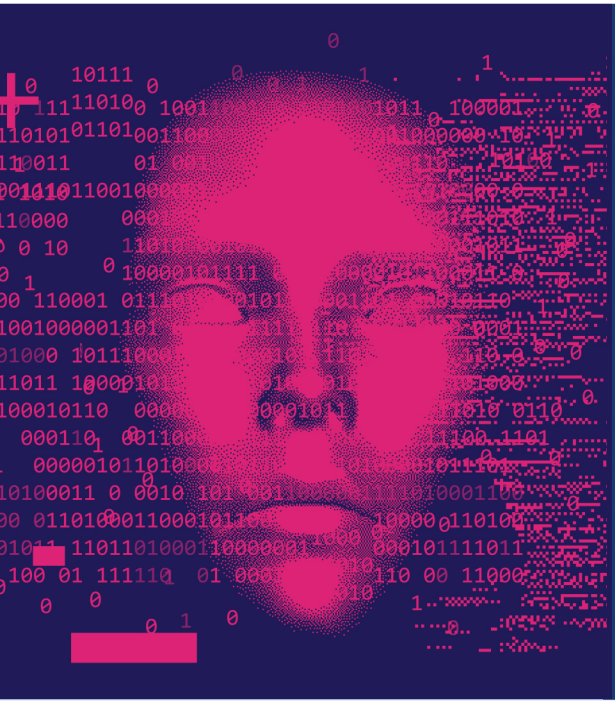
مرکز ملی فضای مجازی
پژوهشگاه فضای مجازی

عصر فضای مجازی
بیست و سوم



درآمدی بر فلسفه هوش مصنوعی

An Introduction to the Philosophy of Artificial Intelligence

جمهوری اسلامی ایران
مرکز ملّی فضای مجازی

درآمدی بر فلسفه هوش مصنوعی

گزارش شماره ۲۳
دی ماه ۱۳۹۸

تهیه شده در: پژوهشگاه مرکز ملی فضای مجازی- گروه مطالات بنیادین فضای مجازی
تهیه‌کننده: دکتر مهدی همازاده (پژوهشگر موسسه پژوهشی حکمت و فلسفه ایران)
ناظر علمی: دکتر حسین مطلبی کربکندی

نشانی: تهران، میدان آرژانتین، خیابان بیهقی، نبش خیابان ۱۶ غربی، پلاک ۲۰، کدپستی ۱۵۱۵۶۷۴۳۱۱
شماره تماس: ۸۶۱۲۱۰۶۱ http://www.majazi.ir

حقوق مادی و معنوی این اثر متعلق به مرکز ملی فضای مجازی است و استفاده از مطالب آن صرفاً با ذکر مأخذ بلامانع است.

محتوای انتشاریافته در این گزارش الزاماً بیانگر دیدگاه مرکز ملی فضای مجازی نیست

سخن تحست

فضای مجازی با شتاب شگرف و رو به تزایدی که در حال بسط و گسترش است تمام ساحات اجتماعی، اقتصادی، سیاسی و فرهنگی زندگی بشر را در نوردیده و هر روز بخش بزرگی از زندگی واقعی را در خود فرو برده و حیات متفاوت و جدیدی به آن می دهد. لذا به نظر می رسد دو نگاه کلان به فضای مجازی وجود دارد: نگاه اول که بالاخص در ابتدای رشد و تکوین فضای مجازی مسلط شده بود، آن را همچون ابزاری کنار سایر ابزارهای بشری تصویر می کرد که تنها طریقیت داشت. اما نگاه دوم، در نتیجه رشد تحولات خیره کننده فضای مجازی و سایه گستری آن در حوزه ها و شئون بشر در یک دهه اخیر آن را چون سکویی می داند که بسیار فراتر از شأن ابزاری حیات انسان ها را سامان جدیدی داده و ادعای تمدن نوینی را دارد. رویکردی که از قضا از چشمان بصیر رهبر انقلاب نیز دور نمانده و انتظاری تمدنی از فضای مجازی در ایران را مطالبه داشته اند.

در همین راستا گزارش های عصر فضای مجازی تلاش می کند تا فهم سازمان ها و دستگاه های مرتبط با حوزه ی فضای مجازی را ارتقاء بخشیده و آن ها را برای مواجهه فعال و خردمندانه با تحولات این عرصه مهیا سازد.

سید ابوالحسن فیروزآبادی

دبیر شورای عالی و رئیس مرکز ملی فضای مجازی

چکیده

این گزارش به هوش مصنوعی از منظر مطالعات مبنایی و وجود شناختی می‌پردازد و بدین منظور، پس از ذکر تاریخچه‌ای از روند شکل‌گیری هوش مصنوعی و اهداف آن، ابتدا دو رویکرد کلاسیک و جدید در ساخت هوش مصنوعی را توضیح می‌دهد.

سپس هوش مصنوعی قوی را به مثابه یک غایت و چشم‌انداز مهم نزد بسیاری از مهندسان این حوزه تبیین کرده و از میان اشکالاتی که فیلسوفان به رویکردها و اهداف هوش مصنوعی وارد کرده‌اند، چهار اشکال مهم را شرح می‌دهد و اشکال مهم دیگری را جداگانه و علاوه بر اشکالات مذکور ذکر می‌کند.

البته از آن‌جا که گزارش صرفاً به جنبه‌های وجودشناختی مربوط می‌شود، از ساحت ادبیات فلسفه ذهن و علم‌النفس اسلامی خارج نشده و مباحث مربوط به اخلاق هوش مصنوعی را – به عنوان یکی دیگر از ساحت‌های فلسفی بحث پیرامون هوش مصنوعی - کنار گذاشته است.

نویسنده در انتها با مراجعه به منابع بومی فلسفی، توضیح می‌دهد که هوش مصنوعی و آگاهی مصنوعی – لااقل در برخی جنبه‌های آن – مطابق علم‌النفس حکمت متعالیه، امکان عقلی دارد و قابل انکار قطعی نیست.

واژگان کلیدی: هوش مصنوعی، ذهن، علوم شناختی، محاسبه‌گرایی، پیوندگرایی

فهرست مطالب

مقدمه ۲

۱. پیشینه ۲

۲. هوش مصنوعی کلاسیک و جدید ۱۱

۲-۱. هوش کلاسیک یا دیجیتال ۱۱

۲-۲. پیوندگرایی در هوش مصنوعی جدید ۱۷

۲-۳. بدن‌مندی/ مکان‌مندی در هوش مصنوعی جدید ۲۵

۳. انواع رویکردها به چیستی و غایت هوش مصنوعی ۳۲

۴. هوش مصنوعی قوی و ضعیف ۳۷

۵. به سمت شبیه‌سازی خاستگاه فیزیولوژیک ۴۱

۶. مسائل فلسفی ۴۲

۶-۱. مسئله حس عمومی یا متناسب بودن رفتار ۴۳

۶-۲. مسئله آدمک چینی ۴۶

۶-۳. مسئله اتاق چینی ۴۷

۶-۴. مسئله سیستماتیک بودن ۵۱

۶-۵. مسئله خصیصه سابجکتیو تجربه پدیداری ۵۳

۷. امکان ظهور آگاهی مصنوعی برپایه فلسفه اسلامی ۵۷

جمع‌بندی ۶۱

منابع ۶۲

مقدمه

هزاران سال است که محقّقان و اندیشمندان تلاش می‌کنند دریابند انسان‌ها چطور دارای آگاهی هستند؟ چطور تعداد محدودی از موجودات فیزیکی می‌توانند احساسات و ادراک حسّی داشته باشند، بفهمند، پیش‌بینی کنند و جهانی که بسیار بزرگ‌تر و پیچیده‌تر از خودشان است را دست‌خوش تغییر سازند. هوش‌مصنوعی از این هم جلوتر می‌رود و سعی می‌کند هویّات هوشمند را نه‌تنها بفهمد، بلکه بسازد.

البته واژه «هوش» در عنوان این حوزه مطالعاتی، گمراه‌کننده است. این واژه را وقتی درباره انسان‌ها سخن می‌گوییم، برای شاهکارهای ذهنی در خلاقیت و استعداد ایشان به کار می‌گیریم. درحالی‌که جذّاب‌ترین مسائل هوش‌مصنوعی، به تلاش‌ها برای تقلید قوای ذهنی مردم عادی (مانند بینایی و زبان طبیعی) برمی‌گردد. مردم عادی کارهایی مانند دیدن و صحبت کردن را راحت می‌یابند و اهمیت زیادی برای آن‌ها قائل نیستند. اما از سوی دیگر، کارهایی مانند ضرب اعداد ده رقمی برای اکثر ایشان، دشوار به نظر می‌رسد. ولی استفاده از کامپیوترها برای مطالعه قوای ذهنی، مشخّص کرد که واقعاً چقدر بازسازی کارهای ساده انسانی مانند دیدن و معاشرت و ...، برخلاف انجام عملیات پیچیده ریاضیاتی، در کامپیوترها چالش‌برانگیز است.

حال سؤال این است که با تمام چالش‌ها و پیشرفت‌های تجربه‌شده، آیا چشم‌انداز امیدوارکننده‌ای برای ساخت موجودات شبیه انسان یا ساخت هرگونه آگاهی مصنوعی (به معنای کامل آن) وجود دارد؟ و در صورت وجود چنین چشم‌اندازی، جایگاه مبانی فکری بومی و اندیشمندان و مدیران داخلی در این چشم‌انداز، کجاست؟

۱. پیشینه

هوش‌مصنوعی یکی از جدیدترین حوزه‌ها در علوم و مهندسی است که بلافاصله پس از جنگ جهانی دوم به نحوی جدّی کلید خورد و این روزها شامل زیر حوزه‌های بسیار متنوّعی می‌شود؛ از موضوعات عمومی (مانند یادگیری و ادراک حسّی) گرفته تا موضوعات تخصّصی‌تر (مانند بازی شطرنج، اثبات تئوری‌های ریاضیاتی، رانندگی در خیابان شلوغ و تشخیص بیماری‌ها). درواقع هوش‌مصنوعی به هر عمل هوشمندانه مربوط می‌شود و حوزه‌ای از پژوهش‌ها را در بر می‌گیرد که به ساخت حیوانات هوشمند (یا دست‌کم، جانداران هوشمندی که ظاهراً حیوان هستند)، اختصاص دارد و در بسیاری موارد، ساخت انسان‌های هوشمند (یا دست‌کم جانداران هوشمندی که ظاهراً انسان هستند) را نیز شامل می‌شود.

ایده ساخت هوش‌مصنوعی البته ایده‌ای قدیمی است. ادبیات اروپایی مملوّ از داستان‌هایی است که در آن‌ها شخصی در حال ساخت موجودی هوشمند است. اما اگر به دنبال تبار علمی این ایده باشیم، شاید بتوانیم به بسیاری منطق‌دانان و فیلسوفان اشاره کنیم که بر روی فرمولیزه‌کردن و نهایتاً ماشینیزه‌کردن «قوانین فکر» کار کرده‌اند. احتمالاً قدیمی‌ترین این فیلسوفان، ارسطو باشد که عقلانیت را خصیصه‌ای ذاتی برای ذهن انسان می‌دانست و قیاس منطقی را نشان این عقلانیت معرّفی می‌کرد. عمیق‌ترین مشارکت ارسطو در هوش‌مصنوعی را می‌توان ایده «صورت‌گرایی» دانست که الگوهای خاص فکری را در قالب فرم نحوی‌شان (مستقل از محتوا)

معتبر می‌دید. این یک ابداع قدرتمند بود که در قلب تئوری محاسباتی ذهن و آن‌چه هوش‌مصنوعی می‌نامیم نیز حفظ شده است. نظر به اهمیتی که در تاریخ فلسفه به قیاس داده شده، ایده ماشین هوشمند غالباً به مثابه ماشینی بوده است که بتواند استنتاج منطقی انجام دهد؛ یعنی نتیجه را به نحوی معتبر از مقدّمات، استخراج نماید. (Arkoudas & Bringsjord, 2014, p. 52.)

بعدها هابز - آن‌گاه که در جمله مشهور خویش، اعلان داشت: «استدلال، محاسبات است»[۱] - هوش‌مصنوعی را در قرن هفدهم پیش‌بینی کرده بود. تقریباً در همان دوره، لایب‌نیتز هم از رؤیای «منطق جهانی»[۲] خبر داد که در آن تمام منازعه‌ها می‌توانند با محاسبات حلّ و فصل بشوند. اما یکی از نام‌های شناخته‌شده متأخّر در این زمینه، جرج بول[۳] (۱۸۶۴-۱۸۱۵) است که "جبر بولی" را ابداع کرد. ایده‌های بول در هر دو حوزه ریاضیات و فلسفه، وارد شده، اما توجّه وی بیشتر به سوی هوش‌مصنوعی معطوف بود. هرچند واقعیت روشن این است که فقط در قرن بیستم بود که ابزارهای هوشمندانه (منطق و مطالعات فرمالیستی) و دستگاه‌های فیزیکی (از تیوب‌های تخلیه تا ترانزیستورها تا تراشه‌ها و ...) در دسترس قرار گرفتند. (Charniak, & McDermott, 1985, p. 10.)

به‌طور مشخّص البته اصطلاح «هوش‌مصنوعی» از سال ۱۹۵۶ و در کنفرانس تابستانی کالج دارموث امریکا کلید خورد. اما حوزه مطالعاتی هوش‌مصنوعی عملاً قبل از ۱۹۵۶ تعریف شده و به راه افتاده بود. مثلاً نقش محوری در خلق آن‌چه که امروزه تئوری کنترل نامیده می‌شود را نوربرت وینر[۴] (۱۹۶۴-۱۸۹۴) ایفا کرد. وی به‌عنوان یک ریاضیدان برجسته، قبل از تشدید علاقه‌مندی‌ها به سیستم‌های کنترل مکانیکی و بیولوژیکی، و اتّصال آن‌ها به شناخت، با برتراند راسل در این زمینه کار کرده بود. لایک کریک[۵] - که او نیز سیستم‌های کنترل را به مثابه مدل‌های روان‌شناختی به کار گرفت – با وینر و همکارانش، رفتار هدفمند را برآمده از یک مکانیسم تنظیم یافتند که تلاش در جهت به حداقل رساندن خطا دارد؛ یعنی به حداقل رساندن اختلاف بین حالت موجود و حالت هدف. وینر و هکارانش، یک‌سری کنفرانس‌های تأثیرگذار سازماندهی کردند که در پی مدل‌های ریاضیاتی و محاسباتی شناخت بود. کتاب پرفروش او در سال ۱۹۴۸ با عنوان *سایبرنتیک*، جلب توجّه همگانی به امکان ساخت ماشین‌هایی با هوش‌مصنوعی را به دنبال داشت.

---

۱. "Ratiocination is computation."

۲. Universal Calculus

۳. George Boole

۴. Norbert Wiener

۵. Like Craik

در همین اثناء، روس اشبای[۱] در ۱۹۴۰ با ایده‌هایی مشابه در کتاب *طراحی یک مغز* (در ۱۹۴۸ و ۱۹۵۲) توضیح داد که هوش می‌تواند با استفاده از دستگاه‌های هومواستاتیک – که در بردارنده چرخه‌های بازخوردگیری متناسب هستند – تولید شود تا به رفتار سازگار و پایدار دست یابد. اما اوّلین کاری که در تاریخچه هوش‌مصنوعی شناسایی شده، به وسیله مک‌کولاک[۲] و پیتز[۳] در سال ۱۹۴۳ به انجام رسید. آن‌ها یک مدل از نورون‌های مصنوعی طراحی کردند که هر نورون در این مدل، می‌توانست روشن یا خاموش باشد و روشن شدن هر نورون از طریق تحریک توسّط تعدادی کافی از نورون‌های مجاور صورت می‌گرفت. آن‌دو نشان دادند که هر کارکرد محاسباتی می‌تواند به‌وسیله چند شبکه از نورون‌های متّصل محاسبه شود. مک‌کولاک و پیتز پیش‌بینی کردند که شبکه‌هایی با تعریف درست، می‌توانند قدرت یادگیری داشته باشند.

