**آنچه که هر مدیر عاملی باید درباره هوش مصنوعی بداند**

**فهرست مطالب**

[۱. مقدمه: چشم انداز استراتژیک هوش مصنوعی 2](#_Toc174779763)

[۱.۱ هوش مصنوعی، استراتژی و مدیرعامل 2](#_Toc174779764)

[۱.۲ عدم قطعیت: بستر استراتژی 4](#_Toc174779765)

[۱.۳ چرا این کتاب؟ 6](#_Toc174779766)

[۲. از ماشین به خدا 8](#_Toc174779767)

[۲.۱: سیستم‌های استدلال نمادین 9](#_Toc174779768)

[۲.۲ سیستم‌های متخصص نمادین 11](#_Toc174779769)

[۲.۳ شبکه‌های عصبی مصنوعی 14](#_Toc174779770)

[۲.۴ اما هوش مصنوعی چیست؟ 18](#_Toc174779771)

[۲.5 همهی اینها چه معنایی دارند؟ 20](#_Toc174779772)

[3 دانش 21](#_Toc174779773)

[3-1 دانش صریح و دانش ضمنی 21](#_Toc174779774)

[3-2 دانش شخصی 24](#_Toc174779775)

[3-3 دانش تنها نیست 27](#_Toc174779776)

[3-4 رسیدن به معنا 30](#_Toc174779777)

[3-5 همهی اینها چه معنی میدهند؟ 32](#_Toc174779778)

[4 یادگیری 33](#_Toc174779779)

[4-1 استعداد و الهامبخشی 33](#_Toc174779780)

[4-2 فراتر از موشهای اسکینر 35](#_Toc174779781)

[4-3 رسوخ کردن 38](#_Toc174779782)

[4-4 عقل سلیم 40](#_Toc174779783)

[4-5 همهی اینها چه معنی میدهد؟ 41](#_Toc174779784)

[5 خلاقیت 42](#_Toc174779785)

[5-1 عملکرد AI 42](#_Toc174779786)

[5-2 ایدههای کاربردی و نوین 44](#_Toc174779787)

[5-3 خلاقیت با AI 46](#_Toc174779788)

[5-4 AI معتبر 48](#_Toc174779789)

[5-5 همهی اینها چه معنی میدهد؟ 50](#_Toc174779790)

[6 مسائل اخلاقی 51](#_Toc174779791)

[6-1 اعتماد به هوش مصنوعی؟ 51](#_Toc174779792)

[6-2 اشتباهات و عواقب آنها 54](#_Toc174779793)

[6-3 اخلاقیات و حقوق 56](#_Toc174779794)

[6-4 تکینگی 59](#_Toc174779795)

[6-5 همهی اینها چه معنی میدهد؟ 61](#_Toc174779796)

[7 سخن آخر 62](#_Toc174779797)

# 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 (2020) 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 1997، 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 20 درصد هزینه مربوط به سخت افزار و نرم افزار است و 80 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 (2015) 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 (1921) دوباره مفهوم سازی کرد و توسط جی سی اسپندر در پنجاه سال گذشته گسترش و غنی سازی شد (اسپندر، 2021). عدم قطعیت با فقدان دانش مشخص می شود (اسپندر، 2014) و این فقدان دانش فضاهای فرصت ایجاد می کند (اسپندر، 2021). 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 100 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 (1954) 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 (3) می شود که زمینه اصلی شوبیک است. (4) عدم قطعیت را به طور محاوره ای می توان به عنوان اختیار اراده در نظر گرفت.

اسپندر (2014) 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 100 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، 2021: 129). 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، 1954: 632).

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 (2004) 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.

پشت‌بندۀ 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 ارائۀ الیور سلفریج (۱۹۵۵) دربارۀ 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 حوزۀ 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).Top of FormBottom of Form

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 حوزۀ 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 مسئلۀ 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: ۲۰:۵۰-۲۱:۴۵).

