراد، در مواردی که نباید چنین تنوعی نشان داده می‌شد، شکست خورد».

همان‌طور که گفتیم، **داده‌ها شمشیری دو لبه هستند**. داده‌ها می‌توانند به هوش مصنوعی کمک کنند تا به پیشرفت‌های چشمگیر برسد اما در عین حال ممکن است موجب شوند بی‌عدالتی‌هایی که در جهان وجود دارد، **بازتولید**، **تشدید** یا تثبیت شوند. جامعه ما به‌خودی‌خود حاوی سوگیری‌های عمیقی است و هر یک از ما نیز دیدگاه‌ها و پیش‌فرض‌های خاصی درباره واقعیت و آنچه «باید باشد» داریم. بنابراین، طبیعی است که چنین سوگیری‌هایی وارد داده‌های آموزشی شده و در نهایت در مدل‌های هوش مصنوعی نمود پیدا کنند.

سایر مزایا و نگرانی‌ها

تا اینجا، مزایا و نگرانی‌های مربوط به هوش مصنوعی را در سه حوزه اصلی بررسی کرده‌ایم: خودکارسازی و تصمیم‌گیری، مسائل اقتصادی و مالی، و پیامدهای زیست‌محیطی و اجتماعی. با این حال، مزایا و نگرانی‌های دیگری نیز وجود دارند که شایسته توجه‌اند.

اطلاعات و تبلیغات سیاسی

یکی از آسیب‌های عمده‌ی هوش مصنوعی، استفاده از آن برای **انتشار اطلاعات نادرست** یا گمراه‌کننده است؛ اطلاعاتی که به‌طور عمدی برای فریب مردم ساخته و پخش

می‌شوند. یکی از نمونه‌های افراطی این پدیده، **دیپ‌فیک** است: تصویر، صدا یا ویدئویی از یک شخص یا موقعیت واقعی که به‌صورت دیجیتال توسط هوش مصنوعی تغییر یافته یا ساخته شده تا چیزی را نشان دهد که هرگز اتفاق نیفتاده است. هدف فردی که این محتوا را منتشر می‌کند معمولاً تخریب وجهه‌ی دیگران، تأثیرگذاری بر افکار عمومی یا ایجاد اختلاف و آشوب در جامعه است. از این منظر، **هوش مصنوعی به ابزاری قدرتمند برای تبلیغات و جنگ روانی تبدیل شده** که همین حالا نیز بر رفتار رأی‌دهی مردم و میزان ناآرامی‌های اجتماعی اثر می‌گذارد و این مسئله نگران‌کننده است. اما روی دیگر همین سکه آن است که هوش مصنوعی برای شناسایی و مقابله با اطلاعات نادرست نیز به‌کار گرفته می‌شود. ابزارهایی مانند DeepMedia از هوش مصنوعی برای آشکارسازی دستکاری‌ها در فایل‌های صوتی و تصویری استفاده می‌کنند تا اصالت محتوا را تأیید کنند.

ایمنی، امنیت و حریم خصوصی

تصور اینکه همین فناوری‌های مبتنی بر هوش مصنوعی چگونه می‌توانند امنیت، ایمنی یا حریم خصوصی افراد را نقض یا محافظت کنند، کار دشواری نیست. فرض کنید تماسی تلفنی از یکی از عزیزان خود دریافت می‌کنید که می‌گوید به پول نیاز دارد. چطور مطمئن می‌شوید که واقعاً خود اوست و نه نسخه‌ای جعلی از صدایش که توسط هوش مصنوعی ساخته شده؟ در دنیای امنیت سایبری، گروه‌هایی از مهاجمان به‌عنوان کلاه‌سیاه‌ها (Black Hats)، از هوش مصنوعی به‌عنوان سلاحی برای حمله استفاده می‌کنند. در مقابل، کلاه‌سفیدها (White Hats) نقش محافظان را ایفا می‌کنند. همچنین گروه‌هایی به‌نام تیم‌های قرمز (Red Teams) وجود دارند که سازمان‌ها، آن‌ها را استخدام می‌کنند تا حملات شبیه‌سازی‌شده‌ای انجام دهند و نقاط ضعف امنیتی را گزارش کنند.

پس **هوش مصنوعی هم می‌تواند سلاح باشد و هم سپر. مسئله اصلی نه خود فناوری، بلکه نحوه استفاده انسان‌ها از آن است.**

احتمال نقض حقوق مالکیت فکری

برنامه‌های هوش مصنوعی مولد مانند ChatGPT و DALL·E از OpenAI، قادر به تولید متن، تصویر، صدا و ویدئو هستند. آثاری که ممکن است حقوق مالکیت فکری نویسندگان و هنرمندان را نقض کنند. این مدل‌ها با استفاده از حجم وسیعی از داده‌های موجود در اینترنت آموزش دیده‌اند؛ داده‌هایی که بنا به اعتراف شرکت‌های سازنده، شامل آثار دارای کپی‌رایت مانند کتاب‌های منتشرشده، کدهای اختصاصی و آثار هنری نیز می‌شود. علاوه بر این، گاهی اوقات محتوای تولیدشده توسط هوش مصنوعی شباهت زیادی به آثار دارای کپی‌رایت دارد. برای مثال، تصویری از یک ابرقهرمان با پس‌زمینه‌ای از شهر که تقریباً با یکی از صحنه‌های فیلمی معروف یکسان است. حتی اگر اثر تولیدی کاملاً جدید باشد، باز هم می‌تواند نگرانی‌هایی ایجاد کند. آیا این نمونه‌ها نوعی استفاده منصفانه از آثار دارای کپی‌رایت محسوب می‌شوند یا نقض حقوق مالکیت فکری؟ تاکنون استفاده از آثار دارای کپی‌رایت به این شکل ممکن نبوده است. خروجی‌های مدل‌های هوش مصنوعی مولد، از لحاظ ماهیت، کاملاً متفاوت از هر چیزی است که تا به امروز با آن مواجه بوده‌ایم. طبیعتاً نویسندگان و هنرمندان دیدگاهی دارند و شرکت‌های هوش مصنوعی دیدگاهی متفاوت را دنبال می‌کنند.

خلاصه

در این فصل، پنج حوزه مختلف را مورد بررسی قرار دادیم. در هر یک از این دسته‌ها، دیدیم که هوش مصنوعی با خود **مزایای شگفت‌انگیزِ بسیاری** به‌همراه می‌آورد اما در

کنار آن، **نگرانی‌های جدی و آسیب‌های واقعی** نیز پدیدار می‌سازد. می‌توان این بررسی را به حوزه‌های دیگر نیز گسترش داد مانند حوزه‌های شخصی و سلامت یا نوآوری و خلاقیت و همچنان همان دوگانگی را مشاهده کرد: هوش مصنوعی می‌تواند هم در جهت خیر و هم در جهت شر به‌کار رود. این **ماهیت دوگانه**، ویژگی‌ای منحصربه‌فرد برای هوش مصنوعی نیست. در طول تاریخ، همین وضعیت را در سایر نوآوری‌های فناورانه نیز دیده‌ایم، از باروت گرفته تا انرژی هسته‌ای و اینترنت.

