آنچه که هر مدیر عاملی باید درباره هوش مصنوعی بداند

فهرست مطالب

۱. مقدمه: چشم انداز استراتژیک هوش مصنوعی

امروزه دیگر جای تردیدی در اهمیت هوش مصنوعی (AI) وجود ندارد، اما اینکه چرا باید در مجموعه ای در مورد استراتژی کسب و کار جایگاهی داشته باشد، چندان واضح نیست. عنوان کتاب نشان می دهد که هر مدیرعاملی باید چیزی در مورد هوش مصنوعی بداند - آیا این به معنای کسانی است که قصد استفاده از هوش مصنوعی را ندارند نیز هست؟ قطعا. و در این کتاب کوتاه توضیح خواهم داد که چرا.

برای جلوگیری از هرگونه سردرگمی، مهم است که از همان ابتدا موضع خود را روشن کنم. من یک علاقه مند به هوش مصنوعی هستم و معتقدم که هوش مصنوعی می تواند در صورت استفاده درست، بسیار مفید باشد. با این حال، من یک تبلیغ‌چی نیستم که صرفاً به دنبال جلب رضایت همه باشم، بدون توجه به اینکه آیا هوش مصنوعی برای موضوع خاص مفید است یا خیر. برعکس، من همچنین پیشگوی نگرانی نیستم که پیش بینی می کند چگونه هوش مصنوعی جهان را تسخیر خواهد کرد و بشریت را به بردگی در خواهد آورد یا نابود خواهد کرد. این کتاب هم محدودیت ها و هم نقاط قوت هوش مصنوعی را به تصویر می کشد و حقایق را از باورها جدا می کند تا مدیران عامل را قادر سازد در شرایط خاص تصمیم بگیرند که آیا به هوش مصنوعی نیاز دارند و چگونه از آن استفاده کنند.

۱.۱ هوش مصنوعی، استراتژی و مدیرعامل

چگونه می دانیم هوش مصنوعی مهم است؟ همانطور که نه تنها با هر فناوری بلکه با هر پدیده ای در جهان، می توانیم به دنبال پول باشیم. به نظر می رسد هوش مصنوعی پرهزینه ترین تلاش بشر تا به حال است. چهار قطب برجسته - ایالات متحده، بریتانیا، کانادا و اسرائیل - با هم بیش از یک تریلیون دلار برای هوش مصنوعی هزینه کرده اند و چین، هند و روسیه احتمالاً چندان عقب نیستند. چنگوی لیو در ای او ام اینسایت (۲۰۲۰) پیشنهاد کرد که اولین فرد تریلیونر جهان از تجارت هوش مصنوعی به وجود خواهد آمد. همچنین می بینیم که هوش مصنوعی در همه جا وجود دارد. در همه زمینه های کار ما و همچنین زندگی خصوصی ما نفوذ می کند. این ما را به همان دلیلی می رساند که چرا هر مدیرعاملی باید در مورد هوش مصنوعی چیزی بداند: امکان انصراف وجود ندارد. در سال ۱۹۹۷، در وب سایت اوراکل، بیانیه ای ظاهر شد: "در پنج سال آینده هیچ کس آن را تجارت الکترونیک نخواهد نامید. این به سادگی تجارت خواهد بود." کمی بیشتر طول کشید، اما اتفاق افتاد. به طور مشابه، هوش مصنوعی اختیاری نیست؛ اگر شرکتی آن را اتخاذ نکند، رقبا آن را اتخاذ خواهند کرد، مشتریان ممکن است آن را انتظار داشته باشند و غیره، بنابراین استفاده نکردن از هوش مصنوعی به این معنی نیست که تحت تأثیر آن قرار نخواهید گرفت.

برخلاف آن چیزی که ممکن است از کتاب های درسی مدیریت (فناوری) یا مقالات مجلات علمی بیاموزیم، به نظر می رسد در سازمان های واقعی، مدیران عامل درگیر تصمیمات فناوری اطلاعات و ارتباطات (ICT) هستند. بنابراین، حتی اگر یک مدیر ارشد فناوری اطلاعات (CIO) وجود داشته باشد، چرا او مسئول تصمیم

گیری در مورد ERP (سیستم برنامه ریزی منابع سازمانی) نیست؟ دلیل این است که چنین تصمیماتی هرگز صرفاً تصمیمات فنی نیستند. مدیرعامل شرکت نفت بریتانیا چند دهه پیش گفت که در سرمایه گذاری های فناوری اطلاعات و ارتباطات، حدود ۲۰ درصد هزینه مربوط به سخت افزار و نرم افزار است و ۸۰ درصد آن صرف یادگیری و تغییر فرهنگی می شود. یعنی چنین تصمیماتی بر هسته اصلی کسب و کار تأثیر می گذارد. این نحوه نگرش آنها به خود و محیط اطرافشان را تغییر می دهد. به عبارت دیگر، بر «نظریه بنگاه» (ToF) آنها تأثیر می گذارد. بنابراین، آنچه هر مدیرعاملی باید در مورد هوش مصنوعی بداند این است که چگونه بر ToF آنها و اقدامات ناشی از ToF آنها - یعنی استراتژی کسب و کارشان - تأثیر می گذارد.

هدف این کتاب دقیقا همین است. در حالی که مدیران عامل نیازی به درک جزئیات فنی هوش مصنوعی ندارند، اما می توانند از درک انواع هوش مصنوعی و انتظارات از هر نوع آن بهره مند شوند. به عبارت دیگر، برای یک مدیرعامل مفید است که بتواند از رویاهای اغراق آمیز محققان هوش مصنوعی و ادعاهای بیش از حد فروشندگان هوش مصنوعی عبور کند تا تصمیمات استراتژیک آگاهانه ای بگیرد - و همچنین به طور کلی در مورد نحوه اجرای آن تصمیم گیری کند. تصمیمات نهایی به طور معنادارتری می توانند در سطوح عمیق سلسله مراتب سازمانی گرفته شوند، جایی که دانش مرتبط در آنجا قرار دارد. این مطابق با اصل تفویض معکوس چارلز هندی (۲۰۱۵) است، به این معنی که تصمیمات به جایی تعلق دارند که دانش در آنجا وجود دارد، و اگر به کمک نیاز باشد، متخصصان به روسای خود تفویض اختیار می دهند.

این کتاب نه تنها برای مدیران عامل، بلکه برای هر کسی که با یک مدیرعامل کار می کند یا برای او کار می کند، مفید است. این بدان معناست که محققان هوش مصنوعی و فروشندگان هوش مصنوعی نیز از آن بهره مند خواهند شد، زیرا آنها در مورد زمینه و ارزش های بالقوه و همچنین تله های هوش مصنوعی اطلاعات کسب خواهند کرد. فروشندگان هوش مصنوعی نمی توانند برای مشتری تصمیم بگیرند که به چه منظور و در کجا به هوش مصنوعی نیاز دارد، زیرا آنها از فرآیندهای اصلی کسب و کار مشتریان خود به طور عمیق درک ندارند. مشتریان نیز با این تصمیمات مشکل دارند، زیرا نمی دانند هوش مصنوعی چه چیزی را می تواند ارائه دهد. این مشکل قدیمی اجرای فناوری اطلاعات و ارتباطات است. چیزی که فروخته می شود با چیزی که خریداری می شود متفاوت است. فروشنده قابلیت های محصول را می فروشد، در حالی که مشتریان برای مشکلات خود راه حل خریداری می کنند. برای یک اجرای موفق، این دو باید با هم ترکیب شوند، که نیاز به همکاری بین یک متخصص موضوعی از طرف مشتری و یک متخصص هوش مصنوعی از طرف فروشنده دارد. برای اینکه یک چارچوب مفهومی و فرهنگی را توسعه دهیم که این امکان را فراهم کند، مفید است که مدیرعامل بتواند بین واقعیت و باور در مورد هوش مصنوعی تمایز قائل شود و بدین ترتیب ارزش واقعی هوش مصنوعی را درک کند.

همانطور که در بالا ذکر شد، علاوه بر مدیران عامل، این کتاب عمدتاً برای کسانی که با مدیران عامل کار می کنند مرتبط است .با این حال، به نوعی، همه ما با مدیران عامل کار می کنیم و بیشتر این کتاب در مورد درک تأثیر استراتژیک هوش مصنوعی بر زندگی ما و نحوه بهترین استفاده از آن است .در این معنا، این کتاب

خلاصه ای از آنچه باید هنگام استفاده از یک راه حل هوش مصنوعی و نحوه استفاده از آن برای بهترین بهره برداری - هر چه که باشد - فکر کنیم، ارائه می دهد .بنابراین، هدف این کتاب ارائه پاسخ به خوانندگان نیست، بلکه نشان دادن نحوه به دست آوردن پاسخ های خودشان است .متأسفانه - یا شاید خوشبختانه - راه‌حل‌های کلی و قابل اعمال جهانی وجود ندارند .همانطور که در اکثر موارد با اشاره به استراتژی، همه باید راه خود را پیدا کنند.

۱.۲ عدم قطعیت: بستر استراتژی

برای ایجاد زمینه تفکر استراتژیک در مورد هوش مصنوعی، باید چشم اندازی را که کسب و کارها در آن وجود دارند به عنوان عدم قطعیت ترسیم کنیم، همانطور که فرانک نایت (۱۹۲۱) دوباره مفهوم سازی کرد و توسط جی سی اسپندر در پنجاه سال گذشته گسترش و غنی سازی شد (اسپندر، ۲۰۲۱) .عدم قطعیت با فقدان دانش مشخص می شود (اسپندر، ۲۰۱۴) و این فقدان دانش فضاهای فرصت ایجاد می کند (اسپندر، ۲۰۲۱) . جوهره کار کارآفرینی و مدیریت، ساخت زبانی است که تعامل با فضاهای فرصت را ممکن می سازد و آنها را قابل اجرا و مدیریت می کند .در این زمینه، هوش مصنوعی در صورتی مفید است که به ما در مقابله با این فقدان دانش کمک کند.

نایت مفهوم عدم قطعیت را برای مقایسه با مفهوم ریسک معرفی کرد .در مورد ریسک، اگرچه ما دقیقاً نمی دانیم که در نتیجه یک تصمیم چه اتفاقی خواهد افتاد، اما از تمام گزینه های احتمالی و همچنین احتمال هر یک از گزینه ها که مجموع آنها ۱۰۰ درصد می شود، دانش کاملی داریم .به عبارت دیگر، توزیع احتمال شناخته شده است .در عدم قطعیت، چنین دانش کاملی نداریم .ما با فقدان دانش مواجه هستیم .این فقدان دانش می تواند از انواع مختلف باشد.

مارتین شوبیک (۱۹۵۴) دو جنبه از عدم قطعیت را از هم تفکیک کرد: جهل و عدم قطعیت. جهل معرف چیزی است که در اصل می توان دانست، فقط مدیرعامل آن را نمی داند. این می تواند اطلاعاتی باشد که وجود دارد اما برای مدیرعامل قابل دسترسی نیست، یا ممکن است به دلیل حجم زیاد اطلاعات، دستیابی به آن غیرعملی یا حتی غیرممکن باشد.

در مقابل، عدم قطعیت بازیگران دیگر را وارد بازی می کند. عدم قطعیت در مورد دانشی نیست که دسترسی یا دستیابی به آن دشوار باشد، بلکه در این مورد، دانش وجود ندارد. نمونه های بارز رقبا هستند که اقدامات خاص خود را دنبال می کنند و رفتار آنها به طور کامل تحت تأثیر قوانین بازی قرار نمی گیرد - که منجر به حوزه نظریه بازی ها (۳) می شود که زمینه اصلی شوبیک است. (۴) عدم قطعیت را به طور محاوره ای می توان به عنوان اختیار اراده در نظر گرفت.

اسپندر (۲۰۱۴) مدل را با افزودن ناسنجیدگی، بیشتر گسترش داد؛ یعنی بازیگران و همچنین موارد دانش قابل مقایسه نیستند، زیرا آنقدر متفاوت هستند که حتی ما از اساس مقایسه نیز محروم هستیم. ناسنجیدگی را همچنین می توان منحصر به فرد بودن نامید.

برای بهبود جهت گیری در مورد مفاهیم عدم قطعیت، من و همکار نویسنده ام آلینا باس (درفلر و باس، منتشر نشده) این مفاهیم را در سه حوزه دسته بندی کردیم: شناخته شده، ناشناخته و غیرقابل شناخت. شناخته شده به دانش کامل اشاره دارد، بنابراین شامل ریسک و همچنین مورد خاص ریسک - قطعیت می شود، جایی که فقط یک گزینه با احتمال ۱۰۰ درصد وجود دارد. مهم است که ریسک از قطعیت کمتر شناخته شده نیست. در هر دو مورد دانش کامل وجود دارد، اما در "قطعیت" دانش کاملی از یک نتیجه خاص است، در حالی که در مورد "ریسک" دانش کاملی از توزیع احتمال است. ناشناخته و غیرقابل شناخت با هم عدم قطعیت را تشکیل می دهند که توسط شناخته شده ها محدود می شود. اسپندر اشاره می کند که شناخته شده ها نیز به خوبی که برخی تصور می کنند نیستند، زیرا می توانند "شگفت انگیز، مبهم، ناهنجار، ناسازگار یا متناقض" باشند (اسپندر، ۲۰۲۱: ۱۲۹). ناشناخته با جهل مطابقت دارد؛ چیزی که دانستن آن ممکن است، اما مدیرعامل در حال حاضر آن را نمی داند.

برعکس، غیرقابل شناخت به معنای غیرممکن بودن دانستن است. این شامل منحصر به فرد بودن، اراده آزاد و دانش ضمنی (یا به اصطلاح اسپندر، ناسنجیدگی، عدم قطعیت و بی‌ربطی) می شود.

مهمتر اینکه، عدم قطعیت به این معنا نیست که مدیرعامل هیچ چیز نمی داند. همانطور که شوبیک نوشت: "کاملاً واضح است که به ندرت شرایط اقتصادی وجود دارد که کارآفرین هیچ اطلاعاتی نداشته باشد. اگر چنین وضعیتی وجود داشته باشد، پس اولین اقدام کارآفرین باید کسب اطلاعات باشد" (شوبیک، ۱۹۵۴: ۶۳۲).

برای مثال، مدیرعامل ممکن است از گزینه‌های A-B-C آگاهی داشته باشد و بداند که احتمال وقوع A بیشتر از B و B بیشتر از C است، بدون اینکه بتواند درصدهای دقیقی را اختصاص دهد، و ممکن است گزینه‌های دیگری فراتر از درک مدیرعامل وجود داشته باشد و برخی از آنها محتمل‌تر از A باشند. در اکثر موارد، فرض بر این است که مقررات مالیاتی، برای مثال، تغییر نخواهد کرد؛ یعنی شناخته شده است. شاید مهمتر از آن، بیش از یکی از چهار جنبه عدم قطعیت می تواند و معمولاً به طور همزمان بر مدیرعامل تأثیر بگذارد: اغلب اطلاعاتی در دسترس نیست یا اطلاعاتی وجود دارد که به اندازه کافی سریع قابل پردازش نیست (ناآگاهی)، اغلب بازیگران متعدد یا حتی زیادی (بی‌همتایی) وجود دارند که تصمیمات خود را می‌گیرند (اختیار) و مدیرعامل نسبت به مسیر صنعت، روحیه رقبا و غیره احساساتی دارد (دانش ضمنی).

در جای دیگری پیشنهاد کردیم که هوش مصنوعی می‌تواند ابزار خوبی برای ناشناخته‌ها باشد، اما مدیرعاملان برای مقابله با ناشناختنی‌ها از شهود خود استفاده می‌کنند (دوفلر و باس، ۲۰۲۰ب). ما می توانیم این استدلال را اصلاح کنیم. هوش مصنوعی در قلمرو ناشناخته ها فوق العاده مفید است، اما به تنهایی نمی تواند کار را انجام دهد. همانطور که اسپندر با بلاغت بیان کرد: «شرکت‌ها زمینه‌هایی هستند که در آن‌ها نظریه‌پردازی

مبتنی بر داده به قضاوت کارآفرینانه تبدیل می‌شود. همانطور که با عدم اطمینان‌هایی که فعالیت ما را متوقف می‌کند برخورد می‌کنیم، به جای عقل با تخیل پاسخ می‌دهیم» (اسپندر، ۲۰۲۱: ۱۲۵).

در نهایت، مدیرعامل همچنان باید تصمیم گیری کند، اما هوش مصنوعی می تواند اطلاعات مفیدی را برای آن تصمیمات ارائه دهد، اگر بفهمیم که چگونه از آن به شیوه ای معقول استفاده کنیم. در قلمرو ناشناختنی ها، مدیران عامل باید به شهود خود تکیه کنند، اما هوش مصنوعی ممکن است مفید باشد، برای مثال، در کمک به شناسایی دامنه شهود، به شرطی که یک هوش مصنوعی منحصر به فرد برای سازمان خاص در دست باشد و تابع قضاوت ارزشی مدیرعامل باشد. در بخش ۷ به این نکات باز خواهم گشت.

۱.۳ چرا این کتاب؟

در سال ۲۰۱۷، بررسی مدیریت اسلونِ مؤسسه فناوری ماساچوست (MIT) به همراه گروه مشاوره بوستون، به‌طور مشترک از ۳۰۰۰ مدیر اجرایی، مدیر و تحلیلگر در صنایع مختلف نظرسنجی کردند و با بیش از ۳۰ متخصص و مدیر فناوری، مصاحبه‌های عمیقی انجام دادند (رنس‌باتهام و همکاران، ۲۰۱۷). یافته‌های آن‌ها به شرح زیر است:

«شکاف بین جاه‌طلبی و اجرا در اکثر شرکت‌ها زیاد است. سه چهارم مدیران اجرایی بر این باورند که هوش مصنوعی به شرکت‌هایشان امکان ورود به کسب‌وکارهای جدید را می‌دهد. تقریباً ۸۵ درصد معتقدند هوش مصنوعی به شرکت‌هایشان اجازه می‌دهد تا مزیت رقابتی به دست آورند یا آن را حفظ کنند. اما تنها حدود یک پنجم از شرکت‌ها هوش مصنوعی را در برخی از محصولات یا فرآیندهای خود ادغام کرده‌اند. تنها یک شرکت از هر ۲۰ شرکت، هوش مصنوعی را به‌طور گسترده در محصولات یا فرآیندهای خود گنجانده است. کم‌تر از ۳۹ درصد از کل شرکت‌ها یک استراتژی هوش مصنوعی دارند. بزرگ‌ترین شرکت‌ها - با حداقل ۱۰۰۰۰۰ کارمند - بیشترین احتمال را برای داشتن یک استراتژی هوش مصنوعی دارند، اما تنها نیمی از آن‌ها چنین استراتژی‌ای دارند. (رنس‌باتهام و همکاران، ۲۰۱۷: ۱)»

این موضوع در این کتاب که تا حدودی اما نه کاملاً منحصر به فرد است، مورد بررسی قرار می‌گیرد. این کتاب در مورد پیشرفته‌ترین فناوری است، با این حال غیرفنی است. این کتاب در درجه اول برای مدیران ارشد (و کلاس MBA من) نوشته شده است تا بتوانند در مورد فناوری بیاموزند، اما برای افراد فنی نیز به همان اندازه مفید خواهد بود، زیرا آن‌ها طرز فکر استراتژی کسب‌وکار را در چارچوب هوش مصنوعی کمی بهتر درک خواهند کرد. کتاب‌ها و مقالات متعددی وجود دارند که استفاده از هوش مصنوعی توسط سازمان‌های تجاری را بررسی می‌کنند، اما آن‌ها معمولاً روی آنچه در دسترس است تمرکز می‌کنند و اغلب در چارچوب یک حوزه عملکردی محدود مانند امور مالی یا بازاریابی. در مقابل، این کتاب به کل سازمان نزدیک می‌شود و بر فرآیند تفکر استراتژیک استفاده از هوش مصنوعی تمرکز می‌کند. کتابی که به این رویکرد نزدیک‌تر است، «مزیت هوش مصنوعی» اثر توماس دِونپورت (۲۰۱۸) است؛ با این حال، آن کتاب بیشتر بر جنبه تحلیلی تمرکز دارد و بنابراین این دو کتاب را می‌توان مکمل یکدیگر در نظر گرفت.

این کتاب تصویری کلی از هوش مصنوعی را در سطح نسبتاً بالایی از انتزاع ترسیم می‌کند و اغلب برای سهولت درک، مسائل را اغراق‌آمیز یا ساده‌سازی می‌کند - اما با دقت این کار را انجام می‌دهد تا از نظر اصولی درست باشد. به همین ترتیب، این کتاب بیشتر شبیه یک "تصویر بزرگ با برس ضخیم و ضربات پهن" است؛ یعنی تصویری جامع است که فاقد جزئیات فراوان است، اما از روی آن می‌توان کل تصویر را تشخیص داد. به همین دلیل، اگرچه این کتاب در سراسر متن در مورد هوش مصنوعی صحبت می‌کند، اما شاید به جای فناوری، بیشتر به روانشناسی و فلسفه پرداخته شده باشد. همانطور که اغلب برای ایده‌های نسبتاً جدید و ناقص درک شده که هنوز در حال توسعه هستند، برای درک ماهیت آن‌ها مفید است که آن‌ها را از آنچه نیستند، تفکیک و حتی متضاد کنیم. از آنجایی که این حوزه پیچیده و به سرعت در حال تغییر است، شاید بسیاری از جزئیات از متن حذف شده باشند و همه جزئیات به روز نباشند؛ با این حال، جزئیاتی که در دسترس قرار می گیرند به راحتی جلب توجه همه خواهند شد، بنابراین این موضوع به اندازه اهمیت ندارد. مهم‌ترین نقش چنین کتاب‌های «تصویر بزرگ» این است که به خواننده کمک کند تا تفکر خود را در مورد این مسائل به روشی که برایشان مناسب است، توسعه دهد. بسیاری از استدلال‌ها ارائه می‌شوند، اما مهم‌تر از آن شیوه‌ی ساختار آن‌ها است؛ نشان دادن این موضوع به خواننده در ساختن استدلال خود کمک می‌کند. با داشتن تصویر کلی به عنوان نقطه مرجع، خوانندگان به راحتی می‌توانند نه تنها از آنچه در حال حاضر در مورد هوش مصنوعی وجود دارد، بلکه از آنچه در آینده به وجود خواهد آمد، سر در بیاورند.

این که این کتاب برای مدیران عامل نوشته شده است، نه تنها به محتوا بلکه به سبک آن نیز اشاره دارد .متن به پنج بخش اساسی تقسیم شده است که هر بخش از پنج زیربخش تشکیل شده است .هر بخش شامل چهار زیربخش با حدود ۱۰۰۰ کلمه است و هر کدام را می‌توان به طور مستقل خواند .این به معنای عدم وجود پیشرفت در کتاب یا عدم ارتباط زیربخش‌ها نیست - آن‌ها با هم مرتبط هستند، اما می‌توان آن‌ها را به صورت جداگانه یا چند مورد از آن‌ها را با هم در یک زمان خواند و سپس بعداً ادامه داد .پنجمین زیربخش در هر بخش با عنوان «پس چه؟» نامیده می‌شود؛ این بخش‌ها به نوعی نکات کلیدی هستند؛ چیزی که می‌توانید اگر قصد بحث در مورد هوش مصنوعی را دارید یا به عنوان خلاصه‌ای کوتاه با خود همراه داشته باشید .دلیل طراحی متن به این شکل این بود که سعی شود با برنامه‌های شلوغ خوانندگانم سازگار شود.

بخش اول، «اِکس مَکینا»، فنی‌ترین بخش است؛ این بخش انواع مختلف هوش مصنوعی را معرفی می‌کند . برای مدیران عامل مهم است که تصویری دقیق و کامل از آنچه در این زمینه وجود دارد داشته باشند .سه بخش بعدی با مهم‌ترین جنبه‌های انسانی آنچه با هوش مرتبط می‌دانیم سروکار دارند؛ دانستن، یادگیری و خلق کردن .هر بخش، روش‌های انسانی را با روش‌های ماشینی مقایسه و متضاد می‌کند .مهم‌ترین نکات قابل برداشت از هر بخش، جنبه‌های تصمیم‌گیری در مورد نحوه استفاده از هوش مصنوعی و اینکه در کجا به متخصصان انسانی نیاز داریم، است .آخرین بخش بر جنبه‌های اخلاقی هوش مصنوعی تمرکز دارد .این پویاترین حوزه هوش مصنوعی در حال حاضر است؛ ما هر روز با معضلات اخلاقی جدید مرتبط با هوش مصنوعی مواجه می‌شویم .بنابراین، تا زمانی که خواننده وارد این کتاب شود، مسائل اخلاقی جدیدی وجود خواهد داشت که پوشش داده نشده‌اند - اما امیدوارم بتوانم مبنایی محکم برای مقابله آگاهانه با آن‌ها ارائه

دهم .همچنین اعتقاد دارم که در طی یک دهه آینده، ما بیشتر در مورد هوش مصنوعی از طریق اخلاق، به جای فناوری، خواهیم آموخت .در بخش ۷، به مسائلی که در بخش ۱ مطرح کردم باز خواهم گشت :یعنی اینکه چگونه هوش مصنوعی می‌تواند به مدیران عامل در مقابله با عدم قطعیت کمک کند.

بخش اول، «از ماشین به خدا»، فنی‌ترین بخش است و انواع مختلف هوش مصنوعی را معرفی می‌کند. برای مدیران عامل مهم است که تصویری دقیق و کامل از آنچه وجود دارد داشته باشند. سه بخش بعدی به مهم‌ترین جنبه‌های انسانی آنچه ما با هوش مرتبط می‌دانیم می‌پردازند: دانستن، یادگیری و خلق کردن. در هر بخش، شیوه‌های انسانی با شیوه‌های ماشینی مقایسه و متضاد می‌گردند. مهم‌ترین نکات کلیدی هر بخش، جنبه‌های تصمیم‌گیری در مورد استفاده از هوش مصنوعی و جایی است که به متخصصان انسانی نیاز داریم.

آخرین بخش بر جنبه‌های اخلاقی هوش مصنوعی تمرکز دارد. این پویاترین حوزه هوش مصنوعی در حال حاضر است و ما هر روز با معضلات اخلاقی جدید مرتبط با هوش مصنوعی مواجه می‌شویم. بنابراین، تا زمانی که خواننده وارد این کتاب شود، مسائل اخلاقی جدیدی وجود خواهد داشت که در این کتاب پوشش داده نشده است - اما امیدوارم بتواند پایه محکمی برای برخورد آگاهانه با آن‌ها فراهم کند. همچنین بر این باورم که در حوزه اخلاق، نه فناوری، است که ما در طول دهه آینده بیشترین چیزها را در مورد هوش مصنوعی خواهیم آموخت. در بخش هفتم، به موضوعاتی که در بخش ۱ مطرح کردم باز خواهم گشت: یعنی اینکه هوش مصنوعی چگونه می‌تواند به مدیران عامل در مقابله با عدم قطعیت کمک کند.

۲. از ماشین به خدا

این بخش بر انواع مختلف هوش مصنوعی تمرکز دارد و هر کدام در چارچوب تاریخی مربوط به خود توضیح داده می‌شود .دلیل گنجاندن این دیدگاه تاریخی این است که به ما کمک می‌کند تا درک کنیم که انواع خاص هوش مصنوعی چه چیزی را باید انجام می‌دادند، چگونه انتظار می‌رفت که کار کنند و دامنه اعتبار مورد انتظار آنها چه بود .به عبارت دیگر، ما باید مفروضات پشت انواع خاص هوش مصنوعی را درک کنیم و ببینیم چه چیزی از زمان پیدایش آنها تغییر کرده است.

این بخش همچنین به ما کمک می‌کند تا واژگان هوش مصنوعی کاربردی را توسعه دهیم، مفاهیمی مانند "یادگیری عمیق" را رمزگشایی کنیم و پیشینه فناوری را با اصطلاحات غیرتخصصی نسبتاً قابل فهم توضیح دهیم .اگرچه این امر برای مدیران عامل برای تفکر در مورد هوش مصنوعی ضروری نیست، اما هنگام صحبت در مورد هوش مصنوعی، به ویژه با افرادی که "متخصصان فناوری بومی" هستند و با فناوری اطلاعات امروزی آشنا به دنیا آمده‌اند، بسیار مفید است .یک مطالعه مکمل مهم برای این بخش، بررسی شخصی پاملا مک‌کورداک (۲۰۰۴) از تاریخ هوش مصنوعی است که ریشه در مصاحبه با بسیاری از گوروهای اولیه هوش مصنوعی دارد .من دیدگاه‌های متفاوتی دارم و در ادامه به آنها می‌پردازم، اما غنای داستان مک‌کورداک قابل

تحسین است .این بخش همچنین برخی از تاریخ را بازگو می‌کند، به تفصیل در مورد خاستگاه انواع مختلف هوش مصنوعی می‌پردازد و به کشف مفروضات اساسی در مورد آنها کمک می‌کند.

۲.۱: سیستم‌های استدلال نمادین

بر اساس افسانه‌های هوش مصنوعی، تاریخچه هوش مصنوعی در ژانویه ۱۹۵۶ آغاز شد، زمانی که هربرت سایمون پس از بازگشت از تعطیلات سال نو، اولین کلاس خود را با عنوان «مدل‌های ریاضی در علوم اجتماعی» در مؤسسه فناوری کارنگی (امروزه دانشگاه کارنگی ملون) تدریس کرد. همانطور که او به دانشجویانش گفت، این یک تعطیلات کاری بود: «در تعطیلات کریسمس، من و آل نیول یک ماشین متفکر اختراع کردیم» (سایمون، ۱۹۹۱: ۲۰۶). ادوارد فایگنباوم یکی از دانشجویان این کلاس بود که تعریف می‌کند این جمله او را چنان نسبت به این موضوع هیجان‌زده کرد که دانشجوی دکترای سایمون شد (فایگنباوم، ۱۹۹۲: ۳). بعداً فایگنباوم را به عنوان پدر نوع دیگری از هوش مصنوعی می‌شناسیم، اما فعلاً با نیول و سایمون پیش می‌رویم.

«ماشین متفکری» که نیول و سایمون به همراه کلیف شاو از شرکت راند ایجاد کردند، نظریه‌پرداز منطق یا ماشین نظریه منطق نامیده می‌شد (نیول و سایمون، ۱۹۵۶). فقط برای اشاره به وضعیت فناوری در آن زمان، نیول و سایمون مجبور بودند برای انجام این کار به لس آنجلس پرواز کنند، زیرا شرکت راند در آنجا به آن‌ها دسترسی به یک کامپیوتر – یک IBM 7017 پرقدرت با کمتر از ۲۰ کیلوبایت حافظه (لپ‌تاپ شما حدود یک میلیون برابر بیشتر دارد) که از لامپ‌های خلأ غیرقابل اعتماد استفاده می‌کرد و میلیون‌ها دلار هزینه داشت – دانشگاه‌ها کامپیوتر نداشتند؛ آن‌ها توانایی خرید آن را نداشتند.

هدف نظریه‌پرداز منطق جاه‌طلبانه بود: «این برای یادگیری چگونگی حل مشکلات دشوار مانند اثبات قضایای ریاضی، کشف قوانین علمی از داده‌ها، بازی شطرنج یا درک معنای نثر انگلیسی طراحی شده بود» (نیول و همکاران، ۱۹۶۳: ۱۰۹). با ادامه‌ی این خط فکری، پروژه ماشین نظریهٔ منطق به حل‌کننده عمومی مسائل (GPS) با جاه‌طلبی بیشتر تبدیل شد، که قرار بود این دامنه را به هر حوزه و همه حوزه‌های حل مسئله انسان گسترش دهد. این نوع هوش مصنوعی اولین تلاش برای چیزی بود که امروزه به عنوان «هوش مصنوعی گسترده» یا در حالت افراطی، هوش مصنوعی عمومی (آگزی) شناخته می‌شود.

پشت‌بندهٔ ماشین نظریه منطق و حل‌کننده عمومی مسائل (GPS) ایده‌ای مستحکم وجود داشت، البته به عنوان یک ایده. هدف اولیه سایمون و نیول درک یا «مدل‌سازی» ذهن انسان بود. آن‌ها از یک شبیه‌سازی رایانه‌ای استفاده کردند. با گوش دادن به ارائهٔ الیور سلفریج (۱۹۵۵) دربارهٔ دستکاری نماد و بازشناسی الگو در رایانه‌ها (نیول می‌گفت که این را می‌توان هوش مصنوعی در نظر گرفت: به مک‌کورداک، ۲۰۰۴ مراجعه کنید)، نقشه‌ای جاه‌طلبانه‌تر در ذهن آلن نیول شکل گرفت. اگر تمام حل‌مسئله‌های انسان را بتوان به عنوان دستکاری نماد نشان داد، و اگر رایانه‌ها بتوانند نمادها را دستکاری کنند و الگوها را شناسایی کنند، پس ماشین‌ها باید بتوانند مشکلات دنیای واقعی را حل کنند، نه فقط مسائل حسابی. این امر با دنبال کردن مراحل

حل مسئله انسان برای تکرار نتایج (یعنی راه‌حل‌های مشکلات) به دست می‌آمد. ماهیت این ایده این بود که با گردآوری مراحل حل مسئله در بسیاری از زمینه‌ها، استخراج اصول کلی استدلال امکان‌پذیر شود و با به کارگیری این اصول و مراحل عمومی، هوش مصنوعی نه تنها در یک حوزهٔ محدود و تعریف‌شده، بلکه در بسیاری از حوزه‌ها کار می‌کرد. برای ثبت مراحل حل مسئله انسان، از متخصصان خواسته شد تا از تکنیک «فکر کردن با صدای بلند» استفاده کنند؛ یعنی بگویند و ضبط کنند که به چه چیزی فکر می‌کنند. قابل توجه است که از آن‌ها انتظار می‌رفت حتی اشتباهاتی را که مرتکب می‌شدند نیز ذکر کنند، زیرا این اشتباهات می‌توانستند بخش‌های مفید یا حتی ضروری کل فرایند حل مسئله باشند.