سپس دونالد هب[۴] در ۱۹۴۹ یک قانون به‌روزشده‌ی ساده برای بهبود قدرت اتّصالات بین نورون‌ها ارائه کرد که امروزه با نام قانون یادگیری هِبین[۵]، همچنان به‌عنوان مدلی تأثیرگذار باقی مانده است. دو دانشجوی فارغ‌التّحصیل هاروارد – مینسکی[۶] و ادموندز[۷] - نیز اوّلین شبکه کامپیوتر نورونی را در ۱۹۵۰ ساختند. بعد از آن، مینسکی در پرینستون بر روی محاسبات جهانی در شبکه‌های نورونی متمرکز شد. (See: Russell, & Norvig, 2010, pp. 15-17.)

این‌ها چند نمونه از کارهای اوّلیه در حوزه هوش‌مصنوعی بودند، اما شاید بتوان گفت دیدگاه آلن تورینگ بیشترین تأثیرگذاری و شهرت را در این دوره داشته است. تا سال ۱۹۴۰، هیتلر بخش معظمی از سرزمین اروپا را به چنگ آورده بود و دولت انگلستان برای دفاع، بهترین مهندسان برق و ریاضیدانان خود را تحت رهبری آلن تورینگ و با مأموریت رخنه به رمزکدهای ارتش آلمان، گرد آورد. مشخّص بود که با توجّه به تفوّق نیروی هوایی آلمان در آسمان، شکست در این مأموریت، هزینه و تلفات سنگینی به دنبال خواهد داشت.

تورینگ و همکارانش اوّلین کامپیوتر عملگر جهان را از رله‌های تلفن ساختند و آن را "رابینسون" نامیدند. کامپیوتر این گروه، موفّقیت درخشانی داشت و نسخه‌ای از تقریباً تمامی پیام‌های نازی‌ها را در اختیار انگلستان گذاشت. سپس آلمان‌ها با افزودن چرخه‌های کدگذاری در ماشین رمزنگار، بر پیچیدگی کدهایشان افزودند و تورینگ نیز هوش الکترومغناطیس رابینسون را با یک نسخه الکتریکی هوش با نام کلسوس جایگزین کرد که

---

۱. Ross Ashby

۲. Warren McCulloch

۳. Walter Pitts

۴. Donald Hebb

۵. Hebbian Learning

۶. Minsky

۷. Edmonds

از دو هزار تیوب رادیویی ساخته می‌شد. کلسوس با ۹ ماشین شبیه خود – که به‌طور موازی کار می‌کردند – یک رمزگشایی غیرمنقطع و حیاتی را برای نیروهای متّفقین فراهم آورد. اطلاعاتی که توسّط رابینسون و کلسوس فراهم آمده بود، با احتیاط زیاد مورد استفاده قرار گرفت؛ ولی رخنه در رمزگذاری کافی بود تا نیروی هوایی سلطنتی را برای پیروزی در نبرد هوایی، توانمند سازد.

بدین ترتیب، بر اساس تنوّع سنّت‌های هوشمند و با تجدید نیرو از ضروریات جنگ، شکل جدیدی از هوش بر روی زمین ظهور کرد. تورینگ، مشابهت فرآیند محاسباتی و فرآیند تفکّر انسانی را به خاطر سپرد و با حجم قابل توجّه بنیادهای نظری محاسبات و اختراع اوّلین کامپیوتر عملگر، تلفیق کرد و زمینه‌ساز اوّلین تلاش‌ها در به‌کارگیری این تکنولوژی جدید برای شبیه‌سازی هوش گردید (Kurtzweil, 1999, p. 57.). وی همچنین سخنرانی‌هایی را دراین‌باره از ۱۹۴۷ آغاز کرد و تفسیری قابل توجّه در مقاله ۱۹۵۰ خویش با عنوان «محاسبات ماشینی و هوش»[۱] ارائه داد و یک دستور کار برای تحقیقات نیم قرن بعدی کامپیوترهای پیشرفته، توصیف نمود: بازی کردن، تصمیم‌گیری، زبان طبیعی، درک و فهم، ترجمه، اثبات قضایا، و البته رمزگذاری و رمزگشایی. وی در این مقاله آزمون تورینگ، یادگیری ماشین، الگوریتم‌های ژنتیکی، و تقویت یادگیری را نیز مطرح و استدلال کرد که پرسش «آیا یک ماشین می‌تواند فکر کند؟» باید با این پرسش جایگزین شود که «آیا یک ماشین به‌لحاظ زبانی از یک انسان غیرقابل تمایز است»؟ منظور تورینگ در این‌جا، همین کامپیوترهای استانداردی است که کارکردهای محاسباتی را انجام می‌دهند.

تورینگ در این‌جا آزمونی را مطرح می‌کند و در حالت دوم آزمون - که به عنوان تفسیر رایج و استاندارد از آزمون تورینگ شناخته می‌شود - یک انسان و یک کامپیوتر در دو اتاق مجزّا و بسته، قرار دارند و یک شخص دیگر هم در شرایطی که هیچ‌گونه شناختی از موجود حاضر در هریک از اتاق‌ها ندارد، سؤالاتی را به‌واسطه ایمیل (یا دورنگار)، از حاضرین دو اتاق می‌پرسد. اگر این شخص ثالث نتواند براساس قوّت پاسخ‌های دریافتی، بهتر از ۵۰ – ۵۰ تشخیص دهد که احتمالاً کدام‌یک از دو پاسخ‌دهنده (انسان یا ماشین) در کدام اتاق حاضر هستند، نتیجه خواهیم گرفت که کامپیوتر این آزمون را پاس کرده است. پاس‌کردن به این معنا، تمایزناپذیری زبانی را عملی می‌سازد.

آزمون تورینگ همچنان و هنوز در قلب مطالعات هوش‌مصنوعی و مباحث بنیادین مربوط به آن قرار دارد. هرچند از منظر فلسفی که تحقیقات درباره هوش ماشین را معنادار و پرفایده، اما جدا از سمبولیسم خاص منطقی می‌بینند، نه مقاله تورینگ و نه کنفرانس سال ۱۹۵۶ دارموث، حتّی به آغاز هوش‌مصنوعی، نزدیک هم نشده‌اند.

---

[۱] . "Computing, Machinery and Intelligence"

کنفرانس دارموث البته نقطه عطف و عزیمت تاریخی در حوزه مطالعاتی هوش‌مصنوعی به حساب می‌آید. زمانی که جان مک‌کارتی[۱] پس از اخذ دکترای خویش در ۱۹۵۱ از پرینستون، به کالج دارموث رفت و مینسکی و شانون و راتچستر و سیمون را متقاعد کرد تا با کمک یکدیگر، پژوهشگران امریکایی علاقمند به تئوری ماشین، شبکه‌های نورونی و مطالعات هوش را گردآورند. آن‌ها یک کارگاه دوماهه را در تابستان ۱۹۵۶ و در کالج دارموث راه‌اندازی کردند و در طرح‌نامه این کارگاه، صراحتاً واژه "هوش‌مصنوعی" را برای ماشینی که بتواند جنبه‌های مختلف هوش را شبیه‌سازی کند، به‌کار گرفتند. در این کارگاه، پدران هوش‌مصنوعی مدرن، برای یک مدرسه تابستانی در کالج دارموث (ایالت نیوهمشایر) و ذیل این فرضیه گرد هم آمدند که: «تمامی جنبه‌های یادگیری یا هر خصیصه هوشمندانه دیگر، علی‌الاصول می‌تواند به دقّت و به گونه‌ای توصیف شود که یک ماشین نیز بتواند آن را شبیه‌سازی کند.» (Lungarella & Iida & Bongard & Pfeifer, 2007, p. 2.)

کارگاه کالج دارموث به هیچ موفّقیت خاصّی ختم نشد، اما همه مؤلّفه‌های عمده را به شرکت‌کنندگان معرّفی نمود. در واقع آن رویداد را می‌توان زمان تولّد هوش‌مصنوعی دانست؛ زیرا پس از آن بود که تحقیقات متعدّدی در سرتاسر دنیا برای به‌کارگیری سیستم‌های مصنوعی با هدف ادّعایی تقلید و برابری و حتّی فراروی از توانایی‌های فیزیکی و ذهنی انسان، آغاز شد. برای ۲۰ سال پس از این کارگاه، حوزه هوش مصنوعی توسّط همین شرکت‌کنندگان معدود (۱۰ نفر) و دانشجویان و همکارانشان در MIT و IBM و CMU تسخیر شد.

پیشرفت در دهه ۵۰ آن‌چنان سریع بود که برخی پیشگامان اوّلیه احساس کردند بالاخره تسلّط یافتن بر عملکرد مغز انسانی نباید دشوار باشد. در ۱۹۵۶، تحقیقات آلن نول[۲] و هربرت سیمون[۳]، برنامه‌ای را به وجود آورد که نظریه منطق[۴] نامیده می‌شد و در ۱۹۵۷ نیز نسخه جدیدتر آن با نام حلّ مسئله عمومی (GPS)، عرضه گردید که تکنیک‌های جستجوگر بازگشتی را برای حلّ مسائل ریاضیاتی به کار می‌گرفت. نظریه منطق و حلّ مسئله عمومی، توانایی یافتن اثبات‌هایی برای بسیاری قضایا در کارهای اوّلیه برتراند راسل و آلفرد وایتهد در نظریه مجموعه‌ها را دارا بود. (See: Kurtzweil, 1999, pp. 58-59.)

موفّقیت GPS و برنامه‌های متعاقب آن به عنوان مدل‌های شناخت، نول و سیمون را در ۱۹۷۶ به فرمولیزه کردن فرضیه معروف سیستم نماد فیزیکی[۵] واداشت که می‌گوید: "یک سیستم نماد فیزیکی، ابزار لازم و کافی برای رفتار عمومی هوشمندانه را داراست." منظور آن‌ها این بود که هر سیستم (انسان یا ماشین) که هوش را به منصه ظهور می‌رساند، باید به وسیله دستکاری ساختارهای داده که از نمادها ترکیب یافته‌اند، عمل نماید.

---

۱. John McCarthy

۲. Allen Newell

۳. Herbert Simon

۴. Logic Theorist

۵. Physical Symbol System

در IBM، روچستر و همکارانش، بعضی از اوّلین برنامه‌های هوش مصنوعی را ایجاد کردند. آرتور ساموئل[۱] از سال ۱۹۵۲، شروع به نوشتن یک سلسله برنامه‌ها برای انجام بازی چِکِرز کرد که در نهایت، کامپیوتر را قادر به بازی در سطح قوی آماتور کرد. همزمان او ایده‌ای که می‌گفت کامپیوترها فقط می‌توانند کارهایی که به آن‌ها گفته می‌شود را انجام دهند، رد کرد. برنامه وی به سرعت توانست بهتر از خالق خودش، بازی کند. این برنامه در فوریه ۱۹۵۶ در تلویزیون به نمایش درآمد و تأثیری شگرف برجای نهاد.

جان مک‌کارتی از دارموث به MIT کوچ کرد و در سال تاریخی ۱۹۵۸، سه پیشرفت مهم هوش مصنوعی را در آن‌جا رقم زد: او زبان سطح بالای LISP را تعریف کرد که به زبان برنامه‌نویسی غالب در ۳۰ سال آینده‌اش تبدیل شد. وی با این برنامه به ابزارهایی نیاز داشت که دسترسی به منابع محدود و گران آن، مشکلی بزرگ به حساب می‌آمد. مک‌کارتی همچنین در همان سال مقاله‌ای با عنوان "برنامه‌هایی با حسّ عمومی"[۲] منتشر کرد که در آن یک برنامه فرضی با نام Advice Taker را توصیف کرده بود که می‌تواند به مثابه اوّلین سیستم کامل هوش مصنوعی لحاظ شود. او نشان داد که چطور به‌کارگیری برخی اصول ساده می‌تواند برنامه را در راستای تولید یک طراح برای رانندگی توانمند سازد. بدین ترتیب برنامه Advice Taker، اصول محوری استدلال و بازنمایی معرفت را پیاده کرد.

در ۱۹۵۸، ماروین مینسکی نیز به MIT کوچ کرد، اما همکاری نخستین او با مک‌کارتی به طول نینجامید. مک‌کارتی بر بازنمایی و استدلال در منطق صوری تأکید داشت؛ حال آن‌که مینسکی علاقمند به برنامه‌های کاربردی و نهایتاً توسعه یک چشم‌انداز ضدّ منطقی بود. در ۱۹۶۳ مک‌کارتی آزمایشگاه هوش مصنوعی را در استنفورد کلید زد. طرح او برای به‌کارگیری منطق جهت ساخت نهایی Advice Taker، به‌وسیله کشف رابینسون در ۱۹۶۵ به پیشرفت در روش‌های تحلیل رسید: یک الگوریتم کامل اثبات قضیه در منطق مرتبه اوّل.

همچنین کارهای اوّلیه ساخت شبکه‌های نورونی توسّط مک‌کولاک و پیتز رونق گرفت. از همان ابتدا، محقّقان هوش مصنوعی از اظهار پیش‌بینی‌ها درباره موفّقیت‌های آینده این حوزه، ابایی نداشتند. جملات ذیل که به‌وسیله هربرت سیمون در ۱۹۵۷ بیان شده، غالباً نقل می‌شود:

«من نمی‌خواهم شما را شوک‌زده یا غافلگیر کنم. ولی به ساده‌ترین روش می‌توانم این‌طور بگویم که الان در جهان، ماشین‌هایی وجود دارند که فکر می‌کنند، یاد می‌گیرند و خلق می‌کنند. علاوه بر این، توانمندی آن‌ها برای انجام چنین کارهایی رو به افزایش است؛ تا این‌که در یک آینده نزدیک، طیف مسائلی که می‌توانند مدیریت کنند، همسطح با طیف مسائلی باشد که ذهن انسانی، قادر به اداره آن‌هاست.»

---

۱. Arthur Samuel

۲. Programs with Common Sense

سیمون همچنین برخی پیش‌بینی‌های جزئی‌تر هم داشته است: «تا ۱۰ سال آینده، یک کامپیوتر قهرمان شطرنج خواهد شد و یک قضیه ریاضیاتی مهم به وسیله ماشین به اثبات خواهد رسید.» این پیش‌بینی‌ها البته پس از ۴۰ سال – و نه ۱۰ سال – تقریباً یا کاملاً محقّق گردید. اعتماد به نفس سیمون، ناشی از عملکرد امیدوارکننده سیستم‌های هوش مصنوعی اوّلیه در مثال‌های ساده بود. هرچند تقریباً در تمام موارد، این سیستم‌های اوّلیه وقتی با مسائل گسترده‌تر و دشوارتر مواجه می‌شدند، ناکام بودند. (Russell, & Norvig, 2010, p. 21.)

در طول دهه ۶۰، حوزه آکادمیک هوش مصنوعی، کار بر روی دستورالعملی را آغاز کرد که تورینگ برای نیم قرن بعدی تحقیقات علوم کامپیوتر، آماده کرده بود. کارهایی که نتایجی بعضاً دلگرم‌کننده و بعضاً ناامیدکننده به دنبال داشت. برنامه دنیل بابرو[۱] می‌توانست مسائل جبر را از زبان طبیعی حلّ کند و آزمون‌های ریاضیات دبیرستان را به خوبی پشت سر بگذارد. مشابه همین موفّقیت برای برنامه توماس ایونس[۲] گزارش شده که می‌توانست مسائل قیاس هندسی را در آزمون آی‌کیو حل نماید. حوزه سیستم‌های کارشناسی نیز با سیستم دندرال[۳] ساخته ادوارد فیجنبام[۴] آغاز شد که می‌توانست پرسش‌ها درباره ترکیبات شیمیایی را پاسخ گوید و فهم زبان طبیعی هم با سیستم SHRDLU ساخته تری وینوگراد[۵] شروع شد که می‌توانست هر جمله معنادار انگلیسی را تا آن‌جا که درباره بلوک‌های رنگی صحبت می‌کردید، بفهمد.