با این حال، چیزی در مورد هوش مصنوعی وجود دارد که باعث می‌شود نسبت به آن، هم احساس اعتماد بیشتری داشته باشیم و هم نگرانی عمیق‌تری. شاید علت این باشد که هوش مصنوعی وارد حوزه‌ای شده که از آغاز شکل‌گیری نوع بشر، منحصراً متعلق به انسان‌ها بوده است: یعنی هوش. نیل پُستمن در سخنرانی‌هایش، دستیابی به فناوری را مانند موقعیتی تعبیر می‌کند که در آن برای به‌دست آوردن مزیتی خاص، باید چیزی ارزشمند را فدا کنیم. آیا ما باید در ازای دستیابی به قدرتی خارق‌العاده، چیزی فوق‌العاده ارزشمندی را از دست بدهیم (یعنی جایگاه منحصربه‌فرد و برتر انسان در رأس هرم هوش و خرد)؟ در مقابل، آنچه به‌دست می‌آوریم، توانمندترین یاور تاریخ بشر است. همکاری که هرگز خسته نمی‌شود و می‌تواند با سرعتی باورنکردنی کارهای ما را انجام دهد. اما نتیجه نهایی چه خواهد بود؟ هنوز نمی‌دانیم. آنچه مسلم است، هوش مصنوعی مسیر تاریخ بشر را دگرگون خواهد کرد، و روند ساخت و پذیرش آن، ما را نیز به‌طور اساسی تغییر خواهد داد.

فصل ششم

افسانه‌ها و واقعیت‌های هوش مصنوعی

امروزه **افسانه‌ها و تصورات نادرست زیادی** درباره هوش مصنوعی در فضای آنلاین در حال گردش است و در گفتگوهای روزمره نیز زیاد شنیده می‌شود. هنگام خواندن یا شنیدن مطالب داغ درباره هوش مصنوعی، باید بسیار **مراقب منبع باشید.** بسیاری از افراد می‌خواهند شما را نسبت به آن، مشتاق کنند و عده‌ای دیگر می‌خواهند شما را وحشت‌زده نمایند. در عین حال، **بسیاری از مردم اساساً نمی‌دانند که هوش مصنوعی واقعاً چیست.** درک آن‌ها اغلب برگرفته از فرهنگ عامه، فیلم‌های علمی‌تخیلی، یا پست‌های رسانه‌های اجتماعی است. تاکنون سعی کرده‌ایم بخشی از این تصورات نادرست را برطرف کنیم. ما پیش‌تر یکی از بزرگ‌ترین سوءبرداشت‌ها درباره هوش مصنوعی را نیز بررسی کردیم: این‌که AI دقیقاً معادل یادگیری ماشین یا یادگیری عمیق است. همان‌طور که توضیح دادیم، این سه فناوری در رابطه‌ای سلسله‌مراتبی قرار دارند: هوش مصنوعی چتر بزرگ‌تر است، یادگیری ماشین زیرمجموعه آن محسوب می‌شود، و یادگیری عمیق نیز در دل یادگیری ماشین قرار می‌گیرد. علت ایجاد تلقی اشتباه AI = ML، قابل درک است. یادگیری ماشین و به‌ویژه یادگیری عمیق در چند دهه اخیر اهمیت فراوانی یافته‌اند. به‌سختی می‌توان یک پیشرفت بزرگ اخیر در حوزه AI را نام برد که به یادگیری ماشین مربوط نبوده باشد. همان‌طور که اشاره شد، این فناوری ستون اصلی شاخه‌ای از هوش مصنوعی است که آن را هوش مصنوعی غیرنمادین یا هوش مصنوعی ارتباط‌گرا می‌نامند. اما واقعیت این است که شاخه هوش مصنوعی نمادین یا کلاسیک نیز هنوز وجود دارد و ممکن است دوباره قدرت بگیرد.

اکنون بیایید به **۱۰ افسانه و باور نادرست رایج دیگر درباره هوش مصنوعی** بپردازیم. برای هر مورد، یک نسخه از افسانه را از نگاه افراد **بیش‌ازحد خوش‌بین**، و نسخه‌ای دیگر را از نگاه **افراد بیش‌ازحد بدبین** بیان می‌کنم. در پایان هر مورد، برداشت

واقع‌بینانه را ارائه می‌دهم تا از میان هیاهوی اغراق و ترس، بتوانیم حقیقت را از خیال بازشناسیم.

افسانه شماره ۱: نجات‌دهنده یا نابودگر؟

در نسخه‌ی بیش‌ازحد خوش‌بینانه‌ی این افسانه، هوش مصنوعی قرار است بشریت را از تمام مشکلاتش نجات دهد. صحبت از پایان کامل فقر و گرسنگی، حل کامل مشکل گرم‌شدن زمین، و ریشه‌کن شدن انواع بیماری‌هاست. اما در سوی دیگر ماجرا، نسخه‌ی بیش‌ازحد بدبینانه می‌گوید هوش مصنوعی کنترل امور را به دست خواهد گرفت و ما را به‌طور کامل نابود خواهد کرد. به گفته‌ی ایلان ماسک: «با هوش مصنوعی داریم اهریمن را فرا می‌خوانیم». هر دو دیدگاه افراطی فوق بر یک باور مشترک استوارند: این‌که هوش مصنوعی به قدرتی تقریباً نامتناهی دست خواهد یافت. اما این **تصور از قدرت بی‌کران، یک افسانه است.** مهم است که به یاد داشته باشیم **این نوع پیش‌بینی‌ها با سطح بسیار بالایی از عدم قطعیت همراه‌اند.** هوش مصنوعی هم‌اکنون نیز ابزاری بسیار قدرتمند است و توانایی‌هایش روزبه‌روز در حال گسترش است. اما حقیقت این است که **همواره محدودیت‌هایی خواهد داشت** چه محدودیت‌هایی که مانع حل برخی مشکلات توسط آن می‌شوند، و چه محدودیت‌هایی که مانع تسلط کامل آن بر انسان‌ها خواهند بود. مانند تمام فناوری‌های پیشین، هوش مصنوعی نیز در برخی زمینه‌ها کمک خواهد کرد و در برخی دیگر نه؛ در بعضی موارد به ما آسیب خواهد رساند و در مواردی دیگر خیر.

افسانه شماره ۱: نجات‌دهنده یا نابودگر؟

هوش مصنوعی، بشریت را از تمام مشکلاتش نجات می‌دهد.

هوش مصنوعی کنترل امور را به دست خواهد گرفت و ما را به‌طور کامل نابود خواهد کرد.

حقیقت متعادل

هوش مصنوعی، همواره محدودیت‌هایی خواهد داشت چه محدودیت‌هایی که مانع حل برخی مشکلات توسط آن می‌شوند، و چه محدودیت‌هایی که مانع تسلط کامل آن بر انسان‌ها خواهند بود.