این ایده درخشان بود و انحرافی رادیکال از رویکردهای حل مسئله رایانه‌ای آن زمان به شمار می‌رفت. رویکرد «عادی» این بود که یک مسئله را به طور کامل و با جزئیات کافی در نظر بگیریم که به نوعی روش بهینه‌سازی یا پژوهش عملیاتی (OR) اجازه می‌داد. در مقابل، رویکرد جدید نوید این را می‌داد که بتواند در فرآیند حل مسئله حتی زمانی که مشخص نبود راه‌حلی وجود دارد یا چند راه‌حل می‌تواند وجود داشته باشد، درگیر شود. عملکرد ماشین نظریه منطق شگفت‌انگیز بود؛ در نهایت، ۳۸ مورد از ۵۲ قضیه اول در اصول ریاضیات را اثبات کرد (سایمون، ۱۹۹۵) و سایمون اثبات قضیه ۲.۸۵ را نسبت به اثباتی که توسط وایتهد و راسل ارائه شده بود، ظریف‌تر یافت. راسل با شنیدن این موضوع بسیار خوشحال شد (مک‌کورداک، ۲۰۰۴: ۱۶۷، پاورقی).

با وجود دستاوردهای شگفت‌انگیز ماشین نظریه منطق، و حتی خالقان تحسین‌برانگیزتر آن، اما برخی از مفروضات و پیامدهای آن مشکل‌ساز هستند. فرض اساسی این بود که اصول کلی حل مسئله وجود دارد؛ برخی خطوط استدلال که تنها در جزئیات با هم تفاوت دارند. با در نظر گرفتن وسعت حوزهٔ حل مسئله انسان، آیا واقعاً می‌توانیم انتظار داشته باشیم که برخی مراحل کلی وجود داشته باشد که از بوسیدن تا پختن خورش و همچنین طراحی ماشین یا آهنگسازی قابل اعمال باشد؟ من باور ندارم که چنین اصولی وجود داشته باشد، و اگر هم وجود داشته باشد، آنقدر کلی و بی‌فایده خواهند بود که واقعاً نتوانیم آن‌ها را به کار گیریم. در واقع، هنگامی که از یادگیری ماشین برای کشف قوانین ترجیحات کتاب یک شخص خاص استفاده می‌شود، مشخص می‌شود که برای دسته‌بندی‌های «فانتزی» و «علمی-تخیلی» قوانین متفاوتی اعمال می‌شود، و این به غیر از قوانین کاملاً متفاوتی است که باعث می‌شود یک رمان جنایی خوب در ذهن همان فرد شکل بگیرد. هیچ چیز نشان نمی‌دهد که حتی برای یک نفر، چه برسد به همه ما، یک روش کلی برای حل مسئله وجود داشته باشد.

در مرحله بعد، در حالی که به نظر می‌رسد استدلال را می‌توان از طریق دستکاری نمادها نشان داد، اما این یک واقعیت نیست. در هر خط استدلال (یا حداقل اکثر آن‌ها) توالی‌ای از مراحل منطقی وجود دارد که می‌توان از این طریق با آن‌ها برخورد کرد، اما قبل از اعمال آن مراحل، ما پیش‌فرض‌هایی را قائل می‌شویم — این پیش‌فرض‌ها اغلب به صراحت بیان نمی‌شوند و بنابراین ممکن است توصیف آن‌ها به همان روشی که مراحل استدلال توصیف می‌شوند، امکان‌پذیر نباشد. با این حال، اگر همه این پیش‌فرض‌ها درست باشند (اگرچه توجیه نشده باشند، باز هم می‌توانند درست باشند)، هیچ دلیلی وجود ندارد که نشان دهد بازی شطرنج یا

اثبات قضایا، که ممکن است استدلال‌های منطقی صریح را شامل شوند، با «درک معنای نثر انگلیسی» یا هر چیزی که شامل درک معنا باشد، وجه اشتراکی داشته باشد. برای انصاف، ما حتی نمی‌دانیم «درک معنا» به چه معناست (مقایسه کنید با وینوگراد، ۱۹۸۰) – چگونه می‌توانیم این کار را به یک رایانه آموزش دهیم؟ این نکته در بخش بعدی (زیربخش ۳.۴) هنگام بحث در مورد مسئلهٔ دانش در هوش مصنوعی دوباره مورد بررسی قرار خواهد گرفت.

ماشین نظریه منطق دستاورد شگفت‌انگیزی از خلاقیت انسان بود. این ماشین عملکرد چشمگیری را ارائه کرد، به غیر از اینکه به اولین نمونه عملیاتی هوش مصنوعی در تاریخ تبدیل شد. با این حال، این دستاورد که اکنون مورد ستایش قرار می‌گیرد، از همان ابتدا به طور واضح مورد تحسین قرار نگرفت. در همان سالی که سایمون اختراع «ماشین متفکر» را گزارش کرد، یک کارگاه آموزشی دو ماهه در کالج دارتموث برگزار شد. این کارگاه به خودی خود، یک نقطه عطف در تاریخ هوش مصنوعی به شمار می‌رفت، جایی که مک‌کارتی اصطلاح «هوش مصنوعی» را ابداع کرد. نیول و سایمون ماشین نظریه منطق را ارائه کردند. واکنش به «ماشین متفکر» آن‌ها کمرنگ بود. سایمون گفت که او و نیول قبلاً کاری را انجام داده بودند که دیگران در کنفرانس فقط در مورد آن صحبت می‌کردند، با این حال به نظر نمی‌رسید که تحت تأثیر قرار گرفته باشند.

۲.۲ سیستم‌های متخصص نمادین

پس از دریافت مدرک دکترا (جزئیات بیشتر در زیربخش ۳.۴)، فایگنبوم دور از تأثیر مستقیم سایمون و نیول به برکلی نقل مکان کرد. او با امید به ساخت ماشینی فوق هوشمند با کارایی برابر یا حتی برتر از کارشناسان انسانی، به علاقه قدیمی خود به عملکرد ماشین بازگشت. بنابراین، او اصول کلی استدلال و مدل‌های پردازش اطلاعات را کنار گذاشت و بر اکتشافات تجربی و حوزه‌های محدود تخصص تمرکز کرد. او دریافت که از آنجایی که کارشناسان هستند که عملکرد بالایی را در حوزه‌های مربوطه خود ارائه می‌دهند، دانش تخصصی باید نقش مهمی ایفا کند.

او نتیجه گرفت که رایانه باید نمایشی از مسئله داشته باشد؛ یک "مدل" داخلی از محیط خارجی که در حال استدلال است. او حدس زد که چنین نمایش‌هایی از دانش باید استقرایی باشد. روش‌های رایج هوش مصنوعی در استدلال قیاسی خوب بودند ولی در استقرایی نه، بنابراین او تصمیم گرفت روی استقراء تمرکز کند. سپس او فکر کرد که باید با دانشمندانی کار کند که آنها را به عنوان "استنتاج کنندگان حرفه ای" در نظر می‌گرفت.

این توصیف ممکن است مانند یک توالی از مراحل منطقی به نظر برسد، اما به نظر می‌رسد بیشتر یک جهش شهودی بوده است: "اینها شهودها و حدس‌های بیش از حد ساده بودند. آنها می‌توانستند کاملاً اشتباه باشند. اما درست بودند. شهود "کار با دانشمندان" یکی از پربارترین‌هایی بود که تا به حال داشته‌ام" (فایگنبوم، ۱۹۹۲: ۷).

با شروع کار با دانشمندان با تمرکز بر استقراء، فایگنبوم پروژه خود را بر مبنای طراحی آزمایش بنا کرد؛ سپس هر ماشینی که او و تیمش تولید می‌کردند می‌توانست به طور مناسب آزمایش شود. او روی مسئله تشکیل فرضیه تمرکز کرد و به دنبال شریک و وظیفه‌ای در دنیای علوم طبیعی گشت. او در گردهمایی علاقه‌مندان به هوش مصنوعی در استنفورد، با جاشوا لدربرگ، برنده جایزه نوبل و رئیس بخش ژنتیک استنفورد، ملاقات کرد. آنها شروع به بحث در مورد اینکه چه کاری می‌توانند با هم انجام دهند کردند، و چند ماه بعد، فایگنبوم به استنفورد نقل مکان کرد.

لدربرگ به جای موضوعی از زمینه تخصصی خودش یعنی ژنتیک، موضوعی از اخترزیست‌شناسی (مطالعه حیات فرازمینی) را پیشنهاد کرد، به طور خاص در مورد استنتاج توپولوژی‌های مولکولی از طیف جرمی. این پروژه از کاوشگر مریخ برای یافتن حیات یا پیش سازهای حیات در مریخ حمایت می‌کرد. داده‌های ورودی قابل اعتماد بودند، محاسبات مورد نیاز فراتر از حد معمول بود اما خارج از توانایی‌های هوش مصنوعی آن زمان نبود، و فرصتی برای آزمایش تجربی یافته‌ها وجود داشت.

پروژه دِندرال دو مرحله را انجام داد. اول، تعداد زیادی از ساختارهای "از نظر توپولوژیکی مجاز" (بر اساس ظرفیت) تولید کرد، و سپس آنها بر اساس طیف جرمی به ساختارهای "از نظر شیمیایی معقول" فیلتر شدند. مرحله دوم نیاز به تخصص قابل توجهی در تفسیر طیف جرمی داشت. این پروژه سال‌های زیادی به طول انجامید و کارشناسان بیشتری و بیشتری را درگیر کرد و در نهایت، ناحیه‌ای وسیع‌تر از دانش هر یک از کارشناسان فردی را پوشش داد. در نهایت، برنامه در سطح کارشناسان برتر عمل می‌کرد و از آنها پیشی می‌گرفت.

شاید فکر کنید دلیل پرداختن به جزئیات شکل‌گیری دِندرال علاقه‌ی من به تاریخچه‌ی این فرآیند است، هرچند خودم هم از این داستان لذت می‌برم. اما دلیل اصلی این است که بتوانم به این نکته اشاره کنم که از زمان اولین سیستم متخصص مبتنی بر دانش، چیز زیادی تغییر نکرده است.

فایگنبوم در انتخاب مسئله‌ای مناسب برای ساخت این نوع هوش مصنوعی بسیار دقیق عمل کرد. مسئله باید از نظر اندازه و پیچیدگی در سطح مناسب می‌بود، کارشناسان در دسترس بودند و امکان بررسی نتایج وجود داشت (مقایسه کنید با ولنسی، ۲۰۱۷). با صرف زمان قابل توجهی در ۲۵ سال گذشته برای پشتیبانی از مدیران با استفاده از یک سیستم متخصص مبتنی بر دانش، می‌توانم تأیید کنم که این دقیقاً همان روشی است که ما هنوز از آن استفاده می‌کنیم.

کارشناسان ضروری هستند؛ بدون کارشناس، سیستم متخصصی وجود ندارد. اگر مشکلی را انتخاب کنیم که بیش از حد پیچیده باشد، کارشناسان نمی‌توانند دانش خود را برای مدل‌سازی ما به کار گیرند. این اتفاق برای ما زمانی افتاد که به دنبال امکان ساخت یک سیستم متخصص برای نگهداری نیروگاه هسته‌ای بودیم. اگر

مسئله خیلی ساده باشد، چیز زیادی برای قرار دادن در پایگاه دانش وجود ندارد، به علاوه اینکه ممکن است اصلاً کارشناسی برای آن وجود نداشته باشد.

علاوه بر این، لدربرگ شریک فوق‌العاده‌ای بود؛ او موضوعی را پیشنهاد کرد که کاملاً مناسب با هدف بود، موضوعی مبتنی بر پروژه‌ی خودش اما نه از حوزه‌ی تخصصی او (ژنتیک). در واقع، سال‌ها بعد او پیشنهاد کرد که ژنتیک به نقطه‌ای رسیده است که ساخت یک سیستم متخصص برای آن منطقی به نظر می‌رسد (فایگنبوم، ۲۰۰۶: ۲۱:۴۵-۲۰:۵۰).

پروژه دِندرال به اولین سیستم متخصص مبتنی بر دانش تبدیل شد.

اصطلاح "مبتنی بر دانش" برای یک سیستم به مدل داخلی به نام "نمایش دانش" اشاره دارد که در قالب یک "پایگاه دانش" ذخیره می‌شود، در حالی که اصطلاح "سیستم متخصص" به این معناست که دانش کارشناسان مدل‌سازی شده است. در شکل اولیه‌ی خود، نمایش دانش در دِندرال به صورت لیست‌هایی در یک زبان تخصصی به نام لیسپ ذخیره می‌شد. با این حال، با اضافه شدن دانش بیشتر و بیشتر، پیچیدگی افزایش یافت و شروع به تهدید پایداری سیستم کرد. در همین حال، نیول و سایمون در مدل‌سازی حافظه پیشرفت بیشتری کردند و مدلی از حافظه به نام "تولیدات" را توسعه دادند. با اتخاذ این رویکرد، بروس بوکانن، همکار دیرینه‌ی فایگنبوم، سیستم را دوباره برنامه‌نویسی کرد و استاندارد پایگاه‌های دانش را ایجاد کرد؛ "قوانین تولید" یا قوانین "اگر ... سپس" که به صورت سلسله مراتبی سازماندهی شده‌اند.

نمایش دانش در یک سیستم متخصص در فرآیند کسب دانش به دست می‌آید که زیرمجموعه‌ای از فرآیند کلی مهندسی دانش است .فایگنبوم (۱۹۷۷) برای تأکید بر نقش مرکزی و همچنین نشان دادن ماهیت مهندسی دانش، آن را «هنر هوش مصنوعی» نامید .فایگنبوم از همان ابتدا بر این باور بود که نمایش دانش مهم‌ترین بخش سیستم متخصص است .او این موضوع را در «فرضیه دانش قدرت است» مطرح کرد؛ بعداً ادعا کرد که شواهد زیادی برای حمایت از این فرضیه وجود دارد به حدی که آن را به «اصل دانش» تغییر داد (لنت و فایگنبوم، ۱۹۹۱).

برخلاف سیستم‌های استدلال، به نظر نمی‌رسد در راه‌اندازی سیستم‌های متخصص هیچ فرضیه‌ی نادرستی وجود داشته باشد .علاوه بر این، به نظر می‌رسد فایگنبوم همه چیز را در همان بار اول درست انجام داده است، زیرا از آن زمان تاکنون تغییرات کمی ایجاد شده است.

فایگنبوم (۱۹۹۲ :۱۶) درباره‌ی آینده‌ی سیستم‌های متخصص به دو محدودیت اشاره می‌کند :شکنندگی و انزوا .مورد اول به این معناست که اگرچه سیستم‌های متخصص در حوزه‌های محدود خود عملکرد بالایی ارائه می‌دهند، اما حتی با یک قدم فراتر رفتن از مرز حوزه‌ی خود کاملاً بی‌استفاده می‌شوند، زیرا در آن صورت هیچ دانش غیرتخصصی‌ای برای تکیه کردن وجود ندارد .مورد دوم به این معنی است که سیستم‌های متخصص هرگز برای همکاری و حل مشکلات بزرگتر و متفاوت گرد هم نیامده‌اند.

بر اساس این دو محدودیت، او حدس زد که عصر دوم سیستم‌های متخصص شامل پایگاه‌های دانش بزرگ، اشتراک دانش و قابلیت همکاری بین پایگاه‌های دانش پراکنده جغرافیایی خواهد بود .با ظهور اینترنت، مشکل پراکندگی جغرافیایی ناپدید شده است .قابلیت همکاری نیز دیگر مشکلی نیست؛ ما می‌توانیم پایگاه‌های دانش را به هم متصل کنیم و حتی از یک پایگاه دانش به عنوان ورودی برای پایگاه دانش دیگر استفاده کنیم .با این حال، همچنان باید به مشکل پایگاه‌های دانش بزرگ و اشتراک دانش توجه داشته باشیم (این موارد در زیربخش ۴.۴ مجدداً مورد بحث قرار خواهند گرفت).

هر دو سیستم نمادین - سیستم‌های استدلال و سیستم‌های متخصص - بر دستکاری نمادها تکیه دارند و برای به دست آوردن مراحل استدلال یا نمایش دانش به ترتیب، به نوعی فرآیند کسب دانش نیاز دارند .شکاف بزرگی بین این دو نوع هوش مصنوعی نمادین وجود دارد .در حالی که هدف سیستم‌های استدلال ایجاد یک سیستم "هوشمند عمومی" مستقل از حوزه (یک هوش مصنوعی عمومی) بود، هدف سیستم‌های متخصص دستیابی به عملکردی استثنایی در حوزه‌های محدود و به خوبی تعریف‌شده از طریق مدل‌سازی دانش خاص-حوزه‌ی کارشناسان است.

سیستم‌های متخصص مبتنی بر دانش تا اواسط دهه ۱۹۸۰ بر عرصه هوش مصنوعی تسلط داشتند و تعداد زیادی از پیاده‌سازی‌های موفق را در طیف وسیعی از زمینه‌ها، از علوم و تولید تا درمان‌های پزشکی، به ارمغان آوردند .در همین زمان، رویکرد استدلال نمادین هیچ پیاده‌سازی‌ای را که بتواند با موفقیت در چندین حوزه کار کند، تولید نکرد، اگرچه برخی از پیاده‌سازی‌ها در حوزه‌های محدود موفق بودند.

۲.۳ شبکه‌های عصبی مصنوعی

در حالی که رویکردهای نمادین به هوش مصنوعی (ابتدا سیستم‌های استدلال و بعداً سیستم‌های متخصص) بر روزهای اولیه هوش مصنوعی تسلط داشتند، رویکرد اتصال‌گرا در نخستین تجسم خود، زودتر توسعه یافته بود. وارن مک‌کالاک و والتر پیتس (۱۹۴۳) با الهام گرفتن از فیزیولوژی و عملکرد نورون‌ها، منطق گزاره‌ای وایتهد و راسل (۱۹۲۷) و نظریه محاسبه تورینگ (۱۹۳۷) (برای جزئیات بیشتر به راسل و نورویگ، ۲۰۲۰ مراجعه کنید)، مدل‌هایی از نورون‌های مصنوعی ارائه کردند. مک‌کالاک و پیتس نشان دادند که هر تابعی را می‌توان با یک شبکه مناسب از نورون‌های مصنوعی محاسبه کرد و همچنین خاطرنشان کردند که چنین شبکه‌هایی در صورت ساخت مناسب می‌توانند «یاد بگیرند».

اولین پیاده‌سازی شبکه عصبی مصنوعی (ANN)، ماشین حساب تقویتی آنالوگ عصبی تصادفی (SNARC)، توسط ماروین مینسکی و دین ادموندز به عنوان بخشی از کارهای کارشناسی خود در پرینستون در سال ۱۹۵۰ انجام شد. هدف شبکه‌های عصبی مصنوعی اولیه نشان دادن این بود که هر تابع ریاضی را می‌توان با یک شبکه مناسب از نورون‌های مصنوعی محاسبه کرد؛ بنابراین، به جای اینکه آنها را به عنوان هوش مصنوعی در نظر بگیریم، شاید بهتر باشد آنها را به عنوان «پیش هوش مصنوعی» توصیف کنیم.

شبکه‌های عصبی مصنوعی تا اواسط دهه ۱۹۸۰ تا حدودی کنار گذاشته شدند، اما در این زمان بازگشت بزرگی داشتند. دلایل متعددی برای زمان‌بندی این بازگشت وجود داشت.

اول اینکه، شبکه‌های عصبی مصنوعی اولیه با فناوری آن زمان بسیار انرژی‌بر بودند. ماشین حساب تقویتی آنالوگ عصبی تصادفی (SNARC) از ۴۰ نورون شبیه‌سازی‌شده و «۳۰۰۰ لامپ خلأ و یک مکانیزم خلبان خودکار اضافی از یک بمب‌افکن B-24» (راسل و نورویگ، ۲۰۲۰) استفاده می‌کرد. تا اواسط دهه ۱۹۸۰، رایانه‌ها به اندازه‌ی کافی سریع شده بودند که بتوانند با شبکه‌های عصبی مصنوعی با اندازه‌ی مناسب کار کنند.

علاوه بر این، تا اواسط دهه ۱۹۸۰، رویکرد نمادین انتظارات را برآورده نمی‌کرد. سیستم‌های متخصص موفق بودند، اما تنها در حوزه‌های محدود. توسعه‌ی آن‌ها زمان زیادی طول می‌کشید (اغلب زمان بسیار زیادی) و بسیاری از آن‌ها تا زمانی که مشکل حل می‌شد، تصمیم گرفته می‌شد، یا هر هدف دیگری که سیستم متخصص خاص برای آن ساخته شده بود، محقق می‌شد، منسوخ شده بودند. امروزه، سیستم‌های متخصص عمدتاً در پشتیبانی از تصمیمات استراتژیک، جایی که برچسب قیمتی اجازه‌ی هزینه‌ی اضافی را می‌دهد، یا برای تصمیمات روتین که به‌طور قابل اعتمادی در طی تعداد بسیار زیادی از تکرارها تغییر نمی‌کنند، استفاده می‌شوند.

سیستم‌های استدلال نمادین، اگرچه با هدف کار در چندین حوزه عمل می‌کردند، ایده‌ی هوش مصنوعی عمومی (AGI) را دنبال می‌کردند، اما تنها در حوزه‌های محدود نتایج مثبتی به دست آوردند، برای مثال، بازی شطرنج.

امروزه، شبکه‌های عصبی مصنوعی پرکاربردترین شکل هوش مصنوعی هستند؛ آن‌ها بخش زیادی از عرصه‌ی هوش مصنوعی را به خود اختصاص داده‌اند، از جمله مفهوم یادگیری ماشین (ML).

بنابراین، در بخش بعدی توضیحی بسیار ساده در مورد اینکه شبکه‌های عصبی مصنوعی چه هستند و چگونه کار می‌کنند (بر اساس دلفلِر، ۲۰۲۰) ارائه خواهد شد.

هر شبکه عصبی مصنوعی (ANN) از سه مجموعه از نورون‌های مصنوعی تشکیل شده است. اولین لایه، لایه ورودی است که سیگنال (تحریک) را دریافت می‌کند - این تحریک می‌تواند تقریباً هر چیزی باشد که به صورت الکترونیکی ارائه می‌شود. سپس یک لایه پنهان وجود دارد که وظیفه ترجمه یا انجام عملیات «جعبه سیاه» بین ورودی و خروجی را بر عهده دارد. در نهایت، لایه خروجی پاسخی را برای محرکی که توسط لایه ورودی دریافت شده است، تولید می‌کند. امروزه اصطلاحاتی مانند «شبکه عصبی عمیق» یا «یادگیری عمیق» را می‌شنویم که مرموز و هیجان‌انگیز به نظر می‌رسند؛ در واقع، این اصطلاحات به سادگی به این معنا هستند که بیش از یک لایه از نورون‌های مصنوعی در لایه پنهان وجود دارد (لکان و همکاران، ۲۰۱۵). از نظر محاسباتی، تعداد لایه‌ها ممکن است تفاوت ایجاد کند، اما از نظر منطقی مفاهیم ساده هستند.

نورون‌های مصنوعی مانند سیناپس‌های مغز انسان (یا سایر موجودات زنده) در سراسر لایه‌ها به هم متصل هستند، از این رو به آن «رویکرد اتصال‌گرا» می‌گویند. هر اتصال دارای یک وزن اولیه است، بنابراین هنگامی که محرک سیستم را «روشن می‌کند»، سیگنال به سمت نورون‌های خروجی منتشر می‌شود و در نتیجه پاسخی را ایجاد می‌کند. پاسخ با آنچه انتظار می‌رفت مقایسه می‌شود، وزن‌ها تنظیم می‌شوند و این فرآیند به صورت تکرارشونده انجام می‌شود. از آنجایی که پردازش تعداد زیادی از نمونه‌های یادگیری (اعم از محرک‌ها و پاسخ‌های مورد نظر) آسان است، شبکه عصبی مصنوعی می‌تواند به نسبت سریع برای تولید پاسخ دلخواه تنظیم شود. به عبارت دیگر، شبکه‌های عصبی مصنوعی از تعداد زیادی نمونه یادگیری نحوه بازتولید فرکانس آماری آنها را یاد می‌گیرند.

از آنجایی که نورون‌های مصنوعی بر اساس نورون‌های بیولوژیکی مدل‌سازی شده‌اند، شبکه عصبی مصنوعی به طور فرضی شبیه مغز انسان است. بنابراین، استدلال بر این است که یک شبکه عصبی مصنوعی به اندازه کافی بزرگ باید قادر به تفکر باشد. علاوه بر این، از آنجایی که نورون‌های دیجیتال سریع‌تر از نورون‌های بیولوژیکی هستند، بر اساس همین منطق، این ماشین‌ها باید از انسان‌ها باهوش‌تر باشند. بیایید حقایق این استدلال را از باورهای آن جدا کنیم (برای آگاهی از موردی با شور و شوق بیشتر اما بسیار شفاف در ترسیم حقایق، باورها و امیدها، به اولمن، ۲۰۱۹ مراجعه کنید).

در مورد شبکه‌های عصبی مصنوعی (ANN)، اندازه اهمیت دارد. مغز انسان از حدود ۸۰ تا ۱۰۰ میلیارد نورون تشکیل شده است، که به طور متوسط هر کدام حدود ۷۰۰۰ اتصال دارند و در مجموع منجر به ۷۰۰ تا ۱۰۰۰ تریلیون (۱۰ به توان ۱۵) سیناپس می‌شود. بزرگترین شبکه‌های عصبی مصنوعی امروزه شامل حدود ۱۶ میلیون نورون مصنوعی هستند که تقریباً با اندازه مغز یک قورباغه مطابقت دارد. از نظر تعداد سیناپس‌ها، ممکن است ۸ تا ۱۰ مرتبه بزرگی از شبکه مغز انسان عقب باشیم.

مورد مهم دیگر این است که آموزش یک شبکه عصبی مصنوعی با ۱۶ میلیون نورون مصنوعی، حتی با سریع‌ترین ابررایانه‌های امروزی، زمان زیادی طول می‌کشد. علاوه بر این، چنین شبکه‌ای آنقدر پیچیده می‌شود که طراح آن دیگر نمی‌تواند آن را درک کند - تنها راه برای تجزیه و تحلیل چنین شبکه عظیمی استفاده از هوش مصنوعی است. بنابراین، بعید به نظر می‌رسد که صرفاً مساله زمان باشد تا ما بتوانیم شبکه‌ای عصبی مصنوعی در اندازه مشابه مغز انسان تولید کنیم، حتی با وجود افزایش ۶ مرتبه بزرگی در طول ۷۰ سال گذشته.

ساختار شبکه‌های عصبی مصنوعی نیز قابل توجه است. در شبکه‌های عصبی مصنوعی، اتصالات معمولاً فقط به جلو می‌روند، به این معنی که هر نورون مصنوعی فقط به نورون‌های لایه بعدی متصل می‌شود (در برخی شبکه‌ها ممکن است اتصالات درون لایه نیز وجود داشته باشد). مشخص نیست که آیا در مغز انسان اتصالات دایره‌ای وجود دارد یا خیر، هرچند ممکن است به زودی متوجه شویم. علاوه بر این، حتی ممکن است لازم

نباشد که مغز انسان یا سایر موجودات را بر اساس «لایه» در نظر بگیریم. به عبارت دیگر، از نظر ساختاری، شبکه‌های عصبی مصنوعی شبیه مغز انسان هستند زیرا با تعداد زیادی از عناصر ساده در شبکه پیچیده‌اند، اما نگاه دقیق‌تر به ساختار نیز تفاوت‌هایی را نشان می‌دهد.

حالا جنبه‌های عملکردی شبکه‌های عصبی مصنوعی (ANN) را در نظر بگیرید. عملکرد نورون‌های مصنوعی بازتاب‌دهنده‌ی دانش ما در اواسط قرن بیستم در مورد نورون‌های بیولوژیکی است: اینکه نورون‌ها یا «فعال» هستند یا نیستند، شبیه به رایانه‌های دیجیتال. امروزه به نظر می‌رسد که حداقل زیرمجموعه کوچکی از نورون‌ها (شاید فقط چند میلیون) رفتار پیچیده‌تری از خود نشان می‌دهند، بیشتر شبیه رایانه‌های آنالوگ؛ یعنی فراتر از اینکه نورون فعال است یا خیر، به نظر می‌رسد شدت تکانه‌ای که ارائه می‌دهد مهم باشد. یک متخصص علوم اعصاب تجربی به من گفت که نورون‌ها به هیچ وجه به اندازه آنچه نشان داده می‌شوند، رفتار ثابتی ندارند. او حداقل دو مرتبه بزرگی تغییرپذیری را در پاسخ به محرک یکسان مشاهده کرد. در حالی که این موارد معمولاً هنگام گزارش در مجلات علمی میانگین‌گیری می‌شوند، اما میانگینی وجود ندارد. همه اینها به این معنی است که یک شبکه عصبی مصنوعی عملکردی در مقایسه با مغز انسان نسبتاً محدود خواهد بود.

اکنون جنبه‌های عملکردی شبکه‌های عصبی مصنوعی (ANN) را در نظر بگیرید .عملکرد نرون‌های مصنوعی بازتابنده‌ی دانش ما در اواسط قرن بیستم در مورد نرون‌های زیستی بود :اینکه نرون‌ها یا «فعال» هستند یا غیرفعال، شبیه به رایانه‌های دیجیتال .امروزه به نظر می‌رسد حداقل زیرمجموعه‌ای کوچک از نرون‌ها (شاید فقط چند میلیون) رفتاری پیچیده‌تر از خود نشان می‌دهند، شبیه به رایانه‌های آنالوگ؛ یعنی فراتر از اینکه نرون فعال است یا خیر، به نظر می‌رسد شدت تکانه‌ای که ارائه می‌دهد نیز مهم باشد .یک عصب‌شناس تجربی به من گفت که نرون‌ها اصلاً به آن خوبی که نمایش داده می‌شوند، رفتار نمی‌کنند .او حداقل دو مرتبه بزرگی تغییر در پاسخ‌ها به یک محرک مشابه را مشاهده کرد .در حالی که این موارد معمولاً در مجلات علمی به صورت میانگین گزارش می‌شوند، اما میانگین همیشه وجود ندارد .همه اینها به این معنی است که یک شبکه عصبی مصنوعی عملکردی در مقایسه با مغز انسان بسیار محدود خواهد بود.

اگر دانشمندان بتوانند چیزی شبیه به یک مغز مصنوعی بسازند، برخی از علاقه‌مندان تأکید می‌کنند که از آنجایی که نورون‌های مصنوعی سریع‌تر از نورون‌های زیستی هستند، مغز مصنوعی باید بهتر یا حداقل از مغز بیولوژیکی سریع‌تر باشد. با این حال، در اینجا با فرض دیگری روبرو هستیم: اینکه مغز مصنوعی یک ذهن مصنوعی ایجاد کند. این ممکن است به نظر منطقی برسد، اما فقط یک فرض است. آنچه تاکنون می‌دانیم این است که مغز و ذهن به نحوی به هم مرتبط هستند و فرآیندهای خاص تفکر را می‌توان با فعالیت در مناطق خاصی از مغز مرتبط کرد. ما واقعاً درک نمی‌کنیم که مغز و ذهن چگونه به هم مرتبط هستند. تشبیه رایج علم کامپیوتر برای درک این است که مغز سخت‌افزار و ذهن نرم‌افزار است - اگرچه نرم‌افزار روی سخت‌افزار اجرا می‌شود، اما سخت‌افزار هرگز نرم‌افزار تولید نکرده است. اگر هر بار که می‌نویسم، همان قسمت پردازنده من فعال باشد، به این معنی نیست که آن قسمت از پردازنده، کتاب را می‌نویسد.