موفّقیت اوّلیه نول و سیمون نیز با برنامه GPS[۶] دنبال شد که برخلاف برنامه تئوریسین منطقی[۷]، با تقلید از روش‌های حلّ مسئله توسّط انسان، آغاز شده بود. در قالب این برنامه سعی شده بود نظمی که برنامه در چارچوب آن، گونه‌ی خاص و محدودی از مسائل را مدیریت می‌کند، شبیه رویکردی باشد که انسان‌ها با آن رویکرد به همان مسائل نزدیک می‌شوند. بدین ترتیب GPS شاید اوّلین برنامه برای پیاده‌سازی رویکرد «تفکّر شبیه انسان» باشد.

برنامه‌های حل مسئله که از دهه اوّل تحقیقات هوش مصنوعی ناشی می‌شد، یک مکانیسم جستجو با اهداف عمومی بود که سعی می‌کرد گام‌های مقدّماتی استدلال را تلفیق کرده تا به حلّ نهایی و کامل مسئله برسد. چنین رویکردهایی را «روش‌های ضعیف» نامیدند؛ زیرا توان ورود به نمونه مسئله‌های دشوار یا بزرگ را

---

۱. Daniel Bobrow

۲. Thomas Evans

۳ . DENDRAL

۴. Edward Feigenbaum

۵ . Terrry Winograd

۶ . General Problem Solver

۷ . Logic Theorist

نداشتند. جایگزین این روش، به‌کارگیری یک معرفت با دامنه اختصاصی بود تا گام‌های بزرگ‌تر استدلال و مدیریت آسان‌تر مسائل تخصّصی را ممکن سازد. برنامه DENDRAL که در ۱۹۶۹ توسّط بیوکنن[۱] و همکارانش در استنفورد نوشته شد، یک نمونه اوّلیه از این رویکرد بود.

R_1 اوّلین سیستم کارشناسی موفّق بود که جنبه تجاری پیدا کرد و در شرکت تجهیزات دیجیتالی[۲] به کار گرفته شد که تا ۱۹۸۶، سالیانه حدود ۴۰ میلیون دلار برای این شرکت، کاهش هزینه به همراه داشت و تا ۱۹۸۸، گروه هوش مصنوعی شرکت دارای ۴۰ سیستم کارشناسی مستقر شده بود و تعداد بیشتری نیز در راه بود. دیگر تقریباً همه شرکت‌های مهم امریکایی، گروه هوش مصنوعی خودشان را داشتند که در حال استفاده از یا تحقیق بر سیستم‌های کارشناسی بودند.

به طور کلّی صنعت هوش مصنوعی از حجم چند میلیون دلار در سال ۱۹۸۰ به میلیاردها دلار در ۱۹۸۸ رسید که شامل صدها شرکت ساخت سیستم‌های هوشمند، سیستم‌های بینایی، روبات‌ها، و نرم‌افزار و سخت‌افزار مخصوص این اهداف بود. اما به زودی پس از آن دوره‌ای فرارسید که «زمستان هوش مصنوعی» خوانده می‌شود؛ دوره‌ای که کمپانی‌ها به‌دلیل ناکامی در تحقّق وعده‌های خارق‌العاده، به سقوط و افول کشیده شدند. (See: Ibid, pp. 23-24.)

انتقادات گسترده‌ای به برنامه‌های اوّلیه وارد شد که ناتوانی آن‌ها برای انجام واکنش‌های هوشمندانه در محیط‌های متنوّع را هدف می‌گرفت. برخی منتقدان - که قابل توجّه‌ترین آن‌ها فیلسوف اگزیستانسیالیست، هابرت دریفوس[۳] بود - پیش‌بینی کردند که ماشین‌ها هرگز با سطح مهارت انسانی برابری نخواهند کرد. دیگر مشخّص شده بود مسائلی که ما فکر می‌کردیم مشکل باشند، از حلّ قضایای ریاضیاتی و بازی آبرومندانه شطرنج گرفته تا استدلال در حوزه‌هایی مانند شیمی و داروسازی، آسان بودند و کامپیوترهای دهه ۶۰ و ۷۰ با چندهزار فرمان در ثانیه، غالباً می‌توانستند نتایج رضایت‌بخشی را در این زمینه‌ها فراهم آورند. مسئله دشوار اما مهارت‌هایی بودند که هر بچّه پنج‌ساله هم داراست؛ مثل بیان تفاوت بین یک سگ و یک گربه، یا درک یک کارتون انیمیشینی. پیش‌بینی‌های خوشبینانه اوّلیه توسط پیشتازان و پایه‌گذاران هوش مصنوعی، اکنون دیگر با واقعیات این حوزه مطالعاتی، فاصله بسیار زیادی را نشان می‌داد.

دهه ۸۰ شاهد اوّلین موج تجاری‌سازی هوش مصنوعی توسّط کمپانی‌های جدید هوش بود که بسیاری از آن‌ها بر زبان تفسیری قدرتمند اما ذاتاً ناکارآمد LISP تمرکز کرده بودند؛ زبانی که در محافل آکادمیک نیز رواج داشت. ناکامی تجاری LISP و این کمپانی‌ها، سبب واپس‌زدن این موج گردید و نظم‌های تشکیل‌دهنده حوزه هوش مصنوعی، شروع به فروپاشی کرد و کمپانی‌های متمرکز بر فهم زبان طبیعی، بازشناسی سخنان،

---

۱. Buchanan

۲. Digital Equipment Corporation

۳. Hubert Dreyfus

روباتیک، بینایی ماشین و دیگر موضوعاتی که اصالتاً بخشی از حوزه هوش مصنوعی دانسته می‌شد، دیگر حاضر به استفاده از عنوان هوش مصنوعی نبودند.

ناکامی پیش‌بینی‌های هوش مصنوعی، محقّقان این حوزه را در دهه‌های ۷۰ و ۸۰ بی‌اعتبار ساخت و بسیاری از دانشمندان علوم نظری کامپیوتر، هدف تحقیرها قرار گرفتند. به طوری که برنده جایزه تورینگ زوریخ و خالق زبان برنامه‌نویسی پاسکال، نیکولاس ورث[۱]، ابایی نداشت که هوش مصنوعی را به طالع‌بینی[۲] نسبت بدهد! هوش مصنوعی کاربردی در آن دوره، به وسیله سیستم‌های تخصّصی قانون‌محور و برنامه‌نویسی منطقی احاطه شده بود. یعنی علی‌رغم تحقیقات بنیادین سولومونف، تمرکز جدّی آن زمان بر منطق و استنباط قطعی واقعیت‌ها از وقایع پیش‌دانسته بود که در برابر استقراء احتمالاتی فرضیه‌ها از تجربیات قرار می‌گرفت. (Schmidhuber, 2007, p. 31.)

بدین ترتیب در اواسط دهه ۱۹۸۰ حداقل چهار گروه مختلف، الگوریتم یادگیری بازگشت به عقب را بازتولید کردند که اوّلین بار در ۱۹۶۹ پایه‌گذاری شده بود. این الگوریتم برای بسیاری مسائل یادگیری در علوم کامپیوتر و روان‌شناسی، و انتشار گسترده نتایج در مجموعه پردازش توزیع‌شده موازی[۳] به کار گرفته شد که موجب شور و هیجان فراوان گردید. این مدل‌های هوش که «پیوندگرا»[۴] خوانده می‌شوند، به عنوان رقیب مستقیم برای دو مدل نمادی (که توسّط نول و سیمون ارتقاء یافت) و رویکرد منطقی (که به‌وسیله مک‌کارتی و دیگران مطرح شد) تلقّی می‌گردید.

البته شاید واضح به نظر برسد که انسان‌ها نیز در برخی سطوح با دستکاری نمادها کار می‌کنند و کتاب ترنس دیکون[۵] با نام گونه‌های نمادی[۶] نیز که در سال ۱۹۹۷ به چاپ رسید، تأکید داشت که این در واقع، تعریف مشخّصه‌های انسانی است. اما طرفداران پیوندگرایی می‌پرسیدند آیا دستکاری نمادها هیچ نقش تبیینی واقعی در مدل‌های جزئی‌شده شناخت دارد؟ در واقع ابهام پیوندگرایان، در مورد این نبود که انسان‌ها لااقل در برخی کارکردهای شناختی، از علائم و نشانه‌ها استفاده می‌کنند؛ بلکه تردید داشتند که دستکاری این علائم و نمادها، در تبیین کارکردهای شناختی (مثل باور و تمایل و حافظه و ...) ضروری باشد یا حتّی نقش خاصّی در تبیین آن‌ها ایفا نماید. در نتیجه چه بسا بتوان بدون استفاده از نمادها هم کارکردهای مذکور را بازسازی کرد. این سؤال در آن زمان، به نحوی جدّی مطرح شده و بدون پاسخ باقی مانده بود. البته دیدگاه معاصر این است

---

۱. Niklaus Wirth

۲. Astrology

۳. Parallel Distributed Processing

۴. Connectionist

۵. Terrence Deacon

۶. Symbolic Species

که رویکردهای نمادین و پیوندگرا، مکمّل همدیگر - و نه در رقیب با هم– هستند. ولی به هرحال، با جداسازی هوش مصنوعی از علوم کامپیوتر، تحقیق در شبکه‌های نورونی مدرن به دو حوزه تقسیم شد که یکی با خلق ساختار مؤثّر شبکه و الگوریتم‌ها و فهم ویژگی‌های ریاضیاتی آن‌ها سروکار داشت، و دیگری با مدل‌سازی دقیق ویژگی‌های تجربی مجموعه نورون‌های واقعی. (See: Russell, & Norvig, 2010, pp. 24-25.)

امروزه دیگر با نقش مهم و روزافزون ماشین‌های هوشمند در تمامی جنبه‌های زندگی (نظامی، پزشکی، اقتصادی، مالی، سیاسی)، عجیب است که درباره حجم گسترده تأثیرات هوش مصنوعی بر حال و آینده زندگی بشر، تردید داشته باشیم. اما جدا از این حجم وسیع و فزاینده «جنبه‌های کاربردی» هوش مصنوعی، تأمّلات و تردیدها درباره ادّعاها در مورد «جنبه روان‌شناختی» هوش مصنوعی ادامه دارد؛ داعیه‌ها و چشم‌اندازی که نه فقط کارکرد و کارآمدی روبات‌ها در تسهیل و ارتقاء زندگی بشر، بلکه ساخت آگاهی مصنوعی و جنبه‌های درونی و پدیداری تجربیات آگاهانه را نیز نوید می‌دهد که طبق اصطلاح فلسفی، «هوش مصنوعی قوی» خوانده می‌شود. پیش از بررسی این دیدگاه از منظر فلسفی و ملزومات آن، ابتدا رویکردهای تاریخی متفاوت در ساخت هوش مصنوعی را مرور می‌کنیم:

۲. هوش مصنوعی کلاسیک و جدید

۲-۱. هوش کلاسیک یا دیجیتال

رویکرد کلاسیک به هوش مصنوعی (با اسم مستعار GOFAI[۱])، اصول یا قوانین منطقی برای گزاره‌ها را به کار می‌گیرد و به طور کلّی، به عناصر و ساختار گزاره وابسته است. وقتی اصطلاح GOFAI به وسیله جان هاگلند فیلسوف (Haugeland, 1985, p. 112) ابداع شد، تعبیر «از مد افتاده»[۲] دلالت بر این داشت که این سبک از هوش مصنوعی، لغو و جایگزین شده است. یکی از دلایلی که سخن فوق می‌توانست داشته باشد، این بود که پیوندگرایی با پردازش‌های موازی توزیع‌شده[۳] در حال نضج‌گیری بود و به فاصله کمی، برخی فیلسوفان (کلارک، ۱۹۸۹) و ژورنالیست‌ها را نیز مشتاق کرد. هرچند ظاهراً خود هاگلند چیزی درباره این طرح جدید نمی‌دانست؛ زیرا در کتابش حتّی به پیوندگرایی اشاره هم نکرده است. انتقاد او به هوش مصنوعی کلاسیک به

[1] . Good Old-Fashioned Artificial Intelligence

[۲] . Old-Fashioned

[۳] . PDP Connectionism

عنوان طرحی کهنه و از مد افتاده، بر اساس فلسفه پدیدارشناسی‌ای بود که هابرت دریفوس بیست سال قبل مطرح کرده بود.[۱]

هوش مصنوعی کلاسیک، دستورالعمل‌های برنامه‌نویسی شده که با بازنمایی‌های نمادین و صوری کار می‌کنند را به خدمت می‌گیرد. این روش به خوبی برای ماهیت سریال و دودویی کامپیوتر دیجیتال وان نیومن مناسب است. از اواسط دهه ۱۹۵۰ تا اواسط دهه ۱۹۸۰، این روش، رویکرد غالب (و نه منحصر) در هوش مصنوعی بود و کارکردگرایی در فلسفه ذهن نیز با این برنامه‌ها توسعه یافت. در آن روزها دیدگاه پردازش اطلاعات، چشم‌انداز جدیدی را درباره مغز ایجاد کرده بود؛ نورون‌ها یا شبکه‌های نورونی به مثابه دستگاه‌هایی لحاظ می‌شدند که قادر به حفظ و انتقال اطلاعات هستند. جالب است که شیفت از دیدگاه الکتریکی–بیوشیمیایی به دیدگاه پردازش اطلاعات، همزمان با تغییر در سایر حوزه‌های زیست‌شناسی اتّفاق افتاد و به طور خاص، زیست‌شناسی ژنتیک و مولکولی کلید خورد تا بر اطلاعات تأکید کند؛ به‌گونه‌ای که ژنوم را به‌صورت یک برنامه در نظر می‌گرفت که به وسیله ماشین سلولی، تفسیر می‌شود. دیدگاهی که تا مدّت‌ها پس از آن غالب شده است. (See: Maynard-Smith, 2000, pp.177-194.)

یک نماد در gofai، آیتمی در یک زبان صوری (زبان برنامه نویسی) است. نمادهای gofai مانند نمادهای ریاضیات یا منطق، می‌توانند به عنوان ساختارهای صرفاً صوری (و فاقد معنا) لحاظ شوند. اگرچه آن‌ها در صحنه عمل معمولاً توسّط کاربران و در قالب برخی محتواهای معناشناختی خاص (کلامی، عددی، بصری، شنوایی، و ...) تفسیر می‌شوند. اجزاء نمادها می‌توانند با یکدیگر ترکیب شده و مطابق با یک‌سری قوانین صوری برای دستکاری نمادها، نمادهایی پیچیده را شکل بدهند. چیزی که در هنگام اجرای یک برنامه هوش کلاسیک اتفاق می‌افتد، اینست که انواع مختلف نمادها که به شیوه‌های مختلف ساختارمند شده‌اند، ساخته و ذخیره و اصلاح و مقایسه و منتقل می‌شوند. به طور خلاصه، محاسبات GOFAI با ساخت و انتقال داده‌های نمادین سر و کار دارد.

در تحقیقات متقدّم هوش کلاسیک، مسئله در یک ساختار سلسله مراتبی (اهداف اصلی، زیر اهداف، زیر مجموعه زیراهداف و ...) تحلیل می‌شد. هدف نهایی در آغاز مشخّص است و وظیفه برنامه اینست که اختلافات بین حالت فعلی و حالت هدف را کاهش بدهد تا این‌که اختلافی باقی نماند. در این دوره از تحقیقات، سلسله مراتب اهداف، به شدّت دقیق بود و برنامه باید در زمان خاص بر یک هدف تمرکز می‌کرد و فقط در صورتی به

---

[۱]. هاگلند استدلال می‌کرد که انتقاد بیست ساله فیلسوفان نشان داده GOFAI غلط است؛ "غلط" و نه فقط غیر کاربردی یا محدود. زیرا تعریف او از اصطلاح هوش مصنوعی صرفاً روش‌شناسی هوش مصنوعی (محاسبات نمادین) را لحاظ نمی‌کرد، بلکه به برخی ادّعاهای طراحان GOFAI نیز نظر داشت که چشم انداز یک کامپیوتر با برنامه‌ریزی درست را واقعاً هوشمند و بالتبع، دارای هیجانات و آگاهی پدیداری می‌دانستند. هاگلند اعتقاد داشت این تلقّی از هوش مصنوعی، دیدگاهی است که به طرق مختلف ردّ شده است؛ هم توسّط جان سرل و دریفوس و هم توسّط استدلال‌های خودش.