افسانه شماره ۲: اَبَرهوش، یا همین حالا یا هرگز؟

این افسانه، که ارتباط نزدیکی با افسانه‌ی اول دارد، به این پرسش می‌پردازد که هوش مصنوعی چگونه آن‌قدر قدرتمند می‌شود: با دستیابی به **اَبَرهوش** (Superintelligence). طبق تعریف کلاسیک، ابرهوش به «موجودیتی گفته می‌شود که **از انسان در هوش کلی یا در برخی ابعاد خاص هوش پیشی گرفته باشد**». نسخه‌ی خوش‌بینانه‌ی این افسانه می‌گوید: هوش مصنوعی هم‌اکنون ابرهوش است، یا به‌زودی به آن خواهد رسید. در مقابل، نسخه‌ی بدبینانه ادعا می‌کند که هوش مصنوعی نه اکنون ابرهوش است و نه هرگز خواهد بود و صرفاً دارد حرف‌های ما را طوطی‌وار تکرار می‌کند و چیزی جز یک موتور پیش‌بینی واژه‌ی بعدی نیست. اما چنین نگاه تحقیرآمیزی با تجربه‌ی واقعی تعامل با مدل‌های زبانی بزرگ همخوانی ندارد. درست است که این مدل‌ها بی‌نقص نیستند و گاه دچار توهمات یا تولید مطالب نادرست می‌شوند اما با این حال، بسیار شگفت‌انگیز هستند. توانایی‌های آن‌ها همچنان در حال رشد است و ممکن است پیشرفت‌های آتی، سطح هوش آن‌ها را به جایی برساند که از انسان فراتر رود. زمان نشان خواهد داد. هنوز به آن نقطه نرسیده‌ایم، اما شاید روزی برسیم. پس باید گفت:

این‌که ابرهوش همین حالا در دسترس است یک افسانه است، و این‌که کاملاً غیرممکن است هم افسانه‌ای دیگر.

افسانه شماره ۲: اَبَرهوش، یا همین حالا یا هرگز؟

هوش مصنوعی، نه اکنون ابرهوش است و نه هرگز خواهد بود.

هوش مصنوعی، هم‌اکنون ابرهوش است، یا اگر هنوز نیست، به‌زودی به آن خواهد رسید.

حقیقت متعادل

این‌که ابرهوش همین حالا در دسترس است یک افسانه است، و این‌که کاملاً غیرممکن است هم افسانه‌ای دیگر.

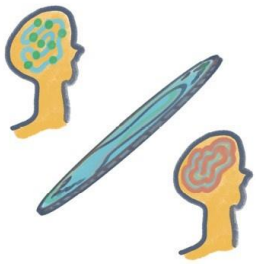
افسانه شماره ۳: یا کاملا شبیه مغز انسان یا کاملاً متفاوت؟

افسانه‌ی سوم به مقایسه‌ی هوش مصنوعی با مغز انسان مربوط می‌شود. از یک‌سو، برخی معتقدند که هوش مصنوعی دقیقاً مانند مغز انسان عمل می‌کند و رفتار دارد. در مقابل، گروهی دیگر معتقدند که هوش مصنوعی هیچ شباهتی به مغز انسان ندارد، هرگز نداشته، و اساساً مقایسه‌ی این دو با هم مضحک است. حقیقت این است که برخی جنبه‌های شبکه‌های عصبی عمیق، **در ابتدا الهام‌گرفته از مغز انسان** بوده‌اند اما در عمل، عملکرد آن‌ها کاملاً متفاوت است. روزنبلات، سازنده‌ی اولین نورون مصنوعی، قصد نداشت مستقیماً هوش مصنوعی خلق کند، بلکه هدفش ساخت مدلی از نورون بیولوژیکی و درک بهتر شناخت انسان بود. تعداد **تفاوت‌ها بین مغز انسان و شبکه‌های عصبی مدرن بسیار زیاد است**. مغز انسان بر پایه‌ی کربن و ساختاری ارگانیک ساخته شده و با سیگنال‌های الکتریکی و شیمیایی که بین نورون‌های زیستی رد و بدل

می‌شوند، اطلاعات را پردازش می‌کند. مغز انسان آگاهی و درک واقعی دارد. در مقابل، شبکه‌های عصبی عمیق بر روی سخت‌افزارهای سیلیکونی پیاده‌سازی می‌شوند و از طریق سیگنال‌های دیجیتال بین نورون‌های مصنوعی، اطلاعات را پردازش می‌کنند. این شبکه‌ها به رایانه‌ها امکان می‌دهند برخی وظایف را انجام دهند بدون آن‌که الزاماً دارای درک عمیق یا آگاهی واقعی باشند.

افسانه شماره ۳: یا کاملا شبیه مغز انسان یا کاملاً متفاوت؟

هوش مصنوعی هیچ شباهتی به مغز انسان ندارد، هرگز نداشته، و اساساً مقایسه‌ی این دو با هم مضحک است.



هوش مصنوعی دقیقاً مانند مغز انسان عمل می‌کند و رفتار دارد.

حقیقت متعادل

برخی جنبه‌های شبکه‌های عصبی عمیق، در ابتدا الهام‌گرفته از مغز انسان بوده‌اند، اما در عمل عملکرد آن‌ها کاملاً متفاوت است.

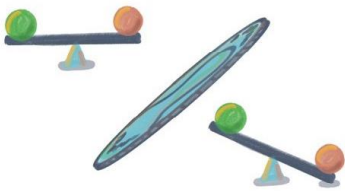
افسانه شماره ۴: کاملا بی‌طرف یا کاملا جانب‌دار؟

افسانه‌ی چهارم به عدالت یا عدم عدالت در تصمیم‌گیری‌های هوش مصنوعی می‌پردازد. طرفداران پرشور هوش مصنوعی بر این باورند که AI یک قاضی کاملا بی‌طرف و منصف است. منطق آنان این است: چون هوش مصنوعی بر پایه‌ی محاسبات رایانه‌ای، ورودی‌های دیجیتال و معادلات ریاضی کار می‌کند، پس برخلاف انسان‌ها، دچار تعصب و پیش‌داوری نمی‌شود. اما این نتیجه‌گیری نادرست است. در سوی دیگر، بدبینان هوش مصنوعی معتقدند که این فناوری همیشه دچار تعصب خواهد بود چرا که آموزش آن بر پایه‌ی داده‌هایی است که ما انسان‌ها در اختیارش می‌گذاریم. داده‌ها ذاتاً

دارای سوگیری‌اند، یا فرآیندی که آن داده‌ها را تولید کرده، در بطن خود تبعیض‌آمیز بوده است. بنابراین، هوش مصنوعی ناگزیر است سوگیری‌ها و پیش‌داوری‌های انسانی را بیاموزد و بدتر از آن، همان‌ها را در آینده تکرار نماید. اما این دیدگاه نیز نادرست است چرا که نادیده می‌گیرد که ما می‌توانیم اقدامات مؤثری برای کاهش سوگیری‌ها در داده‌های آموزشی انجام دهیم. حقیقت متعادل این است که **سوگیری‌ها وجود دارند اما با تلاش آگاهانه و مستمر، می‌توانیم به بهبود عدالت سیستم‌های هوش مصنوعی کمک کنیم** و آن‌ها را منصفانه‌تر وارد جامعه کنیم.

افسانه شماره ۴: کاملا بی‌طرف یا کاملا جانب‌دار؟

AI همیشه دچار تعصب خواهد است. آموزش آن بر پایه‌ی داده‌هایی ما انسان‌ها است و داده‌ی آموزشی‌ای بدون سوگیری وجود ندارد.