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]](#footnote-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]](#footnote-2) و همچنین سیستم‌های خبره­ی نمادین[[3]](#footnote-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]](#footnote-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 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]](#footnote-5) (1996) بود. پس از آن چت­بات­های امروزی بسیاری (مانند Cleverbot: رجوع شود به آرون[[6]](#footnote-6)، 2011) و جدیدترین مورد نیز یوجین گوسمن[[7]](#footnote-7)، شبیه­سازی پسربچه­ی 13 ساله­ی اوکراینی (وارویک و شاه، 2016) 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: «33 درصد از داوران تازه­کار را فریب دهد» (راسل و نورویگ، 2020). بنده به عنوان یک استاد دانشگاه، نمره­ی 33 درصد را برای دانشجویانم نمره­ی قبولی نمی­دانم. البته مشکلات به همین جا ختم نمی­شود؛ به عبارتی یوجین یک بچه­ی 13 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 2014) و نظرتان را بگویید؛ آیا به­نظرتان یوجین واقعا می­تواند یک شخص باشد؟ من شخصا دوست دارم با استوارت راسل و پیتر نورویگ هم­نظر باشم (2020)؛ 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 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 (1998 : 11) 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 3، 4 و 5، 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.

## ۲.5 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 4-5) 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 1950، 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 (2019)، 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.

# 3 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]](#footnote-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 4 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 20 دانشمند برتر، از جمله 17 برنده­ی جایزه­ی نوبل مصاحبه کنم. این مطالعه پس زمینه­ی این فصل و فصل­های 4 و 5 را شکل می­دهد.

## 3-1 دانش صریح و دانش ضمنی

به­منظور درک بهتر نکات ذکر شده در این قسمت، بد نیست ابتدا تصویر جامع­تری از «دانش»[[9]](#footnote-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 7/99 درصد از باکتری­ها را از بین می­برد». به همین دلیل است که جورج کِلی (1955) بر این باور است: «همه­ی ما دانشمند هستیم». دانشمند بودن یک موقعیت شغلی نیست، بلکه نوعی جهان­بینی است. در ادامه باید میان معنای «حقیقت»[[10]](#footnote-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 (1966) 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]](#footnote-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، 1969: 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]](#footnote-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 4 شرح داده خواهند شد؛ فعلا در اینجا بر جنبه­ی صریح[[13]](#footnote-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 در پیش است. همانطور که در بالاتر هم اشاره کردیم، ساخت دانش علمی با توجه به حقیقت، مقدار زیادی دانش ضمنی[[14]](#footnote-14) می­طلبد (پولانی، 1966ب). 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، 2012؛ دورفلر و اِدِن، 2019). دانش شهودی در حیطه­ی دانش­های ضمنی قرار می­گیرد؛ به­عبارتی می­دانیم، ولی دقیق نمی­دانیم که چگونه می­دانیم. این دانستن نوعی دانش مستقیم[[15]](#footnote-15) است (دورفلر و استیراند، 2017). 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، 2017؛ سین­کلیر و آشکاناسی، 2005). 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 1966الف : 7).

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 1959 : 13).

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[[16]](#footnote-16) 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 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[[17]](#footnote-17)» 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، 2012). 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[[18]](#footnote-18)» 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 4 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.

## 3-2 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 1962الف : 312). در این صورت، مفهوم دانش شخصی قصد­مندی[[19]](#footnote-19)، رانه[[20]](#footnote-20) و اشتیاق[[21]](#footnote-21) را نیز شامل می­شود و صرفا بر دوش کشیدن احساساتمان را در بر نمی­گیرد. در هر صورت، منظور ما نوعی ایدیوسنکرازی[[22]](#footnote-22) (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 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، 2019)، به علاوه­ی افراد برجسته در هر حرفه­ای مانند برترین سرآشپزها (استیراند و دروفلر، 2016)، دارای ویژگی به نام پرش­های شهودی است[[23]](#footnote-23). 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، 2005). 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، 2020الف). 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، 2015). 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.

تفکر تمثیلی[[24]](#footnote-24) 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، 1919 : 11-12). 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[[25]](#footnote-25) می­گویند (رومل­هارت و نورمن، 1988). 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[[26]](#footnote-26) 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 (2019)» 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 غرب تمایل بر این بوده است که فرستاده را جایگزین ارباب قرار داده و ارباب را نادیده بگیرند[[27]](#footnote-27)، 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 1986) 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، 2009؛ دریفوس و دریفوس، 1986؛ سایمون، 1996). 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)[[28]](#footnote-28) را طراحی کردند. آن­ها کاری را که این برنامه انجام می­داد در رده­ی کارهای دشوار قرار دادند، چون سطح آن در سطح تدریس­های دانشگاهی بود (نول و همکاران، 1963). 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.