۲.۴ اما هوش مصنوعی چیست؟

جدا از سه نوع هوش مصنوعی اصلی که در این فصل توضیح داده شدند، چند نوع هوش مصنوعی دیگر نیز طراحی شده‌اند که به‌صورت مستقل یا در ترکیب با یکدیگر بکار می‌روند. مهم‌ترین موردی که در این‌جا راجع به جزئیات آن صحبت نخواهیم کرد، منطق فازی (FL)[1] است. منطق فازی که مفهوم آن در اصل توسط لطفی زاده (۱۹۶۵) پایه‌گذاری شد، به طور گسترده در مهندسی به کار می‌رود، زیرا تاثیرگذاری لوپ‌های کنترل را به مراتب بیشتر می‌کند. برای مثال، می‌توانید آن را در سیستم ترمز ضد قفل (ABS) خودروهای وولکس‌واگن، در تنظیم چرخه شستشوی بسیاری از ماشین‌های لباسشویی (اغلب مدل‌های گران‌قیمت‌تر) و غیره پیدا کنید. در زمینه‌ی هوش مصنوعی نیز، گاهی اوقات از آن در ترکیب با شبکه‌های عصبی مصنوعی (ANN)[2] و همچنین سیستم‌های خبره‌ی نمادین[3] نیز استفاده می‌شود.

الگوریتم‌های ژنتیکی و تکاملی انواع دیگری از هوش‌های مصنوعی هستند. این الگوریتم‌ها از جهش‌های تصادفی استفاده می‌کنند که اگر تابع برازندگی[4] آن‌ها به درستی انتخاب شود، راه‌حل‌های بهتری را ارائه خواهند داد. الگوریتم‌های ژنتیکی با ANNها همخوانی خوبی دارند و می‌توان از آن‌ها در هر دو نوع هوش مصنوعی نمادین نیز استفاده کرد؛ البته تنها مثال واضح از این موضوع، بازی‌های رایانه‌ای هستند. موارد فوق رویکردهای قدرتمند و مستقلی در هوش مصنوعی نیستند، بلکه تنها شکل‌های متغیر AIهایی هستند که قبل‌تر معرفی شدند، بنابراین بیش از این درباره‌ی آنها صحبت نخواهد شد. در عوض اکنون بر اساس سه‌نوع هوش مصنوعی اصلی، به تعریف AI می‌پردازیم.

اصطلاح هوش مصنوعی دو معنی دارد. از یک طرف، به ماشین‌های دارای هوش تصنعی و روش‌های ساخت آن‌ها اشاره دارد؛ و از طرف دیگر، هوش مصنوعی یک شاخه‌ی تحصیلی میان‌رشته‌ای برای مطالعه این نوع ماشین‌ها است. نخبگان هوش مصنوعی همچون سایمون، اغلب بر این موضوع تأکید کرده‌اند که مطالعه در زمینه‌ی هوش مصنوعی مطالعه ذهن انسان را نیز شامل می‌شود. از این رو، حوزه هوش مصنوعی جدا از شاخه‌های مختلف علوم سخت و مهندسی، زیست‌شناسی، روانشناسی و فلسفه را نیز در بر می‌گیرد. در هر دو تعریف، هوش مصنوعی به طور کلی به عنوان ماشین‌هایی تعریف می‌شود که می‌توانند کارهایی را که انسان از طریق فکرکردن به انجام می‌رساند، انجام دهند. شایان ذکر است تعریفی که ارائه کردیم نمی‌گوید که هوش مصنوعی این کارها را درست به همان شکلی انجام می‌دهد که انسان‌ها انجام خواهند داد.

اما چگونه می‌توانیم بفهمیم که یک ماشین (به همان شکلی که در انسان وجود دارد یا به طوری متفاوت) فکر می‌کند؟ از کجا بدانیم چه وقت باید یک ماشین را دارای هوش (تصنعی) تلقی کرد؟ اولین معیار، آزمون تورینگ است که توسط آلن تورینگ در سال ۱۹۵۰ پیشنهاد شد و تا به امروز محبوب‌ترین آزمون در این

---

[1]- Fuzzy Logic
[2]- Artificial Neural Network
[3]- Symbolic Expert System
[4]- Fitness function

زمینه است. اصل آزمون تورینگ مبتنی بر این است که اگر ما با یک موجودیت تعامل داشته باشیم و نتوانیم تشخیص بدهیم که آیا یک انسان است یا ماشین (و در واقع یک ماشین باشد)؛ این بدان معناست که ماشین مورد نظر «فکر می‌کند» و باید باهوش در نظر گرفته شود. این مطلب در نگاه اول قانع‌کننده است؛ اگر ماشین قادر به فکر کردن نباشد، طبیعتا ما انسان‌ها متوجه خواهیم شد. اما بگذارید قبل از بحث در مورد فرضیات و نتایج این آزمون، این سوال را بپرسم: به‌نظر شما تا به حال هیچ ماشینی (برنامه‌ی کامپیوتری) در آزمون تورینگ موفق عمل کرده است؟ می‌توان گفت تعدادی از آن‌ها توانسته‌اند در آن قبول شوند. شاید اولین موردی که این موفقیت را کسب کرد، ELIZAی جوزف وایزنبام[5] (۱۹۹۶) بود. پس از آن چت‌بات‌های امروزی بسیاری (مانند Cleverbot: رجوع شود به آرون[6]، ۲۰۱۱) و جدیدترین مورد نیز یوجین گوسمن[7]، شبیه‌سازی پسربچه‌ی ۱۳ ساله‌ی اوکراینی (وارویک و شاه، ۲۰۱۶) بودند که این آزمون را پست سر گذاشتند. نتایج آزمون مذکور را می‌تواند در دو گام تحلیل کرد: یکم از جهت صلاحیت داشتن «قبولی» در آن، و دوم از این حیث که اگر قبولی در آزمون تورینگ صلاحیت داشته باشد، این قبولی یعنی چه؟

در بیشتر برنامه‌هایی که قبول شدند، آزمون توسط افرادی انجام شد که خودشان نمی‌دانستند در حال داوری آزمون تورینگ هستند. اولین آزمایشی که در آن برگذاری آزمون برای افراد توجیه شده بود، مربوط می‌شد به ربات یوجین گوسمن. اما آیا این آزمون از صلاحیت برخوردار بود؟ ربات موردنظر توانست: «۳۳ درصد از داوران تازه‌کار را فریب دهد» (راسل و نورویگ، ۲۰۲۰). بنده به عنوان یک استاد دانشگاه، نمره‌ی ۳۳ درصد را برای دانشجویانم نمره‌ی قبولی نمی‌دانم. البته مشکلات به همین جا ختم نمی‌شود؛ به عبارتی یوجین یک بچه‌ی ۱۳ ساله است که زبان انگلیسی زبان دوم آن است و از کشوری آمده است که آزمون گیرندگان آشنایی چندانی با آن ندارند. همین باعث ایجاد شرایطی می‌شود که نتایج را به نفع او تغییر خواهند داد. با این وجود خودتان نگاهی به چند نمونه از مکالمات در این تست بیاندازید (آرونسون ۲۰۱۴) و نظرتان را بگویید؛ آیا به‌نظرتان یوجین واقعا می‌تواند یک شخص باشد؟ من شخصا دوست دارم با استوارت راسل و پیتر نورویگ هم‌نظر باشم (۲۰۲۰)؛ که گفته‌اند: «آزمون تورینگ در واقع بیشتر شبیه به آزمون ساده‌لوحی انسان هاست».

با توجه به آنچه گفتیم، می‌توان این احتمال را داد که هوش مصنوعی زمانی خواهد توانست که با صلاحیت کامل در آزمون تورینگ قبول شود. اما سوال اینجاست که: این قبولی چه چیزی را ثابت می‌کند؟ آیا قبول شدن در آن واقعا به این معناست که عمل فکر کردن در ماشین به همان صورتی است که ما برای انسان تعریف می‌کنیم؟ خیر، به نظر بنده چنین نمی‌آید. اجازه بدهید بار دیگر نگاهی به آزمون تورینگ بیندازیم. تورینگ آزمون خود را در قالب «بازی تقلید» مدل‌سازی کرد. در این بازی، شما تظاهر می‌کنید که شخص دیگری هستید. شما هر آنچه که می‌توانید را در مورد دیگری یاد می‌گیرید و سپس سعی می‌کند که دیگران را متقاعد کنید همان شخص هستید. اگر نکات را به خوبی یاد بگیرید، ممکن است موفق شوید. آیا برنده شدن در بازی بدین معناست که شما به‌طور معجزه‌آسایی به همان فرد تبدیل شده‌اید؟ قطعا خیر. برنامه‌هایی که در

[5]- Joseph Weizenbaum
[6]-Aron
[7]- Eugene Goostman

آزمون تورینگ قبول شدند برای همین منظور طراحی شده بودند: به منظور قبول شدن در آزمون؛ و فریب دادن و قانع کردن افراد به این‌که فکر می‌کنند، نه این‌که واقعا فکر کنند.

شاید قوی‌ترین استدلال علیه آزمون تورینگ مربوط شود به جان سرل (۱۹۹۸ : ۱۱) که با نام استدلال اتاق چینی شناخته می‌شود. اصل این استدلال بدین صورت است: فرض کنید که من زبان چینی بلد نیستم و در اتاقی پر از کتاب‌های فرهنگ لغات چینی حضور دارم. از خارج از اتاق، پیام‌هایی در قالب نمادهای چینی دریافت می‌کنم. سپس در میان این کتاب‌ها به دنبال پاسخ مناسب به این نمادها می‌گردم و این پاسخ‌ها را به بیرون از اتاق می‌فرستم. اگر این کتاب‌ها از محتویات درستی برخوردار باشند، مردمی که بیرون از اتاق حضور دارند گمان خواهند کرد که فرد داخل اتاق زبان چینی را می‌فهمد؛ ولی در اشتباه خواهند بود؛ چرا که من یکباره زبان چینی را یاد نگرفتم. به طور مشابهی، قبول شدن در آزمون تورینگ ثابت نمی‌کند که کامپیوتر فکر می‌کند، بلکه صرفا نشان می‌دهد که برنامه‌نویسی خوبی دارد.

مسلما این مثال‌ها و استدلال‌ها نمی‌گویند که ماشین‌ها قابلیت فکر کردن ندارند، بلکه تنها شواهد به دست آمده از آن‌ها کافی نیست. و چون دلایل قاطعی وجود ندارد، تنها باور شخصی هر یک از ما برای قبول یا رد آن باقی می‌ماند. در این فصل بیشتر تمرکز بر روی کامپیوترها بود. فصل‌های ۳، ۴ و ۵، کامپیوترها را با انسان‌ها در ابعاد مختلفی مقایسه می‌کنند و مزایای AI و نحوه‌ی دستیابی به آن‌ها را توضیح می‌دهند. فراموش نکنید که من خود طرفدار هوش مصنوعی هستم؛ اما سعی می‌کنم به شیوه‌ی متفاوتی به نحوه‌ی بهره‌گیری بهتر آن نگاه کنم. اما بگذارید پیش از رفتن به فصل بعد، یک سری نکات مهم را جمع‌بندی کنم.

۲.۵ همه‌ی اینها چه معنایی دارند؟

باید خیلی باریک بین باشیم تا در مورد هوش مصنوعی پیش‌باورها را با حقایق اشتباه نگیریم. به حقایق زیر توجه کنید:

- فقط یک نوع AI وجود ندارد. درست است که امروزه دنیای هوش مصنوعی در دستان شبکه‌های عصبی مصنوعی (ANNها) است؛ ولی ما چندین نوع هوش مصنوعی دیگر نیز داریم. شما می‌توانید بسته به هدف خود، یکی از آنها را انتخاب نموده و یا اینکه چند نوع را با هم ترکیب کنید؛ کاری که در مورد بهترین AI امروزی (رجوع شود به مثال AlphaFold در بخش ۴-۵) انجام شده است.
- انسان‌ها و ماشین‌ها هر یک در کارهای متفاوتی به چالش بر می‌خورند. فهمیدن نثر برای انسان‌ها و به ویژه برای آن‌هایی که زبان موردنظر زبان مادری‌شان است، راحت است. با این حال، ضرب دو عدد شش رقمی به کنار، اثبات یک قضیه‌ی ریاضی نیز برای بسیاری از افراد امر دشواری خواهد بود. ولی ماشین‌ها برخلاف انسان‌ها، ضرب اعداد شش رقمی را سریع و بی‌دردسر انجام می‌دهد. تعدادی از AIهای تخصصی حتی توانسته‌اند بعضی از قضایا را اثبات کنند؛ هرچند فهمیدن معنی این قضایا، حداقل در حال حاضر برای ماشین‌ها کاری غیرممکن به نظر می‌رسد.

- علی‌رغم ادعای برخی در مورد ماشین‌هایی که فکر می‌کنند (به آن‌ها AI گسترده و هوش مصنوعی عمومی [AGI] نیز می‌گویند)، تاکنون همه‌ی هوش‌های مصنوعی ایجاد شده تنها در حیطه‌های علمیِ تعریف شده و محدودی عملکرد عالی از خود نشان می‌دهند. همین‌که از این حیطه‌ها فراتر برویم، به‌سرعت نامفهوم می‌شوند. هر چه غیر از این بیان شود، چیزی بیش از وعده، استنباط و امید برخی افراد نیست.

اکنون اجازه بدهید باور شخصی خودم را خدمتتان ارائه دهم. از دهه‌ی ۱۹۵۰، ما شاهد چندین موج از انواع برنامه‌ها بوده‌ایم که بر دنیای هوش مصنوعی غالب شدند. امروز ANNها تقریبا در این زمینه یکه‌تازی می‌کنند. بنده عقیده دارم که ANNها نیز آخرین حکمران‌های دنیای هوش مصنوعی نخواهند بود؛ موج‌هایی از AIهای مختلف پشت سر هم می‌آیند و مانند آنچه که داون‌پورت و اُدِل گفته‌بودند (۲۰۱۹)، به عقیده‌ی من نیز سیستم‌های خُبره دوباره به عرصه بر می‌گردند. همچنین اطمینان دارم که حرکت‌ها همگی به سمت AI هیبرید (مرکب) خواهند بود و ترکیبی از سیستم‌های خبره، ANN، و احتمالا انواع دیگر AI مانند الگوریتم تکاملی و منطق فازی عرضه می‌شوند.

۳ دانش

منطق نقطه‌ی آغاز حکمت است ... نه پایان آن. اسپاک

این فصل نشان می‌دهد که دانش برای انسان و ماشین هرکدام چه معنی‌ای می‌دهد؛ و این دانش گونه با ابعاد دیگر انسان یا ماشین درگیر است. این بخش از تحلیل ما بیشتر در رابطه با سیستم‌های خبره است که از پایگاه‌های دانش[8] برخوردارند. البته متخصصان عقیده دارند که ANNها هم نوعی دانش در مورد جهان حاصل از مثال‌های یادگیری‌شان بدست خواهند آورد؛ اما این موضوع مربوط به فصل ۴ است. علاقه‌ی بنده به برای تحقیق در زمینه‌ی دانش، یادگیری، خلاقیت و بصیرت- به خصوص با کمک گرفتن از نخبگان برجسته- منجر شد که با ۲۰ دانشمند برتر، از جمله ۱۷ برنده‌ی جایزه‌ی نوبل مصاحبه کنم. این مطالعه پس زمینه‌ی این فصل و فصل‌های ۴ و ۵ را شکل می‌دهد.

۳-۱ دانش صریح و دانش ضمنی

به‌منظور درک بهتر نکات ذکر شده در این قسمت، بد نیست ابتدا تصویر جامع‌تری از «دانش»[9] در مرجع خود داشته باشیم. دانش علاوه بر دانش علمی، شامل دیگر اقسام نیز می‌شود؛ مثلا درست کردن یک قهوه، سرودن شعر، حس لذت طعم توت فرنگی و مانند این‌ها. یک تبلیغ تجاری بر روی آنتن تلویزیون را در نظر بگیرید که می‌گوید: «این شوینده ۹۹/۷ درصد از باکتری‌ها را از بین می‌برد». به همین دلیل است که جورج کِلی (۱۹۵۵) بر این باور است: «همه‌ی ما دانشمند هستیم». دانشمند بودن یک موقعیت شغلی نیست، بلکه نوعی جهان‌بینی

[8]- knowledge base
[9]- knowledge

است. در ادامه باید میان معنای «حقیقت»[10] و معنای «دانش» تمایز قائل شویم. ما انسان‌ها عادت داریم فرض کنیم که دانستن یک چیز، با حقیقت داشتن آن یکسان است. کنت بولدینگ (۱۹۶۶) استدلال آورد که در این مساله یک اشکال معرفت‌شناختی مهمی وجود دارد. بدین صورت که در زبان انگلیسی ما نمی‌توانیم فعل دانستن را صرفا در قالب یک محتوای ذهنی دربیاوریم؛ مگر آنکه فرض کنیم آن محتوا حقیقت دارد. این مساله‌ی مهمی است؛ مثلا ما می‌دانیم که زمین تخت نیست، ولی تقریبا تمام فعالیت‌های مهندسی (به جز ساخت موشک) مانند ساختن یک ماشین یا ساختمان، بدون نیاز به در نظر گرفتن این‌که «زمین یک کُره است» قابل انجام هستند. و چنانچه بخواهیم خیلی باریک‌بین باشیم- که البته این کار برای نشان دادن حقیقت لازم است- باید به‌جای کُره می‌گفتیم بیضی، سپس باید عبارت بیضی کج و معوج را به‌میان می‌آوردیم و در نهایت به شکلی می‌رسیدیم که به آن ژئوئید[11] می‌گوییم، که معنی آن می‌شود: «چیزی شبیه به زمین». پس اگر بخواهیم در تعریف شکل زمین دقت بالایی به خرج بدهیم، تنها جسمی (سیاره یا اجسام دیگر) که در دسته‌بندی این تعریف قرار می‌گیرد، خود زمین خواهد بود.

این دو مشکل با هم ادامه پیدا می‌کنند؛ به عبارتی افراد دانش علمی را اصل قرار می‌دهند و باور دارند که حقیقی است. برخی حتی از این هم پا فراتر گذاشته و استدلال می‌آورند که علم تنها دانش حقیقی است. اما علم و حقیقت در واقع با هم متضاد هستند:

این یک پارادوکس معرفت‌شناختی مدرن است، که ما علم را نمونه‌ای از دانش بدانیم، و با این حال اصرار داشته باشیم که علم صورت آشکار حقیقت است. زیرا علم در سایه‌ی یافته‌ها و کشفیات زنده است، و اگر عطش حل مشکلات، دنبال کردن فرضیات و آزمایش ایده‌های جدید و خام نبود، بقای علم ناممکن می‌شد (پولانی، ۱۹۶۹: ix).

اینکه چرا ما ترجیح می‌دهیم علم را والاتر از دیگر انواع دانش بدانیم معلوم نیست. برخی می‌گویند علت آن به کارهای ارسطو برمی‌گردد؛ ولی من به شخصه از ارسطو چنین برداشتی ندارم. مشخصا او طرفدار فرونسیس[12] هم بود (به زبان یونان باستان: φρόνησ ς، که ترجمه‌ی رایج آن حکمت عملی یا فضیلت عملی است). به هر حال اگر کسی بر این باور باشد که علم از همه‌ی دانش‌های دیگر برتر است، باز هم نباید فکر کرد که علم تنها نوع دانش است. به علاوه، علم بر جنبه‌هایی از دانش متکی است که خارج از حیطه‌ی خود علم هستند. برخی از دیگر جنبه‌های دانش در ادامه‌ی این فصل و فصل ۴ شرح داده خواهند شد؛ فعلا در اینجا بر جنبه‌ی صریح[13] دانش تمرکز می‌کنیم.

افراد غیردانشمند معمولا گمان می‌کنند دانش علمی کاملا آشکار و صریح است. این گمان آن‌ها زمانی درست خواهد بود که تمام این دانش در قالب کتاب‌های راهنما دربیاید ؛ اما راه طولانی برای رسیدن به چنین هدفی

[10]- truth
[11]- geoid
[12]- phrónēsis
[13]- explicit

در پیش است. همانطور که در بالاتر هم اشاره کردیم، ساخت دانش علمی با توجه به حقیقت، مقدار زیادی دانش ضمنی[۱۴] می‌طلبد (پولانی، ۱۹۶۶ب). وجود «روش‌های علمی» که دانش علمی را به وجود می‌آورند، افسانه و خیال است؛ زیرا هیچ‌کس نمی‌داند چنین روش‌هایی چگونه‌اند. طبق آنچه که از مصاحبه‌هایم فهمیدم، درون همه‌ی یافته‌های علمی بزرگ، نوعی بینش شهودی پنهان بوده است (دورفلر و آکرمان، ۲۰۱۲؛ دورفلر و اِدِن، ۲۰۱۹). دانش شهودی در حیطه‌ی دانش‌های ضمنی قرار می‌گیرد؛ به‌عبارتی می‌دانیم، ولی دقیق نمی‌دانیم که چگونه می‌دانیم. این دانستن نوعی دانش مستقیم[۱۵] است (دورفلر و استیراند، ۲۰۱۷). آنچه که لازم است درباره‌ی دانش ضمنی متوجه شویم این است که دانش ضمنی دانشی است که نمی‌تواند در قالب کلمات توصیف شود، نه دانشی که هنوز برایمان قابل توصیف نیست. دانش ضمنی «ناگفتنی و غیرقابل توصیف» است (دورفلر و استیراند، ۲۰۱۷؛ سین‌کلیر و آشکاناسی، ۲۰۰۵). چرا این‌قدر بر این نوع دانش تاکید می‌کنم؟ چون هیچ دانشی بدون دانش ضمنی وجود نخواهد داشت:

اگرچه خود دانش ضمنی به دیگری متکی نیست، دانش صریح ناگزیر وابسته به درک و فهم ضمنی است. از آنجا که هر دانشی یا ضمنی است و یا ریشه در دانش ضمنی دارد، تصور دانش کاملا صریح غیر ممکن است (پولانی ۱۹۶۶الف : ۷).

این فرضی که مایکل پولانی ادعا کرده است، شاید قوی‌تری ادعایی باشد که تا به حال علیه باور اکثریت نسبت به دانش و علم ایراد شده است. پولانی با کوشش‌های فراوان کتاب کاملی را نوشت تا نشان دهد چگونه شکل‌های مختلف دانش ناگزیر ریشه در دانش ضمنی پیدا می‌کنند. سپس او نتیجه گرفت: «دانش ضمنی در واقع جزء غالب تمام دانش‌هاست، و رد کردن آن خودبخود مساوی است با رد هر نوع دانشی» (پولانی ۱۹۵۹ : ۱۳).

مدیران عامل و مدیران معمولا به راحتی دانش ضمنی را خواهند پذیرفت؛ آن‌ها به خوبی از اهمیت «احساسات درونی» در تصمیم‌گیری، مذاکرات و همه‌ی حیطه‌های کاری‌شان آگاه هستند. علت اینکه دانستن نقش دانش ضمنی برای ما ضروری است، این است که هر دو نوع هوش مصنوعی نمادین (سیستم‌های استدلال[۱۶] و خبره‌ی نمادین) تماما متکی به دانش صریح هستند. مثال مشابه این مورد این است که بخواهیم وظایف یک مدیر عامل را بنویسیم و در قالب یک کتاب راهنمای برای دانشجویان MBA (ارشد مدیریت بازرگانی) با موضوع تصمیم‌گیری و هدف‌گذاری دربیاوریم. مسلما خواندن این دسته از کتاب‌ها مهم است و بایستی دانشجویان MBA در مورد تصمیم‌گیری و هدف‌گذاری آموزش داده شوند. اما دقت کنید که گفتیم «در مورد»؛ ما نمی‌توانیم قوه‌ی تصمیم گرفتن و هدف گذاری را به دانشجویان یاد بدهیم؛ بلکه صرفا آن‌ها را در مورد این مسائل آگاه می‌سازیم. کریس آرجریس- مدیر برجسته و ممتاز- در یک کنفرانس AoM (آکادمی مدیریت) بیان کرد که اگر قرار بر این باشد تصمیمات خود را طبق کتاب‌های آموزشی تصمیم‌گیری MBA بگیریم، هیچ

[14]- tacit
[15]- direct
[16]- reasoning

وخت از رختخوابمان بیرون نخواهیم آمد؛ چون تا بخواهیم تصمیماتی را که در طول روز قرار است با آن‌ها مواجه شویم بگیریم، آن روز به اتمام می‌رسد و دوباره وقت خواب است.

برای ساخت یک سیستم استدلال نمادین، باید با کمک تکنیک «تفکر گویا[۱۷]» مراحل حل مسئله را تعیین کرد. بدون شک این روش می‌تواند برای درک بهتر فرایندهای شناختی به شدت پیچیده همچون خلاقیت و حل مسئله بسیار مفید واقع شود؛ اما هنوز برای بازتولید این فرایندها کافی نیست و راه زیادی پیش‌رو دارد. مهندس دانش در هنگام طراحی پایگاه‌های دانش، دانش موردنظر را ذره ذره از منابع خبره دریافت می‌کند (باراکسکای و ولنسی، ۲۰۱۲). گاها بخش کوچکی از دانش ضمنی در طی این فرایندها به دانش صریح تبدیل می‌شود. این اتفاق گرانقدر که از باارزش‌ترین جنبه‌های مهندسی دانش است، «کشف دانش[۱۸]» نام دارد. اگرچه سیستم‌های خبره در حیطه‌های تخصصی بسیار محدود و متمرکزی قرار می‌گیرند، تنها درصد کوچکی از دانش ضمنی متعلق به منبع خبره، به دانش صریح تبدیل می‌شود؛ البته همین میزان کم نیز ممکن است رخ ندهد. در حقیقت مهندسی دانش، اطلاعاتی را در مورد حیطه‌ی مربوطه کسب و ساماندهی می‌کند که از اول صریح و آشکار هستند. نه این‌که درصد خاصی را مدنظر قرار دهم، اما دانش صریح تنها بخش جزئی از دانش خبره را شکل می‌دهد. از این رو محدود کردن هوش مصنوعی به داده‌های صریح می‌تواند دلیل کافی بر این موضوع باشد که چرا کامپیوترها نمی‌توانند مهارت‌های واقعی داشته باشند. به هر حال، همه‌ی اینها دانشی هستند که خود ما صریحا در هوش مصنوعی قرار می‌دهیم، نه دانشی که هوش مصنوعی شخصا یاد بگیرد. این مساله در فصل ۴ بررسی می‌شود؛ برای اکنون بهتر است به کاوش در برخی دیگر از جنبه‌های دانش و قلمروهای مرتبط به همان شکلی که در انسان‌ها وجود دارند بپردازیم، تا درک بهتری از ماشین‌ها داشته باشیم.

۲-۳ دانش شخصی

یک مشخصه‌ی مهم دانش این است که دانش اساسا شخصی است. به گونه‌ای می‌توان گفت که دانش لزوما سابجکتیو یا ذهنی است؛ البته پولانی در مفهوم‌سازی ابتدایی خود تاکید کرده بود که دانش شخصی « فراتر از تعارض میان عینیت و ذهنیت است»(پولانی ۱۹۶۲الف : ۳۱۲). در این صورت، مفهوم دانش شخصی قصد-مندی[۱۹]، رانه[۲۰] و اشتیاق[۲۱] را نیز شامل می‌شود و صرفا بر دوش کشیدن احساساتمان را در بر نمی‌گیرد. در هر صورت، منظور ما نوعی ایدیوسنکرازی[۲۲] (ویژگی فردی) است؛ به عبارتی دانش شخصی هر فرد منحصر به همان فرد است. این مساله مشکل بزرگی را مقابل جامعیت‌بخشی به پایگاه‌های دانش (یا هر نوع معرف دانش دیگر) قرار می‌دهد. برای گشودن مساله‌ی ایدیوسنکرازی، ابتدا نگاهی انداختم به ویژگی‌های دانشی که طی

[17]- thinking aloud
[18]- knowledge discovery
[19]- intentionality
[20]- drive
[21]- passion
[22]- idiosyncrasy

مصاحبه با برندگان نوبل به آن‌ها پی برده بودم؛ همچنین میان این ویژگی‌ها و تجربیات روزمره‌ای که همه‌ی ما داریم ارتباط برقرار کردم.

طرز تفکر دانشمندان برتر (دورفلر و اِدِن، ۲۰۱۹)، به علاوه‌ی افراد برجسته در هر حرفه‌ای مانند برترین سرآشپزها (استیراند و دروفلر، ۲۰۱۶)، دارای ویژگی به نام پرش‌های شهودی است[۲۳]. مشکل شهود برای هوش مصنوعی در این است که شهودی بودن غیر الگوریتمی است. به عبارتی ما واقعا نمی‌دانیم که شهودیت چگونه کار می‌کند. شهودیت غالبا به صورت نوعی دانش مستقیم توصیف می‌شود که هیچ پردازش آگاهانه و ترتیبی در آن دخالت ندارد (سینکلر و آشکاناسی، ۲۰۰۵). طبیعتا چنین فرایندی را نمی‌توان با روش تفکر گویا بدست آورد. آن دسته از شهودها (محصولات فرایند شهودیت) که قبلا رخ داده‌اند را می‌تواند درون پایگاه‌های دانش وارد کرد (و البته به دفعات این کار انجام می‌شود)؛ ولی به هر حال این کار باعث نمی‌شود که در ادامه شهودهای بیشتری به‌وجود بیایند. به‌طور خلاصه، ما نمی‌توانیم هوش مصنوعی را وادار به ساخت شهود کنیم، و اگر هم توانسته باشد به روش‌های دیگر به شهود دست یابد، ما در تشخیص آن ناکام بوده‌ایم؛ زیرا مدلی نداریم که فرایند شهودیت را با جزئیات کافی توصیف کند. با وجود همه‌ی اینها، شهودیت باز وجود دارد و بخش مهمی از تفکر ما را شکل می‌دهد، حتی اگر دانش علمی‌مان محدود باشد (دورفلر و باس، ۲۰۲۰الف). شهودها چقدر اهمیت دارند؟ با استناد بر مصاحبه‌هایی که با دارندگان جایزه‌ی نوبل انجام داده‌ام، می‌توانم جسارت به خرج دهم و بگویم که هر دستاورد پژوهشی مهم در تاریخ، تحت تاثیر نقش مهم شهود بدست آمده است.

دیگر جنبه‌ی دانش شخصی، دانستن زیبایی و هماهنگی است. ما می‌توانیم زیبایی رنگین کمان را تحسین کنیم، بدون آن‌که لازم باشد بدانیم کدام اثر نوری باعث ایجاد آن می‌شود. حتی اگر بدانیم این رنگین کمان از یک منشور به وجود آمده است، باز هم برایمان زیبا خواهد بود. دانشمندان، سرآشپزها و دیگر افراد برجسته در هر حرفه در هنگام خلق آثار جدید، از این دانش بهره‌ی فراوانی می‌برند؛ چه این اثر یک نقاشی، یک نوع غذا و یا ایده‌ی علمی جدیدی باشد (رجوع شود به ویلچک، ۲۰۱۵). شاید معروف‌ترین مثال علمی در این مورد، معادلات مشهور جیمز کلارک ماکسول برای توصیف الکترومغناطیس باشد. او شکل اولیه‌ی معادلات خود را با در نظر گرفتن دانش بدست آمده از فیزیک نظری و آزمایشگاهی بدست آورد. سپس نگاهی به معادله‌ها به دست آمده انداخت و با خود فکر کرد که این معادله‌ها زیبایی چندانی ندارند، در نتیجه جزئی را به معادله‌ها اضافه کرد. در آن زمان جزء اضافی اختلاف بسیار ناچیزی ایجاد می‌کرد و از این رو غیر ممکن بود که بگوییم کدام نسخه از معادلات درست هستند. اما بعدا که محاسبات اصلاح شدند، مشخص شد که جزئی که ماکسول برای زیباتر شدن معادلات خود استفاده کرده بود، بخشی ضروری هستند و معادلات بدون آن‌ها نادرست در می‌آمد. همه‌ی ما آنچه را که برایمان زیباست در هنگام خلق یک اثر مدنظرمان قرار می‌دهیم. برای مثال من ساختار این کتاب را به همین شکل تصور کرده بودم.

[23]- intuitive leaps

تفکر تمثیلی[۲۴] به معنای داشتن مدل‌های ذهنی است که تجلی از واقعیت باشند و ما می‌توانیم با آن‌ها کار کنیم. شاید مشهورترین مثال در دنیای علم، مربوط باشد به نیکولا تسلا که در ذهنش ماشین‌های خود را خلق می‌کرد، اصلاحات لازم را انجام می‌داد و مشکلات را رفع می‌کرد و هنگامی که همه‌ی این‌ها در ذهنش به خوبی صورت می‌گرفتند، آن‌ها را در واقعیت می‌ساخت و واقعا نیز بدون نقص کار می‌کردند (تسلا، ۱۹۱۹ : ۱۲-۱۱). تسلا تنها زمانی وادار به طراحی نقشه‌ها روی کاغذ می‌شد که لازم بود حق اختراع آن را ثبت نماید؛ اما برای کار خودش به آن‌ها نیازی نداشت. تفکر تمثیلی به بخشی از ذهن انسان برمی‌گردد که به آن سیستم تمثیل و بازنمایی[۲۵] می‌گویند (رومل‌هارت و نورمن، ۱۹۸۸). همه‌ی ما این سیستم را داریم، اما بسیاری از ما به زیرکی تسلا از آن بهره نمی‌گیریم. اگار شما به خانه‌ای که قبلا در آن زندگی می‌کردید فکر کنید، می‌توانید تعداد پنجره‌های آن را بشمارید، در ذهن خود در آن قدم بزنید، چهره‌ی همسایگان‌تان را به‌خاطر بیاورید و ... . اگر به آخرین صبحانه‌ی داغی که خوردید بیندیشید، نه تنها تصویر، بلکه مزه تخم‌مرغ‌ها، صدای کارد و چنگال و تمام طعم‌های همراه با آن‌ها را نیز به یاد خواهید آورد. با توجه به همه‌ی این‌ها می‌توان گفت که سیستم تمثیل و بازنمایی یک سیستم چند حسی و کل‌نگر است و صرفا حقایقی که به یاد می‌آوریم این کار را به خودی خود انجام نمی‌دهند.