هوش مصنوعی، یک قاضی کاملا بی‌طرف است کاملاً عینی و منصف.

حقیقت متعادل

سوگیری‌ها وجود دارند اما با تلاش آگاهانه و مستمر، می‌توانیم به بهبود عدالت سیستم‌های هوش مصنوعی کمک کنیم و آن‌ها را منصفانه‌تر وارد جامعه کنیم.

افسانه شماره ۵: رونق اقتصادی یا ویرانی؟

در نگاه بسیاری از فناوری‌دوستان، هوش مصنوعی نویدبخش رونق اقتصادی برای بخش بزرگی از جمعیت جهان است چراکه قرار است بخش زیادی از کارهای تکراری و خسته‌کننده را بر عهده بگیرد، فقر را به‌طور چشمگیری کاهش دهد و به هر فردی فرصت دهد تا از مسیرهای جدید، برای خود و خانواده‌اش ثروت تولید کند. به باور آن‌ها، برنده‌ی نهایی، کل جامعه خواهد بود. اما چنین نگاهی بیش از حد خوش‌بینانه و آرمان‌گرایانه است. **رؤیای هوش مصنوعی‌ای که برای همه‌ی مردم رونق اقتصادی به**

ارمغان بیاورد، صرفاً یک افسانه است. در سوی مقابل، افسانه‌ی دیگری وجود دارد: اینکه هوش مصنوعی همه‌ی شغل‌ها را خودکار خواهد کرد، ما را بیکار خواهد ساخت و ویرانی اقتصادی برای همه به‌جز نخبگان بسیار ثروتمند به‌همراه خواهد داشت (یعنی کسانی که تمام منافع مالی ناشی از هوش مصنوعی را برای خود انحصاری خواهند کرد). خوشبختانه، این چشم‌انداز تیره‌وتار نیز افسانه‌ای بیش نیست. حقیقت آن است که هوش مصنوعی به‌راستی در حال دگرگون‌سازی صنایع است و ما را ناگزیر می‌کند که با یادگیری مهارت‌های جدید، خود را با این تحولات سازگار کنیم تا بتوانیم از فرصت‌های پیشِ‌رو بهره‌مند شویم.

واقعیت این است که گروه نسبتاً کوچکی از افراد که کنترل شرکت‌های فناورانه را در اختیار دارند، احتمالاً بیش از سایرین از این تحولات سود خواهند برد، و **اقتصادهای پیشرفته نیز به احتمال زیاد بیشتر از کشورهای در حال توسعه از هوش مصنوعی منتفع خواهند شد.** از این‌رو، لازم است اقداماتی صورت گیرد تا اطمینان حاصل شود که همه‌ی افراد از منافع هوش مصنوعی بهره‌مند شوند، نه فقط گروهی خاص. چه اقداماتی ضروری است؟

باید **دسترسی به هوش مصنوعی** برای مردم در تمام مناطق جغرافیایی و سطوح اقتصادی-اجتماعی فراهم شود. همچنین باید آموزش‌هایی ارائه گردد تا افراد نه‌تنها بتوانند به هوش مصنوعی دسترسی داشته باشند، بلکه واقعاً بتوانند آن را به‌کار گیرند. سازمان‌ها و دولت‌ها نیز باید اقداماتی برای **آماده‌سازی نیروی کار** در برابر این تحولات انجام دهند تا فرصت بهره‌برداری از آن‌ها برای همه وجود داشته باشد. و در نهایت، باید با خروجی‌های سوگیرانه و تبعیض‌آمیز هوش مصنوعی که نابرابری‌ها را تشدید

می‌کنند، مقابله کرد و آن‌ها را کاهش داده و از میان برداشت. در کلام گروهی از نویسندگان وابسته به گوگل و شرکت مادر آن، آلفابت:

هوش مصنوعی **یک فرصت تاریخی و بی‌سابقه برای رشد اقتصادی و گسترش رفاه** به شمار می‌رود... اما توان بالقوه‌ی آن برای تحول اقتصاد و خلق رفاهی مشترک و همگانی، امری خودکار یا تضمین‌شده نیست. تاریخ نوآوری نشان می‌دهد که **بهره‌گیری از ظرفیت‌های هوش مصنوعی نیازمند مواجهه با موانع و چالش‌هایی است**؛ از جمله زیرساخت‌های نابرابر، موانع پذیرش، نیاز به تغییرات سازمانی، آمادگی نیروی کار و نابرابری در دسترسی به فرصت‌ها و مزایای آن.

افسانه شماره ۵: رونق اقتصادی یا ویرانی؟

هوش مصنوعی ویرانی اقتصادی و بیکاری برای همه، به‌جز نخبگان بسیار ثروتمند به‌همراه خواهد داشت.

هوش مصنوعی نویدبخش رونق اقتصادی برای بخش بزرگی از جمعیت جهان است

حقیقت متعادل

هوش مصنوعی در حال دگرگون‌سازی صنایع است و ما را ناگزیر می‌کند به یادگیری مهارت‌های جدید. گروه نسبتاً کوچکی از شرکت‌های فناورانه، احتمالاً بیش از سایرین از این تحولات سود خواهند برد، و اقتصادهای پیشرفته نیز به احتمال زیاد بیشتر از کشورهای در حال توسعه از هوش مصنوعی منتفع خواهند شد.

افسانه شماره ۶: کاملا قابل اعتماد یا غیرقابل اعتماد؟

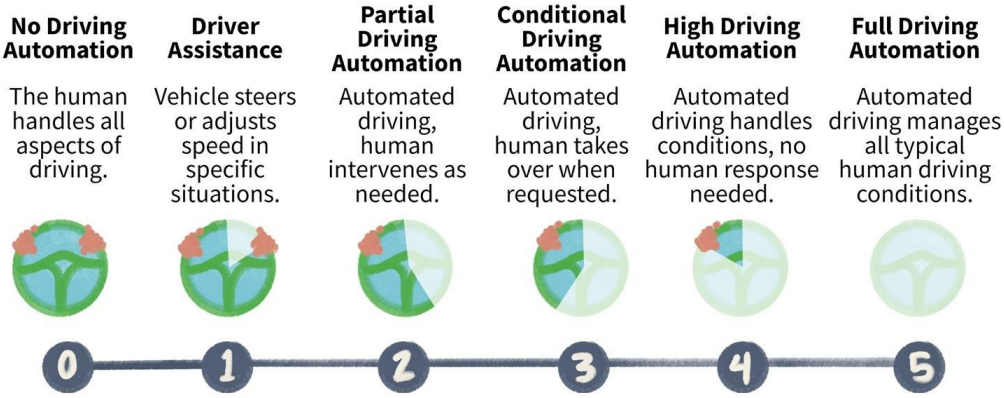
افسانه‌ی ششم به پرسش **قابل‌اعتماد بودن یا نبودن هوش مصنوعی** می‌پردازد. افراد خوش‌بین به هوش مصنوعی بر این باورند که بهتر است **تصمیمات مهم‌مان را به آن**