## 3-3 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[[29]](#footnote-29)، هیجان­ها[[30]](#footnote-30) و ارزش­ها[[31]](#footnote-31) (ارزش­ها در فصل 6 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[[32]](#footnote-32) و در نتیجه با نیازهای اساسی ما دارند (دورفلر و چندری، 2008). 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 زیست­شناسان مانند فون برتالانفی (1981)، و برخی روان­شناسان مانند فروم (1982) 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 (1739 : 415) 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 (1995) 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 (1994) 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، 2021).

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 (2006) 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 برخورداند، 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، 1988 : 163).

## 3-4 رسیدن به معنا

فِین­باوم پس از مواجهه­ی اولیه­ای که با لاجیک تئوریست داشت (بخش 1-2) پروژه­ای را بر عهده گرفت که سیستم مقدماتی برخوردار از ادراک و حافظه (EPAM)[[33]](#footnote-33) نام داشت. هدف این پروژه ساخت مدلی از شناخت انسان بود. ماجرای EPAM از سپتامبر 1956، درست پس از بازگشت فین­باوم از تعطیلات تابستان شروع می­شود (فین­باوم، 2006 : 7:30). سایمون یک مقاله در زمینه­ی یادگیری انسان را از نشریه­ی ساینتیفیک امریکن[[34]](#footnote-34) به او نشان داد که درباره­ی موضوعی صحبت می­کرد که در آن زمان جدید به شمار می­رفت. مقاله مربوط می­شد به یادگیری گفتاری[[35]](#footnote-35)؛ 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[[36]](#footnote-36)» 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، 1984؛ سایمون و فین­باوم، 1964). 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 (1957) 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 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، 1966الف : 16).

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[[37]](#footnote-37) را در قالب شبکه­ای از معانی در ذهن­مان ذخیره می­کنیم (رومل­هارت و نورمن، 1988). 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 (1979)، 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، 1990). 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)[[38]](#footnote-38) 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.

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، 2007 : 457).

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 4-2 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، 1986 : 71). 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، 1986 : 131).

## 3-5 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[[39]](#footnote-39) 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.
* 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، 2022)، 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.

# 4 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 3، 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.

## 4-1 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، 1995)؛ 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، فعالیت­های 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، 1994)؛ به عبارت ساده­تر، همانطور که ژورنالیست­ها اغلب درباره­ی دستاوردهای بزرگ می­گویند که: 10 درصد استعداد و 90 درصد سخت­کوشی باعث رقم زدن آن­ها شده است. به نظر من این عقیده بی­معنی است. در واقع به نظر می­رسد دستاوردهای فوق­العاده از ترکیب 100 درصد استعداد 100 درصد تلاش زیاد به­وجود می­آیند. اما عنوان این کتاب «پدیده­های فوق­العاده» (دورفلر و استیراند، 2019) 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، 2017). مواردی از این دست تغییرات عمیق در یادگیری ما هستند که تحت عنوان «مفاهیم آستانه» نشان داده می­شوند (مِیِر و همکاران، 2010). 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[[40]](#footnote-40)، شکل­گیری فرا الگو[[41]](#footnote-41) گفته می­شود. این پدیده تا حدودی مترادف است با مفاهیم آستانه (رجوع شود به دورفلر، 2010). 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.

## 4-2 فراتر از موش­های اسکینر

در یک رویداد زنده­ی نیو ساینتیست[[42]](#footnote-42) در سپتامبر 2017 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 5 ذکر خواهد شد. نخستین مورد درباره­ی الگوریتم­های یادگیری DeepMind بود؛ به گفته­ی او «DeepMind مانند انسان­ها - از طریق یادگیری تقویتی[[43]](#footnote-43)- 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)[[44]](#footnote-44) 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 (1927) با این روش توانست رفلکس­های پاولووین را بشناسد و بوروس اسکینر (1950) 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 (1950) 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.

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، 1969 : 66). 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 (1946 : 30-29) 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[[45]](#footnote-45) 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 (2000) در کتاب «دانش کارآمد»[[46]](#footnote-46) 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، 2000 : 84). داستان در این جا متوقف شده و خواننده با یک پاسخ نصفه و نیمه تنها گذاشته می­شود؛ به این عبارت که سیستم خبره مذکور با شکست مواجه شد. تا اینکه در 11 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، 2000 : 95)

## 4-3 رسوخ کردن[[47]](#footnote-47)

ایده­ی رسوخ کردن، مانند بسیاری از ایده­های مرتبط با دانش و یادگیری توسط پولانی بنیان­گذاری شد (1962ب). 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 (1962ب) 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، 1962الف : 61).