و بالاخره من به چیزی پی برده‌ام که به آن جوهره‌بینی[۲۶] می‌گویم. نیمه‌ی ابتدایی این فرایند طبق گفته‌ی نخبگان برجسته بسیار راحت است: این افراد می‌توانند «تصویر کلی» را ببینند. یعنی آن‌ها موقعیت را به صورت کلی و همراه با مفهوم به شکل واحد درک می‌کنند؛ به علاوه تمام ارتباطاتی را که این کل ممکن است با هر چیزی داشته باشد، در نظر می‌آورند. همچنین این افراد به طرز استثنائی در دیدن جزئیات توانمند هستند، و نیز ارتباط میان هریک از جزئیات را با کل (به‌طور شهودی) می‌دانند. چنین چیزی فوق‌العاده است؛ چرا که با توجه به علوم سایبرنتیک ما می‌دانیم جزئیات باعث به وجود آمدن کل نمی‌شوند. اغلب به اشتباه می‌گویند که کل چیزی بیشتر از مجموع اجزاست؛ در حقیقت کل، چیزی متفاوت از مجموع اجزای خود است. یک زوج را تصور کنید که شما از قبل آن دو نفر را می‌شناختید. آن دو به عنوان زوج چیز متفاوتی از دو نفر هستند. به علاوه هر یک از آن دو در زوج بودن نسبت به آنچه که هر یک از خودشان هستند، متفاوت عمل می‌کنند. به همین دلیل است که سیستم‌های پیچیده را نمی‌توان از هم جدا کرد و مجددا به هم پیوست. از این رو، اینکه این افراد نخبه می‌توانند «ببینند» که با تغییر جزئیات خاصی، چه اتفاقی برای تصویر کلی می‌افتد، باورنکردنی است. نکته‌ی حیرت‌انگیزتر اینجاست که اگر این افراد بخواهند تصویر کلی را به‌شیوه‌ی خاصی تغییر بدهند (مثلا آن را تصحیح کنند)، خواهند دانست که کدام جزئیات را چگونه بایستی دگرگون کرد. من این را به عنوان شاهدی می‌گیرم که ماهیت تفکر غیرخطی است؛ هوش مصنوعی اما، به طرز اسفناکی خطی کار می‌کند. این توصیف در مدل «ارباب و فرستاده‌اش (۲۰۱۹)» متعلق به ایین مک‌گلیکریست بهتر مشخص می‌شود. در این مدل استاد نیمه‌ی راست مغز است و تصویر کلی را تعیین می‌کند؛ در حالی که فرستاده‌ی او نیمه‌ی چپ مغز است و با جزئیات سروکار دارد و تحت فرمان ارباب است. متاسفانه در دنیای

[24]- analogical thinking
[25]- analogical representational system
[26]- seeing essence

غرب تمایل بر این بوده است که فرستاده را جایگزین ارباب قرار داده و ارباب را نادیده بگیرند[۲۷]، ما نیز در زمینه‌ی هوش مصنوعی همین مساله را می‌بینیم.

هدف بر این است که خوانندگان تخصص بالایی در موضوعات این بخش کسب کنند، زیرا مشخصاتی که در این جا شرح دادیم به ویژه در افراد نخبه به خوبی مشاهده می‌شوند. درحقیقت موارد مذکور نکات جزئی نیستند که نیازمند سطح پایینی از مهارت باشند؛ بلکه اجزای ضروری برای کارکرد بالای تفکر انسان محسوب می‌شوند. بسیاری گمان می‌کنند با بالا رفتن تخصص، دانش بیشتر مجرد (انتزاعی) خواهد بود؛ با این ایده‌ی منطقی که هرچه دانش بیشتری جمع شود جامع‌تر خواهد شد، و جامعیت بیشتر نیز یعنی مجردتر بودن آن. اما در حقیقت برادران دریفوس (هیوبرت و استوارت ۱۹۸۶) نشان دادند که بیشترین تجرد دانش برای نوآموزان رخ می‌دهد، زیرا این دسته از افراد نمی‌توانند مطالبی را که به تازگی یاد گرفته‌اند با تجربه‌های واقعی زندگی ارتباط دهند. با افزایش تخصص، دانش ما مستحکم‌تر از قبل می‌شود. جدا از آن، شهود نیز با افزایش تخصص تجسم بیشتری می‌یابد (دورفلر و همکاران، ۲۰۰۹؛ دریفوس و دریفوس، ۱۹۸۶؛ سایمون، ۱۹۹۶). این مطلب همچنین بدان معناست که دسته‌بندی دشوار و آسان بودن کارها به‌طور یکسان برای ذهن انسان و هوش مصنوعی راحت نیست. نِوِل و سایمون برنامه‌ی لاجیک تئوریست (نظریه‌پرداز منطقی)[۲۸] را طراحی کردند. آن‌ها کاری را که این برنامه انجام می‌داد در رده‌ی کارهای دشوار قرار دادند، چون سطح آن در سطح تدریس-های دانشگاهی بود (نول و همکاران، ۱۹۶۳). گرچه اثبات قضایای ریاضی برای انسان‌ها بسیار سخت است، اما یک هوش مصنوعی نسبتا ساده از عهده‌ی انجام آن برآمد. در مقابل، درک معنای یک متن هرچند که برای انسان کار راحتی است، مانع دشواری بر سر راه هوش مصنوعی محسوب می‌شود.

۳-۳ دانش تنها نیست

جدا از جنبه‌های فراوانی که دانش شخصی دارد و ظاهرا هوش‌های مصنوعی قابلیت تکرار و تقلید آن‌ها را ندارند، لازم است به نحوه‌ی ارتباط دانش با دیگر ابعاد وجودی خودمان نیز پی ببریم. سه تا از مهم‌ترین این ابعاد عبارت‌اند از احساسات[۲۹]، هیجان‌ها[۳۰] و ارزش‌ها[۳۱] (ارزش‌ها در فصل ۶ به تفصیل شرح داده خواهند شد).

احساسات ظاهرا سد راه افکار عقلانی هستند. ما اگر گرسنه باشیم نمی‌توانیم تمرکز کنیم. اگر ترس از افتادن داشته باشیم، نمی‌پریم. ظاهر امر به ما می‌گوید احساسات به ضرر انسان‌ها- در مقایسه با هوش مصنوعی- عمل می‌کنند. اما بحث این نیست که آیا احساسات خوب‌اند یا بد؛ بلکه فقط کافی‌ست بگوییم احساسات در تمایز انسان‌ها نقش دارند. صد البته بشر می‌تواند با کمک دانش خود به‌طور موقت بر احساسات غلبه کنیم، اما این امر گذرا و موقتی است. احساسات ارتباط نزدیکی با غریزه[۳۲] و در نتیجه با نیازهای اساسی ما دارند

---

[۲۷]- اشاره به همان کتاب ارباب و فرستاده‌اش (The Master and His Emissary)

[28]- Logic Theorist
[29]- feelings
[30]- emotions
[31]- values
[32]- instincts

(دورفلر و چندری، ۲۰۰۸). ما گرسنگی، تشنگی، میل به آمیزش جنسی، ترس از مرگ، اشتیاق به دوستی با دیگران و ... را احساس می‌کنیم. ما می‌توانیم پاسخ‌دهی به هر یک از این‌ها را به تاخیر بیندازیم، ولی در نهایت ناچاریم تسلیم احساستمان شویم؛ باید غذا بخوریم، بخوابیم، از عامل ایجاد کننده‌ی ترس دوری کنیم و ... . با این وجود کنترل احساسات از اهمیت بالایی برخودار است. اهمیت آن تا حدی زیاد است که عده‌ای از زیست-شناسان مانند فون برتالانفی (۱۹۸۱)، و برخی روان‌شناسان مانند فروم (۱۹۸۲) باور دارند کنترل احساسات یکی از مشخصه‌های ضروری برای تمایز انسان از حیوانات است. اما اگر بخواهیم راجع به انسان و هوش مصنوعی صحبت کنیم، داشتن احساسات و توانایی رهایی از فشار آن‌ها و برخورد با آن‌ها به شیوه‌ی مناسب خود، با نبود کلی احساسات قابل مقایسه نیست.

برخلاف احساسات، هیجانات در مقام بالاتری از دانش قرار می‌گیرند؛ با توجه به این موضوع که هیجانات می‌توانند بر دانش فائق شوند؛ همانطور که دانش بر احساسات غلبه می‌کند. فیلسوف دیوید هیوم (۱۷۳۹ : ۴۱۵) باور داشت «استدلال خدمتگزار اشتیاق بوده و بایستی که باشد، و هرگز نباید به جز خدمت و تبعیت از آن، وظیفه‌ی دیگری بر عهده بگیرد». در این نوع خط فکری، انسان‌ها بر اساس هیجانات‌شان تصمیم می‌گیرند که چه می‌خواهند و از استدلال برای فهمیدن نحوه‌ی حصول به آن استفاده می‌کنند. پس از او آنتونیو داماسیو (۱۹۹۵) با بیماری سر و کار داشت که او را با نام «الیوت» می‌خواند. الیوت مرکز هیجانات خود را طی یک تصادف از دست داده بود. به عبارتی او نمی‌توانست هیچگونه هیجانی را تجربه کند. با وجود این‌که تمام توانایی‌های ذهنی الیوت بدون مشکل کار می‌کرد، او شغلش را از دست داد، در اجتماع خوب عمل نمی‌کرد و نمی‌توانست مفاهیم انسانی را به خوبی بیان کند، کاری که قبلا جزو مهارت‌های ذاتی‌اش به شمار می‌آمد. او می‌توانست محاسبات بسیار طولانی ریاضی را به درستی انجام دهد، اما در کشیدن یک تصویر عاجز بود. او ساعت‌ها بر سر تصمیم‌گیری برای یک چیز (مثلا تعیین تاریخ ملاقات بعدی با داماسیو) بحث می‌کرد اما قادر به انتخاب نبود. در این پژوهش شواهد محکمی ارائه می‌شوند که بر ضرورت هیجان برای تصمیم‌گیری انسان دلالت دارند.

ارزش مواقعی پدیدار می‌شود که کاری را از روی باور به درست بودن آن- و نه برای رسیدن به یک هدف- انجام می‌دهیم. هر یک از ما به‌خاطر ارزش‌های خاصی که داریم، می‌توانیم از منافع شخصی فراتر رفته و بی‌توجه به آنچه که می‌خواهیم و حتی متضاد آن عمل کنیم؛ اگرچه آگاهی کاملی نسبت به خواسته‌های خودمان داریم. پدری که با دخترش برای تعطیلات به خارج شهر می‌رود شاید رفتن به کوهستان را بیشتر دوست داشته باشد، اما بخاطر دخترش کنار دریا را انتخاب می‌کند. البته این مثال ساده‌ای است؛ اما بسیاری از تصمیم‌های خانوادگی، فرزندپروری، زناشویی و دوستانه در این سطح صورت می‌گیرند. ارزش‌ها همچنین ممکن است از جامعه، دین، جبهه‌های سیاسی، شرکت‌ها و یا هر نوع سازمان دیگری برخاسته باشند، تا ما آنچه را که در اجتماع پذیرفته شده است انجام دهیم و یا مطابق با اظهارات یک کتاب مقدس، معلم، سیاستمدار و یا یک مدیر عمل کنیم. مساله‌ی چالش‌برانگیز این است که ارزش‌ها درون ما مانند یک فهرست خطی قرار نگرفته‌اند و ساماندهی نامرتبی دارند. به علاوه ما برخی از ارزش‌ها را از خانواده‌هایمان به ارث می‌بریم و برخی دیگر را از هم سن و سالانمان، مدرسه، جامعه و ... می‌گیریم. اما ما در پذیریفتن این ارزش‌ها منفعل نیستیم؛

یعنی آن‌ها را تغییر می‌دهیم، از نو ارزش‌های جدید می‌سازیم، چند ارزش را با هم ترکیب می‌کنیم و آن‌هایی را که سنخیتی با ما ندارند دور می‌اندازیم. از آن‌جا که همه‌ی ما تحت تاثیر این قالب‌گیری‌های اجتماعی قرار می‌گیریم، جیمز مارچ (۱۹۹۴) عقیده دارد اولین سوالی که ما باید از خودمان بپرسیم همان است که دُن کیشوت می‌پرسید: «من که هستم؟» همچنین برای تصمیم‌گیری باید همه‌ی جوانب را در نظر بگیریم، حتی آن‌هایی که در نگاه اول ارتباطی با انتخاب ما ندارند؛ مثلا سایر تصمیماتی که پیش رویمان هستند و یا نحوه‌ی تاثیر انتخاب ما بر روی تصمیم‌گیری فرد دیگر (همسر، دوست، همکار) در موضوعی متفاوت؛ با این استدلال که تصمیمات در لابه‌لای این‌ها صورت می‌گیرند نه در قالب ترتیب‌های سازمان‌یافته (دورفلر، ۲۰۲۱).

بنابراین، ما برخی اوقات از تصمیمات عقلانی فاصله گرفته و خود را به دست احساسات می‌دهیم، گاهی اوقات هیجانات بر ما غلبه می‌کنند و برخی مواقع نیز از ارزش‌ها پیروی می‌کنیم. برنامه‌نویسی برای هر یک از این‌ها کاری شدنی است، اما ما این ابعاد وجودی خود را از این طریق بدست نمی‌آوریم؛ قسمت عمده‌ی آن‌ها دقیقا مانند دانش در بعد ضمنی قرار گرفته‌اند. بیشتر محققان هوش مصنوعی از وجود احساسات، هیجانات و ارزش‌ها غافل نیستند، اما اغلب با گفتن جمله‌ی «بهتر است اول دانش کافی را ذخیره کنیم، بعدا بقیه‌ی کار را انجام خواهیم داد» آن را به بعد موکول می‌کنند، و یا فکر می‌کنند که هوش مصنوعی ممکن است این‌ها را در کنار دانش خودبه‌خود یاد بگیرد، و اگر هم نتواند احتمالا آنقدر مهم نیست که خود را با آن مشغول کنند. البته مورد عجیب مینسکی، اندیشمند دقیق و طرفدار حقیقی AGI در اینجا قابل ذکر است.

مینسکی (۲۰۰۶) استدلال اتاق چینی سرل را معکوس کرد و به عنوان یک آزمایش ذهنی، «ماشین زامبی» را ساخت. بنا بر فرض ماشین زامبی دقیقا ظاهری شبیه به انسان دارد و اگر دچار آسیبی بشود- مثلا پایش زخمی شود- مانند انسان از وجود درد شکایت می‌کند. مسلما نالیدن او از درد به برنامه‌نویسی او بر می‌گردد؛ اگر برنامه نویسی این ماشین خوب باشد، در نظر ما پاسخ او به درد از همین پاسخ در انسان‌ها غیر قابل تمایز می‌شود. آیا با آگاهی بر این موضوع باز هم می‌توان گفت که ماشین زامبی درد را حس نمی‌کند؟ در منطق استدلال اتاق چینی می‌توانیم این ادعا را داشته باشیم. اما مینسکی مساله را به همان نکته‌ای که در ابتدا می‌خواست بیان کند برمی‌گرداند: اگر نتوانیم بین انسان و این ماشین زامبی وجه تمایزی پیدا کنیم ولی باز هم بگوییم که زامبی در احساس درد ناتوان است، آیا واقعا می‌توان ادعا کرد انسان خود درد را حس می‌کند؟ مسلما خیر؛ گرچه همه‌ی ما انسان‌ها را مثلا در بازی فوتبال دیده‌ایم که تظاهر به احساس درد می‌کنند. با این حال هدف مینسکی صرفا طرح یک سوال زیرکانه نبود. او کاملا آگاه بود که احساسات در هوش (انسان یا موجودات دیگر) از اهمیت بالایی برخوردارند، بنابراین چنین بیان کرد:

مساله این نیست که آیا ماشین‌های هوشمند می‌توان هیچ گونه احساس و هیجانی داشته باشند یا خیر، بلکه سوال این است آیا می‌توان ماشین‌ها را بدون داشتن احساس هوشمند تلقی کرد؟ به نظر من اگر به ماشین‌ها قابلیت تغییر مهارت‌هایشان را اعطا کنیم، باید در کنار آن انواع سیستم‌های کنترل پیچیده را نیز فراهم نماییم. بی‌خود نیست وقتی که موجودی را به ماشین تشبیه می‌کنیم، دو معنی مخالف هم را می‌شود از آن برداشت کرد. یکی منظور از موجودی بی‌تفاوت، خالی از احساس و بدون هیچ علاقه‌ای؛ دیگری نیز به معنای موجودی

که بدون تغییر فقط یک هدف را دنبال می‌کند. بنابراین نه تنها عدم وجود انسانیت، بلکه نوعی حماقت را نیز نشان می‌دهد. التزام بیش از حد به یک چیز، باعث می‌شود که فقط یک مسیر را در پیش بگیریم؛ بی‌تفاوت بودن هم باعث سرگردانی می‌شود (مینسکی، ۱۹۸۸ : ۱۶۳).

۳-۴ رسیدن به معنا

فِین‌باوم پس از مواجهه‌ی اولیه‌ای که با لاجیک تئوریست داشت (بخش ۱-۲) پروژه‌ای را بر عهده گرفت که سیستم مقدماتی برخوردار از ادراک و حافظه (EPAM)[۳۳] نام داشت. هدف این پروژه ساخت مدلی از شناخت انسان بود. ماجرای EPAM از سپتامبر ۱۹۵۶، درست پس از بازگشت فین‌باوم از تعطیلات تابستان شروع می‌شود (فین‌باوم، ۲۰۰۶ : ۷:۳۰). سایمون یک مقاله در زمینه‌ی یادگیری انسان را از نشریه‌ی ساینتیفیک امریکن[۳۴] به او نشان داد که درباره‌ی موضوعی صحبت می‌کرد که در آن زمان جدید به شمار می‌رفت. مقاله مربوط می‌شد به یادگیری گفتاری[۳۵]؛ یا به طور دقیق‌تر، حفظ کردن فهرست‌هایی از کلمات تصادفی و بدون معنی. سایمون از نتایج این مقاله بسیار خشنود بود، چون داده‌هایی را در خود داشت که او به آن‌ها «یادگیری مقدماتی[۳۶]» می‌گفت و باور داشت این‌ها ساده‌ترین مثال‌های ممکن در محدوده‌ی یادگیری انسان هستند. EPAM به‌طرز فوق‌العاده‌ای موفق عمل کرد. اگر بنا بر این باشد که معنایی پشت واژه‌ها نباشد، نتایج آزمایشی حاصل (بهتر به یاد آوردن ابتدا و انتهای فهرست‌ها) باید به طریقی با ساختارهای حافظه‌ای مرتبط باشد. سایمون مدل کامپیوتری می‌خواست که توضیحات لازم را ارائه کند، و نیز یک پیش‌بینی کمی از نتایج آزمایشی در اختیارش قرار دهد. EPAM از پس این کار بر آمد و عنوان پایان‌نامه‌ی دکترای فین‌باوم را تشکیل داد (فین‌باوم و سایمون، ۱۹۸۴؛ سایمون و فین‌باوم، ۱۹۶۴). نکته‌ی مهم در اینجا این است که در آن زمان، معنا مدنظر آن‌ها نبود؛ بلکه تنها به ساختارهای حافظه‌ای توجه داشتند. گمان نمی‌کنم دست‌یابی به قلمرو معنا فقط با توجه به ساختارهای حافظه‌ای شدنی باشد؛ زیرا بدست آوردن معنا تنها با افزودن واِژگان بدون معنی ممکن نیست.

ظاهرا برای تشخیص این‌که هوش مصنوعی چه کارهایی را می‌تواند انجام دهد و در چه اموری ناتوان است، اینکه آیا هوش مصنوعی می‌تواند به همان شیوه‌ی انسانی هوشمند باشد، و نیز برای تشخیص این‌که هوش مصنوعی می‌تواند فکر کند یا خیر، لازم است مفهوم معنا را بهتر درک کنیم. ابتدا باید = فرض قبلی خود را- اینکه با استفاده از کلمات بی‌معنی نمی‌توان به معنا دست یافت- توجیه نمایم. ساده‌ترین اثبات موجود برای این فرض که از کلمات بی‌معنا معنایی حاصل نمی‌شود، از خود زبان منشا می‌گیرد. نوآم چامسکی (۱۹۵۷) نشان داد که ساختارهای نحوی مستقل از معنا هستند؛ به عبارتی دستور لغات و ساختار جملات بدون رجوع به معانی‌شان قابل درک هستند. این یافته نشان می‌دهد که ساختارهای حافظه‌ای ممکن است با هم مشابه باشند و مطالعه‌ی آن‌ها شاید در باب معنا اطلاعات چندانی به ما ندهد. پولانی استدلال کرد که در اینجا یک سیستم چند سطحی وجود دارد که در آن سطوح بالاتر با سطوح پایینی در تقابل نیستند، هرچند که با

[33]- Elementary Perceiver and Memorizer
[34]- Scientific American
[35]- verbal learning
[36]- elementary learning

استفاده از سطوح پایینی نمی‌توان به سطوح بالا دست یافت؛ درست‌تر این است بگوییم اجزای سطوح بالاتر در فضایی در حال فعالیت هستند که توسط سطوح پایین‌تر ایجاد شده است. به گفته‌ی پولانی: «شما نمی‌توانید با کمک آواها به دایره‌ی لغات دست پیدا کنید؛ نمی‌توانید از دایره لغات به دستور لغات برسید؛ استفاده‌ی درست از دستور لغات زیبا بودن سبک نگارش را تضمین نمی‌کند؛ و درستی سبک نگارش نیز محتوای یک نثر را به وجود نمی‌آورد» (پولانی، ۱۹۶۶الف : ۱۶).

با وجود این، باز هم ممکن است که ساختارهای حافظه‌ای بر اساس معنا باشند. در واقع دیدگاه غالب در روان‌شناسی شناختی این است که ما دانش گزاره‌ای[۳۷] را در قالب شبکه‌ای از معانی در ذهن‌مان ذخیره می‌کنیم (رومل‌هارت و نورمن، ۱۹۸۸). اختلاف‌نظرهای بسیاری در مورد شیوه‌ی ساخت این شبکه‌های معنایی وجود دارد. واضح است که این شبکه‌ها بر اساس معنای مفاهیم مختلف بنا شده‌اند. هر یک از این مفاهیم را می‌توان با تعدای مشخصه‌ی معنایی توصیف کرد. میان این مفاهیم ارتباطاتی وجود دارند و این ارتباطات نیز هر یک دارای مشخصه‌های معنایی هستند. به عنوان نمونه، مفهوم توپ با تمام اشکال کروی که در ذهن ما وجود دارند و نیز با خود مفهوم انتزاعی «کروی بودن» ارتباط پیدا می‌کند. این مفهوم با بسیاری از اجسام دیگر مانند میز نیز ارتباط دارد؛ این نوع ارتباط همچنین باید دارای نوعی توصیف باشد که بگوید: یک توپ می‌تواند بر روی میز گذاشته شود. و یا شاید اصلا این ارتباط وجود نداشته باشد؛ فرد تا زمانی که مثلا شکستن شیشه‌ی پنجره را توسط یک توپ نبیند یا آن را تصور نکند، ارتباط توپ و پنجره ممکن است شکل نگیرد. این یک مدل فرضی است و تا حد زیادی قابل پذیرش است؛ اگرچه عمدتا بر پایه‌ی فرضیات شکل گرفته است. آنچه که آزمایش‌ها نشان می‌دهند به ما می‌گوید که این شبکه‌های مفهومی که در ذهن ما جای گرفته‌اند، عجیب هستند. بنا به گفته‌ی داگلاس هوفشتاتر (۱۹۷۹)، این شبکه‌ها «لوپ‌های عجیبی» دارند که می‌توانند چندین سطوح از سیستم‌های مرتبه‌ای را به هم وصل کنند و باعث ایجاد پارادوکس و خودارجاعی شوند. برای مثال، مشخص شده است که چقدر طول می‌کشد تا ذهن از یک مفهوم به مفهوم دیگری که مستقیما به آن وصل است برسد. بر این اساس، تعیین شد که طی کردن مسیر ارتباط دلفین – ماهی چهار واحد طول می‌کشد، بدین صورت: دلفین – وال – پستاندار – ماهی. اما مشاهده می‌شود که برای اکثر مردم این مسیر تنها یک واحد است؛ به عبارتی ما یک مفهوم «غیر ماهی» داریم که مستقیما به مفهوم دلفین متصل است (رجوع شود به مرو، ۱۹۹۰). این یک مثال بدوی است که نشان می‌دهد چگونه چنین سلسله مراتب مستحکمی فرو می‌ریزد؛ در واقع اگر ما حقیقتا در ذهن‌مان دارای شبکه‌های معنایی باشیم، نمی‌توان گفت که آن‌ها پایدار هستند.

دو نفر در زمینه‌ی توسعه‌ی AI مشارکت‌های فراوانی انجام دادند و از مفهومی حمایت کردند که من آن را معنا می‌خوانم. اولین فرد تِری وینوگراد بود که پروژه‌ی SHRDLU10 متعلق به او نخستین برنامه‌ی پردازش زبان طبیعی (NLP)[۳۸] و نیای تمام NLPها و برنامه‌های تشخیص گفتار امروزی به شمار می‌رود. اما بعدا حیطه‌ی فعالیت وینوگراد تغییر پیدا کرد و او به یک فیلسوف مبدل شد.

[37]- propositional knowledge
[38]- Natural Language Processing

آنچه من به آن پی بردم این بود که موفقیت در برقراری ارتباط به هوش واقعی شنونده احتیاج دارد. با فرض این‌که کامپیوتر فاقد هوش باشد، راه‌های دیگر بسیاری برای ارتباط برقرار کردن با آن وجود دارد. در این‌جا من دیدگاهم را از اندیشه‌ی صرف درباره‌ی هوش مصنوعی، به یک پرسش وسیع‌تر تغییر دادم: «چگونه می‌خواهیم با یک کامپیوتر تعامل داشته باشیم؟» سپس مشتاق شدم بدانم که چه چیزی باعث می‌شود تعامل با کامپیوترها به خوبی صورت بگیرد و یا با شکست مواجه شود، و رسایی ارتباط با آن‌ها از کجا سرچشمه می‌گیرد. این پرسش‌ها جهت فعالیت‌ها من را شکل دادند (ماگریج، ۲۰۰۷ : ۴۵۷).

دیگر فردی که در زمینه‌ی هوش مصنوعی تاثیر شایان ذکری را داشت وایزن‌بام بود. وایزن‌بام پس از ساختن اولین نرم‌افزاری که توانست در آزمون تورینگ قبول شود (ELIZA را در بخش ۴-۲ ببینید)، به این نتیجه رسید که کامپیوترها در مفهوم معنا با مشکل روبه‌رو هستند. در واقع هوبرت دریفوس دستاورد او را پیروزی ذهن بر ماشین می‌داند: «وایزن‌بام نشان داد که چگونه یک نفر می‌تواند کامپیوتر را وادار به نشان دادن مقدار زیادی هوش واضح بکند، بدون آن‌که هیچ معنایی در او قرار دهد؛ با این کار او روش مینسکی پوچ و بیهوده جلوه داد» (دریفوس و دریفوس، ۱۹۸۶ : ۷۱). باید اعتراف کنیم که حداقل تا الان هیچ دستاوردی نداشته‌ایم که حتی شبیه به ایجاد معنی در AI شود؛ دست کم نه به آن صورتی که در ذهن انسان وجود دارد. تئودور روژاک، جامعه‌شناس سیستم‌های اطلاعاتی (IS) که در نظر من یکی از بهترین کتاب‌های حیطه‌ی IS را نوشته است، می‌گوید:

امید داشتن به تعبیر معانی توسط ماشین‌ها تنها یک خیال دلفریب و پوچ است. تعبیر معانی منحصرا به ذهن‌های زنده تعلق دارد، دقیقا همان‌طور که تولد صرفا برای اجسام زنده وجود دارد. اگر بخواهیم «تعبیر معانی» را از ذهن جدا کنیم، و یا «تولد» را به چیزی غیر از جسم زنده نسبت دهیم، تبدیل به یک استعاره می‌شوند (روژاک، ۱۹۸۶ : ۱۳۱).

۵-۳ همه‌ی اینها چه معنی می‌دهند؟

دانش علمی تنها دانش موجود و لزوما تنها نوع راستین آن نیست. یکی از معیارهای ضروری برای یک دانشمند خوب این است که ذهنی گشوده داشته باشد تا احتمال در اشتباه بودن را برای خود بپذیرد. به علاوه، دانش در اشکال و فرم‌های مختلفی ظاهر می‌شود و تنها نوعی که توسط AI قابل پذیرش است، حقایق ساده و صریح هستند. بنابراین به این چند نکته‌ی مهم باید توجه کرد:

- یک مساله‌ی مهم راجع به مدل‌سازی این است که ما تنها در مواردی می‌توانیم مدل‌سازی انجام دهیم که خودمان مرجع آن مدل را بفهمیم؛ اگر چه ما همه فهمیدن را تجربه کرده‌ایم، اما خودِ فهمیدن را نمی‌فهمیم و مدل‌سازی آن برایمان امکان‌پذیر نیست.
- هر گونه بازنمایی دانش[۳۹] به معنای محدود کردن دانش به نوع صریح آن است. هر فردی که در حیطه‌ی هوش مصنوعی فعالیت می‌کند از این موضوع آگاه است. آنچه که اغلب از آن غافل هستند این است که دانش معنادار بدون وجود دانش ضمنی غیرقابل تصور است.

[39]- knowledge representation

- علم تنها نیست. موارد دیگری از جمله احساسات، هیجانات و ارزش‌ها وجود دارند که با علم در تعامل هستند. مراقب باشید فورا از آن‌هایی که می‌گویند این‌ها فقط ناخالصی‌هایی هستند که تفکر عقلانی خالص را آلوده می‌کنند، جانب‌داری نکنید. احساسات، هیجانات و ارزش‌ها در زندگی ما، امور کسب و کار و خود علم نقش ارزشمندی دارند. طی مصاحبه با برندگان نوبل، آنچه که همه‌ی آن‌ها را به هم مرتبط می‌ساخت، عشق و علاقه‌ای بود که هر یک به حیطه‌ی پژوهشی خود داشتند.

به عنوان یک نکته‌ی شخصی نیز باید اضافه کنم که به عقیده‌ی من با ارزش‌ترین ابعاد دانش انسانی، دقیقا همان‌هایی هستند که توسط AI غیرقابل تقلید و تکرارند.این ابعاد شامل، شهود، درک زیبایی، استعاره‌ها و ادغام شدن دانش با قلمروهای دیگر (مانند احساسات، هیجانات و ارزش) هستند. علاوه بر این‌ها، دانش انسان مبنای حسی دارد (باس و همکاران، ۲۰۲۲)، بدین معنا که ما از تجربه‌هایمان درس می‌گیریم. بهتر است بگذاریم هوش مصنوعی کارهای کسل‌کننده را انجام دهد و در همین حین خودمان بر روی موارد هیجان‌انگیزتر متمرکز شویم. با این حال برای تحقق یافتن این موضوع، باید سیستم تحصیلی‌مان بر روی استعدادهای بشر تمرکز کند، نه کارهایی که AI می‌تواند انجام دهد.

۴ یادگیری

من به شخصه هر لحظه آماده‌ی یادگیری هستم، اگر چه همیشه از آن لذت نمی‌برم.

وینستون چرچیل

در این فصل یادگیری در انسان‌ها و ماشین‌ها مقایسه می‌شود. مانند فصل ۳، برای این کار از تجربیات حاصل از مصاحبه‌هایی که با برندگان جایزه‌ی نوبل داشتم، بهره می‌گیرم. البته که یادگیری همه‌ی ما انسان‌ها به همان شیوه‌هایی است که در اینجا توضیح خواهم داد ؛ اما شاید الگوهای یادگیری در افرادی که به بالاترین درجات استادی رسیده‌اند راحت‌تر دیده شوند. این‌که اصلا ما چگونه می‌توانیم یاد بگیریم همیشه مرا متعجب شگفته زده کرده است.

۴-۱ استعداد و الهام‌بخشی

نخست بگذارید نگاهی به برخی مقدمات یادگیری بیندازیم. هر یک از ما می‌تواند درباره‌ی خود بگوید که در چه زمینه‌هایی یادگیری خوبی داشته است و کجاها دچار مشکل می‌شد. برای من دروس ریاضی و فیزیک هیچ‌گاه دردسرساز نبودند. با این وجود همیشه با درس تاریخ مشکل داشتم. بعدها در دبیرستان معلم تاریخ بسیار خوبی داشتم و این درس دیگر برایم مشقت‌بار نبود. در نتیجه ظاهر امر به من می‌گوید که دو چیز در اینجا اثر بسیار بزرگی بر نحوه‌ی یادگیری ما دارند: استعدادی که در یک زمینه داریم و چگونگی احساس ما نسبت به آن.