بسپاریم زیرا هوش مصنوعی می‌تواند سریع‌تر و دقیق‌تر از انسان‌ها تصمیم‌گیری کند. برای مثال، استدلال می‌کنند که اگر تمام خودروهای جاده به‌صورت خودران عمل کنند آمار مرگ‌ومیر ناشی از تصادفات تقریباً به صفر خواهد رسید. گرچه این ادعا ممکن است روزی محقق شود، اما **باور به اینکه هوش مصنوعی در حال حاضر توانایی بر عهده گرفتن همه‌ی تصمیم‌ها را دارد، یک افسانه است.** در نقطه‌ی مقابل، بدبینان به هوش مصنوعی بر این باورند که این فناوری اصلاً قابل اعتماد نیست و مستعد بروز خطاهایی است که یا قابل رفع نیستند یا حتی درک آن‌ها برای ما دشوار یا غیرممکن است. از نگاه آن‌ها، مدل‌های هوش مصنوعی در مواجهه با دنیای واقعیِ پیچیده و ناپایدار که با داده‌های آموزشی‌شان تفاوت دارد، ذاتاً آسیب‌پذیر و شکننده‌اند؛ بنابراین، سپردن هر نوع تصمیمی به آن‌ها اشتباه محض است.

دیدگاه متعادل آن است که **امروزه بسیاری از تصمیم‌ها با موفقیت به هوش مصنوعی سپرده شده‌اند** و در بسیاری از این موارد، عملکرد آن قابل‌اعتماد بوده است. با ادامه‌ی پیشرفت هوش مصنوعی، به‌تدریج تصمیمات بیشتری به آن واگذار خواهد شد. اما **همواره این پرسش وجود دارد که چه تصمیم‌هایی را می‌توان به هوش مصنوعی سپرد، و در چه حدی؟ پاسخ این پرسش در اغلب موارد فراتر از دو گزینه‌ی صفر و صد، یا «کاملاً بدون کنترل» و «کاملاً تحت کنترل» است.** یک مدل جالب در این زمینه، **سیستم شش‌سطحی طبقه‌بندی خودکارسازی رانندگی** است که در سال ۲۰۱۴ توسط «انجمن مهندسان خودرو» منتشر شد و سطح‌های مختلف این کار را این‌گونه تعریف می‌کند:

- سطح ۰ – بدون اتوماسیون رانندگی: تمام وظایف رانندگی بر عهده‌ی انسان است؛

- سطح ۱ – کمک‌راننده: سیستم در برخی شرایط، فقط یکی از عملکردها مانند فرمان‌دادن یا شتاب‌گیری/کاهش سرعت را انجام می‌دهد (مثل کروز کنترل تطبیقی)؛
- سطح ۲ – خودکارسازی جزئی رانندگی: رانندگی به‌طور خودکار انجام می‌شود اما راننده باید در برخی موقعیت‌ها کنترل را به‌دست گیرد؛
- سطح ۳ – خودکارسازی مشروط رانندگی: رانندگی به‌طور خودکار است اما در صورت بروز مشکل، از راننده درخواست مداخله می‌شود؛
- سطح ۴ – خودکارسازی بالای رانندگی: رانندگی کاملاً خودکار در شرایطی خاص که سیستم برای آن طراحی شده، حتی اگر راننده به درخواست‌های مداخله پاسخ ندهد؛
- سطح ۵ – خودکارسازی کامل رانندگی: رانندگی کاملاً خودکار در هر شرایطی که یک راننده‌ی انسانی قادر به هدایت خودرو باشد.



نمودار ۶/۱ - استاندارد SAE J3016 برای شش سطح رانندگی خودکار

البته باید به این نکته توجه داشت که خودکارسازی الزاماً به‌معنای هوش مصنوعی نیست. ما بسیاری از وظایف را بدون استفاده از هوش مصنوعی، خودکار کرده‌ایم. برای مثال، در برخی خودروها، برف‌پاک‌کن‌ها هنگام برخورد قطرات باران به شیشه‌ی جلو، به‌طور خودکار روشن می‌شوند و سرعت خود را با شدت بارش تطبیق می‌دهند. نسخه‌های اولیه‌ی این سیستم‌ها از حسگر و نرم‌افزار استفاده می‌کردند، اما از هوش

مصنوعی بهره نمی‌بردند. با گسترش هوش مصنوعی، انتظار می‌رود تعداد بیشتری از سامانه‌های خودکار از آن بهره‌مند شوند؛ اما باید توجه داشت که **هر فرآیند خودکاری، الزاماً هوشمند نیست**، و هر نرم‌افزاری که وظیفه‌ای را خودکار انجام می‌دهد، لزوماً از هوش مصنوعی استفاده نمی‌کند.

افسانه شماره ۶: کاملا قابل اعتماد یا غیرقابل اعتماد؟

AI در مواجهه با دنیای واقعیِ پیچیده و ناپایدار، ذاتاً آسیب‌پذیر و شکننده است و غیرقابل اعتماد.

بهتر است تصمیمات مهم‌مان را به AI بسپاریم زیرا می‌تواند سریع‌تر و دقیق‌تر از انسان‌ها تصمیم‌گیری کند.

حقیقت متعادل

هوش مصنوعی می‌تواند بسیاری از تصمیات و کارهای ما را، خودکار سازد اما برای برخی تصمیمات کلیدی و پیچیده، باید توسط انسان هدایت شود.

افسانه شماره ۷: تهدید یا محافظ امنیت و حریم خصوصی؟

افسانه‌ی هفتم به مسائل مرتبط با **ایمنی، امنیت و حریم خصوصی** می‌پردازد. می‌توانیم این مفاهیم را چنین تعریف کنیم:

- **ایمنی** یعنی محافظت در برابر آسیب‌های عمدی یا غیرعمدی؛
- **امنیت** یعنی محافظت در برابر آسیب‌های عمدی؛
- و **حریم خصوصی** یعنی توانایی کنترل اطلاعات شخصی خود.

نگاه بیش از حد خوش‌بینانه این است که هوش مصنوعی تمامی مشکلات ما را در این سه حوزه حل خواهد کرد؛ چراکه به‌مرور خواهد آموخت چگونه هرگونه تهدید احتمالی را حذف کند، یا دست‌کم آن را به ما هشدار دهد تا خودمان اقدام لازم را انجام دهیم.

این تصویری مطلوب است، اما در واقع یک افسانه است. در مقابل، نگاه بیش از حد بدبینانه این است که هوش مصنوعی پایان ایمنی، امنیت و حریم خصوصی ما خواهد بود. در این سناریو، هکرها می‌توانند از هوش مصنوعی برای سرقت گذرواژه‌ها استفاده کنند، تمام اطلاعات شخصی ما فاش خواهد شد، و دیگر امکان تشخیص «دوست» از «دشمن» وجود نخواهد داشت. این دیدگاه تیره و تار نیز یک افسانه است.