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 ببیند 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[[48]](#footnote-48) نگاه کنیم (دورفلر و استیراند، 2018). 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[[49]](#footnote-49) خواند. صرف­نظر از نوع سنتی رابطه­ی استاد و شاگردی که در بخش قبل توضیح داده شد، رسوخ اشتراکی در «شاگردان سیار[[50]](#footnote-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]](#footnote-51)» نام گرفته است (دروفلر و ادن، 2017). 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]](#footnote-52) نیز مشاهده کردیم، چیزی که آن را «تفکر گروهی» نامیدیم (پیرکو و همکاران، 2017) و دریافتیم که تفکر گروهی یگانه عامل ضروری در شکل­گیری CoPهاست. به این پدیده «رسوخ متحد[[53]](#footnote-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، 2017 : 390).

شاید گفتن این جالب باشد که پایه­گذاری مفهوم اولیه­ی CoP توسط ژان لاو و اتین ونگر (1991)، 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، 2022). 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، 2011 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. 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.

## 4-4 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[[54]](#footnote-54) 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.

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، 1988 : 22).

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 5 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، 1988 : 72).

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، 1990 : 122). 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، 1992 : 17). ساخت آن در سال 1986 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؟

## 4-5 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]](#footnote-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.

# 5 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».

## 5-1 عملکرد AI

دومین نکته­ای که هابیس در سخنان خود طی رویداد زنده­ی نیو ساینتیست 2017 (آن را در بخش 2-4 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 18 دوره از مسابقات گو- را در طی 5 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 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 1957 بود که سایمون پیش­بینی کرد:

1. طی 10 سال آینده کامپیوترهای دیجیتال در شطرنج قهرمان جهان خواهند شد، مگر آن­که قوانین شرکت کامپیوترها را در مسابقه منع نمایند.
2. طی 10 سال آینده یک کامپیوتر دیجیتال خواهد توانست که یک قضیه­ی ریاضی را کشف و اثبات کند.
3. در 10 سال آینده یک کامپیوتر دیجیتال موسیقی­ای خواهد نوشت که منتقدان می­پذیرند ارزش زیباشناختی بالایی دارد.
4. در 10 سال آینده بیشتر نظریه­های روان­شناسی در قالب برنامه­های کامپیوتری یا عبارت­های کیفی ظاهر می­شوند که مشخصات همین برنامه­ها را توصیف می­کنند.

(سایمون و نول، 1958 : 8-7)

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 1997، چهل سال بعد از پیش­بینی سایمون، یک برنامه­ی کامپیوتری به نام دیپ بلو[[56]](#footnote-56) 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 2016 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 100،000 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 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، 2019). 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، 2012). 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 3 یا 4 را بازی کنند و هیچ­کس خط 5 را بازی نمی­کرد. با این­حال AlphaGo در «دوره­ی آموزشی» خود حدود 300 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 غیره­منتظره، 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 1980) 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.

## 5-2 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 1982 پیشنهاد داد که AI در 100 سال آینده و شاید حتی 50 سال بعد، متن­هایی در سطح ادبیات شکسپیر می­نویسد (آمابیل، 202). 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، 1977 : 6).

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، 1983ب و 1996). 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، 1983ب). 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، 1968 : 8). 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، پست­ایت[[57]](#footnote-57) (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؟

اگر تعریف خلاقیت را به «ایده­، راه حل و دیگر خروجی­ها» بسط بدهیم (آمابیل 2020 : 3)، 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، 1982) و مبنای تکنیک سنجش وفاقی (CAT)[[58]](#footnote-58) می­باشد. این تکنیک یکی از گسترده­ترین روش­های سنجش خلاقیت است (بائر، 2020). 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 ابزازی 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.

## 5-3 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[[59]](#footnote-59) 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 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 (2018 : 44) 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 2-3 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 [[60]](#footnote-60)AI» می­گویند و ایده در مرحله­ی بعد قرار می­گیرد. البته عکس آن نیز ممکن است؛ در حالت «تقدم ایده[[61]](#footnote-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]](#footnote-62) آن­ها تبدیل می­شود (ویتگنشتاین، 1969 : 94). 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.

## 5-4 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، 2011). 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 10 ماه از پاندمی کووید-19 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 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[[63]](#footnote-63).