با استعداد بودن در چیزی یعنی چه؟ صرف نظر از بحث‌های عمیق روان‌شناسی شناختی، ما در حیطه‌های خاصی نسبت به سایر زمینه‌ها متمایل‌تر هستیم و یادگیری ما در این حیطه‌ها سریع‌تر است (گاردنر، ۱۹۹۵)؛

که آن را استعداد و یا - به بیانی زیباتر- موهبت می‌نامیم؛ موهبت از آن جهت که خودمان آن را کسب نکرده‌ایم. اما متمایل بودن به چه معناست؟ چرا ما در برخی زمینه‌ها سریع‌تر یاد می‌گیریم؟ جواب من به این پرسش‌ها جزو دیدگاه غالب در روان‌شناسی شناختی به شمار نمی‌رود؛ زیرا با موضوع بحث‌برانگیزی سروکار داریم. همچنین این جواب برگرفته از مقالات روانشناسی شناختی نیست؛ بلکه سخنان یک دبیر ریاضی است که تعدادی دانش‌آموز را تحت نظر داشت و آن دانش‌آموزان بعدها به ریاضی‌دانان خارق‌العاده‌ای تبدیل شدند. این خانم دبیر می‌گفت که انگار این بچه‌ها از قبل تمام ساختارهای ریاضی را در ذهن خود دارند، و من تنها آن ساختارها را برای آن‌ها نشانه‌گذاری کردم. نمی‌خواهم از دیدگاه قدیمی دانش ذاتی در یک زمینه‌ی خاص طرفداری کنم، اما برای من این‌طور به نظر می‌آمد که دانش روزمره‌ی این دانش‌آموزان همانند دانش حیطه‌ای که در آن مستعد بودند ساماندهی می‌شد. به حیطه‌ای که یادگیری‌تان در آن سریع است فکر کنید: غالبا یادگیری در آن بیشتر شبیه به بازی است تا کاری طاقت فرسا؛ یاد گرفتن برایتان سرگرم‌کننده است، فعالیت-های چالش برانگیز را به راحتی انجام می‌دهید و همه‌ی اینها به گونه‌ای است که انگار اصلا در حال یادگیری هم نیستید. حس می‌کنید کافی است چیزی را نشان‌تان دهند؛ سریعا آن را یاد گرفته و خودبخود شروع به کار می‌کنید. البته تجربه‌ی مخالف این را هم داشته‌ام: زمانی چندین آکورد گیتار را یاد گرفتم و پس از گذشت سال‌ها فقط همان را بلد بودم. یکبار دانشم را به فردی که واضح بود در موسیقی استعداد دارد نشان دادم. پس از گذشت فقط چند ساعت توانست از من بهتر گیتار بزند و پس از یک هفته، می‌توانست هر موسیقی‌ای را تنها با شنیدن یاد بگیرد و بنوازد. در برخی مقاله‌ها در زمینه‌ی رشد مهارت، سعی شده است که نقش استعداد را کم‌رنگ جلوه دهند (اریکسون و چارنِس، ۱۹۹۴)؛ به عبارت ساده‌تر، همانطور که ژورنالیست‌ها اغلب درباره‌ی دستاوردهای بزرگ می‌گویند که: ۱۰ درصد استعداد و ۹۰ درصد سخت‌کوشی باعث رقم زدن آن‌ها شده است. به نظر من این عقیده بی‌معنی است. در واقع به نظر می‌رسد دستاوردهای فوق‌العاده از ترکیب ۱۰۰ درصد استعداد ۱۰۰ درصد تلاش زیاد به‌وجود می‌آیند. اما عنوان این کتاب «پدیده‌های فوق‌العاده» (دورفلر و استیراند، ۲۰۱۹) نیست، بنابراین از ذکر جزئیات در مورد افراد دارای موهبت‌های خاص خودداری می‌کنم. آنچه اهمیت دارد این است که همگی ما در برخی از زمینه‌ها نسبت به بقیه استعداد بیشتری داریم و این وابستگی انسان به استعداد اغلب به عنوان یک ناکارآمدی در مقایسه به AI تلقی می‌شود. AI به استعداد نیازی ندارد. تا حالا ندیده‌ام جایی این را بگویند، ولی به‌طور ضمنی می‌توان برداشت کرد که عقیده بر این است: «هوش مصنوعی در هر چیزی استعداد دارد». حتی اگر بتوانیم استعداد را برای هوش مصنوعی تعیین کنیم، یقینا می‌گفتم که AI در هر زمینه‌ای به یک میزان استعداد دارد، که این معادل با این است که بگوییم AI در هیچ زمینه‌ای استعداد ندارد. برای اثبات این موضوع این نکته را تذکر می‌دهم که استعداد یعنی شهود در یک حیطه‌ی خاص زمانی شروع به فعالیت بکند که هنوز سطح مهارت لازم برای بروز شهود در آن کسب نشده است؛ و همانطور که قبلا بحث شد، هیچ مدرکی نداریم که نشان دهد شهود در AI امکان‌پذیر است.

فارغ از حیطه‌ای که در آن یاد می‌گیریم، می‌توان گفت که فرد آموزش‌دهنده نیز تفاوت بسیاری ایجاد می‌کند. مسلما در این حد ساده نیست که بگوییم برخی معلمان خوب و برخی بد هستند. البته چنین گزاره‌ای درست است، اما همه‌ی معلمان خوب به یک اندازه و برای هر فرد مستعد خوب نیستند. سال‌ها قبل یک دانشجوی کارشناسی داشتم که به او توصیه کرده بودم تا گرفتن PhD به تحصیل ادامه دهد، چراکه عملکرد او عالی بود؛

ولی نه در نزد من. نمی‌توانستم علت آن را توضیح دهم، فقط به طور شهودی می‌دانستم که ما با هم سازگاری خوبی نداشتیم. ما تا چند سال بعد آن‌ها با هم در ارتباط بودیم و او تصدیق کرد زمانی که نحوه‌ی کار کردن من با دانشجویان دکتری را دید، فهمید که نمی‌توانست نزد من شاگردی کند. بنابراین همیشه راه‌های متفاوت و متعددی برای یادگیری یک چیز وجود دارد، و برخی راه‌ها با بعضی افراد همخوانی بیشتری دارند. همچنین ممکن است لازم باشد که در مراحل مختلف یادگیری‌مان، سبک‌های یادگیری متفاوتی را دنبال کنیم. در مورد AI هیچ یک از این‌ها صدق نمی‌کنند؛ برای AI فقط یک سبک یادگیری وجود دارد، که در بخش بعدی محدویت‌های آن ذکر خواهند شد.

اما قبل از آن باید مفهوم الهام گرفتن را آن‌طور که طی مصاحبه با برندگان نوبل به آن پی‌بردم در این‌جا مطرح نمایم. در زندگی این افراد یک الگوی انگیزشی بزرگ وجود داشته است؛ الگویی که تصمیم آن‌ها را در مورد این‌که می‌خواهند چه کسی باشند رقم زده است؛ یک معلم الهام‌بخش بزرگ. این معلمان الهام‌بخش بسیار معدودند؛ همه‌ی برندگان نوبل فیزیکی که با آن‌ها مصاحبه داشتم (به جز یک مورد) از ریچارد فاینمن یا انریکو فرمی الهام گرفته بودند (دورفلر و ادن، ۲۰۱۷). مواردی از این دست تغییرات عمیق در یادگیری ما هستند که تحت عنوان «مفاهیم آستانه» نشان داده می‌شوند (مِیر و همکاران، ۲۰۱۰). مفاهیم آستانه همچنین بیان می‌کنند که یادگیری یک تمرین ساده و جمع‌شونده نیست؛ یعنی علاوه بر آنکه ما طی یادگیری دانش بیشتری به دانش قبلی می‌افزاییم، معنای مفاهیمی را که قبلا فرا گرفته بودیم هم گسترش می‌دهیم، دسته‌ای از مفاهیم را مجددا تعبیر می‌کنیم، و آن‌هایی را که توسط مفهوم‌های بهتر جایگزین شده‌اند و یا صرفا غلط هستند حذف می‌کنیم (این‌ها معمولا فراموش نخواهند شد؛ و به‌خاطر خواهیم سپرد که این مفاهیم برای ما نادرست بودند). از این رو یادگرفتن در اصل تعامل یادگیری و سلب یادگیری است؛ این تعامل در یک سیستم پیچیده اتفاق می‌افتد که در آن محتویات دانش فقط اضافه شدنی نیستند، بلکه با هم تعامل دارند. گاها در طول فرایند یادگیری، درک عمیق‌تری پیدا می‌کنیم و این نشان می‌دهد. بسیاری از عناصر (دانش) از پیش موجود و بعضا نیز برخی از عناصر جدید، در یک الگوی پیچیده آرایش پیدا می‌کنند. در روان‌شناسی شناختی به چنین نمونه‌هایی از یادگیری تحولی[۴۰]، شکل‌گیری فرا الگو[۴۱] گفته می‌شود. این پدیده تا حدودی مترادف است با مفاهیم آستانه (رجوع شود به دورفلر، ۲۰۱۰). بدون شک AI از این پدیده‌های یادگیری تحولی برخوردار نیست؛ یادگیری هوش مصنوعی از مدل هرم پیروی می‌کند (اطلاعات جدید همواره با پذیرش اطلاعات قدیمی تر و با تکیه بر آن‌ها ساخته می‌شوند). اگر یادگیری ما انسان‌ها هم فقط از مدل هرم پیروی می‌کرد، به پیشرفت‌های زیادی دست نمی‌یافتیم.

۲-۴ فراتر از موش‌های اسکینر

در یک رویداد زنده‌ی نیو ساینتیست[۴۲] در سپتامبر ۲۰۱۷ در لندن، دیمس هاسابیس، بنیانگذار و مدیر عامل DeepMind در سخنان خود به دو مورد اشاره کرد: یکی از این موارد مربوط به همین فصل است و دیگری، در فصل ۵ ذکر خواهد شد. نخستین مورد درباره‌ی الگوریتم‌های یادگیری DeepMind بود؛ به گفته‌ی او

[40]- transformational learning
[41]- meta schema
[42]- New Scientist Live

«DeepMind مانند انسان‌ها - از طریق یادگیری تقویتی[۴۳]- یاد می‌گیرد». بگذارید توضیح دهم که منظور او چه بوده است. حدودا در نیمه‌ی اول قرن بیستم، دوران تاریخی در رشته‌ی روان‌شناسی به‌وجود آمد که معمولا آن را با نام «روان‌شناسی رفتارگرا» می‌شناسند. همانطور که از نامش پیداست محور این روان‌شناسی رفتارهای انسان است، نه آنچه که در ذهن رخ می‌دهد. در واقع این رویکرد روان‌شناختی ذهن را یک «جعبه‌ی سیاه» می‌بیند؛ در نتیجه به رشته‌ای تبدیل شد که عنوان پژوهشی خود را نقض می‌کند؛ حقیقتا این دوره، «عصر تاریک» روان‌شناسی است. در بطن رویکرد یادگیری تقویتی، مدلی وجود دارد که به آن مدل محرک – پاسخ (S-R)[۴۴] می‌گویند. طبق این مدل، ما می‌توانیم با بررسی محرک دریافتی توسط موجود (انسان، حیوان یا ماشین) و پاسخی که به آن محرک می‌دهد، رفتار او را مطالعه کنیم. چنین شرایطی را در یکی از انواع یادگیری‌های تقویتی خواهیم داشت؛ در این نوع یادگیری به پاسخ‌های مطلوب پاداش می‌دهیم و (یا) برای پاسخ‌های نامطلوب مجازات تنظیم می‌کنیم. اگر موجود توانایی یادگیری داشته باشد، همواره دیر یا زود رفتاری را نشان می‌دهد که منجر به دریافت پاداش (یا اجتناب از مجازات) شود. این چارچوب نتایج زیادی به همراه داشت؛ ایوان پاولوف (۱۹۲۷) با این روش توانست رفلکس‌های پاولووین را بشناسد و بوروس اسکینر (۱۹۵۰) نیز به موش‌هایش یاد داد که در هنگام گرسنگی اهرمی را فشار دهند. اگرچه شرطی‌کردن فعال و دیگر انواع یادگیری تقویتی بر روی موش‌ها خوب کار می‌کنند؛ اما در اینکه انسان‌ها نیز با این روش یاد می‌گیرند جای شک هست.

منطقی است که تورینگ در آن زمان (۱۹۵۰) یادگیری تقویتی را مدنظر قرار داده باشد. اما امروزه روان‌شناسی راه زیادی را از زمان «عصر تاریک» طی کرده است؛ امروزه می‌دانیم که ما انسان‌ها به مقدار بسیار کمی از یادگیری تقویتی بهره می‌گیریم. البته مواردی هستند که در آن انسان از این طریق یاد می‌گیرد؛ به عنوان مثال، کودک سریعا یاد می‌گیرد که نباید به شعله‌ی گاز دست بزند. اما به هر حال، ما از داستان‌ها (گفتاری یا نوشتاری) هم یاد می‌گیریم، دیگران را مشاهده و رفتار آن‌ها را تقلید می‌کنیم، رابطه‌ی استاد- شاگردی با دیگران برقرار می‌کنیم، خود راه‌های جدید ابداع نموده و به خودمان یاد می‌دهیم و ... . یادگیری تقویتی تنها قسمت کوچکی از یادگیری کلی ما را تشکیل می‌دهد. شرط گذاشتن برای انسان‌ها هم ممکن است، با این تفاوت که اکثر ما از شرط گذاشتن خوشمان نمی‌آید، و پس از یادگیری کارهای بسیار بیشتری انجام می‌دهیم.

همین مواد آموزشی را می‌توان با چندین روش دیگر نیز یاد گرفت؛ به‌خصوص اگر دانش موردنظر خیلی پیچیده نباشد. برای مثال ریاضیات دانشگاهی را در نظر بگیریم؛ عده‌ای ترجیح می‌دهند توضیحات انتزاعی کتاب را بخوانند، برخی مایل هستند سخنرانی استاد را به همان شیوه‌ی انتزاعی گوش بدهند، و بقیه نیز دوست دارند همین مفاهیم مجرد (انتزاعی) را در قالب حل مساله ببینند، و بر روی حل مسائل کار می‌کنند تا زمانی که به اصول کلی‌تر و مجردتر برسند. حتی دیده‌ام دسته‌ای با نوشتن قضایا و برهان آن‌ها به خوبی یاد می‌گیرند، اما همه‌ی اینها دانش واضحی هستند و در دایره‌ی دانش پیچیده قرار نمی‌گیرند.

[43]- reinforcement learning
[44]- stimulus-response

شاید پیچیده‌ترین شکل یادگیری در روابط استاد و شاگرد رخ می‌دهد. جالب است ذکر کنم که ما در واقع نمی‌دانیم رابطه‌ی استاد – شاگردی چگونه کار می‌کند. آنچه که بر آن واقفیم، این است که ظاهرا این رابطه تنها شیوه‌ی انتقال دانش ضمنی است و همه‌ی کسانی که به بالاترین درجات استادی رسیده‌اند- از جمله دارندگان نوبل که با آن‌ها مصاحبه داشتم- نوعی رابطه‌ی استاد – شاگردی را پشت سر گذاشته‌اند. «روش‌های تحقیق علمی را نمی‌توان به‌صورت صریح بیان کرد و به همین دلیل تنها از راه ارتباط شاگردان با استاد منتقل می‌شود؛ درست مانند آموزش هنر» (پولانی، ۱۹۶۹ : ۶۶). یک فرد نظاره‌گر عادی خواهد گفت که شاگرد دارد از استاد تقلید می‌کند؛ با این وجود، این کار یک تقلید ساده نیست؛ بلکه چیزی است که پولانی (۱۹۴۶ : ۳۰-۲۹) آن را «تقلید هوشمندانه» می‌خواند. در تقلید هوشمند برخلاف تکرار طوطی‌وار و بدون فکر، یادگیری ما به‌شیوه‌ای قابل تطبیق انجام می‌گیرد، و ما می‌توانیم آنچه را که می‌آموزیم تغییر دهیم، گسترده‌تر کنیم و ... . برای مثال ما زبان مادری خود را به همین شیوه یاد می‌گیریم و تا اینکه در نهایت از والدین و معلمانی که زبان را به آموزش داده‌اند بهتر می‌شویم. این مشابه همان مساله است که یک شاگرد می‌تواند در نهایت از استادش پیشی بگیرد. در عمق رابطه‌ی استاد شاگرد یک پارادوکس نهفته است. شاگرد باید از استادش تقلید کند، چرا که استاد از او داناتر است. در همین حال، تقلید از استاد کار اشتباهی است، چرا که در آن صورت شاگرد تنها به یک کپی ضعیف از استادش تبدیل می‌شود. این پارادوکس که کار را برای شاگرد دشوار می‌کند. اینکه شاگرد بتواند به عنوان یک استاد جدید – یک نسخه‌ی تقویت‌یافته از استاد خود- ظهور کند، و نه اینکه تنها به یک نسخه‌ی ضعیف از استادش مبدل شود، کوشش زیادی می‌طلبد. بخشی از این یادگیری در سطح فراشناختی رخ می‌دهد، یعنی استاد با بهره‌گیری از مثال‌ها، نه تنها نحوه‌ی انجام کارها را می‌آموزد، بلکه نشان هم می‌دهد که هر یک از این آموزه‌ها در جای خود چقدر با ارزش هستند.این ارزش توسط شاگرد درک و احساس خواهد شد اگر، و تنها اگر، شاگرد دارای موهبت و انگیزه باشد. در آن صورت شاگرد می‌تواند به یک نسخه‌ی پیشرفته‌تر «ارتقای سطح» پیدا کند. رابطه‌ی استاد و شاگرد اگر به‌صورت یک تمرین کاملا ذهنی صورت بگیرد، شدنی نخواهد بود. برای به وقوع پیوستن این رابطه هیجانات نیز به اندازه‌ی اندیشه‌ها لازم هستند؛ باور داشتن و شجاعت در فراتر رفتن از آنچه که هستیم ضروری است. این یک شیوه‌ی کهن در یادگیری است. سقراط و مریدان او را در آگورا[۴۵] تصور کنید؛ یا کارگاه‌های هنری را که لئوناردو یا مایکل آنجلو در آن مشغول به تلمذ بودند در نظرتان مجسم نمایید. متاسفانه امروز روابط استاد – شاگردی به دلیل نامتقارن و طولانی بودن آن‌ها محبوبیت چندانی ندارند. من در تمام سخنرانی‌ها و نوشته‌هایم از این روابط دفاع و حمایت می‌کنم تا این شیوه‌ی ارزشمند یادگیری همچون گنجینه‌ای گرانبها حفظ شود.

توماس داون پورت و لری پروساک (۲۰۰۰) در کتاب «دانش کارآمد»[۴۶] داستان یک تحلیل‌گر تصاویر ماهواره‌ای را بازگو می‌کنند که در کار خود به بهترین جهان معروف بود. با نزدیک شدن زمان بازنشستگی او، شرکت نفتی که در آنجا مشغول به کار بود، یک طراح سیستم‌های خبره را استخدام کرد تا تخصص این فرد را کد نویسی کند (داون‌پورت و پروساک، ۲۰۰۰ : ۸۴). داستان در این جا متوقف شده و خواننده با یک پاسخ نصفه

[۴۵]- Agora: میدان‌هایی در شهرهای یونان باستان

[46]- working knowledge

و نیمه تنها گذاشته می‌شود؛ به این عبارت که سیستم خبره مذکور با شکست مواجه شد. تا اینکه در ۱۱ صفحه بعد داستان این‌بار با شفافیت کامل بیان می‌شود:

برنامه نویس مانند یک شاگرد، طی فرایند طولانی مدتی دانش فرد خبره را استخراج کرده و یاد گرفت. گفتگوهای طولانی مدت، بررسی دو نفره‌ی تصاویر، پرسش سوالات و جستجو برای درک بیشتر همگی باعث شدند که برنامه نویس مهارت او را بیاموزد. پروژه تمام شد و سیستم خبره‌ی حاصل کاملا به درد نخور بود، اما طراح سیستم اکنون دومین تحلیل‌گر برتر تصاویر هوایی در جهان محسوب می‌شد! (داون‌پورت و پروساک، ۲۰۰۰ : ۹۵)

۳-۴ رسوخ کردن[۴۷]

ایده‌ی رسوخ کردن، مانند بسیاری از ایده‌های مرتبط با دانش و یادگیری توسط پولانی بنیان‌گذاری شد (۱۹۶۲ب). او این ایده را از طرافداران هنر برگرفت؛ آن‌ها از این مفهوم برای توضیح مهارت هنری استفاده می‌کردند: ما هنر را فقط از طریق رسوخ کردن می‌توانیم یاد بگیریم، یا به عبارتی از طریق گنجاندن خود در درون شیء مورد نظر. پولانی (۱۹۶۲ب) این مفهوم را به دانش تجربی و حتی به کل محدوده‌ی دانش ضمنی گسترش داد: «ما خود را به درون آن جاری می‌کنیم و آن را تبدیل به بخشی از وجود خود می‌سازیم» (پولانی، ۱۹۶۲الف : ۶۱).

اینکه آشپزهای ماهر چنان چاقو را در دست می‌گیرند که انگار بخشی از جسم‌شان است، برای بیشتر ماها به راحتی قابل درک است. اما تصور این‌که چگونه خلبانان جت جنگی جت‌شان را به عنوان بخشی از بدن خود ببینند و یا اخترفیزیک‌دان‌ها به درون «کهکشان‌های بسیار بسیار دور» رسوخ بکنند اندکی دشوارتر است. و اینکه یک فیزیک‌دان نظری به درون تقارن و عدم‌تقارن‌های کوارکی رسوخ کند و یا ریاضی‌دانی خود را در مباحث انتزاعی توپولوژی بگنجاند حقیقتا غیر قابل درک است. اما آنچه که خود این افراد می‌گویند بسیار شبیه همان نگاه آشپز به چاقوهایش است. از این رو ما می‌توانیم به رسوخ کردن به عنوان یک مفهوم جامع‌تر از تجسم دانش[۴۸] نگاه کنیم (دورفلر و استیراند، ۲۰۱۸). این جنبه‌ی بسیار مهمی از یادگیری است و اهمیت یادگیری تجربی را نشان می‌دهد؛ با این وجود رسوخ کردن به مباحث یادگیری تنها یکی از ابعاد رسوخ است که می‌توانند در یادگرفتن به ما کنند.

جنبه‌ی مهم دیگر رسوخ مربوط به افراد پیرامون است. شاید به همین دلیل باشد که معلمان الهام‌بخش تا این حد دارای اهمیت هستند؛ البته شواهد خاصی برای اثبات این موضوع وجود ندارد. آنچه که واضح‌تر است، برقرار بودن رسوخ در روابط استاد و شاگرد است. شاگردان در ابتدای کار، در سطحی قرار ندارند که بتوانند به راحتی به مبحث مورد علاقه‌شان رسوخ نمایند. با این وجود، آن‌ها می‌توانند به درون استادشان رسوخ پیدا کرده و استاد نیز به درون مبحث ورود پیدا کند. و بدین شیوه، شاگردان به کمک استادشان وارد مبحث مورد نظر می‌شوند. این شکل از رسوخ را می‌توان رسوخ اشتراکی[۴۹] خواند. صرف‌نظر از نوع سنتی رابطه‌ی استاد و

[47]- Indwelling
[48]- embodiment of knowledge
[49]- shared indwelling

شاگردی که در بخش قبل توضیح داده شد، رسوخ اشتراکی در «شاگردان سیار[50]» - که استادشان را پی در پی عوض می‌کنند- هم بسیار دیده می‌شود. این رسوخ شاید نه به علت جنبه‌های خاصی از دانش، بلکه به دلیل تنوع ابعاد رسوخ ایجاد شود. سومین نوع رابطه استاد – شاگردی که شناسایی کرده‌ایم (من و یکی از مصاحبه‌شوندگان به نام اریک ویشوس)، «شاگردی متقابل[51]» نام گرفته است (دروفلر و ادن، ۲۰۱۷). برای برقراری این رابطه لازم است دو نفر که مهارت بالایی دارند ولی هنوز به درجه‌ی استادی دست نیافته‌اند در زمان مناسب همدیگر را ملاقات کنند و شاگرد و استاد یکدیگر شوند. مثالی از شاگردی کردن متقابل، رابطه‌ی دنیل کانمن و عاموس تورسکی است؛ دیگران این دو را «یک ذهن در قالب دو جسم» توصیف می‌کنند. پدیده‌ی ذهن مشترک را در انجمن‌های خبرگی (CoP)[52] نیز مشاهده کردیم، چیزی که آن را «تفکر گروهی» نامیدیم (پیرکو و همکاران، ۲۰۱۷) و دریافتیم که تفکر گروهی یگانه عامل ضروری در شکل‌گیری CoPهاست. به این پدیده «رسوخ متحد[53]» می‌گوییم: فرایند رسوخ به‌طور متحدی بر یک موضوع متمرکز است و افراد می‌توانند همدیگر را به سمت فهمیدن مساله‌ی واقعی که همه از آن اطلاع دارند هدایت کنند و در نتیجه، دانش ضمنی را با هم به اشتراک بگذارند (پیرکو و همکاران، ۲۰۱۷ : ۳۹۰).

شاید گفتن این جالب باشد که پایه‌گذاری مفهوم اولیه‌ی CoP توسط ژان لاو و اتین ونگر (۱۹۹۱)، نتیجه‌ی مطالعه بر روابط شاگرد و استاد بود. آنچه که مفهوم CoP را متفاوت می‌کند این است که در موارد قبل از آن، افراد به‌صورت دو نفره در یکدیگر رسوخ می‌کردند (البته برای شاگردان سیار این دو نفر متغیر بودند)، اما در CoP افراد به‌صورت یک اجتماع عمل رسوخ را انجام خواهند داد. در این‌صورت شاید تعجب برانگیز نباشد که بگوییم نوع چهارم رابطه‌ی استاد – شاگردی (که ما آن را «کانون گرم» نامیده‌ایم- این نام توسط روی گلابر، یکی از مصاحبه‌شوندگان پیشنهاد شد)، نوعی CoP با عملکرد بسیار قوی محسوب می‌شود. باید توجه داشته باشیم که در فرایند یادگیری و گاها طی فرایند عمل، دانش از شخص فراتر می‌رود؛ برای همین بهتر است در مورد دانش بین‌شخصی نیز صحبت شود. اهمیت رسوخ در تمام اشکال یادگیری مذکور، از حیث تجربی بودن آن است؛ این جنبه‌ی تجربی به دنیای قابل لمس پیرامون محدود نمی‌شود و می‌تواند تمام یک مبحث علمی، حتی ابعاد دوردست و مجرد آن را در بر بگیرد، بدون آنکه از حیطه‌ی تجربه خارج شود (رجوع کنید به باس و همکاران، ۲۰۲۲). اهمیت تجربی بودن رسوخ به ماهیت تجربی مفاهیم آستانه بر می‌گردد؛ به عبارتی، یادگیری تحولی نیازمند تجربه و همچنین توانایی بازتاب دادن این تجارب است (کانمن، ۲۰۱۱ را ببینید).

خوب است اکنون نگاهی به شبیه‌سازی‌ها بیندازیم. زمانی که من دانشجوی MBA بودم، یک بازی شبیه‌سازی کسب و کار در کلاس من نمایش داده شد. در پایان تمرین متوجه شدم آن‌هایی که دانشجویان PhD بودند و تجربه‌ای در زمینه‌های سازمانی نداشتند، بهترین نتایج را کسب کردند، در حالی که مدیرعاملان مشهور و باتجربه ضعیف‌ترین عملکرد را داشتند. چرا چنین رخ داده بود؟ شبیه‌سازی صرفا طبق دیدگاه برنامه‌نویسان بازی به کسب و کار عمل می‌کرد؛ شبیه‌سازی آن‌ها مزخرف بود. البته این برای هر شبیه‌سازی صدق می‌کند.

[50]- wandering apprentices
[51]- mutual apprenticing
[52]- communities of practice
[53]- interlocked dwelling

در برخی موارد، شبیه‌سازی تا حد زیادی به واقعیت نزدیک است؛ مثلا به بازی شطرنج یا گو نگاه کنید. اما در مواقع دیگر مانند ازدواج، سیاست و یا مدیریت کسب و کار، هیچ شباهتی با دنیای واقعی ندارد. پیدا کردن این مثال‌های متضاد راحت است، اما ساخت یک اتومبیل خودران به کدام دسته از مثال‌ها نزدیک‌تر است؟ این نمونه‌های مشابه را در هر سازمانی پیدا خواهید کرد.

۴-۴ عقل سلیم

بیشتر پژوهشگران هوش مصنوعی، خواه در حوزه هوش مصنوعی ضعیف و خواه در زمینه هوش مصنوعی قوی فعالیت داشته باشند، بر یک مورد توافق دارند: آنچه که یک هوش مصنوعی حقیقی- ماشینی با تفکر واقعی- کم دارد، ایجاد یک مدل از عقل سلیم[۵۴] است. من (هوش مصنوعی ضعیف) این مطلب را نخستین بار در نوشته‌های مینسکی (هوش مصنوعی قدرتمند) خواندم و کاملا با آن موافق بودم. آنچه که در آن اختلاف نظر داریم میزان دشواری ساخت چنین مدلی است. طرفداران AI قوی باور دارند که این مدل تنها یک گام پیش‌رو است که بزودی آن را خواهند یافت. این در حالی است که حامیان AI ضعیف باور دارند که این کار غیرممکن است. اما این عقل سلیم چیست که چنان موجب سردرگمی است و احتمالا کلید درک توانایی‌های AI در آینده خواهد بود؟

عقل سلیم جنبه‌ای مسلم از دانش روزمره است و ما هر لحظه، بی‌توجه به میزان تخصص امری که با آن روبرو هستیم و حتی در زندگی شخصی، از آن بهره می‌گیریم؛ به‌طوری که اغلب حتی به چشم دانش نیز به آن نمی‌نگریم. با این حال بدون عقل سلیم، هر دانشی غیر ممکن می‌نماید. همه‌ی تعابیر، عقل سلیم را به‌عنوان مرجعی برای قضاوت درست و استفاده از ادراک حسی و فرایندهای تفکر عقلانی می‌دانند؛ البته در اینجا به تعریف دقیق آن نخواهیم پرداخت. در گفتار روزمره، توضیحی برای عقل سلیم وجود ندارد؛ البته به دلیل ساده بودن آن نیازی به توضیح دیده نمی‌شود. اما اگر واقعا سعی کنیم چیزی را در قلمرو عقل سلیم توضیح دهیم، غالبا به این نتیجه می‌رسیم که کار دشواری در پیش داریم. مهم‌تر اینکه عقل سلیم، همواره با دنیای واقعی (شرایط عملی) مرتبط است و هرگز در فلسفه‌بافی‌های انتزاعی وارد نمی‌شود. و به همین دلیل فلسفه‌بافی درباره‌ی خود آن بسیار دشوار است.

اگرچه بسیاری از جنبه‌های عقل سلیم ساده به نظر می‌رسد، خود عقل سلیم به دور از سادگی است. به نظر می‌رسد عقل سلیم عموم مردم به نحوی در یک جهت موافق هستند، با این حال ظاهرا در ذهن افراد مختلف ساختار مختلفی دارند. به همین دلیل عقل سلیم بسیار شخصی است، حتی اگر بسیاری از پیامدهای آن در میان مردم هم جهت باشند. به علاوه، این هم‌جهت بودن تا حدودی در اجتماع‌های کوچک رخ می‌دهد. اگر نگاهی به تفاوت‌های فرهنگی میان ملت‌ها و نواحی مختلف و یا حتی سازمان‌ها بیندازیم، خواهیم دید که عقول سلیم این افراد می‌تواند بسیار متفاوت از هم باشند. امری که در یک‌جا «مثل روز روشن است»، در جای دیگر و توسط عقل سلیم دیگر شاید کاملا متفاوت تلقی شود. از این رو عقل سلیم هم فردی و هم بین‌فردی است، و مرتبط و وابسته به مفهوم است.

[54]- common sense

عقل سلیم چیز ساده‌ای نیست؛ بلکه اجتماعی عظیم از ایده‌های عملی و سخت به دست آمده، و انبوهی از قواعد زندگی و استثنائات، استعدادها و تمایلات، و کنترل و معیارهاست. اگر عقل سلیم به راستی چنین پیچیده و بی‌کران است، چه چیزی موجب شده است آنقدر واضح و طبیعی به نظر بیاید؟ این توهم ساده بودن از فراموش شدن وقایع دوران طفولیت ما برمی‌خیزد؛ زمانی که در طول عمرمان می‌خواهیم از این دوران یاد کنیم، تنها می‌توانیم بگوییم «به‌خاطر نمی‌آورم» (مینسکی، ۱۹۸۸ : ۲۲).