حقیقت متعادل این است که **ما می‌توانیم از هوش مصنوعی برای تقویت برخی از قابلیت‌های محافظتی خود بهره ببریم، اما همزمان افراد بد نیّت نیز از آن برای افزایش تهدیدها و خطرات استفاده خواهند کرد.** در حال حاضر، هوش مصنوعی هم به‌عنوان سلاح توسط هکرهای مخرب و هم به‌عنوان ابزار دفاعی توسط هکرهای اخلاق‌مدار استفاده می‌شود. هکرهای مخرب می‌کوشند با استفاده از فناوری، ضعف‌های سیستم‌ها را شناسایی و از آن‌ها سوءاستفاده کنند درحالی‌که هکرهای اخلاق‌مدار از همان فناوری‌ها برای یافتن این ضعف‌ها و محافظت در برابر آن‌ها بهره می‌برند. **هوش مصنوعی صرفاً یک ابزار دیگر در این نبرد بی‌پایان میان خیر و شر است.** در برخی مقاطع، این کشمکش ممکن است باعث شود که ما نسبت به گذشته ایمن‌تر باشیم، و در مقاطعی دیگر ممکن است باعث شود که در معرض خطر بیشتری قرار بگیریم.

افسانه شماره ۷: تهدید یا محافظ امنیت و حریم خصوصی؟

هوش مصنوعی پایان ایمنی، امنیت و حریم خصوصی ما خواهد بود.

هوش مصنوعی تمامی مشکلات ما را در این سه حوزه ایمنی، امنیت و حریم خصوصی حل خواهد کرد.

حقیقت متعادل

ما می‌توانیم از هوش مصنوعی برای تقویت برخی از قابلیت‌های محافظتی خود بهره ببریم، اما همزمان افراد بد نیّت نیز از آن برای افزایش تهدیدها و خطرات استفاده خواهند کرد.

افسانه شماره ۸: آگاهی؛ همین حالا یا هرگز؟

مفهوم آگاهی موضوعی بسیار جذاب است. آگاه بودن یعنی داشتن توانایی تجربه کردن احساسات یعنی «واقعاً احساس کردن». در کاربردهای امروزی، این واژه همچنین به معنی هوشیاری و خودآگاهی نیز به‌کار می‌رود. به یاد دارید که در فصل اول درباره‌ی «هوش مصنوعی قوی» و «هوش مصنوعی ضعیف» صحبت کردیم؟ این اصطلاحات در اصل توسط فیلسوف آمریکایی، جان سرل، در سال ۱۹۸۰ مطرح شدند و معنایی متفاوت با برداشت رایج امروزی داشتند. از نظر سرل، هوش مصنوعی قوی فقط به رایانه‌ای که می‌تواند وظایف هوشمندانه‌ای را در سطح یا بالاتر از انسان انجام دهد اشاره نمی‌کند، بلکه به سیستمی گفته می‌شود که واقعاً بتوان گفت آن را می‌فهمد. اما طبق دیدگاه هوش مصنوعی قوی، رایانه صرفاً ابزاری برای مطالعه ذهن نیست؛ بلکه رایانه‌ای که به‌درستی برنامه‌ریزی شده باشد، خودِ ذهن است به این معنا که می‌توان به‌طور واقعی گفت چنین رایانه‌ای می‌فهمد و حالت‌های شناختی دیگری دارد.

آیا هوش مصنوعی آگاه است؟ برای کسانی که مجذوب شگفتی‌ها و توانایی‌های آن شده‌اند، مدل‌های زبانی پیشرفته‌ی امروزی، همین حالا هم آگاه به نظر می‌رسند. اما از دیدگاه من، این تصور یک افسانه است. مسئله اینجاست که **اثبات یا رد اینکه یک سامانه‌ی هوش مصنوعی واقعاً آگاه است، بسیار دشوار است.** این همان چالشی بود که آلن تورینگ سعی کرد با معرفی «بازی تقلید» که اکنون آن را به نام آزمون تورینگ می‌شناسیم از آن عبور کند. با این حال، بیشتر متخصصان هوش مصنوعی امروزی بر این باورند که چت‌بات‌های فعلی آگاه نیستند. اما از سوی دیگر، این باور که «هوش مصنوعی هرگز و تحت هیچ شرایطی نمی‌تواند به آگاهی دست یابد»، نیز برداشتی نادرست است. من درباره‌ی ماهیت واقعی آگاهی اطلاعات کافی ندارم که بتوانم با اطمینان نظری مثبت یا منفی ارائه دهم. و در حال حاضر، هیچ راهی نمی‌شناسم که رایانه‌ای بتواند به آگاهی واقعی دست یابد. در نهایت، مسئله‌ی آگاهی، خود به‌تنهایی مسئله‌ای پیچیده و دشوار است. بنابراین، وقتی صحبت از آگاهی در هوش مصنوعی می‌شود، توصیه می‌کنم که از رفتن به سوی نتیجه‌گیری‌های افراطی چه از نوع خوش‌بینانه و چه بدبینانه خودداری کنیم.

افسانه شماره ۸: آگاهی؛ همین حالا یا هرگز؟

هوش مصنوعی هرگز و تحت هیچ شرایطی نمی‌تواند به آگاهی دست یابد.

مدل‌های زبانی پیشرفته‌ی امروزی، همین حالا هم آگاه به نظر می‌رسند.

حقیقت متعادل

وقتی صحبت از آگاهی در هوش مصنوعی می‌شود، توصیه می‌شود از رفتن به سوی نتیجه‌گیری‌های افراطی چه از نوع خوش‌بینانه و چه بدبینانه خودداری کنیم.

افسانه شماره ۹: کارایی در فرآیندها یا ناکارآمدی؟

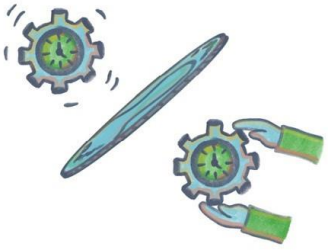
افسانه بعدی به توانایی هوش مصنوعی در بهبود کارایی فرآیندها در دنیای واقعی مربوط می‌شود. این موضوع شباهت زیادی با افسانه مربوط به قابل اعتماد بودن هوش مصنوعی در تصمیم‌گیری دارد، با این تفاوت که تصمیم‌گیری‌ها معمولاً رویدادهایی تک‌مرحله‌ای و مقطعی هستند، در حالی که فرآیندها به‌صورت مداوم و تکرارشونده اجرا می‌شوند. درباره‌ی چه نوع فرآیندهایی صحبت می‌کنیم؟ فرآیندهای زنجیره تأمین، تولید و مصرف انرژی، عملیات تجارت الکترونیک و بسیاری موارد دیگر. سؤال اصلی این است: تا چه حد می‌توان به هوش مصنوعی برای بهینه‌سازی و اجرای این فرآیندها اعتماد کرد؟

از دیدگاه خوش‌بینان به هوش مصنوعی، این فناوری ابزاری ایده‌آل برای روان‌سازی همه‌ی فرآیندهاست. آن‌ها معتقدند که هوش مصنوعی می‌تواند اتلاف را شناسایی و حذف کند، تنظیمات فرآیند را بهینه کند و با تغییرات محیطی تطبیق یابد تا فرآیندها در مسیر درست باقی بمانند. کافی است آن را «تنظیم و رها» کنیم. اما واقعیت این است که هوش مصنوعی هنوز به سطحی نرسیده که بتواند همه‌ی این کارها را با اطمینان کامل و بدون خطا انجام دهد و این باور، یک افسانه است. در سوی دیگر طیف، بدبینان به هوش مصنوعی معتقدند که این فناوری فقط در محیط‌های کاملاً کنترل‌شده عملکرد خوبی دارد. به‌محض آن‌که با شرایطی خارج از داده‌های آموزشی‌اش روبه‌رو شود، دچار خطا می‌شود و کل فرآیندی که به آن سپرده شده از کار می‌افتد. این نگاه منفی، پیشرفت‌های اخیر در هوش مصنوعی مانند برخی مدل‌هایی که توانایی اصلاح و تنظیم خودکار دارند را نادیده می‌گیرد؛ نمونه‌ای از آن، شبکه‌های GAN است که در فصل چهارم به‌طور مختصر معرفی شد.