هدفی دیگر منتقل می‌شد که: الف) یک زیر هدف جدید برای رسیدن به هدف باید تنظیم می‌شد، یا ب) هدف اصلی حاصل آمده و برنامه به سوی سطح دیگری از هدف می‌رفت یا در صورت رسیدن به بالاترین سطح اهداف، خارج می‌شد.

روش متأخّر GOFAI اما سلسله مراتب اهداف را درون یک دسته بزرگ از قوانین مستقل منطقی شرط-عمل می‌گنجاند. قوانین منحصر به فرد را عمل‌آورها[۱] می‌نامیدند و برنامه‌های ساخته شده از آن‌ها را سیستم‌های عمل‌آوری می‌خواندند. شرط و عمل درون یک قانون می‌توانند هر کدام یک آیتم جداگانه یا عطفی یا فصلی از آیتم‌های بسیار باشند. (See: Boden, 2014, PP. 112-114.)

طبیعی به نظر می‌رسد که قوانین در «برنامه»ها پیاده‌سازی بشوند؛ در حقیقت فرم "اگر شرایط x برآورده بشود، y را انجام بده و در غیر این‌صورت، z را انجام بده"، می‌تواند به عنوان قانونی توصیف بشود که: "y را انجام بده اگر x، و z را انجام بده اگر نه x". با ادامه دادن این مفاهیم، می‌توان داده‌های ورودی را به مثابه بازنمایی‌های ساختارمند از واقعیت‌ها و برنامه‌ها را نیز شیوه‌های پیاده‌سازی قوانین برای دستکاری این بازنمایی‌ها در نظر گرفت. (Robinson, 2014, P. 94.)

البته قبلاً کامپیوترها می‌توانستند داده‌های ساختاریافته را ذخیره‌سازی کنند و با دستکاری آن‌ها به روشی سیستماتیک و از طریق اجرای دستورالعمل‌های مرتبط، قادر به حل مسائل بودند. ولی این برای تبیین تأثیر علّی بازنمایی‌های ذهنی کافی نبود؛ چگونه این بازنمایی‌ها علّت رفتار و تفکّر خردمندانه می‌شوند؟ تئوری محاسباتی دقیقاً برای مواجهه با این نیاز تئوریک مهم، ارائه گردید که با عنوان تئوری محاسباتی ذهن۲ (CTM) شناخته می‌شود. نظریه‌ای که قطعاً با دیدگاه هوش مصنوعی قوی، ارتباط داشته است.

اوّلین ایده مرکزی CTM این بود که حالات ذهنی التفاتی را با ایجاد یک چرخش محاسباتی در تحلیل راسل درباره جملات التفاتی، تبیین نماید. مثلاً جمله «تام باور دارد که ۷ یک عدد اوّل است» را به عنوان یک گرایش گزاره‌ای[۳] در نظر بگیرید که در بردارنده یک گرایش روانی (در این مثال: باور داشتن) به یک گزاره خاص (در این مثال: که ۷ عدد اوّل است) می‌باشد. خود این فرآیند بازنمایی، از طریق نمادها انجام می‌گیرد. به عبارت دیگر جعبه باور شما، در بردارنده مصداقی از یک ساختار نمادین است که معنا یا «محتوا»ی آن اینست: ۷ عدد اوّل است. بدین ترتیب بازنمایی‌های ذهنی دارای هر دو جنبه نحو/ ساختار و معنا می‌شود؛ بسیار شبیه به جملات زبان طبیعی. این‌ها زبان فکر[۴](LOT) را شکل می‌دهند که بدان Mentalese هم می‌گویند.

---

۱. Productions

2. Computational Theory of Mind

3. Propositional Attitude

۴. Language of Thought

به وسیله دستکاری فرمول‌ها که منحصراً بر مبنای ویژگی‌های نحوی آن‌ها صورت می‌گیرد، امکان استخراج فرمول‌هایی دیگر که به نحوی منطقی از آن‌ها تبعیت می‌کنند، وجود دارد. بدین ترتیب نحو و ساختار می‌تواند معنا و محتوا را منعکس سازد، یا آن‌گونه که هاگلند می‌گوید: «اگر از نحو و ساختار مراقبت نمایید، معنا هم از خودش محافظت خواهد کرد.» (Haugeland, 1985, p. 106.) بر اساس این مدل، فرآیند ذهنی همان توالی بازنمایی‌های ذهنی است که محتوای گزاره‌ای تفکّرات متناظر را اظهار می‌دارد. تأثیر هر بازنمایی ذهنی (کاری که در واقع می‌تواند انجام دهد)، به وسیله ساختار آن مشخّص می‌شود؛ بسیار شبیه به شیوه‌ای که هندسه یک کلید، مشخّص می‌کند چه قفلی را باز خواهد کرد (See: Fodor, 1987.). این دومین ایده مرکزی CTM است. بدین ترتیب ذهن به مثابه یک موتور نحوی[۱] لحاظ می‌شود که موتور معناشناختی را راه می‌اندازد و عملکرد آن – دست‌کم علی‌الاصول – می‌تواند توسّط کامپیوتر بازسازی شود.

یک دیدگاه بسیار نزدیک به ctm و در واقع قوی‌تر از آن، فرضیه سیستم نمادین فیزیکی[۲] (PSSH) است که توسّط نول و سیمون در ۱۹۷۶ مطرح شد. طبق این فرضیه، یک سیستم نمادین فیزیکی «تمام ابزار لازم و کافی برای عملکرد هوش عمومی را داراست» (Newell & Simon, 1976, p. 116.). سیستم نمادین فیزیکی در این فرضیه، ماشینی است که مجموعه‌ای از ساختارهای نمادین را در فاصله‌های زمانی تولید می‌کند. CTM و PSSH، دو نظریه‌ای بودند که در آن زمان، هوش مصنوعی نمادین یا کلاسیک یا آن‌چه که هاگلند، GOFAI می‌نامد را تئوریزه می‌کردند.

هر دو دیدگاه مذکور به عنوان تئوری‌هایی اصیل لحاظ می‌شوند که CTM درباره ذهن انسانی بود و PSSH درباره هوش عمومی. این دو دیدگاه، پارامترها و اهدافی برای اکثر محقّقان هوش مصنوعی در سه دهه اوّل فعالیت این حوزه مطالعاتی فراهم آوردند و همچنان نیز – اگرچه دیگر تنها گزینه‌های موجود نیستند – تأثیری غالب دارند و البته از شکست‌های سنگینی که حاصل هجمه‌های قدرتمند معضلات تجربی و مفهومی رویکرد GOFAI است، رنج می‌برند.

برای فهم بهتر رویکرد کلاسیک یا نمادین، بهتر است با یک مثال پیش برویم. بازنمایی‌های ذهنی یک تفکّر مانند «همه انسان‌ها اخلاقی هستند»، در بردارنده بازنمایی‌های ذهنی از مفاهیم «همه»، «انسان‌ها»، «اخلاقی» و «هستند» است. این اجزاء به گونه‌ای مونتاژ می‌شوند که تفکّر مذکور را شکل بدهند. داستانی شبیه به این در صورتی مقبول خواهد بود که فهرستی از نمادهای اوّلیه در اختیار داشته باشیم که بتواند در نهایت بلوک‌های بازنمایی‌های پیچیده‌تر را بسازند. پرسش محوری در برابر تئوری محاسباتی ذهن اینست که چگونه نمادهای اوّلیه زبان ذهن – درون مغز ما یا درون واحد پردازشگر یک روبات – درباره اشیاء بیرون مغز

---

۱. Syntactic Engine

۲. Physical Symbol System Hypothesis

ما خواهند بود؛ ابژه‌هایی که شاید حتّی وجود خارجی هم نداشته باشند. این مسئله را معضل زمینه‌سازی نماد[۱] هم می‌نامند. (See: Harnad, 1990.) این فقط یک معمّای فلسفی درباره ذهن انسانی یا یک پرسش علمی در روان‌شناسی نیست. بلکه دلالت‌های مستقیم مهندسی برای هوش مصنوعی دارد؛ چرا که یک پاسخ مقبول به آن، ممکن است به روشی برای ساخت یک روبات ختم بشود که برخی از سنگین‌ترین اشکالات علیه CTM را دفع نماید.

طبق برخی دیدگاه‌های فلسفی، چنین روباتی از طریق محاسبات بر روی ساختارهای نمادین صوری، «فکر خواهد کرد». البته با این اوصاف، در جهان واقعی نیز برای آن زمینه‌سازی شده است؛ به نحوی که طبق نظر طرفداران این ایده، می‌توانیم بگوییم (مانند ما) به درک فرا نمادی دست می‌یابد. تعدادی از تئوری‌ها در فلسفه ذهن به منظور پاسخ به اشکال فوق مطرح شده که همگی تحت عنوان «طبیعی‌سازی محتوا» یا «طبیعی‌سازی حیث التفاتی» می‌گنجند. در واقع هدف آنست که یک نظریه فیزیکالیستی در این‌باره فراهم آورند که چطور مصادیق نمادهای زبان ذهن در مغز ما، درباره اشیاء بیرون ما هستند؟ یا (فارغ از دیدگاه محاسباتی ذهن) چطور حالات ذهنی ما می‌توانند معنا را به دست آورند؟

در ادامه به طور خلاصه، دو تلاش فلسفی برجسته برای نیل به چنین هدفی را مرور می‌کنیم: تئوری‌های اطلاعاتی، معناشناسی نقش مفهومی[۲]. لبّ کلام تئوری‌های اطلاعاتی، مفهوم Covariance است؛ یعنی این ایده که اگر کمیت الف به نحوی سیستماتیک با کمیت ب، هم‌تغییری دارد، آن‌گاه الف حامل اطلاعاتی درباره ب است. سرعت‌سنج ماشین به نحوی سیستماتیک با سرعت ماشین، هم‌تغییری دارد و بدین‌ترتیب حامل اطلاعاتی درباره آن است و بر این اساس می‌توانیم سرعت‌سنج را به عنوان یک سیستم التفاتی ببینیم؛ چراکه خواندن آن درباره سرعت ماشین است. همچنین دود را «به معنای» آتش می‌بینیم؛ چرا که حامل اطلاعاتی درباره آتش است. این گونه‌ای از «معنادهی» است که گریس[۳] آن را معنادهی طبیعی می‌نامد؛ مفهومی که بر پیشانی تئوری‌های معناشناختی اطلاعاتی نشسته است. در معناشناسی زبان ذهن نیز ایده مرکزی اینست که معنای یک نماد به وسیله چیزی که مصادیق این نماد با آن به نحوی سیستماتیک و قانون‌مند[۴] هم‌تغییری دارند، تعیین می‌شود.

معناشناسی نقش مفهومی (CSR) از تئوری مشهور ویتکنشتاین در باب معنا با عنوان تئوری کاربرد (Use Theory) الهام می‌گیرد که مطابق آن، معنای یک عبارت زبانی، شیوه‌ای است که کاربران این زبان، عبارت مذکور را به کار می‌گیرند. تز اساسی CRS اینست که معنای نمادهای زبان ذهن به وسیله نقشی که این نمادها

---

۱. Symbol Ggrounding Problem

۲. Conceptual-role Semantics

۳. Grice

۴. Nomologically

در حیاتِ شناختی شخص ایفا می‌کنند، فیکس می‌شود. به طور خاص به وسیله ارتباطی که یک نماد با سایر نمادها و با ادراکات حسّی و با رفتار دارند. بنابراین دیدگاه فوق بسیار شبیه به کارکردگرایی درباره حالات ذهنی است و شباهت بسیاری به معناشناسی عملکردی[۱] زبان برنامه‌نویسی در علوم نظری کامپیوتر دارد. در علوم شناختی، CRS را به عنوان معناشناسی رویه‌ای[۲] می‌شناسند. این دیدگاه ابتدا توسّط جانسون لیرد[۳] مطرح شد که می‌گفت: «زبان‌های هوش مصنوعی که برنامه دستورالعمل‌ها را به کامپیوترها مرتبط می‌سازد، دارای هر دو جنبه نحوی و معناشناختی است. نحو آن‌ها شامل قوانینی برای نوشتن برنامه‌هایی با شکل درست است تا یک کامپیوتر بتواند تفسیر و اجرا نماید. معنای آن‌ها نیز شامل رویه‌هایی است که به کامپیوتر دستور داده شده تا اجرا کند.» (Johnson-Laird, 1977, p. 189.)

اما یکی از مناقشات فلسفی مرتبط با دیدگاه فوق، بحث برون‌گرایی است. CRS معنای یک نماد را به شیوه‌ای کاملاً درونی تبیین می‌نماید که فقط منوط به روابطی است که با دیگر نمادها و حالات (شامل حالات ادراکی و رفتاری) دارد. ولی بنا به دیدگاه برون‌گرایی در باب محتوا[۴] - که دیدگاه پرطرفدار معاصر در فلسفه

---

۱. Operational Semantics

۲. Procedural Semantics

۳. Johnson-Laird

۴. برون‌گرایی درباره محتوا، دیدگاهی است که اعتقاد دارد محتوای حالات التفاتی ذهن، به‌وسیله محیط حال یا گذشته تعیّن می‌یابد. حال آن‌که درون‌گرایی در محتوا، دیدگاهی است که اعتقاد دارد محتوای حالات ذهنی منحصراً به‌وسیله حالات درونی شخص، تعیّن می‌یابد. برون گرایان به یکی از مهم‌ترین تحوّلات درباره محتوای التفاتی که توسّط هیلاری پاتنم در مقاله کلاسیک معنای معنا۴ رقم زده شد، اشاره می‌کنند. البته هدف پاتنم در آن مقاله، معطوف به معنای واژگان دالّ بر انواع طبیعی بود و مستقیماً به مسئله محتوای التفاتی نمی‌پرداخت. ولی بعدها برون گرایی در محتوا هم براساس همین استدلال (برون گرایی در معنا)، مورد بحث و پذیرش وسیع قرار گرفت. پاتنم یک آزمون فکری با استفاده از همزاد زمین مطرح می‌کند؛ زمینی که از هر جهت دقیقاً شبیه زمین واقعی است؛ فقط با این تفاوت که مایع شفّاف و بی‌رنگ و بی‌بویی که از ابرها می‌بارد و دریاها را پر می‌کند و «آب» نامیده می‌شود، دارای ترکیب مولکولی XYZ است و نه H2O. اسکار در زمین واقعی و همزاد اسکار در همزاد زمین، به لحاظ درونی مولکول به مولکول شبیه هم هستند و فرض کنید که حتّی تاریخچه و سوابق علّی یکسانی هم دارند. پاتنم در آن‌جا به‌نحوی مقبول استدلال می‌کند که وقتی همزاد اسکار در همزاد زمین، واژه «آب» را به کار می‌برد، همان معنایی را نمی‌دهد که برای اسکار در زمین واقعی دارد. «آب» برای همزاد اسکار، به معنای مایعی با ترکیب مولکولی XYZ است و برای اسکار به معنای مایعی با ترکیب مولکولی H2O (ولو این‌که هیچکدام از آن‌ها اطلاعی از ترکیب‌های مولکولی نداشته باشند). بنابراین با این‌که تمامی وضعیت‌های درونی اسکار و همزاد اسکار، یکسان و مشترک بود، معنای واژه «آب» برای آن دو متفاوت گردید، و این تفاوت از اختلاف بیرونی و محیطی ناشی می‌شود. وقتی همزاد اسکار باور دارد که «آب، شفّاف است»، باور او درباره همان مایع بیرونی فرضی با ترکیب مولکولی XYZ است، و باور اسکار به این‌که «آب، شفّاف است»، درباره مایع واقعی با ترکیب H2O است. (See: Putnam, 1975.) نتیجه استدلال پاتنم آن است که باورهای ما درباره انواع طبیعی، محتوای وسیع دارند و صرفاً به آن‌چه در درون ذهن‌ها و سرهای ما رخ می‌دهد، وابسته نیستند و نمی‌توانند صرفاً بر حالات فیزیکی – روان‌شناختی

ذهن است - چیزی بیش از این برای معنادهی، مورد نیاز است. حداقل به نظر می‌رسد معنا، نمادها را به جهان خارج متّصل می‌سازد و نه فقط به سایر نمادها. معنای واژه سگ باید سر و کاری با مرجع آن – یعنی سگ‌های واقعی – داشته باشد. چالش‌های پیچیده دیگری نیز از سوی پاتنم و برج اقامه شده که به ترتیب استدلال می‌کنند معنا کارکردی از محیط فیزیکی و اجتماعی‌ای است که شخص در آن واقع شده است. آن‌چه که معناشناسی «نقش مفهومی دو فاکتوری» شناخته می‌شود، در پاسخ به همین اشکالات مطرح شد که تلاش می‌کرد بین محتوای ذهنی محدود و وسیع، تمایز بگذارد. معنای محدود در سر ماست و به شرایط پیرامونی بستگی ندارد، در حالی که معنای وسیع به مرجع‌هایش اتّصال می‌یابد و دارای شرایط صدق است. (See: Arkoudas & Bringsjord, 2014, pp. 61-63.) اتّفاقاً فقدان اتّصال بین زبان ذهن و جهان (یا بین نمادهای کامپیوتر و جهان)، اعتراض اصلی فودور علیه معناشناسی رویه‌ای بود که توسّط جانسون لیرد ارائه می‌شد.