البته قرار نیست در لابه‌لای متن این کتاب مساله‌ی عقل سلیم را بشکافیم. اما خوب است سعی کنیم معنای عقل سلیم را در هوش مصنوعی یاد بگیریم. اگر عقل سلیم را در مقابل دانش تخصصی قرار دهیم، خواهیم دید موفقیت‌های AI همگی در بخش دانش تخصصی بوده است به ویژه‌ی در حیطه‌های تخصصی محدود. در این جا، در بعضی از نواحی دیده‌ایم که AI از افراد خبره بهتر عمل می‌کند (فصل ۵ بیشتر به آن می‌پردازد) و در دیگر حیطه‌ها AI عملکرد برابری با افراد دارای سطح تخصصی متوسط دارد. در این جا مجددا جانب مینسکی را می‌گیرم: «چرا برنامه‌نویسی برای انجام کار متخصصان، از برنامه‌نویسی برای انجام کار کودکان سخت‌تر است؟ (مینسکی، ۱۹۸۸ : ۷۲).

این موضوع سوالات بیشتری برایم بوجود می‌آورد: چه نوع دانشی با عقل سلیم مترادف است؟ تخمین زده می‌شود که عمده‌ی عقل سلیم ما حول سطوح خبره قرار می‌گیرد، نه کارهایی مانند صحبت کردن به زبان مادری و برخورد با موقعیت‌های روزمره (مرو، ۱۹۹۰ : ۱۲۲). مسلما اگر فردی نویسنده باشد، زبان مادری نیز در حیطه‌ی استادی قرار می‌گیرد و می‌تواند جزو سطوح بالاتر محسوب شود، ولی نه برای همگی ما. ما تجربه‌ی محدودی از AI در زمینه‌ی عقل سلیم داریم. تلاش‌های برای طراحی عقل سلیم مصنوعی صورت گرفته‌اند؛ یکی از مشهورترین آن‌ها، «پایگاه دانش CYC» است. CYC شاید بزرگ‌ترین پایگاه دانش تا به کنون است که با مشارکت هزاران نفر به وجود آمده است. این پایگاه شامل «میلیون‌ها حقایق، باورها و دیگر اجزای دانش است» (فین‌باوم، ۱۹۹۲ : ۱۷). ساخت آن در سال ۱۹۸۶ شروع شد و طولانی‌ترین پروژه‌ی درحال اجرای سیستم مبتنی بر دانش به شمار می‌رود که هم اکنون نیز فعالیت دارد؛ اما هنوز هم عقل سلیم را به وجود نیاورده است. در حقیقت، تلاش‌ها برای مدل‌سازی عقل سلیم تا به امروز بیهوده بوده‌اند، و هیچ پیشرفتی که بتوانم آن را نوید بخش بدانم نشان ندادند. به عقیده‌ی من، طراحی عقل سلیم مصنوعی امکان پذیر نیست. ما نمی‌توانیم آن را از یک نفر بگیریم، زیرا این کارزمان بسیار طولانی می‌طلبد. همچنین به دلیل ماهیت فردی، بین‌فردی، حسی و مفهومی عقل سلیم، گرفتن همزمان آن از چندین نفر نیز ناممکن است. شاید زمان آن فرا رسیده است که پرسش متفاوتی مطرح کنیم: آیا می‌توانیم هوشی را پایه‌گذاری کنیم (انسانی یا مصنوعی) که نیازی به عقل سلیم نداشته باشد؟

۵-۴ همه‌ی اینها چه معنی می‌دهد؟

عقل سلیم حکم می‌کند که از عقل سلیم استفاده کنیم. اینکه بخواهیم عقل سلیم را درک کنیم، کاری به دور از عقل سلیم است. همین می‌تواند کافی باشد تا نشان دهد AI تا چه حد می‌تواند پیشرفت کند: یادگیری عقل سلیم برای او غیرممکن است، درحالی که ظاهرا دانش تماما وابسته به همین عقل سلیم می‌باشد. به این چند نکته در مورد یادگیری دقت کنید:

- محدود کردن یادگیری به یادگیری تقویتی بدین معناست که تنها به دنبال بخش جزئی از یادگیری باشیم.
- ما برخلاف ماشین‌ها در حیطه‌هایی که استعداد داریم راحت‌تر یاد می‌گیریم و کوشش در آن‌ها برایمان آسان‌تر است. و اگر از کسی الهام بگیریم، یادگیری ما بهتر هم می‌شود (بهتر در جاهای مختلف معانی مختلفی می‌دهد، مانند: عمیق‌تر، راحت‌تر یا سریع‌تر). درست است مطالعاتی موجودند که نشان می‌دهند الهام‌بخش بودن یا نبودن معلم تفاوتی ایجاد نمی‌کند؛ اما این مطالعات آزمون‌ها بازتاب مستقیم یادگیری در نظر گرفته شده‌اند.
- یادگیری یک فرایند جمع‌شونده و ساده نیست. لحظاتی از یادگیری تحولی وجود دارند که مشابه لحظات یورکا[55] در زمینه‌ی خلاقیت هستند. در این لحظات به یک درک ناگهانی دست می‌یابیم، نوعی بینش که دانسته‌هایمان را بازآرایی می‌کند و حتی رویکرد و ارزش‌هایمان را هم می‌تواند تغییر دهد.

من شخصا به رابطه‌ی استاد و شاگرد باور دارم؛ این رابطه یکی از بهترین گنجینه‌های فکری است که بشر دارد. تنها راه آموزش است که رسیدن به بالاترین درجه (درجه‌ی استادی) را ممکن می‌سازد. این شکل از یادگیری احتمالا دوباره محبوبیت کسب خواهد کرد، شاید تا حدودی دقیقا به‌خاطر همین AI.

۵ خلاقیت

خلاقیت همان هوش است که خود را سرگرم می‌کند. (ناشناس)

در این فصل عملکرد هوش مصنوعی را در قلمروهای انسانی که در آن‌ها خلاقیت وجود دارد بررسی می‌کنیم. تاثیر احتمالی AI در کمک به خلاقیت بشر نیز مورد بحث واقع می‌شود؛ موردی را برای AI می‌سازیم تا بتواند الگوها را شناسایی کند ولی معنی‌دار بودن آن‌ها را تشخصی ندهد. گرچه خود من می‌گویم AI نمی‌تواند به خلاقیت درست به همان معنی که در انسان بکار می‌رود دست پیدا کند، اما مجال را برای احتمال ساخت خلاقیت مصنوعی با مفهوم‌سازی متفاوت (تا به امروز نامشخص) باز می‌گذارم. به نظر من هرچند هم که تعجب برانگیز باشد، AI می‌تواند از روش‌های غیر منتظره‌ای به خلاقیت انسان کمک کند: ما را یاری می‌دهد تا «خارج از چارچوب فکر کنیم».

۵-۱ عملکرد AI

دومین نکته‌ای که هابیس در سخنان خود طی رویداد زنده‌ی نیو ساینتیست ۲۰۱۷ (آن را در بخش ۲-۴ ذکر کرده بودیم) خاطر نشان شد، این بود که برنامه‌ی AlphaGo مربوط به دیپ‌مایند و گوگل توانست شهود و خلاقیت از خود به نمایش بگذارد و لی سدول- قهرمان جهانی ۱۸ دوره از مسابقات گو- را در طی ۵ بازی شکست بدهد. بررسی عملکرد در زمینه‌هایی که انسان در آن خلاقیت به خرج می‌دهد می‌تواند کمک بسیاری به ما بکند، تا انتظارات‌مان از AI چه اکنون و چه در آینده معلوم شوند. اما نخست اجازه بدهید دوباره با

[55]- Eureka moments

گذشته‌های دور نگاه بیندازیم و به دنبال نشانه‌های اولیه بگردیم. تاریخچه‌ی هوش مصنوعی را می‌توان یک کمدی اغراق دانست. یکی از این موارد اغراق در سال ۱۹۵۷ بود که سایمون پیش‌بینی کرد:

۱. طی ۱۰ سال آینده کامپیوترهای دیجیتال در شطرنج قهرمان جهان خواهند شد، مگر آن‌که قوانین شرکت کامپیوترها را در مسابقه منع نمایند.
۲. طی ۱۰ سال آینده یک کامپیوتر دیجیتال خواهد توانست که یک قضیه‌ی ریاضی را کشف و اثبات کند.
۳. در ۱۰ سال آینده یک کامپیوتر دیجیتال موسیقی‌ای خواهد نوشت که منتقدان می‌پذیرند ارزش زیباشناختی بالایی دارد.
۴. در ۱۰ سال آینده بیشتر نظریه‌های روان‌شناسی در قالب برنامه‌های کامپیوتری یا عبارت‌های کیفی ظاهر می‌شوند که مشخصات همین برنامه‌ها را توصیف می‌کنند.

(سایمون و نول، ۱۹۵۸ : ۸-۷)

حتی یکی از این پیش‌بینی‌ها نیست در آن بازه‌ی زمانی درست از آب درنیامد؛ و فقط اولین مورد به وقوع پیوست، آن‌هم پس از چهل سال نه یک دهه. جالب است بدانیم که سایمون از ذهنی استثنائی برخوردار بود و به ندرت پیش می‌آمد در ارزیابی و پیش‌بینی‌های خود اغراق کند. با نگاه کردن به تنها مورد از پیش‌بینی‌هایی که بالاخره به وقوع پیوست و عملکردهای خارق‌العاده‌ی مشابه آن، می‌تواند باعث شود بهتر درک کنیم که چرا این پیش‌بینی‌ها ناکام ماندند و در آینده چه انتظاراتی می‌توان داشت.

سال ۱۹۹۷، چهل سال بعد از پیش‌بینی سایمون، یک برنامه‌ی کامپیوتری به نام دیپ بلو[۵۶] طراحی شده بود که طبق هدف خود، گری کاسپاروف- که شاید بهترین شطرنج‌باز (انسان) تمام اعصار باشد- را شکست داد. در صحت این بُرد می‌توان تردید روا داشت چرا که دیپ بلو بین بازی‌ها تغییر داده می‌شد و بنابراین غیرممکن بود که سبک یکسانی از بازی را به نمایش بگذارد. با این حال باز هم می‌توانیم برد او را بپذیریم. دومین عملکرد مشابه همان بود که در ابتدای این فصل به آن اشاره شد؛ که در سال ۲۰۱۶ AlphaGo، لی سدول قهرمان جهانی گو و پس از او استادهای بزرگ بسیاری را در این بازی شکست داد. گرچه دیپ بلو برای بازی شطرنج برنامه‌ریزی شده بود، AlphaGo بازی Go را بر پایه‌ی یادگیری حدود ۱۰۰،۰۰۰ نمونه (بازی‌هایی که از اینترنت دانلود شده بودند) و نیز چندمیلیارد بازی که مقابل خود انجام داده بود، یاد گرفت. بنابراین AlphaGo دیتابیس عظیمی از حرکات را جمع‌آوری کرده بود و می‌توانست فرکانس آماری حرکات خاصی را که در یک موقعیت خاص منجر به برد یا باخت می‌شدند، تقلید کند. اینکه AlphaGo بهتری بازیکنان گو را توانست شکست دهد دستاورد شگفت‌انگیزی است؛ اما نه برای خود ماشین، بلکه برای سازندگان آن. AlphaGo نه خلاق بود و نه شهود داشت. اما خالقان AlphaGo هر دوی این خصوصیات را داشتند. AlphaGo تنها در جای خود به فرکانس آماری حرکات بدست آمده از صدها میلیارد بازی نگاه می‌کرد و حرکاتی را انجام می‌داد

[56]- deep blue

که توسط عادت‌هایی که انسان‌ها در این بازی دارند محدود نشده بود. در اصل او به جای بهره‌گیری از شهود، محاسبه می‌کرد و به جای خلاقیت، از احتمال استفاده می‌کرد.

حالا اجازه بدهید این عملکردهای حیرت‌انگیز را با هم مرور کنیم (برای جزئیات بیشتر رجوع کنید به بوری، ۲۰۱۹). هر دو مورد از دیدگاه بازیکنان انسان شاهکار به شمار آمدند و از تمجید زیادی برخوردار شدند. طراحان دیپ بلو بعدها گفتند که در سیستم دیپ بلو یک بدلیل وجود باگ، اشکالی به وجود آمده بود و دیپ‌بلو نمی‌توانست یکی از گام‌های ارزیابی شده را انتخاب کند؛ در عوض او گام‌ها را به طور تصادفی برمی‌گزید (فینلی، ۲۰۱۲). همانند استدلال اتاق چینی، این کاسپاروف بود که به این گام‌ها معنی می‌بخشید و آن‌ها را در قالب استراتژی بلندمدت درمی‌آورد. AlphaGo حرکت‌هایی انجام می‌داد که با همه‌ی قوانین مطابقت داشتند اما هرگز انسان‌های استاد در بازی گو آن را انجام نمی‌دادند. برای محدود کردن شمار عظیم حرکات ممکن در گو، قانون نانوشته‌ای میان انسان‌ها شکل گرفت که در موقعیت‌های خاصی، خطوط ۳ یا ۴ را بازی کنند و هیچ‌کس خط ۵ را بازی نمی‌کرد. با این‌حال AlphaGo در «دوره‌ی آموزشی» خود حدود ۳۰۰ میلیارد بازی مقابل خود انجام داد (بسیار بیشتر از بازی‌هایی که استادان بزرگ در طول زندگی انجام می‌دهند) و دیتابیسی از حرکات ایجاد کرد، و در آن موقعیت‌های خاص حرکت غیرمنتظره، احتمال برد بیشتری را به ارمغان می‌آورد. با نگاه اول پی می‌بریم که در این دو مورد تاریخی هیچ خلاقیتی -دست‌کم به همان معنای انسانی- در AI وجود نداشت.

به گمانم باتوجه به آن‌چه تاکنون مطرح کرده‌ام، گفتن این به اندازه‌ی کافی قانع‌کننده باشد که هوش مصنوعی (به معنای انسانی) فکر نمی‌کند، یادنمی‌گیرد و خلاقیت ندارد. با این حال این لزوما بدان معنی نیست که AI نمی‌تواند این کارها را به روش‌های متفاوت یا تا حدودی مشابه با انسان انجام دهد. هوش مصنوعی لازم نیست یک کار خاص را مانند انسان‌ها انجام دهد؛ و دیدیم که توانست ما را در زمینه‌هایی مانند بازی Go شکست دهد. با این وجود تمام این نمونه‌های موفق در حیطه‌های نسبتا کوچکی انجام گرفته‌اند؛ صرف‌نظر از تعدادی استثنا، حیطه‌هایی که تنها با یک مساله روبه‌رو هستند. این استثنائات عبارت‌اند از: دیپ‌مایند که بازی‌های کامپیوتری مختلف آتاری (نوعی PC اولیه در دهه ۱۹۸۰) را بازی می‌کند، و آلفازیرو (AlphaZero) که گو، شطرنج و شوگی را بازی می‌کند. اگرچه پشتیبانی از چندین حیطه توسط یک سیستم گام بزرگی رو به جلو محسوب می‌شود، اما انجام چندین بازی مختلف هنوز از رسیدن به حل طیف کاملی از مسائل انسان مانند رهبری یک سازمان، نوشتن شعر، پرستاری از بیمار یا لذت بردن از بازی فوتبال (مانند یک طرفدار) فاصله‌ی بسیاری دارد. به‌طور خلاصه، شواهد قانع‌کننده‌ای وجود دارند که بازتولید تفکر، یادگیری و نیز خلاقیت به طور مصنوعی، در حالت انسانی آن دست‌کم در حال حاضر غیر ممکن است. بخش بعدی به تعریف خلاقیت در رابطه با عملکرد AI می‌پردازد تا بررسی شود آیا AI می‌تواند به‌معانی دیگر خلاق باشد یا خیر.

۲-۵ ایده‌های کاربردی و نوین

ادعاهای بسیاری وجود دارد که می‌گویند هوش مصنوعی بالاخره بهترین سطح از عملکرد خلاقانه انسانی را خواهد ساخت. برای مثال مینسکی در سال ۱۹۸۲ پیشنهاد داد که AI در ۱۰۰ سال آینده و شاید حتی ۵۰ سال بعد، متن‌هایی در سطح ادبیات شکسپیر می‌نویسد (آمابیل، ۲۰۲). عده‌ای باور دارند که خلاقیت AI قرار

نیست در آینده رخ بدهد چرا که قبلا به وقوع پیوسته است: «من باور دارم که کامپیوترهای زمان کنونی هر کاری که یک فرد می‌تواند انجام دهد، انجام خواهد داد. به نظر من کامپیوترها می‌توانند بخوانند، فکر کنند، خلاقیت به خرج دهند و ...» (سایمون، ۱۹۷۷ : ۶).

در مورد دانش و یادگیری نشان دادم که AI تنها با قسمت کوچکی از هر کدام از آن‌ها (در معانی انسانی‌شان) سروکار دارد؛ به عبارتی فقط با دانش صریح و یادگیری تقویتی. در مورد خلاقیت چطور؟ تعریف سنتی خلاقیت عبارت است از ساخت ایده‌های جدید و کاربردی (آمابیل، ۱۹۸۳ب و ۱۹۹۶). برای این دو صفت جایگزین‌ها دیگر با تفاوت جزئی هم می‌توان بکار برد: مبتکرانه، بدیع، نوین و (یا) تعجب برانگیز به جای بدیع؛ مفید، باارزش، بالقوه باارزش، مقتضی و یا صحیح به جای کاربردی (آمابیل، ۱۹۸۳ب). اما AI تا بحال چیز جدید و کاربردی به وجود آورده است؟ البته که این کار را کرده است. اگر بخواهیم دقیق‌تر نگاه کنیم، مثال‌های عملکردی خارق‌العاده در بخش قبل شامل این جواب نمی‌شوند، زیرا مثلا در دیپ بلو، حرکات تصادفی اجرا می‌شدند و در AlphaGo نیز حرکات برای استادان بازی گو جدید بود، نه برای خود AlphaGo؛ زیرا اگر آن حرکت احتمال (فرکانس آماری) برد بالایی داشت، حتما قبل از آن به دفعات بازی شده بود. به‌علاوه خود من به طرز شهودی شک دارم این دستاوردها را خلاقانه بخوانم؛ نحوه‌ی بدست آمدن آن‌ها تا حدودی واضح و بر طبق الگوریتم است، ولی خلاقیت به شهود نیاز دارد که در زمینه‌ی دانش ضمنی است: «چیزی به اسم روش منطقی برای ساخت ایده و یا بازسازی منطقی از آن وجود ندارد. شاید نظر خودم را بتوانم این‌گونه بیان کنم که هر دستاوردی شامل یک «عنصر غیرعقلانی» و یا یک «شهود خلاقانه»- بنابر تعریف برگسون- بوده است» (پاپر، ۱۹۶۸ : ۸). البته اگر قضیه به همین سادگی می‌بود، نیازی به توضیح اضافه نداشتیم؛ اما ملاحظات بسیاری وجود دارند. اگر بازگردیم و نگاهی به دو صفت ذکر شده در تعریف خلاقیت بیندازیم، متوجه می‌شویم AI چیزهایی بوجود آورده است که نوین و کاربردی محسوب می‌شوند. این نکته علاوه بر AI، برای خطای تجهیزات آزمایشگاهی، ماشین‌های خراب و صد البته اشتباهات انسانی هم صادق است. مثلا مایکروویو، پست-ایت[۵۷] (برند کاغذهای یادداشت رنگی چسب‌دار)، پنی‌سیلین یا دستگاه اشعه ایکس نمونه‌هایی از اختراعات و یافته‌های مهم هستند که تصادفی به دست آمدند. ما این تصادفات و تجهیزات دارای خطا را خلاق نمی‌نامیم. منظورم این است که در تعریف خلاقیت یک الزام سومی نیز پنهان شده است: خلاقیت باید یک ایده هم باشد. مایکروویو، پست‌ایت، پنی‌سیلین و دستگاه اشعه ایکس ایده نیستند؛ آن‌ها صرفا زمانی به ایده تبدیل شدند که یکی آن‌ها را دید و پی برد که چگونه روش‌های فکری قدیمی را پاک کند و اندیشه‌های جدید و کاربردی را جایگزین آن‌ها نماید. به عبارت دیگر، آن‌ها یک «چیز» جدید و کاربردی را به یک ایده‌ی جدید و کاربردی تبدیل کردند. هرچه را که ما در انسان‌ها خلاقانه می‌خوانیم، یا نخست یک ایده بوده و سپس صورت خارجی به خود گرفته است، و یا اینکه الگو (یا چیزی) در واقعیت بوده که بعدا به یک ایده مبدل شده است. باز هم تاکید می‌کنم که ما فقط می‌خواهیم در خلاقیت انسان یک نقطه‌ی مرجع پیدا کنیم. آیا امکان دارد که نوع متفاوتی از خلاقیت برای AI داشته باشیم؟

[57]- Post-it

اگر تعریف خلاقیت را به «ایده، راه حل و دیگر خروجی‌ها» بسط بدهیم (آمابیل ۲۰۲۰ : ۳)، کمک بهتری به ما می‌کند. در این تعریف الزام سوم حذف شده است. آیا حرکت مشهور AlphaGo زمانی که آن را برای اولین بار انجام داد، جدید به حساب می‌آید؟ شاید؛ ولی در این امر بیشتر تصادف دخیل بود تا خلاقیت. حرکت دیپ‌بلو قطعا جدید بود اما باز هم به نظر من درست در نمی‌آید؛ این حرکت یک خطا بود. خطا هم مانند تصادف برای من در رده‌ی خلاقیت قرار نمی‌گیرند.

آیا به میان آوردن ایده‌ی «داوران مناسبی» که تشخیص دهند پیامد یک AI خلاقانه است یا خیر و اگر هست تا چه حد، می‌تواند مفید واقع شود؟ در مورد خلاقیت انسان‌ها، این کار به مدت چندین دهه به خوبی جواب داده است (آمابیل، ۱۹۸۲) و مبنای تکنیک سنجش وفاقی (CAT)[۵۸] می‌باشد. این تکنیک یکی از گسترده‌ترین روش‌های سنجش خلاقیت است (بائر، ۲۰۲۰). با این‌حال این من را قانع نمی‌کند. عدم کفایت آزمون تورینگ (همانطور که مثلا با استدلال اتاق چینی نشان داده شد) به ما می‌گوید که موضوع چیز دیگری است. سایمون و پس از او راسل، برهان جدیدی را که لاجیک تئوریست ارائه داده بود حتی از برهان وایت‌هد و راسل هوشمندانه‌تر تلقی کردند؛ در نتیجه باید خلاقانه محسوب شود. جدا از جزئیات کار لاجیک تئوریست که نیازمند دانش فنی است، می‌توان به وضوح گفت فرایندی که طی می‌کند آن شفاف و کاملا آشکار و حتی الگوریتمیک است؛ برخلاف آنچه که در فرایند خلاقیت به وقوع می‌پیوندد. همچنین نباید فراموش کنیم که حرکاتی که دیپ‌بلو و آلفاگو انجام می‌دادند توسط افراد خبره، خلاقانه تشخیص داده شدند.

قصد ندارم مهر خلاقیت را فقط بر روی کارهایی بزنم که انسان‌ها انجام می‌دهند. با این وجود سعی می‌کنم کارهایی که واقعا خلاقانه نیستند را خلاقانه تلقی نکنم. در مورد خلاقیت دو جنبه در نظر من وجود دارد که نمی‌توانم نسبت به آن‌ها سخت‌گیر نباشم: خلاقیت باید یک ایده و فرایند آن غیر الگوریتمی باشد. البته این بدان معنی نیست که هوش مصنوعی در زمینه‌هایی که خلاقیت انسان را می‌طلبند، نمی‌تواند با ابزاری غیر از خلاقیت عملکرد بهتری داشته باشد. در این صورت ما می‌توانیم از آن نفع ببریم؛ برخی از جنبه‌های این کار را در بخش بعدی بررسی خواهم کرد.

۳-۵ خلاقیت با AI

در دو بخش قبلی چند نمونه‌ی تحسین برانگیز از موفقیت‌های AI نشان داده شدند. اگرچه با خلاقانه دانستن این موارد مخالفت کردم، اما قبول دارم که این عملکردها دستاورد فوق‌العاده‌ای به شمار می‌روند و نمونه‌های مذکور بدون شک بینشی عمیق از کاربردی بودن AI در خلاقیت ما انسان‌ها به ارمغان می‌آورند. شاید عجیب نباشد که به جای جایگزین کردن افراد خلاق با هوش مصنوعی، استفاده از هوش مصنوعی برای کمک به افراد خلاق را پیشنهاد بدهم. شاید دیدگاهم نوع‌دوستانه به نظر بیاید، ولی در حقیقت هدف من مدیریت بهتر است. آنچه که باید به آن برسیم، همان است که کاسپاروف به خوبی در این روزها انجام می‌دهد: او با کمک AI شطرنج بازی می‌کند. دیپ‌بلو در مقایسه با استاک‌فیش[۵۹] و دیگر ماشین‌های شطرنج معاصر، شباهت بیشتری با یک ماشین‌حساب جیبی دارد. کاسپاروف شانس مقابله با آن را ندارد، ولی آن‌طور که من فهمیده‌ام او با

[58]- Consensual Assessment Technique
[59]- stockfish

کمک گرفتن از AI به طرز استثنائی مقابل ماشین‌های شطرنج موفق عمل می‌کند. می‌توان امید داشت که چنین عملکردی را در زمینه‌های مختلف زندگی و کسب و کار بدست آورد.

فحوای داستان این است که انسان‌ها و هوش مصنوعی ذاتا در زمینه‌های مختلفی مهارت دارند. AI در آنالیز سریع حجم عظیمی از داده‌ها و شناسایی الگوها در مجموعه‌های انبوهی از اطلاعات واقعا خوب عمل می‌کند؛ کارهایی که نیاز به دقت دارند. برای همین داون‌پورت (۲۰۱۸ : ۴۴) می‌گوید که AI «آنالیزورهایی هستند که دوپینگ کرده‌اند». این تعریف یکی از توصیف‌های موردعلاقه‌ی من در مورد هوش مصنوعی است. انسان‌ها در این جور کارها خیلی خوب نیستند، و اگر هم بتوانند آن‌ها را انجام دهند، زمان بسیار زیادی طول می‌کشد. جای تعجب است که تقریبا تمام تحصیلات ما بر همین زیرمجموعه‌ی کوچک از ظرفیت ذهنی انسان متمرکز شده است؛ و بیشتر ما در آن خوب عمل نمی‌کنیم. البته فهمیدن علت آن خیلی دشوار نیست؛ و می‌تواند نکته‌ی آموزنده‌ای باشد. علت آن این است که ما سعی می‌کنیم به دانسته‌هایمان عینیت ببخشیم؛ می‌توانم حتی بگویم که دلمان میخواهد ذهنیت را کامل از دانش حذف کنیم. و ظاهرا AI در مسائل عینی از ما بهتر است.

AI در تشخیص اینکه کدام الگو در یک موقعیت مد نظر، معنی دار و یا بالقوه معنی دار است؛ اما انسان‌ها در این زمینه عالی عمل می‌کنند. برای مثال یک آنکولوژیت و یک شیمی‌دان در حال کار با چند ماده‌ی شیمیایی هستند تا یک درمان موثر برای سرطان تولید کنند. AI ممکن است ترکیبات بی‌شماری از این مواد شیمیایی را ارائه دهد، اما AI جدا از بررسی چند معیار از پیش تعیین شده، نمی‌تواند تشخیص بدهد کدام ماده‌ی شیمیایی حاصل می‌تواند مطلوب واقع شود. با این وجود آنکولوژیست و شیمی‌دان با نگاه کردن به مشخصات مواد شیمیایی که AI ساخته است، می‌توانند بفهمند که برخی خصوصیات جدید (چه هدفمند و چه تصادفی) آن ماده ممکن است برای درمان سرطان مفید باشد. به علاوه شاید متوجه نکاتی بشوند که مورد هدف آن‌ها نبود؛ مثلا این‌که آن خصوصیات بگویند یکی از این مواد شیمیایی حاصل برای درمان آنفلوانزا موثر است. علت اینکه AI نمی‌تواند این‌ها را متوجه بشود این است که این راه حل در یک قلمروی متفاوت از قلمروی مساله‌ی مورد نظر قرار می‌گیرد. دانش کامل این دو فرد خبره برای تشخیص معنی دار بودن الگوی بدست آمده لازم است. به‌طرز مشابهی، طراحان مد، مبلمان ماشین و ... امروزه برای ساخت مدل‌های جدید از AI کمک می‌گیرند. با وجود اینکه AI می‌تواند مدل‌های جدید ایجاد کند (از طریق بازآرایی مدل‌های فعلی)، این طراح است که تصمیم می‌گیرد کدام یک از آن‌ها، زیبا، مطبوع و یا زننده است. از این رو AI می‌تواند یک پردازش اولیه در مورد شناسایی و ساماندهی الگوها انجام دهد، و با این کار ارزش بسیاری به فرایند خلاقانه‌ی انسان می‌افزاید (مثال تاخوردگی پروتئین را در بخش بعد ببینید). پس ظاهرا آنچه که انسان‌ها در آن خوب هستند بیشتر در حیطه‌ی ارزش‌ها و داوری‌ها قرار دارد و نه اطلاعات و تحلیل‌ها؛ بیشتر شهود است تا محاسبات. داوری، به ویژه داوری ارزش‌ها همواره شخصی است و بنابراین ذهنی و ارادی محسوب می‌شود (همانطور که در بخش ۲-۳ دیده شد).

اگر به موضوع نوین و کاربردی بودن یک ایده در تعریف خلاقیت برگردیم، می‌توانیم بگوییم که در مورد کمک AI به یک فرد خلاق، AI یک چیز جدید و کاربردی می‌سازد و فرد خلاق ایده را به وجود می‌آورد. به این

حالت «تقدم AI[60]» می‌گویند و ایده در مرحله‌ی بعد قرار می‌گیرد. البته عکس آن نیز ممکن است؛ در حالت «تقدم ایده[61]»، فرد خلاق به مفهوم آنچه که می‌خواهد بسازد دست می‌یابد، ولی مراحل زیادی درپیش هستند تا این مفهوم صورت مادی به خود بگیرد. در هر دو حالت نقش‌ها یکی هستند.

AI به یک شیوه‌ی دیگر- که شاید ساده‌تر باشد- هم به خلاقیت انسان ارزش می‌بخشد. داشتن سطح بالایی از خلاقیت انسانی ارتباط نزدیکی با رسیدن به درجه‌ی استادی در حیطه‌ی (رشته یا مساله) موردنظر دارد. رسیدن به چنین درجه‌ی استادی هم در سطح حیطه‌ی مورد نظر و هم در سطح فردی زمان می‌برد. یعنی رشته یا مساله‌ی مربوطه ساز و کار دانش خود را شکل می‌دهد و فرد مبتدی این ساز و کار را طی برخورد با آن حیطه فرا می‌گیرد و به «پس زمینه‌ی اندوخته‌ی»[62] آن‌ها تبدیل می‌شود (ویتگنشتاین، ۱۹۶۹ : ۹۴). این کار به برقراری ارتباط درون حیطه کمک می‌کند و با مواردی مانند راهکارهای مشترک منجر به تشکیل یک اجتماع می‌شود؛ در عین حال مسئول ایجاد تفکر «درون چارچوب است» و محدودیت‌های غیرضروری و قراردادی را به وجود می‌آورد. به همین دلیل است که پیروزی آلفاگو جانی دوباره به جهان بازی گو بخشید؛ استادان انسان در این بازی شروع به امتحان حرکت‌ها و استراتژی‌های جدید کردند و از رویکردهای مدارس سنتی گامی فراتر نهادند. به علاوه شاید عجیب به نظر برسد اما در بهترین معنی، AI به ما کمک می‌کند تفکری «خارج از چارچوب» داشته باشیم.

۴-۵ AI معتبر

تا به اینجا ادعاهای گزافی که در مورد آینده وجود دارند را ذکر نموده‌ام و دستاوردهای خارق‌العاده‌ی امروز را هم بررسی کردیم، همچنین محدودیت‌های AI و ذهن انسان را با مقایسه نشان دادیم. با این وجود هیچ کدام از اینها پاسخ پرسش ما نیستند: چه انتظاری می‌توان از AI داشت. برای این پیش‌بینی‌ها غیر از نظرات شخصی از روی چیز دیگری نمی‌توانیم قضاوت کنیم. آیا ممکن است که تفکر، یادگیری، داشتن احساس و هیجان در قالب دستکاری‌های داده‌ای دربیاید؟ جواب من خیر است اما این تنها یک باور شخصی است. بسیاری در پاسخ خواهند گفت که بله ممکن است. در هر حالت چه راهکاری باید نسبت به AI در پیش گرفت؟ به نظر من خیلی مهم نیست که جواب این «سوال حیاتی» بله یا خیر باشد، آنچه که مهم است درک قابلیت‌های AI و گسترش بیشتر آن‌ها می‌باشد. به هر حال ما دائما باید جهت پیشرفت‌های آینده را تعیین کنیم. و البته تصمیم دشوارتر این است که چه میزان سرمایه‌ای در پیشرفت AI قرار می‌دهیم؛ هم از جهت پول و هم زمان، انرژی، زحمت، دانش، اشتیاق و ... (رجوع کنید به آکرمان و ادن، ۲۰۱۱). اگر AI راه حل ما برای مشکلات فقر، تغییرات آب و هوایی، توسعه‌ی پایدار، پاندمی‌ها، حوادث طبیعی، صلح جهانی و این گونه موارد باشد، در آن صورت باید هرچه در توان داریم در میان بگذاریم. اما من طور دیگری به قضیه نگاه می‌کنم. اکنون که این جملات را می‌نویسم ۱۰ ماه از پاندمی کووید-۱۹ گذشته است؛ واکسن‌ها به تازگی عرضه شده‌اند و شایعاتی مبنی بر سویه‌ی جدیدی از این ویروس وجود دارد. اما ظاهرا هنوز AI هیچ نقشی در حل مشکل پاندمی ایفا نکرده است. البته نمی‌توان تقصیر را بر گردن AI انداخت؛ بلکه مقصر آن‌هایی هستند که در مورد AI و

60- AI-first mode
61- idea-first mode
62- inherited background

چیستی آن اغراق به خرج می‌دهند نمای نادرستی از آن ارائه می‌کنند. ما نتوانستیم داده‌هایی با کیفیت خوب تولید کنیم تا AI با آن‌ها کار کند. کاستی از ما انسان‌هاست، نه هوش مصنوعی. برای پاسخی کامل‌تر به سوال موردنظر باید تا فصل بعد منتظر بمانید که مسائل اخلاقی هوش مصنوعی را بررسی کنیم. اما مطرح کردن این سوال برای بررسی عنوان این بخش لازم است: مساله‌ی AI معتبر[۶۳].