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، 2020). 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 50 سال تحقیق بر روی پروتئین توسط صدها دانشمند، با کمک تیم نخبگان زیست­شناسی، فیزیک و علوم کامپیوتر دیپ­مایند، و فراهم کردن حدودا 170،000 پروتئین برای آموزش ANN حاصل شد. این دستاورد از جهات مختلف شبیه پروژه­ی DENDRAL فین­باوم بود (بخش 2-2 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 50 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 (2009 : 244-245) در مورد برنامه­ای می­گوید که در دهه 1980 توسط آهنگساز دیوید کوپ طراحی 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، 1998)، 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 2-4 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.

## 5-5 همه­ی اینها چه معنی می­دهد؟

یک کاریکاتور از شخصیت کارتونی گارفیلد وجود دارد که آن را بسیار دوست دارم. جان[[64]](#footnote-64) 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 2019، 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 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.

# 6 مسائل اخلاقی

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

از پاییز 2017 و به مدت به مدت دو سال و نیم، 20 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؟»

## 6-1 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 19 از استنفورد بود که باعث حذف شدن پزشکان خط مقدم شد(گوئو و هائو، 2020). 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[[65]](#footnote-65) است؛ در این رویکرد ما به افرادی که آن­ها را بافضیلت می­دانیم اعتماد داریم. اگر ما طرفدار اخلاقیات پیامدگرایانه[[66]](#footnote-66) باشیم، تنها نتیجه­ی کار را مد نظر قرار خواهیم داد. پیروان وظیفه­گرایی[[67]](#footnote-67) (به آن اخلاق مبتنی بر قاعده هم می­گویند) هم باور دارند که عمل، و در نتیجه نیتی 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، 1959 : 22).

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: 1- AI فقط اطلاعاتی که در دیتابیس هستند را در نظر می­گیرد و امکان دارد متغیرهایی باشند که در دیتابیس ثبت نشده­اند؛ و 2- ممکن است ارزش­هایی در دیتابیس باشند که ما نیازی به رونویسی آن­ها نداشته باشیم (برای دیدن نمونه رجوع کنید به هائو، 2020 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 4-4). 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.

## 6-2 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 447 ایر فرانس» (اولیور و همکاران، 2017الف، 2017ب)، 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 228 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 (2019 : 164) 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-3 استفاده می­کرد، به یک بیمار غیرواقعی پیشنهاد خودکشی داد (داوس، 2020). 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، 2019 : 165).

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 62 سالگی برای اولین بار به پرو سفر کرد. طی 5 سال بعدی، او هر روز پیشنهاد سفر به لیما[[68]](#footnote-68) 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 4-4) از این 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، 1966ب : 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 4-3) ممکن است «لوپ­های عجیبی» را هم شامل شوند (هوفشتاتر، 1979). 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، 2019 : 164). تشخیص «اهم موارد» اصلی­ترین کاری است که AI از انجام آن عاجز است. در اینجاست که عقل سلیم و حکمت[[69]](#footnote-69) با هم شباهت پیدا می­کنند. وایزن­بام سازنده­ی ELIZA (اولین هوش مصنوعی که در آزمون تورینگ قبول شد: بخش 4-2 و 4-3 را ببینید)، نخستین کتابی را که مساله­ی اخلاقیات در AI را مستقیما مطرح کرد نوشت (وایزن­بام، 1976). 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[[70]](#footnote-70) هستند.

## 6-3 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 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 4-5 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[[71]](#footnote-71)، 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[[72]](#footnote-72) 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، 2019)، 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، 2020) اما به عنوان یک سیستم هشداردهی 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 12 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، 1968: 22). 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 (2019) 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 (2020) 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، 2020 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.

## 6-4 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[[73]](#footnote-73) 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 (2005) می­گوید، تکینگی در کمین­مان است؟ فکر نکنم. نکته­ی مهم در این جا مفهوم خودآگاهی است. برخی وعده داده­اند که در طی 5 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 (1998) 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 1-2) 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 (1990) از پیشتازان AI این موضوع را به زمان توماس هابز (1651) 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 (2018) 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، 2015).

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 2-6 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.

## 6-5 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 اتومبیل­های 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، 2019)، 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.

# 7 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، 1995 : 54).

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 (2019 : 165) 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 اَن کانلیف (2009) 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 (2018): 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 2-1): 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 (1956)، 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 موقعیت­های 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.