۲-۲. پیوندگرایی در هوش مصنوعی جدید

معضلات فنّی و مفهومی در برابر هوش مصنوعی کلاسیک، با سرخوردگی‌های پس از دوره کوتاه هیجانات در دهه ۸۰ تلفیق گردید و زمینه را برای پس زدن رویکردهای کلاسیک (gofai)، هم در هوش مصنوعی و هم در علوم شناختی فراهم آورد. این رویگردانی تا حدّ زیادی در پیشرفت پیوندگرایی[۱] در دهه هشتاد نمود یافت. بنیان پیوندگرایی البته لااقل از دهه ۴۰ و توسط مک‌کالوچ و پیتز گذاشته شده بود، اما فقط در دهه هشتاد بود که به عنوان یک جایگزین جدّی برای gofai ظهور یافت. این جایگزینی البته بیشتر مدیون تلاش‌های راملهرت[۲] و مک‌کللند[۳] و گروه تحقیقاتی PDP بود. (Rumelhart & McClelland & The PDP Research Group. 1986.)

خلاصه آن‌که مدل‌های پیوندگرایانه – یعنی مدل‌هایی شامل شبکه‌هایی از واحدهای پردازش ساده که از طریق الگوهای متنوّع اتّصال، به هم پیوند دارند – در دهه ۸۰ به عنوان پارادایم اصلی هوش مصنوعی و علوم شناختی و پس از یک دوره رکود در اواخر دهه ۶۰ و دهه ۷۰، احیاء شد. بعد از این نقطه عطف تاریخی در دهه هشتاد، پیوندگرایی و شبکه‌های نورونی، به تکیه‌گاه اصلی هوش مصنوعی و علوم شناختی تبدیل شد که تا به اکنون هم ادامه دارد. امروزه کنفرانس‌ها درباره شبکه‌های نورونی از چشم‌انداز هوش مصنوعی، به طور منظّم برگزار می‌شود و معمولاً مشارکت خوبی را نیز بر می‌انگیزد. در کنفرانس‌های مهم علوم شناختی، کارهای بر محور مدل‌های پیوندگرایانه، بخش عمده‌ای را به خود اختصاص می‌دهند. در بسیاری کنفرانس‌ها و

درونی یکسان‌مان سوپروین (مبتنی) شوند. البته پاتنم خود در برون گرایی معنا متوقّف می‌ماند، اما تایلر بِرج چند سال بعد، به سادگی استدلال وی را به برون گرایی محتوا هم تعمیم داده است. (Burge, 1982)

[1] . Connectionism

[۲]. Rumelhart

[۳]. McClelland

ژورنال‌های مهندسی نیز تحقیقاتی که از مدل‌های شبکه‌های نورونی بهره برده، شایع و رایج هستند و عمومیت و جذابیت این بحث، به یک وضعیت نسبتاً پایدار رسیده است. (See: Ron, 2014, pp. 134-135.)

پیوندگرایی به طور کلّی یک روش برای تصویر و فهم مکانیسم‌ها و فرآیندهای شناخت از طریق ساخت مدل‌هایی است که شبکه‌هایی از واحدهای ساده پردازش شبیه نورون استفاده می‌کند و هرکدام محاسبات ساده متعدّدی را به انجام می‌رسانند. مدل‌های پیوندگرا برای وظایف متفاوتی به کار گرفته شده‌اند؛ مثلاً ادراک ابژه‌ها و رویدادها، تلفّظ متون انگلیسی، ذخیره و بازیابی اطلاعات از حافظه، تولید و فهم زبان، یادگیری مهارت، استدلال و ... . در واقع، ابزار اساسی فنّی و مفهومی در دیدگاه پیوندگرایی، شبکه نورونی است. یک شبکه نورونی شامل تعدادی از گره‌ها (یا واحدها)ست که مانند نورن‌های مغز هستند. هر گره، تعدادی از سیگنال‌های ورودی را دریافت و یک سیگنال خروجی تحویل می‌دهد. گره‌ها با یکدیگر اتّصال دارند؛ به گونه‌ای که خروجی یک گره، ورودی گره دیگر خواهد بود. در واقع، دستگاه‌های پیوندگرا – چه فرآیندهای موازی توزیع‌شده و چه شبکه‌های نورون مصنوعی - نوعاً شامل «واحدها»یی هستند که خروجی آن‌ها به واحدهای دیگر، کارکردی از مجموع بارها (وزن‌ها)ی ورودی است که از واحدهای دیگر دریافت کرده‌اند.

مقادیر ورودی و خروجی نوعاً به وسیله اعداد حقیقی بازنمایی می‌شوند. بنابراین پیوندها دارای بار (وزن) هستند که آن نیز از طریق اعداد حقیقی بازنمایی می‌شود. وزن یک پیوند، تأثیری که یک گره بر خروجی گره دیگر دارد را بازنمایی می‌کند. خروجی هر گره، کارکرد ساده و خطی ورودی‌هاست؛ معمولاً مجموع مقادیر ورودی‌ها محاسبه می‌شود و خروجی‌ای از یک یا صفر تولید می‌کند؛ بسته به این‌که آیا مجموع مذکور از یک آستانه خاص بالاتر رفته یا نه؟ اگر خروجی یک باشد، گره فعّال می‌شود (شلیک می‌کند) و در غیر این‌صورت، غیر فعّال می‌ماند.

به عبارت دیگر، در یک مدل پیوندگرایانه، یک «بازنمایی» غالباً الگویی از فعالیت‌ها در یک‌سری واحدهای پردازش است (هرچند تغییرات و جایگزین‌های بسیاری هم برای آن وجود دارد). پردازش از طریق انتشار فعالیت‌ها در میان واحدهای پردازشگر و اتّصالاتی که بین‌شان برقرار است، انجام می‌شود. آن‌چه واسطه این انتشار فعالیت‌هاست، «بارها»ی متعدّد اتّصال بین هر دو واحد پردازشگر است. یادگیری از طریق تغییرات (معمولاً تدریجیِ) بارهای اتّصال – به عنوان کارکردی از فعالیت در شبکه - اتّفاق می‌افتد.

سیستم‌های نورونی قادر به انواع خاصّی از یادگیری هستند. آن‌ها می‌توانند تمرین کنند که یک کارکرد هدف را محاسبه کنند یا به آن نزدیک بشوند. الگوریتم‌های یادگیری با اهداف عمومی وجود دارند – مانند الگوریتم پردازش پشتیبان[۱] – که با وزن‌های تصادفی آغاز می‌کنند و مکرّراً سیستم را به صورت تمرینی در معرض ورودی‌های مختلف قرار داده و وزن‌ها را تنظیم می‌کند تا این‌که بالاخره خروجی را هرچه نزدیک‌تر به

---

۱. Backpropagation

مقدار صحیح برساند. شبکه‌های نورونی ساخته شده‌اند تا وظایف مختلف شناختی (مانند یادگیری زمان گذشته افعال انگلیسی یا برساختن صحبت از متن مکتوب) را به خوبی به انجام برسانند.

دستگاه‌های پیوندگرا خصوصیات متعدّدی دارند مثلاً آن‌ها قادر به تعمیم هستند؛ بدین معنا که هرگاه یک الگوی ورودی جدید را - که شبیه الگوی تمرین‌شده قبلی است – تمرین می‌کنند، خروجی‌ای تولید خواهند کرد که شبیه خروجی الگوی قبلی باشد. همچنین حافظه‌های آن‌ها دارای محتواهای آدرس‌پذیر است؛ یعنی الگوی ورودی از طریق بارهایی که دارد، مستقیماً خروجی متناظر را تولید می‌کنند و هیچ‌گونه فرآیند «جستجو»ی اطلاعات که وابسته به روابط ورودی – خروجی باشند، وجود ندارد. حال آن‌که در رویکرد دیجیتال، نیازمند فرآیندی جهت محاسبه و رسیدن به خروجی درست بودیم. همچنین الگوهای پیوندگرا با موفقیّت خاصّی تقلیل رتبه می‌دهند؛ یعنی آسیب به بخش‌هایی از واحدهای شبکه، بلافاصله ارتباط ورودی – خروجی را خراب نمی‌کند و عملکرد آن واحد از اتصالات را می‌توانند سایر واحدهای اتصالات شبکه تا حدّی به انجام رسانند و جبران کنند. بنابراین آسیب‌ها در یک روند طیفی و سلسله مراتبی، عملیات را به تدریج ضعیف‌تر و نه کاملاً بی‌فایده می‌سازند. این در حالیست که در کامپیوترهای دیجیتال، خراب‌شدن یک واحد می‌توانست کلّ عملیات را با اختلال مطلق مواجه سازد.

این پیشرفت‌ها سبب انگیزش برخی علاقمندی‌های فلسفی شد؛ چرا که شبیه ویژگی‌های روان‌شناختی ماست و در نتیجه برای برخی محقّقان، این را داشت که مطالعات پیوندگرایانه می‌تواند برخی بصیرت‌ها درباره هوش مصنوعی قوی را فراهم آورد. این دلالت در قالب یک ویژگی (تا حدّی مناقشه برانگیز) دیگر نیز در دستگاه‌های پیوندگرا وجود دارد؛ بدین ترتیب که آن‌ها ظاهراً قادر به فراهم‌آوردن روابط شبه قانونی بین الگوهای ورودی و خروجی هستند بدون آن‌که بازنمایی‌های درونی متناظر از قانون‌ها را داشته باشند.

دیگر مزیت حاصل از خصیصه قابل توجّه فوق - که در واقع مرهون فقدان واحد پردازشگر مرکزی یا هر دستورالعمل کدگذاری صریح است- اینست که رفتار سیستم را متعیّن و ناگزیر نمی‌سازد. فقط تک گره‌ها وجود دارند و هر گره هم فقط مقدار کمی از مجموعه اطلاعات موضعی را داراست. به دلیل همین موضعی‌بودن و پیوندهای همبسته، شبکه‌های نورونی قابلیت افت ظریف را دارند؛ بدین معنی که اگر برخی بخش‌های شبکه دچار آسیب بشود، شبکه به عنوان یک کلّ به کارکرد خود ادامه می‌دهد و البته یک افت عملکرد متناسب با میزان آسیب را نیز خواهد داشت. ولی در مقابل، سیستم‌های دستکاری نمادها در هوش کلاسیک، معمولاً شکننده هستند؛ یک انحراف کوچک از ناحیه رویدادهای برنامه‌نویسی شده می‌تواند به شکست‌های فاجعه‌بار بینجامد. این قبیل شکنندگی در هوش انسانی، نابهنجار به حساب می‌آید. شناخت انسانی – مانند عملکرد شبکه‌های نورونی – یک افت ظریف و مستمرّ را (به جای یک شکست کلّی سنگین) در شرایط ناسازگار، به نمایش می‌گذارند.

دومین خصیصه مهم اینست که هر بازنمایی، توزیع‌شده در کلّ شبکه است؛ بدین معنا که یک بخش از اطلاعات به وسیله ساختارهای نمادین خاص، کدگذاری نمی‌شود؛ بلکه اساساً به عنوان الگویی از فعالیت‌ها در سطح تمامی شبکه بازنمایی می‌شود: شلیک گره‌های مختلف. و «معرفت» کلّی که به وسیله یک شبکه نورونی کدگذاری می‌شود، اساساً در اوزان پیوندهای مختلف واقع شده است: زیرنمادین و به شدّت توزیع‌شده. یک

نتیجه مهم از بازنمایی توزیع‌شده، اینست که پرسش آزاردهنده محتوا که علیه نظریه محاسباتی (CTM) کلاسیک اقامه می‌شد را به پایان می‌رسانند.

این خصیصه‌های جالب توجّه در شبکه‌های نورونی، در ترکیب با این واقعیت که به نظر می‌رسد مقبولیت زیست‌شناختی بیشتری نسبت به کامپیوترهای دیجیتال دارند، موجب استفاده مداوم توسّط بسیاری از دانشمندان علوم شناختی و مهندسان هوش مصنوعی گردید. (See: Arkoudas & Bringsjord, 2014, pp. 71-72.)

اگر بخواهیم فهرستی از ویژگی‌های جالب و منحصر به فرد در مدل‌های پیوندگرا (که مفید هم هستند) را گوشزد کنیم، عبارت خواهند بود از:

- موازی‌بودن[۱]؛ یعنی این واقعیت که ممکن است محاسبات به طور همزمان توسّط واحدها انجام شود و نقاط اتّصال ممکن است اطلاعات را در اکثر این مدل‌ها، به طور همزمان منتشر کنند.
- سازگاری[۲]؛ یعنی توانایی یادگیری درون‌ساخت[۳] در اکثر این مدل‌ها، به عنوان مثال از طریق اصلاح و تعدیل بارها در نقاط اتّصال واحدها.
- افت ظریف عملکرد[۴]؛ یعنی قادر بودن به جلوگیری از شکست‌های فاجعه‌بار در مواجهه با خطاهای پردازش یا ورودی.
- تکمیل خودکار الگوهای جدید یا آشنا، به عنوان ویژگی ذاتی بسیاری از این مدل‌ها.
- تعمیم‌های خودجوش[۵]؛ یعنی توانایی به کارگیری خودجوش دانش موجود در موقعیت‌های مختلف.
- ستبر بودن[۶]؛ یعنی تاب تحمّل ذاتی خطاها که ناشی از پنج ویژگی قبلی و نتیجه ساختار این مدل‌هاست.
- آدرس پذیری محتوا[۷]؛ یعنی توانایی بازیابی اطلاعات براساس دانش جزئی (ناقص) درباره آن محتوا.
- بهینه‌سازی و برآورده‌کردن محدودیت‌ها[۸]؛ یعنی توانایی یافتن راه حل‌های بهینه یا نزدیک به بهینه که قیود متعدّد را برآورده می‌سازد. (Ron, 2014, p. 136.)

---

۱. Parallelism
2. Adaptivity
۳. Built-in Learning
۴. Graceful Degration
۵. Spontaneous Generalization
۶. Robustness
۷. Content Addressability
۸. Optimization and Constraint Satisfaction

البته امروزه با توسعه مدل‌های اوّلیه، مدل‌های پیوندگرای ترکیبی[۱] می‌توانند به عنوان تلفیقی از مدل‌های پیوندگرا و مدل‌های سنّتی نمادین لحاظ شوند. بدین ترتیب مدل‌های مذکور می‌خواهند ما را از بحث قدیمی پیوندگرایی علیه نمادگرایی، به سوی یک اتّحاد هم‌افزا سوق بدهند. طرفداران این دیدگاه اعتقاد دارند که ترکیبی از انواع بازنمایی و پردازش – چه پیوندگرا و چه نمادین – دارای رسایی و قدرت و تأثیر بیشتری است و در نتیجه فایده بیشتری در مدل‌سازی شناختی و کاربردهای صنعتی دارد. چرا که برخی فرآیندها و بازنمایی‌های شناختی به وسیله مدل‌های نمادین و برخی دیگر به وسیله مدل‌های پیوندگرا، بهتر تصویر می‌شوند. بنابراین نیاز به تکثّرگرایی و ترکیب در مدل‌کردن شناخت انسان داریم که طبیعتاً به توسعه مدل‌های پیوندگرای ترکیبی ختم شد.