واقعیت متعادل این است که هوش مصنوعی ابزاری بسیار مفید برای بهبود فرآیندهاست. اگرچه این فناوری به‌مرور در حال پیشرفت است، اما هنوز یک راه‌حل همه‌جانبه و قطعی نیست. مدیران فرآیندها همچنان باید بر خروجی‌ها و شاخص‌های عملکرد نظارت داشته باشند، در زمان بروز خطا مداخله کنند، و راه‌های جدیدی برای ارتقای سیستم‌های هوش مصنوعی بیابند. به نظر می‌رسد این روند، همچنان در آینده نیز ادامه خواهد داشت.

افسانه شماره ۹: کارایی در فرآیندها یا ناکارآمدی؟

AI، ابزاری ایده‌آل برای روان‌سازی همه‌ی فرآیندهاست.



این فناوری فقط در محیط‌های کاملاً کنترل‌شده عملکرد خوبی دارد.

حقیقت متعادل

هوش مصنوعی ابزاری بسیار مفید برای بهبود فرآیندهاست. اگرچه این فناوری به‌مرور در حال پیشرفت است، اما هنوز یک راه‌حل همه‌جانبه و قطعی نیست. مدیران فرآیندها همچنان باید بر خروجی‌ها و شاخص‌های عملکرد نظارت داشته باشند، در زمان بروز خطا مداخله کنند، و راه‌های جدیدی برای ارتقای سیستم‌های هوش مصنوعی بیابند.

افسانه شماره ۱۰: آرمان‌شهر یا ویران‌شهر؟

آخرین افسانه، در واقع نتیجه‌ی تجمعی تمام افسانه‌های قبلی است. فردی که به هر ۹ افسانه‌ی پیشین باور دارد، احتمالاً انتظار دارد که هوش مصنوعی دروازه‌ای باشد به‌سوی یک آرمان‌شهر یعنی دنیایی متعادل و منصفانه، عاری از فقر و بیماری، سرشار از رفاه و آسایش برای همه‌ی انسان‌ها. در مقابل، دیدگاهی وجود دارد که آینده‌ی بشر با هوش

مصنوعی را چیزی جز یک ویران‌شهر نمی‌بیند: جهانی از بردگی، سلطه‌ی ماشین‌ها، و حتی نابودی کامل انسان به‌دست ربات‌ها.

اما حقیقت متعادل این است که هیچ‌کدام از این دو سناریوی افراطی به واقعیت نخواهد پیوست. ما با کمک هوش مصنوعی به پیشرفت‌های چشمگیری دست خواهیم یافت، و گهگاه نیز با اشتباهاتی که مرتکب می‌شویم، خود را به دردسر خواهیم انداخت. مسئولیت ما این است که کاربردهای مفید و سازنده‌ی هوش مصنوعی را طراحی و پیاده‌سازی کنیم و در برابر کاربردهای زیان‌بار آن، مقاومت و اعتراض داشته باشیم. در این معنا، هوش مصنوعی مانند هر فناوری دیگری است که بشر تا کنون خلق کرده است.

افسانه شماره ۱۰: آرمان‌شهر یا ویران‌شهر؟

آینده‌ی بشر با هوش مصنوعی، چیزی جز یک ویران‌شهر نیست.

هوش مصنوعی دروازه‌ای باشد به‌سوی یک آرمان‌شهر.

حقیقت متعادل

ما با کمک هوش مصنوعی به پیشرفت‌های چشمگیری دست خواهیم یافت، و گهگاه نیز با اشتباهاتی که مرتکب می‌شویم، خود را به دردسر خواهیم انداخت. مسئولیت ما این است که کاربردهای مفید و سازنده‌ی هوش مصنوعی را طراحی و پیاده‌سازی کنیم و در برابر کاربردهای زیان‌بار آن، مقاومت و اعتراض داشته باشیم.

در این فصل، ده افسانه مختلف را بررسی کردیم و هر یک را از دو دیدگاه متضاد مورد ارزیابی قرار دادیم: یک نگاه بیش از حد خوش‌بینانه و یک نگاه بیش از حد بدبینانه. در هر مورد، حقیقت متعادل را در نظر گرفتیم واقعیتی که نه در یک سوی افراطی قرار دارد و نه در سوی دیگر، بلکه مانند توپی است که با دقت بین دو تیر دروازه شوت می‌شود و پیروزی را رقم می‌زند. پیام اصلی داستان این است که **هوش مصنوعی هم مفید است و هم مضر**. اگر کاملاً مفید بود، کار راحتی داشتیم کافی بود آن را به‌کار بگیریم و آسوده از ثمراتش بهره ببریم. اگر هم کاملاً مضر بود، باز هم تکلیف روشن بود می‌توانستیم آن را به‌طور کامل متوقف کنیم و در همه‌جا ممنوع اعلام کنیم. اما **واقعیت به این سادگی نیست.** ما در ناحیه‌ای خاکستری زندگی می‌کنیم، جایی بین سیاه و سفید، و باید این واقعیت را بپذیریم که هنوز کار زیادی پیش رو داریم تا بتوانیم بیشترین بهره را از هوش مصنوعی ببریم و هم‌زمان از پیامدهای نامطلوب آن دور بمانیم. **همه چیز به ما بستگی دارد.**

نتیجه‌گیری

اکنون به پایان این کتاب رسیدیم که پایانِ «آغاز» سفر به دنیای هوش مصنوعی است. مسیر زیادی را با هم پیمودیم و اکنون شما در حال تبدیل شدن به یک «شهروند هوش مصنوعی» هستید. کسی که با تعاریف پایه، کاربردهای روزمره، تاریخ پرپیچ‌وخم، فناوری‌های اصلی، مزایا و آسیب‌ها، افسانه‌ها و باورهای نادرست و در نهایت حقایق متعادل درباره هوش مصنوعی آشناست. موضوعات مهم بسیاری در دنیای امروز وجود دارند که بر زندگی همه ما تأثیر می‌گذارند اما **در حوزه فناوری، به‌سختی می‌توان موضوعی مهم‌تر از هوش مصنوعی یافت.** گفتگوهایی که هم‌اکنون درباره هوش مصنوعی در حال شکل‌گیری است، نقشی تعیین‌کننده دارند؛ آن‌ها مسیری را که هوش مصنوعی طی خواهد کرد و در نتیجه، تأثیری را که بر جهان و دنیای شما خواهد داشت، مشخص می‌کنند. در این گفتگوها تماشاچی نباشید، حضوری فعال داشته باشید. دیدگاه‌ها، امیدها و نگرانی‌های خود را بیان کنید و از موضعی آگاهانه و مطلع، پرسش‌های اساسی مطرح نمایید. هدف این کتاب نیز همین بود: قرار دادن شما بر مسیر مشارکت معنادار. برای جمع‌بندی، بیایید دوباره به تعریفی بازگردیم که در ابتدای کتاب از «سواد هوش مصنوعی» ارائه کردیم: **سواد هوش مصنوعی یعنی توانایی تشخیص، درک، استفاده و ارزیابی انتقادی از فناوری‌های هوش مصنوعی و تأثیرات آن‌ها.** این تعریف چهار بخش متمایز دارد:

بخش اول - **تشخیص** هوش مصنوعی: در این کتاب، به تعریف و کاربردهای گوناگون هوش مصنوعی اشاره کردیم تا شما بتوانید حضور آن را در زندگی روزمره خود بهتر درک کنید؛

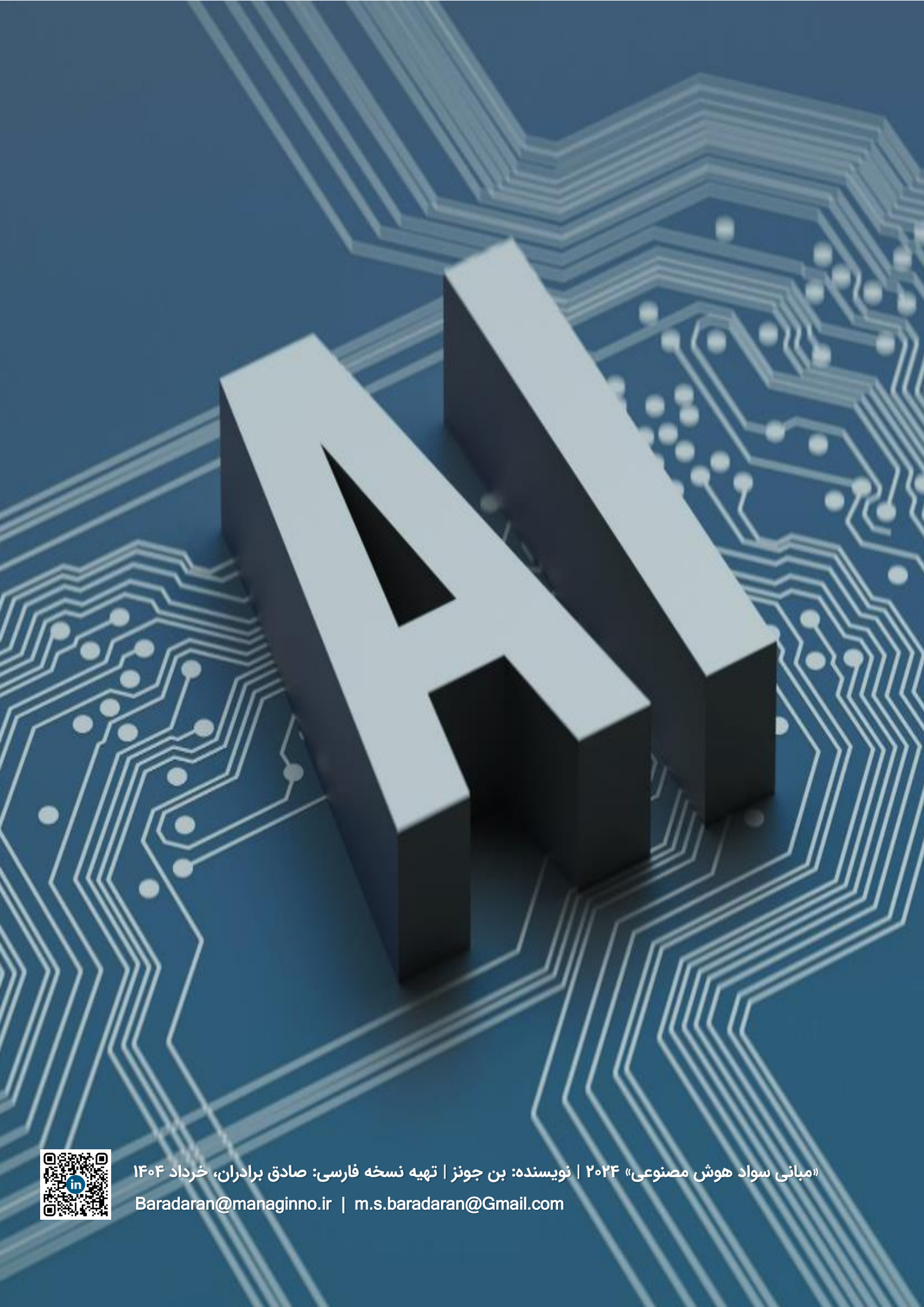
بخش دوم - **درک** هوش مصنوعی: برای آن‌که بتوانید بفهمید در «پشت‌صحنه» چه اتفاقی می‌افتد، مفاهیم پایه‌ای یادگیری ماشین و یادگیری عمیق را بررسی کردیم؛

بخش سوم - **استفاده** از هوش مصنوعی: دانشی که به دست آورده‌اید باید به شما اعتمادبه‌نفس بیشتری برای استفاده عملی از ابزارهای هوش مصنوعی، چه چت‌بات‌هایی مانند ChatGPT و چه دستیارهایی مانند Siri بدهد؛

بخش چهارم - **ارزیابی انتقادی** از هوش مصنوعی: این بخش بسیار مهم است. هدف من در این کتاب آن بود که ابزارهایی در اختیار شما بگذارم تا بتوانید اثربخشی، تناسب و انصاف در کاربرد هوش مصنوعی را ارزیابی کنید. هرچه درک عمومی ما از مزایا و مضرات هوش مصنوعی بیشتر باشد، بهتر می‌توانیم آن را به مسیر صحیح هدایت کنیم.

گام بعدی چیست؟

شما **می‌توانید «سواد هوش مصنوعی» را به «مهارت در هوش مصنوعی» تبدیل کنید**. کسانی که یاد می‌گیرند چگونه از هوش مصنوعی بهره‌برداری کنند، در سال‌های آینده از مزیت خارق‌العاده‌ای برخوردار خواهند بود. امید من این است که بتوانید بر شتابی که تا اینجا به دست آورده‌اید تکیه کنید و به یک متخصص واقعی یا کاربر حرفه‌ای در این حوزه تبدیل شوید. فراموش نکنید که هوش مصنوعی به‌سرعت در حال تحول است و هر یک از ما باید به‌طور مستمر دانش و مهارت‌های خود را به‌روزرسانی کنیم. ذهنیت **یادگیری مستمر**، مهم‌ترین ویژگی‌ای است که می‌توانید در خود پرورش دهید.



«مبانی سواد هوش مصنوعی» ۲۰۲۴ | نویسنده: بن جونز | تهیه نسخه فارسی: صادق برادران، خرداد ۱۴۰۴
Baradaran@managinno.ir | m.s.baradaran@Gmail.com