1. - Fuzzy Logic [↑](#footnote-ref-1)
2. - Artificial Neural Network [↑](#footnote-ref-2)
3. - Symbolic Expert System [↑](#footnote-ref-3)
4. - Fitness function [↑](#footnote-ref-4)
5. - Joseph Weizenbaum [↑](#footnote-ref-5)
6. -Aron [↑](#footnote-ref-6)
7. - Eugene Goostman [↑](#footnote-ref-7)
8. - knowledge base [↑](#footnote-ref-8)
9. - knowledge [↑](#footnote-ref-9)
10. - truth [↑](#footnote-ref-10)
11. - geoid [↑](#footnote-ref-11)
12. - phrónēsis [↑](#footnote-ref-12)
13. - explicit [↑](#footnote-ref-13)
14. - tacit [↑](#footnote-ref-14)
15. - direct [↑](#footnote-ref-15)
16. - reasoning [↑](#footnote-ref-16)
17. - thinking aloud [↑](#footnote-ref-17)
18. - knowledge discovery [↑](#footnote-ref-18)
19. - intentionality [↑](#footnote-ref-19)
20. - drive [↑](#footnote-ref-20)
21. - passion [↑](#footnote-ref-21)
22. - idiosyncrasy [↑](#footnote-ref-22)
23. - intuitive leaps [↑](#footnote-ref-23)
24. - analogical thinking [↑](#footnote-ref-24)
25. - analogical representational system [↑](#footnote-ref-25)
26. - seeing essence [↑](#footnote-ref-26)
27. - اشاره به همان کتاب ارباب و فرستاده­اش (The Master and His Emissary) [↑](#footnote-ref-27)
28. - Logic Theorist [↑](#footnote-ref-28)
29. - feelings [↑](#footnote-ref-29)
30. - emotions [↑](#footnote-ref-30)
31. - values [↑](#footnote-ref-31)
32. - instincts [↑](#footnote-ref-32)
33. - Elementary Perceiver and Memorizer [↑](#footnote-ref-33)
34. - Scientific American [↑](#footnote-ref-34)
35. - verbal learning [↑](#footnote-ref-35)
36. - elementary learning [↑](#footnote-ref-36)
37. - propositional knowledge [↑](#footnote-ref-37)
38. - Natural Language Processing [↑](#footnote-ref-38)
39. - knowledge representation [↑](#footnote-ref-39)
40. - transformational learning [↑](#footnote-ref-40)
41. - meta schema [↑](#footnote-ref-41)
42. - New Scientist Live [↑](#footnote-ref-42)
43. - reinforcement learning [↑](#footnote-ref-43)
44. - stimulus-response [↑](#footnote-ref-44)
45. - Agora: میدان­­هایی در شهرهای یونان باستان [↑](#footnote-ref-45)
46. - working knowledge [↑](#footnote-ref-46)
47. - Indwelling [↑](#footnote-ref-47)
48. - embodiment of knowledge [↑](#footnote-ref-48)
49. - shared indwelling [↑](#footnote-ref-49)
50. - wandering apprentices [↑](#footnote-ref-50)
51. - mutual apprenticing [↑](#footnote-ref-51)
52. - communities of practice [↑](#footnote-ref-52)
53. - interlocked dwelling [↑](#footnote-ref-53)
54. - common sense [↑](#footnote-ref-54)
55. - Eureka moments [↑](#footnote-ref-55)
56. - deep blue [↑](#footnote-ref-56)
57. - Post-it [↑](#footnote-ref-57)
58. - Consensual Assessment Technique [↑](#footnote-ref-58)
59. - stockfish [↑](#footnote-ref-59)
60. - AI-first mode [↑](#footnote-ref-60)
61. - idea-first mode [↑](#footnote-ref-61)
62. - inherited background [↑](#footnote-ref-62)
63. - authentic AI [↑](#footnote-ref-63)
64. - یکی از شخصیت­های همین کارتون [↑](#footnote-ref-64)
65. - virtue ethics [↑](#footnote-ref-65)
66. - consequentialist ethics [↑](#footnote-ref-66)
67. - deontology [↑](#footnote-ref-67)
68. - پایتخت کشور پرو [↑](#footnote-ref-68)
69. - wisdom [↑](#footnote-ref-69)
70. - empathy [↑](#footnote-ref-70)
71. - case-law [↑](#footnote-ref-71)
72. - codified law [↑](#footnote-ref-72)
73. - singularity [↑](#footnote-ref-73)