مزایای مدل‌های پیوندگرا – همان‌طور که اشاره شد - شامل موازی‌گرایی شدید، توانمندی‌های یادگیری، و تاب تحمّل خطاست. مزایای مدل‌های نمادین شامل بازنمایی و پردازش سفت و سخت، سهولت تشخیص مراحل پردازش نمادین، و در نتیجه دقّت پردازش‌هاست. با این مزایای نسبی، ترکیب مدل‌های پیوندگرا و نمادین می‌توان تقریباً به سادگی توجیه شود. اکثر پژوهشگران اعتقاد به ضرورت ترکیب هر دو سوی این دوگانه‌ها دارند؛ چرا که هرکدام یک کارکرد انحصاری را ایفا می‌کنند و در نتیجه اجتناب‌پذیرند. معماری شناخت که هر دو تکنیک پیوندگرایی و نمادگرایی را با هم ترکیب می‌کند، حول و حوش برخی از همین دوگانه‌ها ساختار یافته است. (See: Sun & Slusarz & Terry, 2005.)

ولی با تمام این پیشرفت‌ها، هنوز پرسش‌های مهم فلسفی پیشین، در برابر شبکه‌های پیوندگرا نیز برجا باقی مانده است. دیدگاه سنّتی که توسّط GOFAI پذیرفته و به کار گرفته می‌شد، این بود که هوش انسانی، تعداد بسیاری زیادی از وقایع را ذخیره کرده و سپس برنامه‌ریزی می‌کنند که چطور آن‌ها را استفاده نمایند. حال آن‌که تجربیات انسانی این‌چنین می‌گویند که وقتی کسی دارای سابقه‌ای کافی از تجربیات در یک موضوع خاص باشد، خود به سادگی می‌فهمد که در هر موضعیت، چه باید انجام بدهد. وقتی کسی تجربیاتی کافی برای کارشناس شدن در یک حوزه را داشته باشد، تجربیاتش به گونه‌ای ساختارمند می‌شوند که مستقیماً می‌تواند تجربه نماید چه رویداد و اشیائی و به چه نحو، مربوط به این موقعیت خاص هستند. مثلاً هیدگر و مرلوپونتی خواهند گفت ابژه‌ها برای شخصی که درون یک کانتکست خاص حضور و مشارکت دارد، به صورت ویژگی‌هایی جدا و رها از کانتکست به نظر نمی‌رسند؛ بلکه در قالب اشیائی ظاهر می‌شوند که پاسخ‌ها و واکنش‌های لازم را با جایگاه و معنای خاصّی که یافته‌اند، مشخّص می‌کنند.

به بیان دریفوس، بازیکن خوب شطرنج، به ذخیره یادداشتها مراجعه نمی‌کند تا ارزیابی کند و بسنجد که چه حرکتی را در یک موقعیت خاص باید انجام بدهد. بلکه او بر یک جنبه خاص از موقعیت فعلی تمرکز می‌کند و از آن‌جا تشخیص می‌دهد که چه حرکتی را پیش ببرد. یک استاد بزرگ شطرنج، فقط با نگاه به تخته شطرنج

---

۱. Hybrid Connectionist Models

و براساس تجربیات پیشین، حرکت مقتضی را انجام می‌دهد. انسان‌ها در جهان روزمرّه‌مان، شأن همان استاد شطرنج را دارند. به طور کلّی انسان‌هایی که دارای تجربیاتی وسیع در جهان طبیعی و اجتماعی هستند، یک دریافت مستقیم از این دارند که کارها چطور باید انجام شوند و چه رویدادهایی را باید انتظار داشت. بدین ترتیب آشنایی ما، این توانایی را به ما می‌دهد که به گونه‌ای مربوط واکنش داشته باشیم و از اعمال نامربوط بپرهیزیم، بدون آن‌که براساس بازنمایی‌های واقعیّات رها از کانتکست، طرح و برنامه بریزیم. به همین دلیل است که انسان‌ها بسیار راحت‌تر و ماهرانه‌تر با جهان مواجهه می‌کنند و یاد می‌گیرند بین جنبه‌های متعدّد یک موقعیت، تفاوت و تمایز بگذارند. در حالی که برای پایگاه داده‌های قوانین و واقعیت‌ها، بازیابی آن‌چه که مربوط به موقعیت است، بسیار بیشتر از آن‌چه می‌گویند، دشوار است.

حال به نظر می‌رسد همین معضل معرفت به حس عمومی[1] در رویکرد پیوندگرایی هم اعاده شده و پیشرفت آن را مورد تهدید قرار می‌دهد؛ اتّفاقی که بعینه برای GOFAI افتاد. تمامی طراحان شبکه‌های نورونی چندلایه، توافق دارند که یک شبکه هوشمند باید قادر به تعمیم باشد؛ مثلاً برای یک وظیفه طبقه‌بندی شده، با توجّه به نمونه‌های کافی از ورودی‌های مرتبط با یک خروجی خاص، باید بتواند ورودی‌های دیگر از همین نوع را با همین خروجی مرتبط سازد. ولی «از یک نوع بودن» را چه چیزی مشخّص می‌کند؟ به طور کلّی دستگاهی که نتواند تعمیم‌های ما را یاد بگیرد و تمرین‌های ما را در موقعیت‌های جدید به پیش ببرد، نمی‌تواند «هوشمند» نامیده شود. در واقع، هوشی که توسّط یک شبکه نورونی به نمایش در می‌آید، به وسیله مدل‌ساز انسانی - که شبکه را عملیاتی می‌کند - درون سیستم آن تعبیه شده است. ولی این کافی نیست تا جلوی تعمیمی که از منظر کارآموزی داده‌ها مشروع و از منظر انسانی ناجور و بیجاست را بگیرد.

خود هابرت دریفوس، پس از اشاره به تلاش‌هایی که طراحان شبکه‌های نورونی برای مشخّص کردن شباهت‌ها و تعمیم‌ها داشته‌اند، می‌گوید مسئله اینست که استفاده از معماری شبکه‌های نورونی مشخّص کرده که تعمیم مذکور، هیچ‌گاه پدید نخواهد آمد. حال آن‌که در موقعیت‌های جهان خارج، بخش مهمّی از هوش انسانی، دربردارنده تعمیم به شیوه‌ای مرتبط با کانتکست است. اگر طراح شبکه، آن را محدود به یک سطح از پیش تعریف شده برای پاسخ‌های مرتبط کند، آن شبکه هوش را درون همان کانتکست به نمایش خواهد گذاشت، اما دارای حسّ و عرف عمومی نیست تا او را – مانند هوش انسانی - قادر به سازگاری با دیگر کانتکست‌ها نماید.

برای انجام تعمیم، به شیوه‌ای که انسان‌ها انجام می‌دهند، ساختار شبکه باید به نحوی طراحی شود که به موقعیت‌های مختلف، در قالب آن‌چه که برای انسان‌ها «مرتبط و متناسب» به شمار می‌آید، پاسخ بدهد. این مرتبط و متناسب بودن در مورد انسان‌ها، بر اساس آن‌چه که تجربیات گذشته و حال نشان می‌دهد، استوار می‌شود. اما شبکه‌های نورونی مصنوعی، چنین توانایی‌ای را نشان نمی‌دهند و کسی هم در حال حاضر حتّی

---

[1] . Common Sense

نمی‌تواند حدس و گمانی ارائه بدهد که مغز ما چطور این‌ها را تولید می‌کند. (Dreyfus, 1999, p. xxxviii.) و این مسئله‌ای است که همچنان و همین امروز روی میز باقی مانده است.

همچنین یک مسئله اساسی دیگر با استراتژی هوش مصنوعی از طریق کارورزی تحت نظارت در شبکه‌های پیوندگرا وجود دارد. در GOFAI به خوبی روشن شده بود که هر هوشی که سیستم به نمایش می‌گذارد، صراحتاً توسّط طراح سیستم تعریف و برنامه‌ریزی شده بوده است. سیستم هیچ توانایی یادگیری مستقلی ندارد که دریابد قوانین آموزش داده شده در کدام موقعیت‌ها به کار نمی‌آیند و نیازمند قوانین جدیدی است. شبکه‌های نورونی هم ظاهراً دارای توانایی یادگیری هستند، اما در موقعیت‌های یادگیری تحت نظارت. این در واقع شخص دیگری است که تصمیم می‌گیرد کدام موارد، مثال‌هایی خوب برای تحقّق هوش هستند. آن‌چه که شبکه یاد می‌گیرد، صرفاً توانایی ترسیم این هوش در قالب قوّت پیوندهاست. بنابراین شبکه‌های پیوندگرا نیز – مانند سیستم‌های GOFAI – توانایی شناسایی موقعیت‌هایی که در آن‌ها، آن‌چه آموخته‌اند نامربوط به حساب می‌آید را ندارند. این بستگی به کاربر انسانی دارد که شکست‌ها و ناکارآمدی‌ها را تشخیص بدهد و یا خروجی موقعیت‌هایی که قبلاً شبکه آموخته را بهبود ببخشد یا مواردی جدید فراهم آورد که به اصلاحات در رفتار ختم شود.

سخت‌ترین موقعیت‌ها وقتی پدید می‌آید که محیطی که شبکه در آن قرار دارد، دستخوش تغییراتی ساختاری قرار می‌گیرد. از این منظر، شبکه‌های نورونی تقریباً همان‌قدر وابسته به هوش انسانی هستند که سیستم‌های GOFAI، و خوش‌بینی‌ها و پیش‌بینی‌ها درباره توانمندی یادگیری آن‌ها از منظر دریفوس و بسیاری دیگر از فیلسوفان، تقریباً غیر واقعی است. آن‌چه ما نیاز داریم، سیستمی است که خودش بتواند یاد بگیرد چطور با محیط مواجهه داشته باشد و واکنش‌های خودش را با تغییرات محیط، اصلاح و بهینه کند. (See: Ibid, p. xxxix.)

البته محقّقان برای برآوردن این نیاز، به سراغ رویکردی رفته‌اند که گاه «تقویت یادگیری»[۱] خوانده می‌شود. این رویکرد نسبت به یادگیری تحت نظارت، دو مزیت دارد؛ اوّل آن‌که یادگیری تحت نظارت، نیازمند این است که عمل درست برای هر موقعیت، به دستگاه گفته شده باشد. در حالی که تقویت یادگیری، چنین می‌انگارد که جهان، یک سیگنال تقویتی فراهم می‌آورد که ارزش و فایده یک عمل را نشان می‌دهد. این رویکرد سپس به دنبال کم و زیاد کردن مجموع تقویت‌هایی است که دریافت کرده تا در نهایت به حلّ مسئله منجر شود. سیستم در این روش، به تدریج از تجربه یاد می‌گیرد که رفتار بهینه در موقعیت‌های مختلف، چیست تا به اهداف بلند مدّت خویش دست یابد. بنابراین دستگاه برای یادگیری مواجهه‌های ماهرانه، نیاز به معلّم دانای کلّ ندارد و فقط بازخوردهای محیط را بررسی می‌کند.

---

۱. Reinforcement Learning

دومین مزیت آنست که در یادگیری تحت نظارت، هرگونه تغییر در محیط کارورزی، نیازمند هدایت جدید توسّط کارشناسی است که بداند در محیط تازه باید چه کار کرد. در تقویت یادگیری اما شرایط جدید به طور خودکار به تغییراتی در تقویت می‌انجامد که موجب می‌شود دستگاه به نحوی متناسب خود را سازگار گرداند.

اما با فرض پذیرش این‌که تقویت یادگیری، به نحوی صحیح برخی ذاتیات هوش انسانی مرتبط با یادگیری مواجهه‌های ماهرانه را ترسیم می‌کند، این سؤال به طور طبیعی مطرح می‌شود که آیا کسی می‌تواند دستگاهی بسازد که مانند انسان‌های کارشناس، پدیدارشناسی را به عنوان ماهیت حداقلی و مقبول تقویت یادگیری، (لااقل در حوزه مهارت‌های خاص) به کار گیرد؟ به نظر می‌رسد دست‌کم دو تکمله در عملکردهای فعلی نیاز داریم که هیچ‌کدام از آن‌ها بر اساس دانش فعلی، دست‌یافتنی نیستند؛ نخست آن‌که نیاز است تقویت یادگیری در مسائلی به کار گرفته شود که تعداد موقعیت‌هایی در آن مسائل که می‌تواند مورد مواجهه قرار گیرد، بسیار بیشتر از تعداد موقعیت‌هایی باشد که قبلاً در خلال کارورزی، مواجهه شده است. در نتیجه روش‌هایی به منظور تعیین اعمال و ارزش‌های درست برای موقعیت‌های جدید، نیاز است. دوم آن‌که اگر تقویت یادگیری به دنبال تولید چیزی شبیه هوش انسانی است، دستگاه تقویت یادگیری باید حساسیت عمومی را از طریق مواجهه همراه با مآل‌اندیشی و آینده‌سنجی و از طریق جستجوی فعّال ورودی‌های مرتبط، به نمایش بگذارد.

حال آن‌که رویه‌های جاری، تلاش می‌کنند مرتبط و متناسب بودن رفتارها را از طریق ردگیری مستمرّ آمارهای خاص در خلال یادگیری حدس و خطا، بیاموزند. به عنوان مثال رویه‌ای که چپمن[۱] و کالبلینگ[۲] مطرح کردند، با این پیش‌فرض شروع می‌شود که یک عمل می‌تواند در هر موقعیتی انجام شود و یک ارزش می‌تواند به هر موقعیتی انتساب یابد. در نتیجه، برای هر خصیصه‌ای که احتمال دارد مرتبط و متناسب با یک موقعیت باشد، نحوه عملکرد اشیاء را در هریک از ارزش‌های احتمالی آن خصیصه - که معمولاً دو ارزش حاضر و غایب است - بررسی می‌کنند. اگر براساس این بررسی‌ها، به نظر برسد که ارزش خصیصه بر اعمال یا ارزش‌ها تأثیر قابل توجّهی می‌گذارد، به عنوان یک خصیصه مرتبط معرّفی می‌شود.

ولی همان‌طور که دریفوس بیان کرده، مشکلاتی جدّی با رویه‌ای که ذکر شد وجود دارد. اوّل آن‌که ممکن است یک خصیصه به خودی خود مرتبط و متناسب با رفتار نباشد، اما وقتی با یک یا چند خصیصه دیگر ترکیب می‌شود، مرتبط گردد. برای تدبیر این مسئله، ناگزیر هستیم بررسی‌های مرتبط بودن را بر روی ترکیب ویژگی‌ها انجام دهیم که به حجم سرسام‌آوری از بررسی‌های مهم و محتمل می‌انجامد. دوم آن‌که این رویکرد فرض می‌گیرد که مرتبط بودن یک خصیصه، ویژگی‌ای متعلّق به مجموعه حیطه عملکردی است؛ یعنی آن‌چه سنجیده می‌شود، مربوط بودن خصیصه در تمام موقعیت‌هایی است که با آن مواجه می‌شود. حال آن‌که ممکن است یک خصیصه در برخی موقعیت‌ها مربوط باشد و در دیگر موقعیت‌ها این‌طور نباشد. بنابراین ما نیازمند

---

۱. Chapman

۲. Kaelbling

جمع‌آوری داده‌های مرتبط بودن به‌طور جداگانه برای هر موقعیت خاص هستیم که باز به رشدی سهمگین در کمیت بررسی‌ها می‌انجامد. بنابراین به نظر می‌رسد بررسی‌های آماری، شیوه‌ای کاربردی و عملی برای تعیین مربوط بودن در کامپیوترهای فعلی نیست. با توجّه به اندازه و ساختار مغز، این‌که کسی برای حلّ این مشکل بدون جمع‌آوری حجمی غیرعملی از آمارهای بررسی نمی‌تواند راه حلّی ارائه دهد، تصادفی نیست. مسئله سوم این است که هیچ محدودیتی درباره تعداد خصیصه‌هایی که ممکن است به نحوی تصوّرپذیر در برخی موقعیت‌ها مربوط باشند، وجود ندارد. مشخّصاً ما نمی‌توانیم با تمام خصیصه‌هایی که ممکن است مربوط باشند، آغاز کنیم و بررسی‌های آماری را درباره همه آن‌ها جمع‌آوری نماییم. (See: Ibid, pp. xli-xliv.)