به منظور این که این مساله را دریابید، خوب است قابلیت‌های انسان در مقابل AI را مجددا با استفاده از جدیدترین مثال آلفافولد مرور کنیم (هِوِن، ۲۰۲۰). آلفا فولد یک شاخه از دیپ‌مایند است که ساختار یک پروتئین را بر اساس توالی آمینواسیدی آن پیش بینی می‌کند، و ساختار پروتئین مشخصات آن پروتئین را تعیین می‌نماید. نتیجه‌ی آن شگفت انگیز بود؛ پیش‌بینی‌های آلفافولد برای چندین پروتئین در حد اتم شکل می‌گرفت. این دستاورد با تکیه بر ۵۰ سال تحقیق بر روی پروتئین توسط صدها دانشمند، با کمک تیم نخبگان زیست‌شناسی، فیزیک و علوم کامپیوتر دیپ‌مایند، و فراهم کردن حدودا ۱۷۰،۰۰۰ پروتئین برای آموزش ANN حاصل شد. این دستاورد از جهات مختلف شبیه پروژه‌ی DENDRAL فین‌باوم بود (بخش ۲-۲ را ببینید) از این داستان می‌توانیم نتیجه بگیریم که بین انسان و AI تقسیم وظایف وجود دارد؛ آلفافولد محاسبات عددی فراوان را انجام می‌دهد و پژوهشگران آزمایش می‌کنند و توالی‌های آمینواسیدی را که انتظار دارند مطلوب واقع شوند ارائه می‌نمایند. همچنین خوب است بدانیم که گروه نخبگان متغیرهای لازم برای این کار را از تحقیقات ۵۰ سال گذشته بدست آوردند. اگر مطالعات پروتئین در نمونه‌های قبل به صورت آزمایش‌ها و محاسبات دستی انجام می‌شد؛ صدها برابر زمان طول می‌کشد تا نتایج آلفافولد بدست بیاید. با این وجود، این بدان معنی نیست که راحت به دیپ‌مایند بگوییم برود و ساختارهای پروتئینی را پیدا کند. گرچه محققان می‌توانستند بدون AI و با کمک خلاقیت خود نحوه‌ی رسیدن به ساختارها را بیابند، ولی خود AI کار خلاقانه‌ای انجام نمی‌داد. ظاهرا اگر سیستم مناسب بر پا شود، ساختارها را می‌توان با محاسبات تصادفی هم بدست آورد. البته قابل تحسین برانگیز است که گروه نخبگان دیپ‌مایند از محاسبات بی هدف استفاده نکردند؛ بلکه آن‌ها از محاسبات هوشمند بهره جستند و خلاقانه کار می‌کردند تا مواد لازم برای آلفا فولد فراهم شوند. این نکته را می‌شود آموخت که اگر یک کاری به شکل خلاقانه انجام‌پذیر باشد، بدون خلاقیت هم می‌توان آن را انجام داد؛ می‌شود با محاسبات بی هدف هم به آن رسید. آیا تمام کارهایی که انسان‌ها انجام می‌دهند را با محاسبات تصادفی هم می‌توان انجام داد؟ گمان نمی‌کنم، و چنانچه حق با من باشد، بسیار مهم است که بدانیم کدام کارها را می‌توان از خلاقیت به محاسبات تبدیل کرد.

در اینجا مساله‌ی دیگری در زمینه‌ی خلاقیت مربوط به AI پیش می‌آید، که اکنون برای پاسخ به آن آماده هستیم. من تا همین لحظه تمام حیطه‌های هنری و خلاقانه‌ی صنایع AI را نادیده گرفته بودم. با خواندن متونی که AI نوشته‌اند و نگاه کردن به دستورالعمل‌های AI و سپس بحث در مورد آن با مارک استیراند، چیزی که نشان‌دهنده‌ی خلاقیت باشد دستگیرم نشد. من تخصص کافی برای داوری در زمینه‌ی موسیقی و نقاشی ندارم، ولی درک من از نخبگان این است که عملکرد AI در این زمینه‌ها بهتر داوری می‌شود. مارگارت بودِن (۲۰۰۹ : ۲۴۴-۲۴۵) در مورد برنامه‌ای می‌گوید که در دهه ۱۹۸۰ توسط آهنگساز دیوید کوپ طراحی

[63]- authentic AI

شد: «این برنامه می‌توانست کار آهنگسازهای مشهور مختلف را به طرز قانع‌کننده‌ای تقلید کند». همچنین او یک برنامه‌ی طراحی معماری را مثال می‌آورد که خانه‌هایی را در سبک معماری فرانک لوید رایت می‌ساخت که پیش از آن دیده نشده بودند. بودِن با تفاوت قائل شدن میان خلاقیت ترکیبی، اکتشافی و تحولی و نیز با مطالعه‌ی دقیق برنامه‌های سازنده‌ی اثرات هنری (بودن، ۱۹۹۸)، مشاهده کرد که AI در ساخت چیزهای جدید و کاربردی توانا است اما در تشخیص و داوری آن‌ها به مشکل بر می‌خورد؛ که به همان چیزی برمی‌گردد که من می‌گفتم، البته از یک دیدگاه متفاوت.

با توجه به اینها، بگذاریم اکنون به آزمون تورینگ برگردیم. در بخش ۲-۴ گفتیم که آزمون تورینگ برپایه‌ی «بازی تقلید» ساخته شده است. و باور دارم که این همان دلیل اصلی در عدم توانایی ساخت یک AI معتبر است. بحث ما ساخت هوش مصنوعی‌هایی است که بتوانند انجام دهند، خلق کنند و راه حل ارائه بدهند. بحث ما در مورد همکاری انسان و AI است. ما داریم یک شخصیت را به AI نسبت می‌دهیم. AI (همانند یک برنامه) اجرا می‌شود، ولی خلاقیت ندارد. برنامه تولید می‌کند، اما راه حل ارائه نمی‌دهد. با این انسان‌هایی که خلاق هستند و مشکلات را حل می‌کنند می‌توانند از AI بهره بگیرند. چیزی به نام همکاری انسان و هوش مصنوعی وجود ندارد، زیرا همکاری باید با قصدمندی طرفین همراه باشد. AI چیزی مانند یک خودر و یا پیچ گوشتی که از آن مانند هر ابزار یا محصول دیگری استفاده می‌کنیم. این را نمی‌گویم که نشان دهم با AI مخالف هستم؛ بلکه از آن جهت می‌گویم که من حامی AI می‌باشم. ما باید دست از بازی تقلید برداریم و AI را از روی اینکه چقدر می‌تواند ادای انسان را دربیاورد قضاوت نکنیم. AI معتبر مانند انسان نیست. AI معتبر مثل یک AI است. تنها زمانی می‌توانیم به پتانسیل حداکثری و فوق‌العاده‌ی AI برسیم که این نکته را دریابیم.

۵-۵ همه‌ی اینها چه معنی می‌دهد؟

یک کاریکاتور از شخصیت کارتونی گارفیلد وجود دارد که آن را بسیار دوست دارم. جان[۶۴] یک غذای «شبیه به تن‌ماهی» را جلوی گارفیلد می‌گذارد و گارفیلد هم با طعنه به او می‌گوید: «می‌دانی جان که چه غذای دیگری هم مزه‌ی تن‌ماهی می‌دهد؟ خود تن‌ماهی!» من نیز در مورد AI معتبر همین حس را دارم. با چنین نگاه به AI، نکات زیر را جمع‌بندی می‌کنیم:

- ما امروزه چیزی نداریم که خلاقیت ماشین را به همان معنای انسانی اثبات کند. آنچه که در اختیار داریم، شواهدی دل بر عملکرد حیرت انگیز AI است، که در اصل سازندگان آن را باید تمجید کرد. مابقی همه ادعاست و شبیه این جمله می‌باشد که «هوش مصنوعی تا الان به خوبی شطرنج و گو بازی کرده است، پس به زودی می‌تواند شعر بنویسد».
- در جشنواره‌ی جهانی علم (WSF) در نیویورک در سال ۲۰۱۹، یان لکون گفت که سخت‌ترین کار تعیین هدف درست برای هوش مصنوعی است. مشکل AI این نیست که از آنچه به آن بگویی تبعیت نکند، بلکه این است که دقیق همان کاری را می‌کند که به آن دستور می‌دهی. اما این کار اغلب

[۶۴]- یکی از شخصیت‌های همین کارتون

چیزی که فرض می‌کردی نیست و ما اغلب فرضیات بسیاری می‌سازیم بدون آنکه متوجه شویم. همین ناکامی‌های مربوط به AI را توضیح می‌دهد.

- در حالی که شواهدی مبنی بر خلاقیت AI وجود ندارد، مدارک فراوانی هستند که بیان می‌کنند AI می‌تواند الگوهای کاربردی را شناسایی و (یا) تولید کند تا افراد خلاق از آن استفاده نمایند. آنچه که AI در انجام آن ناتوان است، تشخیص معنی دار بودن یا ارزش الگوهای یافته یا تولید شده است. برای همین است که ما باید سعی کنیم افراد را تشویق کنیم که در خلاقیت به خرج دادن از AI کمک بگیرند.

به نظر من خلاقیت باید در یک ایده باشد. جدا از اینکه یک چیز چقدر جدید یا کاربردی تلقی شود، تا زمانی که ایده نباشد از خلاقیت ریشه نمی‌گیرد. ولی نیازی نیست هرکس و هرچیزی خلاق باشد؛ بسیاری از امور ایده، تشخیص ارزش، شهود یا ملاحظات اجتماعی و احساسی لازم ندارند. هوش مصنوعی دیر یا زود در این زمینه‌ها می‌تواند جای ما را بگیرد.

۶ مسائل اخلاقی

همه‌ی حیوانات با هم برابرند، اما برخی برابرترند. (جورج اورول)

از پاییز ۲۰۱۷ و به مدت به مدت دو سال و نیم، ۲۰ سخنرانی درباره‌ی AI و ذهن انسان اقامه کردم. بسیاری از این سخنرانی ها در قالب کنفرانس بودند. در این کنفرانس‌ها محققان و نیز فروشندگان AI سخنرانی می‌کردند و در ادامه نیز تماشاگران سوالات خود را می‌پرسیدند. هر بار، رویه‌ی کار به همین شکل بود: پژوهشگر در مورد مطالعات خود صحبت می‌کرد، فروشنده ویژگی‌های محصول خود برمی‌شمرد و تماشاگران نیز بدون استثناء، در مورد اخلاقیات می‌پرسیدند. آیا داده‌های ما به شیوه‌ی اخلاقی استفاده می‌شوند؟ آیا در امان خواهیم بود؟ چیزی هست که از AI جلوگیری کند تا آسیبی به ما نرساند؟ آیا AI شغل من را تصاحب خواهد کرد؟ آیا AI مشکلات تبعیضی را تشدید می‌کند؟ یا به عبارت کلی: «قبول! AI شما می‌تواند این کارها را انجام دهد، ولی آیا ما می‌توانیم به آن اعتماد کنیم؟»

۶-۱ اعتماد به هوش مصنوعی؟

در این کنفرانس‌ها و همچنین در کمیته‌ها و آزمایشگاه‌های مختلف، اغلب کار من به بحث کردن با نخبگان (هم آکادمیک و هم عملی) در مورد «اعتماد به AI» می‌کشد. در بسیاری از مواقع، سوال را اینگونه فرض می‌کنند: «چگونه اعتماد خود را به AI بیشتر کنیم؟» این پرسش نادرست است و من به دفعات این را به آن‌ها تذکر داده‌ام. اعتماد باید کسب شود. آنچه که شما باید انجام دهید، ساخت یک AI شایسته‌ی اعتماد است. اگر بخواهیم منصفانه حرف بزنیم، بیشتر این افراد از ته دل می‌خواهند کار درست را انجام دهند. اما اگر یک نفر از آن‌ها مثلا در مورد اتومبیل خودران سوال می‌کرد که این خودرو بین کشته شدن مسافران درون خود و یا زیرگرفتن کودکانی که در خیابان توپ‌بازی می‌کنند کدام را انتخاب می‌کند. آن موقع پاسخ آن‌ها اینگونه

می‌بود «شما خودتان بگویید تا مطابق آن برنامه‌نویسی کنیم». آن‌ها نمی‌دانند که مردم به دنبال این نیستند که AI کدام پیامد را انتخاب خواهد کرد. مردم از این بابت که ماشین‌ها چنین وظایفی را بر عهده بگیرند نگرانند؛ میخواهند بدانند که آیا AI آن‌ها را به کشتن می‌دهد یا خیر. این مثال همچنین بحث مسئولیت‌پذیری را پیش می‌کشد: مردم نمی‌خواهند که چنین مسئولیتی را به ماشین‌ها بسپارند. مسئولیت‌پذیری بحثی است که مدیرعاملان نیز به راحتی آن را درک می‌کنند. افرادی که این سوالات را می‌پرسیدند معمولا اطلاعات زیادی در مورد AI، ذهن انسانی یا اخلاقیات نداشتند. افرادی که سعی می‌کردند به این پرسش‌ها پاسخ دهند به جز مواردی اندک، همگی دانش فنی بالایی در مورد AI داشتند، اما در مورد ذهن انسان و اخلاقیات در حد پرسش‌کنندگان آگاه بودند. همین موضوع باعث شکل گیری گفت‌گوهای بیهوده‌ای می‌شد. ما برای اندیشیدن در مورد سوالات اخلاقی هوش مصنوعی به یک فیلسوف نیازمندیم. در اینجا سعی می‌کنم مساله را بگشایم و کار را با بحث درباره‌ی اعتماد آغاز می‌کنم و سپس به سراغ مسائل اخلاقی عمومی حول مساله‌ی هوش مصنوعی خواهم رفت.

در میان انسان‌ها، دو نوع اعتماد وجود دارد: اعتماد به شایستگی و اعتماد به فرد. ما نوع اول را می‌توانیم نسبت به ماشین‌ها هم تعمیم دهیم. اغلب ما به لپ‌تاپمان اعتماد می‌کنیم که فایل‌های ما را ذخیره می‌کنیم و به جز در استثنائات بسیار بسیار کمی، این اعتماد ما به جاست. وقتی که کارش را به درستی انجام نمی‌دهند، به این «ماشین احمق» دشنام می‌دهیم و به دنبال فایل پشتیبان می‌گردیم. به طور کلی این برای ما جا افتاده است؛ به هرحال بسیاری از افراد نسبت به عدم تطابق داستان‌ها کنونی با امیدهایی که به AI وجود دارد و نیز نسبت به تجربه‌ی کار با آن اگاهی دارند. «اشتباهات» AI در همه‌ی موارد به راحتی قابل حل هستند، پس چرا از قبل حل نشده‌اند؟ یک نمونه‌ی اخیر درباره‌ی الگوریتم توزیع واکسن کووید ۱۹ از استنفورد بود که باعث حذف شدن پزشکان خط مقدم شد(گوئو و هائو، ۲۰۲۰). البته این اشتباه به راحتی قابل حل بود؛ اما اگر امکان رخ دادن چنین اشتباهات پیش پا افتاده‌ای وجود دارد، چگونه می‌توان به تکنولوژی هوشمند اعتماد کرد؟ به علاوه شفافیت که اصولا در هر ابزار آنالیزی وجود دارد، در AI صادق نیست. ما نمی‌توانیم بفهمیم که AI در هر گام چه کاری انجام می‌دهد. اما می‌توانیم به آن اعتماد نکنیم و کاری را که انجام می‌دهد مخالف عقل سلیم بدانیم. ولی آیا به‌راستی این کار رضایت بخش است؟

نوع دوم اعتماد، اطمینان خاطر از این است که فرد کار درست را انجام خواهد داد (یا دست‌کم تلاش می‌کند). البته کار درست به دیدگاه اخلاقی و سیستم ارزش‌گذاری ما بستگی دارد. ایده‌ی اعتماد به فرد تا حد زیادی شبیه به رویکرد اخلاقیات فضیلت‌گرایانه[۶۵] است؛ در این رویکرد ما به افرادی که آن‌ها را بافضیلت می‌دانیم اعتماد داریم. اگر ما طرفدار اخلاقیات پیامدگرایانه[۶۶] باشیم، تنها نتیجه‌ی کار را مد نظر قرار خواهیم داد. پیروان وظیفه‌گرایی[۶۷] (به آن اخلاق مبتنی بر قاعده هم می‌گویند) هم باور دارند که عمل، و در نتیجه نیتی

[65]- virtue ethics
[66]- consequentialist ethics
[67]- deontology

که منجر به عمل شده است مهم هستند نه پیامد آن‌ها؛ زیرا پیامدها به جز عمل ما به بسیاری از موارد دیگر وابسته هستند. اگر بخواهیم تا زمان دیدن پیامدها صبر کنیم، معمولا دیر می‌شود؛ پس منطقی است که به دنبال نیت باشیم. این نیز باز برای AI خالی از مشکل نیست، چرا که AI نیت ندارد. نمی‌تواند هم داشته باشد، زیرا نیت‌ها ریشه در بعد ضمنی دارند: «اکنون فهمیدن را به دانستنِ نیت، معنا و عملی که انجام می‌دهیم تعمیم داده‌ام. برای این منظور باید اضافه کنم که هیچ چیزی که گفته، نوشته و یا چاپ می‌شود نمی‌تواند درون خود معنایی داشته باشد؛ زیرا فقط فردی که آن را بر زبان می‌آورد، می‌شنود و یا می‌خواند می‌تواند با این کار معنایی را برساند» (پولانی، ۱۹۵۹ : ۲۲).

شاید بتوانیم از AI انتظار داشته باشیم که با یادگیری از تصمیماتمان، یک بازنمایی از سیستم ارزش‌گذاری ما ایجاد کند. با این وجود، گرفتن ارزش‌ها از درون دیتابیسی از تصمیمات که توسط افراد مختلف صورت گرفته‌اند، با دو محدودیت قابل توجه مواجه است: ۱- AI فقط اطلاعاتی که در دیتابیس هستند را در نظر می‌گیرد و امکان دارد متغیرهایی باشند که در دیتابیس ثبت نشده‌اند؛ و ۲- ممکن است ارزش‌هایی در دیتابیس باشند که ما نیازی به رونویسی آن‌ها نداشته باشیم (برای دیدن نمونه رجوع کنید به هائو، ۲۰۲۰ سیستم سیاست‌گذاری پیشگویانه‌ای که نرخ فساد را بالا می‌برد). علاوه بر این‌ها، این رویکرد بسیار شبیه پروژه‌ی CYC است که عقل سلیم را مدل‌سازی می‌کرد (بخش ۴-۴). این پروژه انرژی زیادی صرف کرد ولی بی‌نتیجه ماند.

اضافه بر این، ما ظاهرا نمی‌توانیم سیستم ارزش‌گذاری را همزمان از بیش از یک نفر بگیریم، زیرا هر فرد شاخص‌های اخلاقی مختص به خود را دارد؛ به عبارتی سیستم ارزش‌گذاری هر فرد منحصر به فرد است. در این‌صورت باید چه‌کسی را انتخاب کنیم تا یگانه تصمیم‌گیرنده‌ی نهایی باشد؟ فردی که ارزیابی او از ارزش‌ها بر سرنوشت همه تاثیر می‌گذارد؟ اگر واقعا بتوانیم چنین فردی را پیدا کنیم که قضاوت اخلاقی او مورد پذیرش همه‌ی ما باشد، آنگاه اگر کار ما در حد تشخیص ساختار یک پروتئین باشد، صدها هزار تصمیم لازم داریم که باید گرفته شوند؛ اگر هم بیشتر شبیه به بازی گو باشد، به چیزی حدود چند صد میلیارد تصمیم نیاز خواهیم داشت. از این بابت هم خبرهای بدی داریم: تصمیمات اخلاقی بسیار پیچیده‌تر و چندجانبه‌تر از بازی گو هستند و طیف وسیع‌تری را هم پوشش می‌دهند. در نتیجه نمی‌توانیم تعداد نمونه‌ی کافی برای آموزش ANNمان فراهم کنیم.

محققان و فروشندگان AI معمولا ادعا می‌کنند که استفاده از AI در زمینه‌های ارزش‌گذاری بسیار مفید واقع می‌شود: AI سوگیری ندارد، همواره و بدون استثناء به سیستم ارزش‌گذاری پایبند است، خستگی نمی‌پذیرد، هیچ‌گاه دچار حال نامساعد نمی‌شود، و ... . این ادعا، مثال بسیار خوبی از بازنمایی غلط است. در حقیقت، سوگیری تنها زمانی معنی پیدا می‌کند که ما جواب درست را بدانیم، و سپس در این جواب انحراف ایجاد کنیم. اگر یک چیز آنقدر امر مسلم باشد که جواب درست آن را بدانیم، دیگر به AI نیازی نداریم و صرفا یک پردازشگر خودکار می‌خواهیم؛ زیرا دیگر تصمیمی برای گرفتن باقی نمی‌ماند. و مساله‌ی اخلاقی نیز وجود ندارد که بخواهیم مد نظر قرار دهیم. در واقع مسائل اخلاقی زمانی ظاهر می‌شوند که تشخیص درستی یک

چیز، انتخاب بین بد و بدتر، نکات مثبت و منفی گزینه‌های موجود و ... بر ما چندان آشکار نباشد؛ به عبارت دیگر، در موقعیت‌هایی که پیوسته در روزمرگی‌مان با آن برخورد می‌کنیم و تا حدودی یادگرفته‌ایم چگونه مسیر خود را پیدا نماییم.

۲-۶ اشتباهات و عواقب آن‌ها

یک روبات جراحی را در ذهن‌تان تجسم کنید که طراحی فوق‌العاده‌ای دارد: دقت عمل آن نسبت به دست یک انسان بسیار بیشتر شده است؛ این ربات می‌تواند تصاویری هزاران برابر دقیق‌تر از قدرت دید انسان ثبت کند؛ و بخیه‌ها را کاملا صاف می‌زند. البته این ربات فقط می‌تواند عمل‌های ساده را انجام دهد، زیرا طی یک عمل بسیار پیچیده باید همزمان چیزهای زیادی را مورد نظر قرار داد و از آن‌ها سر درآورد؛ اغلب حتی مواردی که با توجه وضعیت، انتظار نمی‌رفت دیده شوند. پس این ربات فقط موارد ساده را عمل می‌کند و عمل‌های بسیار دشوار را انجام نمی‌دهد. جراحان برجسته همچنان سربلندند، زیرا هنوز برای عمل‌های سخت به آن‌ها نیاز هست. با این حال نسل بعدی جراحان در حال آموزش هستند و این افراد هرگز نمی‌توانند تمرین عملی داشته باشند، زیرا عمل‌های ساده توسط روبات‌های جراح انجام می‌شود. آیا جراح‌های تازه‌کار انسانی باید کار خود را با جراحی مغز شروع کنند؟ البته که خیر؛ هیچ‌کس با این کار موافق نخواهد بود. با این وجود این دقیقا همان اتفاقی‌ست که ممکن است رخ بدهد. در «حادثه‌ی پرواز ۴۴۷ ایر فرانس» (اولیور و همکاران، ۲۰۱۷الف، ۲۰۱۷ب)، خلبان خودکار به دلیل شرایط غیرعادی کنترل پرواز را به خلبانان انسان داد. طبیعتا یک خلبان عادی باید بتواند این شرایط را مدیریت کند. اما خلبانان آن پرواز که همیشه به سیستم خلبان خودکار متکی بودند، نتوانستند از پس این کار بربیایند. تمام ۲۲۸ مسافر و خدمه‌ی آن پرواز جان خود را از دست دادند. طبیعتا توصیف من از وقایع رخ داده بیش از حد ساده است اما اصل آن نادرست نیست. و به راحتی می‌توان مسائل مشابه را در بسیاری از زمینه دید حتی اگر عواقب آن- دست‌کم در ظاهر امر- تا این حد وخیم نباشند.

درست این است علیه استدلال‌هایی که مثلا از داده‌های آماری تصادفات جاده‌ای ناشی از خطای انسانی استفاده می‌کنند تا ادعای خود را مبنی بر امنیت اتومبیل‌های خودران تثبیت کنند، احتیاط به خرج دهیم. این استدلال‌ها می‌گویند اتومبیل خودران همچنان تصادفات کمتری از رانندگی توسط انسان‌ها دارند و با این‌کار سعی می‌کنند ترس از ناموفق عمل‌کردن AI را از بین ببرند. مشکل این‌جاست که داده‌ها فقط آمار تصادفی که به دلیل خطای انسانی پیش می‌آیند را بررسی می‌کنند. شواهد آماری که به ما بگویند چند تصادف به لطف مهارت و تصمیم‌گیری‌های انسانی متوقف شده‌اند وجود ندارد، زیرا این تصادفات اصلا رخ نداده‌اند. همچنین نباید فراموش کنیم که اگر AI را وارد خیابان‌ها کنیم، باید خود را از میان رانندگان بی‌نظم انسانی حرکت دهد؛ سیستم حمل و نقل یک شبه کامل جایگزین نمی‌شود و به انحصار اتومبیل‌ها خودران در نمی‌آید.

ما عادت داریم که دقت بالا در عمل را (مثلا دقت در تصاویر ورودی) همیشه خوب تلقی کنیم. با این حال، هِوِن (۲۰۱۹ : ۱۶۴) فهرستی از مثال‌های متعدد «فریب AI» را ایجاد نموده است. با تغییر تصادفی یک (یا شماری معدود) پیکسل می‌توان به‌طرز عجیبی AI را فریب داد. ما تغییر در این تعداد پیکسل را با چشم مسلح

حتی متوجه هم نمی‌شویم. مساله این نیست که AI از قوه‌ی اشتباه کردن گه‌گاه برخوردار است؛ انسان‌ها هم اشتباه می‌کنند. مشکل اینجاست که ما اصلا نمی‌دانیم چرا با تغییر دادن چند پیکسل، تصویری که یک نرم‌افزار آن را «شیر وحشی» شناسایی کرده بود، ناگهان توسط همان نرم‌افزار به یک «کتاب‌خانه» تبدیل می‌شود؛ در حالی که برای ما همان تصویر است. از یک سو ما نمی‌توانیم این اشتباهات را پیش‌بینی کنیم و همین موضوع می‌تواند بسیار جدی باشد. مثالی از این موضوع، زمانی بود که یک چت‌بات پزشکی متعلق به نابلا که از OpenAI جی‌پی‌تی-۳ استفاده می‌کرد، به یک بیمار غیرواقعی پیشنهاد خودکشی داد (داوس، ۲۰۲۰). از سوی دیگر، هکرهای ماهر شاید از این اشتباهات برای حمله کردن به AI استفاده کنند و در آن اختلال ایجاد نمایند؛ اگر بفهمند تغییر در کدام مجموعه از پیکسل‌ها باعث بروز اشتباه می‌شود، شاید حتی بتوانند برنامه‌ی آن را کامل دوباره‌نویسی کنند. در این مورد مثال واقعی نمی‌شناسم، ولی این موضوع به راحتی می‌تواند به یک «بازی جنگ» تبدیل شود. کافی است تصور کنید یک هکاتون راه بیندازید و به چند هکر با استعداد بگویید که این یک بازی کامپیوتری شبیه‌سازی شده است.

با این وجود، مساله‌ی هشدار برانگیز اصلی همان تشخیص نادرست است. هون سخنی از جورگن اشمیدهوبر پژوهشگر موسسه‌ی دال مال از بخش تحقیقاتی هوش مصنوعی را به میان می‌آورد: «او می‌گوید شناخت الگو بسیار حائز اهمیت است؛ به آن اندازه که باعث شده است علی‌بابا، تنسنت، آمازون، فیس‌بوک و گوگل به باارزش‌ترین شرکت‌ها تبدیل شوند» (هون، ۲۰۱۹ : ۱۶۵).

ولی خیر، این تشخیص الگو نبود که چنین ارزش فوق‌العاده‌ای را به این شرکت‌ها بخشید؛ بلکه ایده‌های بی‌نظیر آن‌ها بود؛ ایده‌های بی‌نظیر و زیاد؛ و رهبری بی‌سابقه. من حتی خواهم گفت دست‌کم برخی از این شرکت‌ها علی‌رغم تشخیص الگوی بسیار بدی که دارند، باز هم باارزش محسوب می‌شوند؛ چیزی که من معمولا به آن‌ها «دستگاه‌های هوشمند احمق» می‌گویم. برای نمونه، پس از آنکه ویرایش کودکانه‌ی هری‌پاتر و سنگ جادو را با جلد کاغذی از آمازون خریداری کردم، این سایت باز هم سه نوع ویرایش دیگر از همان کتاب را به من پیشنهاد می‌کرد. آمازون نمی‌توانست تشخیص بدهد که این کتاب شبیه به هم نیستند، بلکه در اصل یک کتاب‌اند که جلد متفاوتی دارند. یکی از دوستان در سن ۶۲ سالگی برای اولین بار به پرو سفر کرد. طی ۵ سال بعدی، او هر روز پیشنهاد سفر به لیما[۶۸] را دریافت می‌کرد. او مسافرت رفتن را دوست داشت، و مانند بسیاری از جهانگردان پرشوق، از دیدن مکان‌های جدید لذت می‌برد. این شیوه از مسافرت بسیار گسترده و میان خیلی از مردم رایج است؛ با این حال توصیه‌های مسافرتی خلاف آن عمل می‌کنند.

این مثال‌ها همگی اشتباهاتی را نشان می‌دادند که یافتن آن‌ها کار سختی نیست(پس چرا این همه سال درست نشده‌اند؟)؛ اما یک چیز میان این اشتباهات مشترک است: همه‌ی این‌ها به موضوع ارزش‌گذاری بر می‌گردند، یعنی مستقیما با مسائل اخلاقی ارتباط دارند. از بسیاری جوانب، ارزش‌ها و عقل سلیم (بخش ۴-۴) از این

[۶۸]- پایتخت کشور پرو

حیث که ما نمی‌توانیم سرچشمه‌شان را پیدا کنیم شبیه همدیگرند؛ این‌ها الگوهای به شدت پیچیده‌ای می‌سازند و به دست آوردن آن‌ها توسط AI غیر ممکن می‌نماید. به علاوه ما نه تنها از ارزش‌هایی که مشاهده می‌کنیم یاد می‌گیریم و از میان ارزش‌های متضاد معلمان‌مان، والدین و ... برخی را انتخاب می‌کنیم، بلکه ما خود نیز ارزش‌ها را می‌سازیم. «لازم است نشان دهم مثلا هنگامی که منشا، ارزش‌های جدیدی به وجود می‌آورد، آن‌ها را به طور ضمنی و از طریق دلایل غیرمستقیم القا می‌کند: ما نمی‌توانیم یک سری ارزش‌های جدید را صریحا انتخاب کنیم، بلکه باید با ساخت و بکار بردن آن‌ها، این ارزش‌ها را بپذیریم» (پولانی، ۱۹۶۶ب : xix).