خلاصه آن‌که رویه‌های تقویت یادگیری، رفتار به نحو محرّک و پاسخ تولید می‌کنند؛ به این معنا که ورودی (یک توصیف موقعیت)، مستقیماً به سوی خروجی (ارزش یک عمل یا موقعیت) هدایت می‌شود. مغز اما مشخّصاً حالاتی درونی دارد که در هنگام دریافت ورودی‌ها، به نحو حسّ و حال‌ها[۱] و پیش‌بینی‌ها و آشنایی‌های وابسته به فعالیت‌های فعلی تجربه می‌کنیم. این تجربیات با ورودی فعلی و همچنین با قوّت اتّصالات سیناپسی – که بر اساس تجربیات طولانی گذشته شکل گرفته – مشخّص می‌شوند. رویکرد تقویت یادگیری برای تعیین و تولید رفتار، نیازمند قانونی جهت تعیین پاسخ فوری ناشی از هر عمل ممکن در هر موقعیت ممکن می‌باشد. ولی انسان‌ها چنین قوانینی ندارند و نیاز هم ندارند. نیازها، تمایلات و هیجانات ما، حسّی از مربوط‌بودن رفتارهایمان برایمان فراهم می‌آورد. حال اگر این نیازها و تمایلات و هیجانات نیز به نوبه خود وابسته به توانایی‌ها و آسیب‌پذیری‌های یک بدن بیولوژیکی باشند که درون یک فرهنگ خاص، اجتماعی شده است، آن‌گاه دستگاه‌های تقویت یادگیری هم راه زیادی برای رسیدن به آن دارند.

۲-۳. بدن‌مندی/ مکان‌مندی در هوش مصنوعی جدید

در دهه ۵۰، مغز به‌شکل گونه‌ای از کامپیوترهای قوی توصیف می‌شد و هوش نیز به عنوان حاصل برنامه‌های کامپیوتری نمادمحور تصوّر می‌گردید که جایی در مغز ما قرار گرفته است. این چشم‌انداز با نام «مغز به مثابه یک استعاره کامپوتری» یا «دیدگاه پردازش اطلاعات» نیز شناخته می‌شود. از آن هنگام تا کنون البته رویکردهای کلاسیک یا کامپیوتری، رشته‌هایی بزرگ با کارکردهایی متعدّد پیدا کرده و موفقیّت‌های بسیاری را به دست آورده است. اما امروزه دیگر اهمیّت هوش مصنوعی بدن‌مند و مکان‌مند تقریباً به نحوی جهانی پذیرفته شده و مکرّراً اظهاراتی از این دست که «هیچ مغزی بدون یک بدن وجود ندارد» رواج یافته است.

این البته مرهون سرخوردگی‌هایی بود که نسبت به هوش مصنوعی کلاسیک (GOFAL) در دهه ۸۰ ظهور کرد و نگرش جدید را در چند دهه اخیر به وجود آورد که مفهومی‌سازی هوش به مثابه یک فرآیند صرفاً

---

۱. Moods

کامپیوتری را برای تبیین انواع طبیعی هوش، کافی نمی‌داند؛ انواعی از هوش که به یک ارگانیسم کامل فیزیکی – و به عبارت دیگر، به یک بدن - برای تعامل با جهان خارج احتیاج دارند.

در واقع استدلال شد که به‌جای انجام محاسبات بسیار پیچیده برای کنترل رفتار، ممکن است مؤلفه‌ها و کارکردهای شناخت را از طریق ویژگی‌های فیزیکی و ساختار به کار گرفته شده (مثل بدن ارگانیسم) و تعامل دوسویه عامل با محیط، تأمین کرد و به رفتار بی‌اشکال و به‌جا، دست یافت. به جای تلاش برای ساخت مدل‌های پیچیده جهان سمبولیک - که به هرحال، استخراج قابل اتّکای آن‌ها از سیگنال‌های جهان واقعی، دشوار است – ممکن است بتوان رفتارهای واکنشی ساده را از طریق جنبه‌های مختلف بدن عامل (مانند سیستم حرکتی، سیستم ادراک حسّی، تعامل مکان‌مند و دو سویه با محیط پیرامون، و ...) بازسازی کرد و در نتیجه، رفتارهایی جهش یافته را شاهد بود. ایده‌هایی شبیه به این، اصول عقاید رویکرد رفتارگرا به هوش مصنوعی را تشکیل می‌داد که در سرتاسر آزمایشگاه‌های مهم و برجسته هوش مصنوعی در اوایل دهه ۹۰، به چشم می‌خورد.

این‌چنین بود که به ناگاه پژوهشگران شروع به ساخت روبات‌های شبه حیوان کردند که یادآور تحقیقات سایبرنتیکی بود که تولّد هوش مصنوعی را در دهه ۵۰ کلید زد. تعاملات شدیدی با زیست‌شناسی – و به‌طور خاص با رفتارشناسی حیوانی[۱] و زیست‌شناسی تکاملی – شکل گرفت و این موج فعالیت، دوباره منجر به توسعه ابزارهای نرم‌افزاری و سخت‌افزاری، فهم‌های بنیادین جدید، و نسل اوّل روبات‌های شبه حیوان گردید. عملکردهای جالب توجّه در مسابقات روبوکاپ، زمینه را برای ساخت روبات‌های جهان واقعی بر مبنای دیدگاه رفتارگرا فراهم کرد و اواخر دهه ۹۰ حتّی شاهد ظهور اوّلین روبات‌های انسان‌واره بودیم.

این جریان همچنین کشف کرد که هوش به ندرت در انزوا پدید می‌آید. حیوانات و انسان‌ها در گروه‌هایی زندگی می‌کنند که معرفت عمومی و سیستم‌های ارتباطی در آن‌ها، از طریق فعالیت‌های جمعی ظهور می‌یابد. در نتیجه دهه ۹۰ شاهد تولّد رویکرد چند عاملی[۲] به هوش مصنوعی بود که بر چگونگی تولید و توزیع هوش در رهگذر تعاملات با دیگران تمرکز می‌کرد. (Steels, 2007, pp. 23-24.) بدین ترتیب رویکرد بدن‌مند به هوش، منجر به یک شیفت پارادایمی شد که بر دلالت‌های فیزیکی رفتار قابل انطباق و بدن‌مند، تأکید دارد که در مقابل دیدگاه‌های غیر بدن‌مند چارچوب محاسباتی قرار می‌گیرد. تمرکز پارادایم جدید بر روی سیستم‌هایی است که در جهان واقعی فیزیکی و اجتماعی، کار می‌کنند. با شیفت پارادایمی بنیادین از چشم‌انداز محاسباتی به بدن‌مندی، انواع حوزه‌های تحقیقاتی، موضوعات مهندسی و تئوریک، و رشته‌های درگیر با هوش مصنوعی نیز دستخوش تغییر جوهری شده‌اند. به‌عنوان مثال، تلاش‌های پژوهشی در این حوزه، به سوی فهم مکانیسم‌های سطح پایین‌تر و فرآیند زیرساخت رفتار هوشمندانه، و نیز محقّق‌ساختن فرم‌های بالاتر هوش به وسیله کلیدزدن انواع ساده آن شیفت کرده است. امروزه شناخت و عمل، نتیجه یک ظهور جهشی و رو به رشد

---

۱. Ethology

۲. Multi-Agent

لحاظ می‌شود و نه چیزی که بتواند مستقیماً در یک روبات ساخته (یا برنامه‌ریزی) شود. چنین رویکردی عمیقاً بر بصیرت‌های دریافتی از سیر رشد کودک، تکیه دارد. در واقع، روش‌های طراحی اتوماتیک، ایده‌هایی را به‌کار می‌گیرند که از فرآیند تکامل (در زیست‌شناسی) وام گرفته شده و رشدشناسی[۱] نیز نکات قابل توجهی را برای فهم ماهیت عمومی هوش، فراهم آورده است.

به هرحال، روند پژوهش‌های ساخت و توسعه هوش مصنوعی، به سمت مطالعاتی وسیع‌تر در مورد تعامل بین بدن، مغز و محیط پیش رفت و این‌که چطور هوش از بستر چنین تعاملی ظهور می‌کند. جامعه علمی به تحرّکات[۲] و تصرّفات[۳] و به‌طور کلّی این‌که چطور یک عامل می‌تواند در محیطی دینامیک و پیچیده، عملکردی موفّق داشته باشد، علاقمند است. مفاهیمی مانند «بدن‌مندی»، «خودمختاری»[۴]، «مکان‌مندی»[۵]، «ظهور تعاملی»[۶]، «دینامیک نشانه‌شناسی»[۷]، «رشد» و «تعامل اجتماعی»، جلودار مباحث شده و طیفی از رویکردهای بین رشته‌ای و جدید را در مطالعات هوش طبیعی و مصنوعی، پایه‌ریزی کرده‌اند. در رویکرد کلاسیک به هوش مصنوعی، رشته‌های علوم کامپیوتر، روان‌شناسی، فلسفه، و زبان‌شناسی، نقش محوری و عمده ایفا می‌کردند. حال آن‌که امروزه در رویکرد بدن‌مندی، علوم کامپیوتر و فلسفه به‌وسیله روباتیک، زیست‌مکانیک، علوم مواد، زیست‌شناسی و علوم اعصاب تکمیل می‌شوند. (Lungarella & Iida & Bongard & Pfeifer, 2007, p. 3.)

یکی از مهم‌ترین دلالت‌های رویکرد بدن‌مندی، این بود که هوش فقط جعبه‌ای نشسته درون مغز نیست؛ بلکه در سرتاسر ارگانیسم توزیع شده و نیاز به این دارد که ارگانیسم با محیطش تعامل داشته و آن را ردگیری کند. امروزه عموماً پذیرفته شده که رفتار هر سیستم که به نحوی فیزیکی در یک محیط خاص قرار دارد، صرفاً خروجی یک ساختار کنترل درونی (سیستم عصبی مرکزی) نیست؛ بلکه تحت تأثیر شکل و ویژگی‌های مادّی سیستم نیز واقع می‌شود. در واقع، شواهد روزافزون نشان می‌دهد که ارتباطی مستقیم بین بدن‌مندی و اطلاعات وجود دارد: فعالیت موتور حسّی و شکل بدن، نظم‌هایی در ورودی حسّی و درون ساختار کنترلی ایجاد می‌کند و در نتیجه، فرآیند پردازش اطلاعات را تسهیل می‌نماید.

---

۱. Ontogeny

۲. Locomotion

۳. Manipulation

۴. Autonomy

۵. Situatedness

۶. Interactive Emergence

۷. Semiotic Dynamic

در نتیجه همین یافته‌ها بود که محقّقان هوش مصنوعی مانند راد بروکز[۱] در MIT به این نتیجه رسیدند که سیستم‌هایی که بر تولید عمل از طریق طراحی‌های مفصّل منطقی تکیه می‌کردند، عملیاتی نبودند و مقبولیت شناختی هم نداشتند. آن‌ها عطف توجّه از وظایف نمادین مرتبه بالاتر (مانند استدلال قیاسی) به سمت وظایف حسّی حرکتی مرتبه پایین‌تر – که ظاهراً ساده‌تر بودند – را خواستار شدند؛ وظایفی مانند احساس‌کردن، حرکت کردن، دور زدن، به دست گرفتن، پرهیز از موانع و ... . ولی اعتقاد آن‌ها بر این بود که فقط عامل‌های کاملاً بدن‌مند که بتوانند وظایف مذکور را به انجام رسانند، به عنوان عامل‌های مصنوعی اعتبار حقیقی خواهند داشت.

هوش کلاسیک اهمیت چنین فعالیت‌هایی را نادیده می‌انگاشت یا تقلیل می‌داد. قوای حسّی و حرکتی صرفاً به مثابه مبدّل‌هایی[۲] لحاظ می‌شدند که کارآمدی محیطی آن‌ها تا جایی بود که بازنمایی‌های نمادین جهان را به فرآیندهای مرکزی تفکّر تحویل می‌دادند یا واحدهای پاسخ‌دهنده تحریک را برای ترجمه خروجی این فرآیندها به حرکات بدنی، مستقرّ می‌ساختند.

بروکز و همکارانش استدلال کردند که توانایی‌های بدنی ما فراتر از موارد ابتدایی است و در حقیقت، GOFAL ثابت کرده که در ساختن سیستم‌هایی با توانایی‌های مذکور، ناتوان است. علاوه بر این، آن‌ها می‌گفتند مطالعه توانایی‌های مذکور می‌تواند بینش ارزشمندی به ما بدهد که چطور شناخت مرتبه بالاتر می‌تواند از چنین فعالیت‌هایی ظهور نماید. اگر ما کارهای ظاهراً ساده و روتین سیستم‌های حسّی - حرکتی را بفهمیم، معمّای هوش می‌تواند شروع به حل‌شدن نماید. زبان و استدلال ساده خواهد شد، وقتی بدانیم چطور یک روبات بسازیم که بتواند به‌طور موفّق جهان فیزیکی را هدایت کند. زیرا طبق نظر بروکز، «مؤلفه اصلی هوش یک روبات»، نه از طریق استدلال بلکه «در تعاملات دینامیکی روبات با محیط اطرافش» یافت می‌شود. برنامه‌ای که توسّط بروکز و پیروانش دنبال شد، به عنوان هوش مصنوعی مکان‌مند[۳] شهرت یافت.

نکته کلیدی‌ای که بروکز و تیم وی بر آن تأکید داشتند، این بود که بازنمایی‌های نمادین پیچیده از جهان برای حلّ طیف وسیعی از مسائل، غیر ضروری‌اند. بسیاری مسائل می‌توانند به نحوی مؤثّرتر با کنار گذاشتن بازنمایی‌ها و استفاده از ساختار محیط پیرامونی، حلّ و فصل بشوند؛ ایده‌ای که به صورت این تکّه کلام در آمده: «بهترین بازنمایی جهان، خودش است». تصوّر می‌شد حسّ و تعامل مستمرّ با محیط در یک چرخه بازخورد، رویکرد بسیار امیدوارکننده‌تری است نسبت به ساخت یک مدل ساکن نمادین از محیط و استدلال در مورد آن. بروکز و تیمش، رویکرد خویش را با ساخت روباتی که هربرت نامیده می‌شد، نشان دادند. وظیفه هربرت این بود که در سالن‌های آزمایشگاه هوش مصنوعی MIT چرخ بزند و قوطی‌های خالی نوشابه را

---

۱. Rod Brooks

۲. Transducers

۳. Situated AI

شناسایی و جمع‌آوری کند. هربرت بر ساختار رده‌بندی استوار بود؛ شامل تعدادی قسمت[۱] که هرکدام برای انجام یک وظیفه مشخّص (مانند حرکت به جلو)، اختصاص یافته بودند. در هر زمان خاص، یکی از این قسمت‌ها – بسته به محرّک‌های دریافت‌شده توسّط هربرت - می‌توانست فعّال یا متوقّف شود. سیستم کلّی بر مقدار اندکی (یا هیچ) بازنمایی درونی و دستکاری نمادها، تکیه داشت ولی مدیریت شده بود تا رفتارهایی قدرتمند و شگفت‌انگیز را به نمایش بگذارد.

با عطف توجّه از بازنمایی‌ها و فرآیندهای درونی به سوی رفتار بیرونی و تعامل مستمرّ با محیط، کار بروکز و دیگر محقّقان هوش مصنوعی مکان‌مند، به مثابه یک تغییر جهت معکوس از انقلابِ شناختی و بازگشت به رفتارگرایی قلمداد شده است که برخی از آن به «انقلاب معکوس» یاد می‌کنند. البته چنین ادّعاهایی، اغراق‌آمیزند؛ اکثر محقّقان این حوزه‌ها نمی‌خواهند مشروعیت علمی بازنمودها را در تبیین ذهن، یا فایده‌مندی آن‌ها را به عنوان ابزار مهندسی ردّ کنند. نکته‌ای که تئورسین‌های بدن‌مندی می‌گفتند، این است که هوش مصنوعی، توجّه بیشتری به کانتکست محیط و بدن‌مندی داشته باشد. ولی وجود بازنمایی‌های ذهنی، امروز هم مثل همیشه غیر قابل انکار به نظر می‌رسد.