در اغلب مواقع گفتن اینکه چه زمانی یک ارزش خاص ایجاد شده است ممکن نیست. ما غالبا فقط می‌دانیم که این ارزش‌ها از پیش وجود داشته‌اند؛ و مدت زمانی هست که آن‌ها را دنبال می‌کنیم. یا اینکه فرد دیگر این ارزش‌ها را به ما می‌شناساند. ارزش‌های ما مانند معانی که داریم (بخش ۴-۳) ممکن است «لوپ‌های عجیبی» را هم شامل شوند (هوفشتاتر، ۱۹۷۹). با این وجود سیستم پیچیده و نامنظم ارزش‌ها، هر بار که بر سر یک دوراهی اخلاقی یا هر دوراهی دیگر با جنبه‌ی اخلاقی قرار می‌گیریم، به سمت قطب‌های درست و غلط جهت می‌گیرند. اما AI برخلاف این سیستم «مدلی ندارد که بتواند موارد مهم را به خوبی برگزیند» (هون، ۲۰۱۹ : ۱۶۴). تشخیص «اهم موارد» اصلی‌ترین کاری است که AI از انجام آن عاجز است. در اینجاست که عقل سلیم و حکمت[۶۹] با هم شباهت پیدا می‌کنند. وایزن‌بام سازنده‌ی ELIZA (اولین هوش مصنوعی که در آزمون تورینگ قبول شد: بخش ۴-۲ و ۴-۳ را ببینید)، نخستین کتابی را که مساله‌ی اخلاقیات در AI را مستقیما مطرح کرد نوشت (وایزن‌بام، ۱۹۷۶). او به جای دقت در دستاوردهای احتمالی پژوهش‌های AI، در مورد حیطه‌هایی صحبت کرد که AI فارغ از قابلیتی که دارد نباید در آن‌ها استفاده شود. صحبت‌های او بر پایه‌ی اخلاق بود، نه زمینه‌های فنی. منظور اصلی او این بود که AI جایی در قضاوت‌های اخلاقی ندارد، زیرا این حیطه‌ها نیازمند حکمت و همدلی[۷۰] هستند.

۳-۶ اخلاقیات و حقوق

رابطه‌ی پیچیده‌ای میان اخلاق و مقررات هنجار شده وجود دارد. مقررات اخلاقی (یا نَرم) بر کاری دلالت دارند که با توجه به یک قاعده اخلاقی خاص، انجام آن درست است و عواقب خطا کردن در آن، احساس بد شخصی‌مان خواهد بود. در مقابل مقررات هنجار شده، مانند مقررات قانونی، استانداردها و رویه‌های سازمانی و ... . مسیرهای قانونی برای انجام کارها توصیه می‌کنند و در صورت انحراف از این مسیرها، جریمه‌هایی برای آن تعیین کرده‌اند. علم اخلاقیات و قانون شامل رشته‌هایی هستند که به ترتیب مسائل اخلاقی و سیستم‌های قانونی را مطالعه می‌کنند. بسیاری از مقررات قانونی در مقررات اخلاقی ریشه دارند؛ ولی همین که آن‌ها را در قالب قانون تدوین می‌کنیم، دیگر بحث اخلاقیات در میان نیست. برای مثال، کشتن یک نفر خلاف قانون

[69]- wisdom
[70]- empathy

است؛ بنابراین مهم نیست که از لحاظ اخلاقی درست باشد یا غلط. حتی امکان دارد دیدگاه اخلاقی و حقوقی در یک چیز با هم متضاد باشند. برای نمونه دزدیدن یک تکه نان توسط یک فرد به عنوان آخرین تلاش برای سیر نگه‌داشتن شکم کودکانش، شاید همزمان هم غیرقانونی و هم اخلاقا درست باشد. هدفم از بیان این نکات مربوط به مقررات اخلاقی و قانونی این است که نشان دهم AI می‌تواند مقررات قانونی را همانطور که ساختارهای پروتئینی را دسته‌بندی می‌کند (مثال آلفا فولد را در بخش ۴-۵ ببینید) مدیریت نماید اما برای مقررات اخلاقی نمی‌تواند این کار بر بکند.

طبیعتا اگر مقررات قانونی به همان سادگی بودند که در تعریف فوق دیده می‌شود، تنها AI برایمان کافی بود و نیازی به وکیل‌های بسیار گران قیمت نداشتیم. چرا وکیل‌ها، خصوصا وکیل‌های ماهر، پول زیادی می‌گیرند؟ زیرا آن‌ها می‌توانند مفاهیم ضمنی در قوانین را دریابند و تعابیر معناداری از آن‌ها ارائه کنند، و صد البته چون که می‌توانند با سیستم قانون دست و پنجه نرم کنند. استفاده از AI در حرفه‌ی حقوقی بسیار هیجان‌انگیز است. زیرا یک مرز واضح و منطقی میان تدوین (یا امتیازدهی) قانون و تفسیر آن وجود دارد. به عبارت دیگر، در اینجا هم مانند ابعاد دانش، یادگیری و خلاقیت، مرز میان دامنه‌ی فعالیت AI و ذهن انسان مشخص است. شاید به همین دلیل باشد که رابطه‌ی استاد و شاگرد در شرکت‌های حقوقی بیش از سایر سازمان‌ها (حتی دانشگاه) تداوم یافته است. اهمیت تعبیر مفاهیم در حقوق و متعاقبا در اخلاق، بیش از اندازه است. در حقوق عرفی[۷۱]، یک پرونده بر پایه‌ی پرونده‌های پیشین قضاوت می‌شود. ولی سوال چالش‌برانگیز این است که کدام یک از پرونده‌های قبلی با این پرونده مطابقت دارد و چگونه این پرونده‌ها را پیدا کرد. AI در یافتن پرونده‌های بالقوه مشابه عالی عمل می‌کند؛ ولی تعیین اینکه از کدام پرونده‌ی متقدم باید استفاده شود توسط انسان قضاوت می‌گردد. در حقوق مدوّن[۷۲] هم قضیه مشابه ولی کمی پیچیده‌تر است. قوانین در حقوق مدون معمولا واضح هستند ولی وسعت زیاد این قوانین باعث روا داشتن تعبیر و قضاوت بیشتری می‌شود.

علاوه بر نکاتی که پیش‌تر ذکر شدند، یک منطق آماری هم برای این موضوع وجود دارد: «نمونه» اغلب معرف «جمعیت» نیست. این می‌تواند منجر شود موارد غیرعادی را درنظر نگیریم(دورفلر و استیراند، ۲۰۱۹)، و در صورتی که هدفمان پیدا کردن موارد غیرمعمول باشد، جستجویمان بیهوده خواهد بود. و در مورد اخلاقیات و حقوق، در نظر AI ممکن است یک نابغه به همان اندازه‌ی یک مجرم از عرف جامعه انحراف داشته باشد.

همین منطق در تشخیص جرم هم مشکل ایجاد می‌کند. ما می‌توانیم مثبت‌های کاذب (افرادی که مجرم تلقی شده‌اند اما بی‌گناه هستند) را مشخص کنیم ولی منفی‌های کاذب را پیدا نخواهیم کرد (مجرمانی که شناسایی نشده‌اند). با همه‌ی این‌ها، بی‌دلیل است بگوییم از این ابزار قدرتمند AI در کمک به اجرای قانون استفاده نخواهیم کرد؛ مادامی که به ماموران پلیس کمک کنیم نه این‌که آن‌ها را جایگزین نماییم. سیاست‌گذاری پیشگویانه‌ی خودکار می‌تواند به یک فاجعه مبدل شود (هائو، ۲۰۲۰) اما به عنوان یک سیستم هشداردهی

[71]- case-law
[72]- codified law

اولیه، می‌تواند به افراد خبره کمک کند تا بدانند در مورد چه چیزی باید قضاوت نمایند. این هم باز به تفسیر قانون بر می‌گردد.

جزئی در همه‌ی سیستم‌های قانونی وجود دارد که بر اهمیت تفسیر اصرار می‌ورزد؛ این بخش همان سیستم هیئت منصفه است. هیئت منصفه قرار نیست عینی باشد، بلکه باید ذهنیت گروهی داشته باشند. به عبارتی دیگر، تفسیر حکم در اختیار این ۱۲ نفر است؛ این افراد موضوع را نه تنها برای خودشان، بلکه باید برای یکدیگر نیز تفسیر نمایند، با یکدیگر بحث و جدل کنند، دیدگاه‌های روایی مختلفی را امتحان کنند و با همدیگر به موضوع بیندیشند. همین رویکرد در زمینه‌های علمی هم به خوبی نتیجه می‌دهد. پاپر توصیه می‌کند این ذهنیت جمعی در نبود عینیت دومین دیدگاه برتر است (پاپر، ۱۹۶۸: ۲۲). تا کنون سعی کردم به شما نشان بدهم هنگام تعیین نقش انسان و AI در دوراهی‌های اخلاقی، مشکلات کجا و چگونه گریبان‌گیرمان خواهد شد و چگونه این نقش‌ها را باید در سیستم قانونی جای داد. اما تاکنون پاسخی مبنی بر چگونگی برخورد با این مشکلات ارائه نداده‌ام؛ زیرا این مشکلات بسیار پراکنده هستند. آنچه که باید بدانیم این است که این مسائل نیاز به توجه فراوانی دارند و عواقب آن‌ها می‌تواند جدی و گسترده باشد. بحث‌های یک طرفه راه به جایی نخواهند برد. با بکار بردن AI پیشگیری و خود تقویتی در مقررات قانونی از لحاظ فنی ممکن خواهد شد. البته این مساله برای آزادی فردی مضر است و سیستم دموکراتیک را زیر سوال می‌برد. اگر بخواهیم این کار را بکنیم، AI به آرزوی دیکتاتورها تبدیل می‌شود. حیطه‌ی دیگری از مشکلات اخلاقی و حقوقی حول هوش مصنوعی مربوط می‌شود به حفاظت از AI در برابر انسان‌ها. شاید اکنون در نظرتان قانع‌کننده نباشد؛ اما کیت دارلینگ (۲۰۱۹) از MIT این موضوع را مطرح می‌کند که بایستی از محصولات بر پایه‌ی هوش مصنوعی مانند روبات، چت‌بات و دستیارهای دیجیتال محافظت نمود. نه به خاطر آنکه ماشین‌ها به این حفاظت نیاز داشته باشند، بلکه چون انسان‌ها نیازمند آن هستند. ویکلوند (۲۰۲۰) یک قدم هم فراتر می‌گذارد، به طوری که می‌گوید: «بهتر است تا زمانی که AI بتواند لذت و رنج را تجربه کند، همان وظایف اخلاقی را که نسبت به حیوان‌ها داریم، نسبت به آن‌ها هم داشته باشیم» (چند رویکرد احتمالی دیگر را هم در کوکلبرگ، ۲۰۲۰ بخوانید).

امروزه به ماشین‌ها زور گفته می‌شود، البته ماشین‌ها عواقبی از این رفتار نمی‌بینند. ولی همین ممکن است باعث ترویج رفتار زورگویی شود، و افرادی که روبات‌ها را آزار می‌دهند و رفتار مخربی با چت‌بات‌ها دارند، ممکن است نسبت به دیگر انسان‌ها هم مخرب شوند. استدلال دیگر این است که اگر زورگویی باعث آسیب رسیدن به ماشین می‌شود، شاید ماشین‌ها به طور هدفمند برای ارضای همین نیاز ساخته شده‌اند، و فرد زورگو رفتار مخرب کمتری نسبت به افراد دیگر نشان خواهد داد. واقعا نمی‌دانیم که کدام یک از این استدلال‌ها صحیح‌اند؛ مثل همیشه می‌توان گفت هر یک تا حدی درست می‌گویند. من خود مایلم طرف دارلینگ را بگیرم. همانطور که حفاظت از سگ‌ها در مقابل رفتار مخرب انسان برای خود انسان‌ها اهمیت بیشتری دارد، حفاظت از AI نیز به همین علت منطقی به نظر می‌رسد. به هر حال، قطعا این یک مساله‌ای است که ارزش اندیشیدن

دارد. بگذارید با مثالی این مساله را تشدید کنم. شاید در مورد ربات‌های جنسی شنیده باشید. امروزه این ربات‌ها بدون هیچ محدودیت آزار داده می‌شوند، چرا که این روبات‌ها فقط تعدادی ماشین هستند. ولی از زمان ارسطو، شواهد نشان داده است که عادت به یک رفتار منفی، آن را توجیه می‌کند.

۶-۴ تکینگی

گمان می‌کنیم تا به اینجا شما را قانع کرده باشم که مسائل اخلاقی یکی از مهم‌ترین انواع مسائلی هستند که امروزه در زمینه‌ی هوش‌های مصنوعی با آنها روبه‌رو هستیم، و چندین دهه‌ی دیگر برای کنار آمدن با آن‌ها زمان لازم است. تمام بحث‌های دیگری که لازم است در مورد دانستن، یادگرفتن و خلاق بودن بدانیم نیز در کنار بسیاری از موارد دیگر، همگی در مسائل اخلاقی گرد آمده‌اند. به علاوه ما در جایگاهی قرار گرفته‌ایم که در آن هر دوراهی اخلاقی مربوط به AI منحصر به فرد است و هیچ جواب کلی وجود ندارد. بایستی سوالات را به طریق بهتری بپرسیم و چند ابزار فکری را برای یافتن پاسخ‌های شخصی کنار بگذاریم. احتمالا طی چند دهه‌ی بعد بیشترین میزان یادگیری در حوزه‌ی هوش مصنوعی، مربوط به مسائل اخلاقی باشد تا پیشرفت‌های فنی (البته نباید اینگونه برداشت شود که پیشرفت در تکنولوژی را باید متوقف کرد).

مساله‌ی نهایی در زمینه‌ی اخلاق، تکینگی[۷۳] نام دارد. این مساله همان است که به سناریوی آخرت بشر توسط هوش مصنوعی منتهی می‌شود. ما با همه‌ی این‌ها به کدام سمت حرکت می‌کنیم؟ آیا همانطور که رِی کورزویل (۲۰۰۵) می‌گوید، تکینگی در کمین‌مان است؟ فکر نکنم. نکته‌ی مهم در این جا مفهوم خودآگاهی است. برخی وعده داده‌اند که در طی ۵ سال آینده می‌توانیم خودآگاهی را به درون کامپیوتر بارگذاری کنیم (قبلا هم این پنج سال وعده داده شده بود). چقدر از این موضوع فاصله داریم؟ از لحاظ تکنولوژی شاید خیلی دور نباشیم، همانطور که توانسته‌ایم یک میکروچیپ را به سیستم عصبی انسان متصل کنیم. مشکل در اینجا از طرف دیگر است: ما تقریبا هنوز نمی‌دانیم که خودآگاهی یعنی چه. این را می‌دانیم که تجربه‌های ما از پدیده‌ها به خودآگاهی مربوط می‌شود، و دیوید چالمرز (۱۹۹۸) به آن مساله‌ی خودآگاهی می‌گوید. قرار است چه چیزی را به درون کامپیوتر بارگذاری نماییم و آن چیست؟

مساله‌ای که در مرکز بحث‌های من قرار می‌گیرد، در بسیاری از حیطه‌های AI دیده شده است. نول و سایمون (بخش ۱-۲) از حل شطرنج و اثبات قضایای ریاضی توسط هوش مصنوعی، به درک نثر انگلیسی رسیدند. ایلان ماسک استدلال (و نمایش) قانع‌کننده‌ای را در مورد نورالینک ارائه کرده است؛ اینکه چگونه نورالینک می‌تواند مثلا به افراد ناتوان کمک کند تا قوه‌ی حرکتی یا شنوایی و یا بینایی خود را بازیابند؛ چرا که چیپ نورالینک می‌تواند مسیرهای عصبی آسیب دیده را جایگزین و یا بای‌پس کند. البته این‌ها توضیحی در این مورد نیستند که چگونه قرار است خاطراتمان را در این چیپ‌ها ذخیره کنیم؛ و اگر بتوانیم باز نخواهیم توانست از خاطرات به خودآگاهی برسیم. تری وینوگراد (۱۹۹۰) از پیشتازان AI این موضوع را به زمان توماس هابز

[73]- singularity

(۱۶۵۱) نسبت می‌دهد. او فکر کردن (استدلال) را محاسبه‌ی نمادها می‌دانست؛ گویی یک زمین‌شناس انتظار داشته باشد با مطالعه و کسب داده‌های (و یا دانش‌های) بیشتر در مورد سنگ‌ها، به درک عمیق از ماهیت درخت و قورباغه و انسان برسد. بسیاری در جهان هوش مصنوعی انتظار دارند که AI می‌تواند از نظر هوش از انسان‌ها پیشی بگیرد؛ زیرا هوش مصنوعی در انجام کارهایش سریع‌تر شده است. با این حال فکر کردن تنها یکی از کارهایی است که انسان می‌تواند انجام دهد، و اتصال نورون‌های بیشتر یا پردازش داده‌های حجیم‌تر و یا افزایش سرعت محاسبات، لزوما و محتملا منجر به پدید آمدن تفکر نخواهد شد؛ چه برسد به خودآگاهی.

تعبیر ماسک (۲۰۱۸) در مورد تکینگی کمی متفاوت است. او از میان هوش مصنوعی‌ها خصوصا دیپ‌مایند را مد نظر قرار می‌دهد، زیرا دیپ مایند به تمام سرورهای گوگل دسترسی دارد و همه‌ی اطلاعات بر روی آن سرورها قرار گرفته اند. هدف دیپ‌مایند این است که مصرف انرژی سرورها را در حد بهینه نگه دارد اما ماسک اینگونه استدلال می‌کند که یک به‌روزرسانی کوچک در دیپ‌مایند، می‌تواند کاری کند که او به تمام داده‌های این سرورها دسترسی پیدا کند؛ تمامی داده‌های ما، و هرکاری که دلش بخواهد با آن‌ها انجام دهد. آیا لازم است از این بابت نگران باشیم، دسترسی به داده‌ها از لحاظ فنی، نباید آن چنان برای یک هکر که می‌تواند به سرورهای گوگل نفوذ کند سخت باشد. ماسک همچنین سناریوی متفاوتی از روز آخرت ارائه می‌دهد: در نسخه‌ی او، AI از ما متنفر نیست، بلکه امکان دارد که ما بر سر راه رسیدن او به هدف‌هایش قرار بگیریم و در نتیجه، ما را نابود کند. همانطور که ما از مورچه‌ها بدمان نمی‌آید، اما اگر بر سر راه ما قرار بگیرند، له می‌شوند. البته باید ذکر کنم ما (انسان‌ها) تمام مورچه‌های روی زمین را از بین نبرده‌ایم. درست است که باعث انقراض چندین گونه جاندار شده‌ایم، اما هرچه که آگاهی بیشتری در این زمینه کسب می‌کنیم، بیشتر در صدد بر می‌آییم که از آن‌ها مراقبت کنیم. پس اگر حق با آن‌هایی باشد که سخن از ابرهوش‌ها می‌رانند، شاید لازم نباشد بترسیم. شاید این کامپیوتر ابر هوش یک دیدگاه اخلاقی برتر نیز دارد. در هر صورت، به عقیده‌ی اسپندر نیازی نیست که نگران تکینگی باشیم:

اگر تکینگی واقعا در حال وقوع باشد، فراتر از توانایی درک ماست. ماشین‌ها ممکن است روزی خودآگاهی کسب کنند- شاید هم قبلا موفق شده باشند- اما احتمال زیادی هست که ما متوجه آن نشویم. اگر هم تکینگی در شرف رخ دادن نباشد، در آن صورت تنها با باورهای کوته‌فکرانه روبه‌رو هستیم. بنابراین کار ما بیشتر عملی خواهد بود؛ ماشین‌ها را آنطور که می‌دانیم به عملکرد وادار کنیم، پروژه‌هایمان را انجام دهیم و بر جهان تاثیر بگذاریم و به دنبال پاسخ «این‌ها برای ما چه معنی می‌دهند؟» باشیم، نه اینکه خود را درگیر کنیم که ما برای آن‌ها چه معنایی داریم (اسپندر، ۲۰۱۵).

این سخنان من را به سناریوی شخصی خودم در مورد آخرت بشر به دستان هوش مصنوعی می‌رساند. از آنجا که من باور ندارم ماشین‌ها بتوانند به معنای انسانی فکر، احساس، هیجان و خودآگاهی داشته باشند، نگران این موضوع نیستم که روزی ماشین‌ها به بیداری برسند و علیه ما قیام کنند. اما یک احتمال واقع‌گرایانه وجود دارد که ما انسان‌ها روزی ماموریتی را بر عهده‌ی AI نه چندان باهوش بگذاریم و این کار به دلیل خطای

ماشنی (اشتباهات تشخیص تصاویر را در بخش ۲-۶ مطالعه کنید) به یک فاجعه منتهی شود. این خطای AI نیست که چنین چیزی را رقم می‌زند، بلکه خطای انسان‌ها خواهد بود که به دلیل اشتیاق بیش از حد کاری را به هوش مصنوعی واگذار می‌کنند که از توانایی آن خارج است.

این مساله لازم است تا در یک مقیاس کوچک‌تر، به طور دقیقی برای همه‌ی موارد استفاده از AI مد نظر قرار بگیرد. باید در مورد وظایفی که بر دوش AI می‌گذاریم محتاط باشیم. مسئولیت AI در قبال افرادی است که درباره‌ی نحوه‌ی استفاده از آن تصمیم می‌گیرند: یعنی مدیران عامل شرکت‌ها.

۵-۶ همه‌ی اینها چه معنی می‌دهد؟

اخلاق و مطالعه‌ی علم اخلاق ممکن است در نظرتان فلسفی جلوه کنند، ولی مسئولیت‌پذیری و وظیفه مفاهیمی هستند که مدیران عامل هر روز با آن‌ها سرو کار دارند. اگر ماشینی فکر بکند، باید مسئولیت این اعمال بر عهده بگیرد؛ چرا که سازندگان این ماشین زیر بار مسئولیت آن نخواهند رفت. چندین نکته در رابطه با اعتماد، اخلاقیت و مسئولیت‌پذیری را در ذیل خواهید دید:

- اصطلاح «ماشین متفکر» برای آن‌هایی که درک درستی از AI ندارند، ترسناک و غیرقابل اعتماد به نظر می‌رسد. ولی باید بدانیم که درک AI امری شدنی است؛ حتی اگر نتوانیم اعمال آن را مرحله به مرحله و با شفافیت کامل ببینیم. آنچه هنوز از توان ما خارج است، درک ذهن انسان‌هاست.
- بسیاری از افراد نمی‌خواهند این را قبول کنند که برای تشخیص بهتر حقایق از تخیلات و تشخیص درستی جملات، به یادگیری آمار نیازمندیم. چرا که لازم است ادعاهای مختلف را بشناسیم و همین‌که فردی با اعداد و ارقام با ما حرف زد گیج نشویم. یک بار دانشجویی از شیوه‌ی نمره دادن من ناراضی بود؛ او اشاره کرد که نصف کلاس من عملکرد متوسط به پایینی داشتند. البته او در حقیقت با تعریف متوسط مخالف بود.
- یکی از مضرترین پیامدهای اعتماد به AI، این است که برخی اوقات اشتباهات ابتدایی از AI سر می‌زند. مثلا تصویر یک شیر را به دلیل تغییر چند پیکسل، با تصویر یک کتابخانه اشتباه می‌گیرد.آنچه که ترسناک می‌باشد این است که تصور اینکه چگونه چنین خطایی می‌تواند رخ دهد غیر ممکن است. چگونه می‌توان به چیزی که اینگونه به خطاهای ابتدایی دچار می‌شود اعتماد کرد و هدایت اتومبیل-های خود ران را به او سپارد؟

زمانی که با برندگان نوبل مصاحبه می‌کردم، به وضوح پی بردم که این افراد به طور خارق‌العاده‌ای می‌توانند از هر دو روش کل نگر و جزء نگر را بهره بگیرند (رجوع کنید به مک‌گیلکرایست، ۲۰۱۹)، من به این مهارت «دیدن جوهره» می‌گویم. گرچه به نظر می‌رسد ما می‌توانیم ارزش‌هایی (جزئیات) که می‌خواهیم را برنامه نویسی کنیم، استفاده از این ارزش ها در موقعیت‌های خاص (تصویر کلی) به شدت امری پیچیده است و به طور خاص منحصر به انسان ها می‌باشد.

۷ سخن آخر

هیچ ذهنی آنقدر خوب ساخته نشده است که فاقد حس شوخ‌طبعی باشد (ساموئل تیلور کولریج).

این کتاب کوچک، به‌جای نتیجه‌گیری با سخن آخر نویسنده به پایان خواهد رسید تا نشان دهیم این داستان هنوز به پایان نرسیده است و باید هر از چندگاهی دوباره بیایم و آن را تصحیح کنیم. هدف من ارائه‌ی پیشنهاداتی مبنی بر داشتن دیدگاه موثر نسبت به هوش مصنوعی بود تا با کمک آن‌ها بتوان داستان آینده‌ی بشریت را بهتر رقم زد. گرچه از بسیاری جزئیات چشم پوشی شد، اما باور دارم که توانسته باشم تصویر کلی از وضع عمومی ارائه بدهم. به عبارتی، این تصویر جامع است و در سطح بالایی نشان داده شد و گرچه همه‌ی جزئیات را در بر نمی‌گیرد اما برای فهمیدن هر جزئیاتی- فعلی یا جدید- کاربردی محسوب می‌شود. امیدوارم نظر بنده به عنوان یک طرفدار شدید هوش مصنوعی، محدودیت‌ها و قابلیت‌های AI را نشان داده باشد. و بتوانم به مدیران شرکت‌ها کمک کنم AI را درک کنند و حقایق را از باورهای شخصی تشخیص دهند و سازمان‌شان بهترین بهره را از AI ببرد.

یک ربع قرن پیش، مدیر ممتاز پیتر دراکر گفته بود:

از سی چهل سال قبل که ابزارهای پردازش داده‌ای نوین پدیدار شدند، افراد شاغل در کسب و کارها اهمیت اطلاعات را در سازمان‌هایشان دست کم و یا بیش از حد مهم گرفته اند. ما- حتی خود من- بیش از حد به امکانات کامپیوترها بها دادیم، تا جایی که ایده‌ی «مدل‌های کسب و کاری» کامپیوتری در ذهنمان شکل گرفت؛ مدل‌هایی که بتوانند تصمیم‌گیری کنند و حتی بخش اعظم کسب و کار را راه بیندازند. اما ما همچنین این ابزارهای نوین را کامل دست کم گرفته‌ایم؛ ما در آن‌ها راهی برای عملکرد بهتر مدیران اجرایی دیدیم تا بتوانند سازمان‌هایشان را مدیریت کنند (دراکر، ۱۹۹۵ : ۵۴).

به عقیده‌ی من ما نیز امروزه در مورد هوش مصنوعی، در شرایط یکسانی قرار گرفته‌ایم. ما هم AI را دست کم گرفته ایم و هم بیش از حد انتظارمان را از آن‌ها بالا برده‌ایم. هر دوی این اشتباهات سبب شده‌اند استفاده‌ی مفید کمتری از AI داشته باشیم. به علاوه تاکید زیادی بر روی بازه‌ی کوتاه مدت داریم. همانطور که هون (۲۰۱۹ : ۱۶۵) از گری مارکوس از دانشگاه نیویورک نقل می‌کند: «یادگیری عمیق آن قدر در کوتاه مدت موثر است که انسان‌ها بلندمدت را به فراموشی سپرده‌اند». با این وجود او باور دارد سیستم‌های مرکب آینده‌ی دراز مدت پیش‌رویمان هستند. به نظر من نیز ترکیب کردن دستاوردهای و آموزه‌های مرتب با AI در بخش‌های متعدد این رشته، منفعت‌های بسیاری برایمان خواهد داشت. اخلاقیات در زمینه‌ی AI نه تنها داغ ترین موضوع است، بلکه به گمان من مطالعه‌ی این موضوع بیشترین اطلاعات را درباره‌ی چند دهه‌ی آینده‌ی هوش مصنوعی در اختیارمان می‌گذارد. ما مقدار بسیاری پرونده‌های اخلاقی داریم و هر روز نیز موارد جدیدی اضافه می‌شوند. و باید با دقت بسیار زیادی بر روی این پرونده‌ها کار کنیم و هر یک را منحصر به فرد بدانیم. همانطور که مشغول پیدا کردن راه حل‌های منحصر به فرد هستیم، یادگیری عمیقی از این پرونده‌ها حاصل می‌شود که

شاید پاسخ جامعی به ما ندهد، ولی به ما در ساخت سریع‌تر و بهتر راه‌حل‌ها کمک می‌کند. درست مانند یک تحلیلگر روانشانسی و یا یک کُوچ اجرایی که برای مشتریان خود راه یکسانی در پیش نمی‌گیرند، ولی هر یک از موارد به بهتر و بهتر شدن آن‌ها کمک می‌کند.

اگرچه این کتاب از حیث فنی تقاضای زیادی از خواننده ندارند، اما بار روانشناسی و فلسفی آن سنگین است. به گمانم چاره‌ای نیست، زیرا یک مدیر اجرایی باید به یک فیلسوف اخلاقی و فیلسوف ذهن تبدیل شود. همانطور که آن کانلیف (۲۰۰۹) می‌گفت: مدیر عامل باید یک «رهبر فیلسوف» باشد.

همانطور که در خلال این کتاب نشان دادیم و بحث کردیم، AI هنوز صرفا متکی بر داده است، حتی در پردازش داده‌های بسیار پیچیده. AI مبتنی بر یافتن و ترکیب کردن الگوهاست. اما این ذهن انسان است که این الگوها را می‌فهمد و معنای (احتمالی) آن‌ها را تشخیص می‌دهد. اگرچه AI در زمینه‌های سرشار از داده- مربوط به دانش صریح و یادگیری تقویتی- بسیار کمک کننده است، اما نمی‌تواند با دانش ضمنی و یا احساسات، هیجانات و باورهایی که بر دانش و یادگیری و اشکال مختلف یادگیری مانند الهام بخشی، استعداد و رابطه‌ی استاد- شاگردی تاثیر گذارند، مواجه شود. AI می‌تواند الگوهای جدیدی بسازد؛ به ویژه با ترکیب کردن الگوهایی که از پیش وجود داشتند. اما نمی‌تواند تشخیص بدهد کدام الگوها خوب، زیبا، خوشایند و یا جذاب هستند. مطابق گفته‌ی داون‌پورت (۲۰۱۸): AI تصمیمات بهتری نمی‌گیرد؛ بلکه باعث می‌شود تصمیمات ما آگاهانه‌تر شوند.

با توجه به موارد فوق، برای من واضح است که AI استراتژی خلق نمی‌کند. اما این که AI را برای چه و چگونه استفاده کنیم یک تصمیم استراتژیک مهم است؛ و برای همین است که هر مدیرعاملی باید اندکی در مورد AI آگاهی داشته باشد. با توجه به قلمروهای شناخته‌ها، ناشناخته‌ها و ناشناختنی‌ها (بخش ۲-۱): برای قلمروی شناخته شده، نیازی به AI نداریم، بلکه فقط به دیتابیس‌هایی با طراحی مناسب و تبادل داده نیازمندیم. در دنیای ناشناخته‌ها، AI می‌تواند با یافتن الگوها در داده و فیلتر کردن داده‌ها و نیز خود الگوها، بسیار مفید واقع شود؛ اما باز هم افراد خبره هستند که معنی همه‌ی این‌ها را درک کنند. در قلمروی ناشناختنی‌ها، ما باید تنها بر شهودمان تکیه کنیم، البته هوش مصنوعی همچنان می‌تواند به ما کمک کند. AI می‌تواند محیط‌های داده‌ی پیرامون را اسکن کند و الگوهایی را به ما ارائه دهد، و ما نیز این الگوها را تشخیص دهیم. همچنین می‌تواند داده‌های افق‌های دوردست را اسکن کند و الگوهای کاربردی احتمالی را بیابد. در اینجا AI عمدتا مکمل شهود مدیر عامل است. فرد مدیر به دنبال نشانه‌های از پیش مشخص شده نیست، بلکه او نشان‌های تازه ظهور کرده را مشاهده می‌کند و معنی آن‌ها را در می‌یابد. AI نیز با گسترش وسعت رادار مدیر عامل، به او کمک می‌کند. با توجه به سخنان کولریج که ابتدای این بخش ذکر کردیم، باید مراقب شکل‌گیری هوشی باشیم که از شوخ طبعی برخودار نباشد. با کمک چند همکار، ما هم به روش سخت‌تری به همین نتیجه‌ی او دست پیدا کردیم: ما سالها تلاش کردیم با بررسی سطوح پیچیدگی سیستمیک بولدینگ (۱۹۵۶)، راه حل استواری برای تفکیک انسان از حیوان پیدا کنیم. نتیجه‌ی کار بدین صورت بود: انسان‌ها می‌توانند جک بسازند

و آن را بفهمند. شاید این نتیجه مضحک و بیهوده به نظر برسد اما این جمله مجموعه‌ی پیچیده‌ای از موقعیت-های شناختی را در بر می‌گیرد. برای اینکه بتوانیم یک جک بسازیم، اول اینکه باید دارای حافظه باشیم و این حافظه باید از نوع تاریخی باشد، همچنین باید توانایی انتزاع، استدلال، و خودآگاهی فراتر از صرف هوشیاری داشته باشیم. اگر من آزمون تورینگ را برگذار می‌کردم، به عنوان اولین کار جوک‌گویی را به میان می‌آوردم.

توصیه‌ی نهایی من این است که نباید فکر کنیم AI ما را هوشیارتر می‌کند. AI آنچه را که داریم تقویت می‌کند. در نتیجه اگر احیانا دچار حماقت باشیم، آن را هم تشدید خواهد کرد. این یعنی یک مدیرعامل با انتساب هوش مصنوعی برای کمک کردن به افراد خبره احتمالا بیشترین منفعت را از AI خواهد برد، نه اینکه از آن برای جایگزین کردن نیروهای عادی استفاده کند. همانطور که من چند سال است که این را می‌گویم، بزرگترین دستاوردها در آینده از افراد هوشمندی حاصل می‌شود، که تکنولوژی‌های هوشمند دارند.