تفاوت‌های عمده بین GOFAI و هوش مصنوعی بدن‌مند / مکان‌مند

بدن‌مند	کلاسیک
غیر بازنمودی	بازنمودی
اجتماعی	فردگرا
عینی و محسوس[۲]	انتزاعی
وابسته به کانتکست	مستقل از کانتکست
دینامیک	ساکن
کل‌گرا[۳]	جزءگرا[۴]
الهام‌گرفته از زیست‌شناسی	الهام‌گرفته از کامپیوتر
معطوف به عمل	معطوف به تفکّر

(See: Arkoudas & Bringsjord, 2014, pp. 71-72.)

۱. Module

۲. Concrete

۳. Holistic

۴. Atomostic

بنا به استدلال طرفداران بدن‌مندی / مکان‌مندی، دینامیک مغز، بدن و محیط به هم پیوسته است. سیستم‌های نورونی در بدن‌ها تجسّم یافته که آن‌ها نیز به نوبه خود در محیط، واقع شده‌اند و تعاملی قوی بین این مؤلفه‌های سیستمی برقرار است. سیستم دوگانه مغز – بدن، اصطلاحاً «عامل»[۱] نام دارد. چفت‌کردن جریان از محیط به عامل اصطلاحاً «حسّی»[۲] خوانده می‌شود و چفت‌کردن جریان در جهت معکوس (از عامل به محیط) هم اصطلاحاً «حرکتی»[۳] به شمار می‌آید. بنابراین «عامل» برای به کارانداختن مهارت‌ها و توانایی‌های حسّی-حرکتی، در معرض قیود و محدودیت‌های زیست‌پذیری است و اگر این قیود نقض شوند، دیگر عامل نمی‌تواند به عنوان هویّتی مستقل باقی بماند و تعامل رفتاری با محیطش داشته باشد. (Sun, 2014, p. 167.)

البته برخی محقّقان هوش مصنوعی نیز با مشاهده نتایج روبات‌های بدن‌مند، تلفیق رویکرد جدید با محاسبات نمادین و کلاسیک را پیشنهاد دادند. طبق نظر ایشان، پرسش‌های مهم برای طراحی نسل جدید روبات‌ها اینست که ببینیم چطور سیستم‌های برمبنای نمادها می‌توانند در تعاملات ارتباطی بین عامل‌های بدن‌مند، ظهور نمایند. با وجود این‌که هوش مصنوعی نمادین اوّلیه، بدون شک تأکید بیش از حدّی بر نمادها داشت، اما طبق دیدگاه این گروه از محقّقان، هوش مصنوعی بدن‌مند نیز آن را به کلّی نادیده می‌گیرد و «نوزاد را همراه با آب تشت حمام، به دور می‌ریزد»! در حالی که به اعتقاد ایشان، دیدگاهی متعادل‌تر و ترکیبی، امکان‌پذیر و ضروری است. (Steels, 2007, p. 25.)

در واقع عملاً بسیاری از پژوهشگران کنونی هوش مصنوعی بر روی روبات‌ها در محیط‌های فیزیکی واقعی تمرکز کرده‌اند و در عین حال، امتداد هوش مصنوعی به قلمرو بدن فیزیکی را گامی بیرون از فرمالیسم (هوش نمادین) تلقّی نمی‌کنند. بلکه تئوری ریاضیاتی هوش مصنوعی بدن‌مند، علوم کامپیوتر نظری را با تئوری قدیمی احتمالات، ترکیب می‌کند تا رفتار بهینه را برای عامل‌های بدن‌مند ومکان‌مند - که در محیط‌های ناشناخته اما قابل آموزش زندگی می‌کنند - استخراج نماید. (see: Schmidhuber, 2007, p. 33.)

هرچند با وجود اهمیت ایده‌هایی که درباره بدن‌مندی در مقالات هوش مصنوعی ارائه شده، همه نویسندگان این مقالات قبول ندارند که بدن‌مندی در آگاهی ماشین، محوریت دارد. اندیشه دیگر این است که یک ماشین آگاه، علاوه بر بدن‌مندی یا شاید به جای آن، نیازمند یک «عالَم درونی» یا سابجکتیویته است که غالباً در تحقیقات هوش مصنوعی، مترادف با «تخیّل داشتن»، لحاظ می‌شود. البته سابجکتیویته و منظر اوّل شخص در تحقیقات فلسفی در مورد آگاهی، ابعادی به مراتب وسیع‌تر و بعضاً مهم‌تر از «تخیّل داشتن» دارد؛ بلکه حسّ درونی از تجربیات پدیداری که نوعی اختصاص و «برای من بودن»[۴] را به همراه دارد، ویژگی اساسی آگاهی

---

۱. Agent

۲. Sensory

۳. Motor

4. For-me-ness

پدیداری است که قابل درک و تبیین از منظر سوم شخص نیست. این یک چالش اساسی در هر نوع مطالعه علمی آگاهی به شمار می‌رود.

اما به هرحال توافقی نسبی در مورد محوریت تصوّر و تخیّل[۱] برای پروژه آگاهی ماشین، در بین مهندسین هوش مصنوعی وجود دارد. در نتیجه پرسش‌هایی که در نسبت با این موضوع مطرح شده - و البته همچنان باقی مانده – این است که تخیّل و تصوّر از چه مؤلفه‌هایی تشکیل شده و چطور با آگاهی ارتباط دارد؟

یکی از تلاش‌های مفصّل برای استدلال راجع به این‌که یک روبات آگاه باید چگونه دارای یک عالَم درونی باشد، در کار هسلو[۲] و جیرنهد[۳] دیده می‌شود. رویکرد آن‌ها بر فرضیه شبیه‌سازی هسلو تکیه دارد که استدلال می‌کند معنای این جمله که می‌گوییم یک عامل، دارای عالَمی درونی است، آنست که می‌تواند تعاملاتش با جهان خارج را تا حدّی کافی شبیه‌سازی کند. این دو اعتقاد دارند می‌توان استدلال آورد که روبات ایده‌آل که توانایی استفاده از ظرفیت‌های شبیه‌سازی را جهت تصوّر تعاملاتش با جهان خارج دارد، حیات درونی و بنابراین یک آگاهی ابتدایی هم دارد.

کریسلی[۴] و پارث‌مور[۵] چشم‌انداز دیگری در مورد نقش محوری تصوّر در آگاهی ماشین، ارائه می‌دهند. برنامه این‌دو (با عنوان پدیدارشناسی نحوی[۶]) شاید بتواند با عنوان متّحد کردن بدن‌مندی و تصوّر هم لحاظ شود. رویکرد ایشان از بدن‌مندی و مکان‌مندی روبات برای مشخّص‌سازی حالات تجربه‌ای بهره می‌گیرد. آن‌ها با اتّخاذ یک رویکرد حسّی حرکتی، آگاهی را در بردارنده توانایی پیش‌بینی یا تصوّر ورودی‌های حسّی فرض می‌کنند که باید برای حرکت به هر سمتی، دریافت بشود.

حداقل سابجکتیویتی در نظریه کیورستین[۷] با عنوان "دیدگاه حسّی حرکتی دینامیک": DSM[۸] اما فقط فقط نیازمند تمرین مهارت‌های حسّی حرکتی ماشین است. او اعتقاد دارد آگاهی ماشین از تمرین غلبه نظم‌های حسّی حرکتی ناشی می‌شود و در صدد است با استدلال‌هایی پیچیده نشان دهد چگونه یک عامل با تمرین معرفت حسّی حرکتی متناسب، دارای یک منظر اوّل شخص و بنابراین، آگاهی از خودش (به عنوان دارنده تجربه) خواهد بود. در حالی‌که روبات هِسلو، علاوه بر آن، تعاملات حسّی حرکتی با جهان خارج را هم شبیه‌سازی

---

۱. Imagination

۲. Hesslow

۳. Jirenhed

۴. Chrisley

۵. Parthemore

۶. Synthetic Phenomenology

۷. Kiverstein

۸. Dynamic Sensory Motor

می‌کند. و در نگاه هولند، خود-شبیه‌سازی جامع[۱]، لازم است. هرچند هیکانن پیشنهاد می‌دهد که نیاز به توسعه دقیق ذهن‌های شبه انسان داریم؛ به‌طور خاص ذهن‌هایی که تصوّرات بازتابی، صحبت درونی، و شکل‌های پیچیده‌ای از خودآگاهی داشته باشند.

با تمام این تلاش‌ها به منظور ارائه یک معیار فنّی و صوری برای بازسازی و ارزیابی آگاهی در ماشین، تردیدهایی جدّی درباره خروجی این پروژه وجود دارد؛ به عنوان مثال برینگسجورد[۲] اعتقاد دارد هنوز هیچ معیار صوری شفّافی برای آگاهی ارائه نشده و او البته شک دارد که اساساً چنین معیاری بتواند فراهم آید. تورنس[۳] هم اکثر محقّقین آگاهی مصنوعی را متهم می‌کند که در تصدیق و اذعان به محوریت پدیدارشناسی (و منظر اوّل شخص از حسّ و تجربه درونی) برای آگاهی، قصور می‌ورزند. (See: Clowes & Torrance & Chrisley, 2007.) در این‌باره توضیح بیشتر در بخش مسائل فلسفی خواهد آمد.

۳. انواع رویکردها به چیستی و غایت هوش مصنوعی

راسل و نورویگ در *آیما (رویکردی مدرن به هوش مصنوعی)* - که یکی از مهم‌ترین و جامع‌ترین منابع آکادمیک در این حوزه محسوب می‌شود – عمده رویکردها درباره چیستی هوش مصنوعی را در چهار تعریف خلاصه کرده‌اند که براساس یک جدول دو بُعدی استخراج شده است. تعاریف ردیف اوّل متناسب با فرآیندهای تفکّر و استدلال لحاظ شده؛ حال آن‌که تعاریف ردیف پایین، معیارهای رفتاری را دنبال می‌کنند. همچنین تعاریف ستون اوّل، موفقیّت هوش مصنوعی را در وفاداری به کارآیی‌های انسانی جستجو می‌کند؛ در حالی که ستون سمت راست، کارآیی ایده‌آل را مدّ نظر دارد. در ادبیات بحث، هر چهار رویکرد به هوش مصنوعی، طرفدارانی با روش‌های مختلف داشته‌اند.

تفکّر عقلانی (ایده‌آل)	تفکّر شبیه انسان
«مطالعه قوای ذهنی با استفاده از مدل‌های محاسباتی. ...» (Charniak, & McDermott, 1985, P. 6.)	«تلاش‌های هیجان‌برانگیز جدید برای ساخت تفکّر کامپیوتری ... *ماشین‌هایی دارای ذهن* به معنای حقیقی و تامّ کلمه. ... ما خودمان نیز کامپیوتر هستیم.

[۱] . Comprehensive Self-Simulation

[۲] . Bringsjord

[۳] . Torrance

مطالعه محاسباتی که ادراک حسّی، استدلال و عمل طبق آن را ممکن می‌سازد. (Winston, 1992, pp. 5-6.)	این ایده که تفکّر و محاسبه، یک چیز است ...» (Haugeland, 1985, p. 2. «هدایت خودکار فعّالیت‌هایی که ما به تفکّر انسانی مربوط می‌دانیم؛ فعّالیت‌هایی مانند تصمیم‌گیری، حلّ مسئله، یادگیری، ...» (Bellman, 1978.)
رفتار عقلانی (ایده‌آل)	**رفتار شبیه انسان**
«هوش محاسباتی، مطالعه بر روی طراحی عامل‌های هوشمند است ... یک عامل هوشمند، سیستمی است که ... مرتبط با شرایط و اهدافش عمل می‌کند، نسبت به تغییر محیط و اهداف، انعطاف‌پذیر است، از تجربیاتش درس می‌گیرد، و انتخاب‌های متناسبی را براساس محدودیت‌های ادراک حسّی و محاسبات، انجام می‌دهد». (Poole, 1998, p. 1.) «هوش مصنوعی به عنوان رفتار هوشمندانه در مصنوعات، لحاظ می‌شود.». (Nilsson, 1998, pp. 1-2.)	«هنر خلق ماشین‌هایی که کارکردهایی را به اجراء می‌گذارند که وقتی به‌وسیله مردم انجام می‌گیرند، نیازمند هوش هستند.» کورزویل در نسخه ۱۹۹۹ عیناً تعریف ریچ را تکرار کرده است: «من تعریف اِلین ریچ[1] از هوش مصنوعی را یادآوری می‌کنم: "تحقیق درباره ساخت کامپیوترهایی است که اعمالی را به انجام می‌رسانند که فعلاً مردم در انجام آن‌ها بهتر هستند"». (Kurzweil, 1999, p. 59.) «مطالعه بر روی این‌که چطور کامپیوترهایی بسازیم که کارهایی انجام دهند که فعلاً انسان‌ها در انجام آن کارها بهتر هستند.» (Rich, E. & Knight, K. 1991.

(Russell & Norvig, 2010, p. 1.)

راسل و نورویگ در خانه بالا سمت راست، "تفکّر ایده‌آل" را قرار داده‌اند و برای نمونه، تعاریف کارنیاک و وینستون را آورده‌اند. هرچند مراجعه به متن کارنیاک، نشان می‌دهد که او از بازسازی برخی عملکردهای ساده و رایج انسانی- مانند بینایی و زبان طبیعی - در سیستم محاسباتی سخن می‌گوید و هدف نهایی تحقیقات هوش مصنوعی را ساخت یک انسان یا حیوان معرّفی می‌نماید. (pp. 6-7.) چنین تعریفی بیشتر در ستون

[1] Elaine Rich

سمت چپ قرار می‌گیرد. همچنین اگرچه نیلسون از تعبیر «رفتار هوشمندانه» استفاده کرده، اما او در ادامه متن خویش تصریح دارد که منظورش از رفتار هوشمندانه، «ادراک حسّی و استدلال و ارتباطات و عملکرد در محیط‌های پیچیده» است. به نظر وی، یکی از اهداف بلندمدّت هوش مصنوعی، ساختن ماشین‌هایی است که بتواند این کارها را به‌خوبی انسان و بلکه احتمالاً بهتر از او انجام دهد. در نتیجه تعریف وی می‌تواند در بخش تعاریف ایده‌آل (و نه شبیه انسان) لحاظ شود؛ اما قراردادن او در ردیف پایین (تعاریف رفتارگرا و نه تفکّرگرا) قابل خدشه به نظر می‌رسد؛ چرا که «استدلال» طبق توضیحات نیلسون، در زمره «رفتار» هوشمندانه قرار گرفته و مانند تعریف وینستون، استدلال در ردیف ادراک حسّی و عملکردهای محیطی، هدف تعریف و ساخت هوش مصنوعی است. بنابراین تفاوت چندانی از این جهت بین دو تعریف نیلسون و وینستون وجود ندارد و تفکیک آن‌ها در دو ردیف تفکّرگرا و رفتارگرا، واضح و تردیدناپذیر نیست. هرچند مناقشات فوق صرفاً از جهت مصداق تعاریف مختلف بوده و اصل تقسیم‌بندی را زیر سؤال نمی‌برد.

به نظر نویسندگان *آیما*، رویکرد انسان‌محور باید در بخشی از مباحث خویش، علمی تجربی باشد که درگیر مشاهدات و فرضیات در باب رفتارهای انسانی است. اما رویکرد عقلانی - که منظور نویسنده از عقلانی، همان ایده‌آل و بدون اشتباه است - درگیر مهندسی و ریاضیات خواهد بود. مروری بر توضیحات بیشتر در مورد چهار رویکرد فوق، به طور خلاصه از این قرار است:

یکم. رویکرد رفتار شبیه انسان (Acting Humanly) یا همان رویکرد آزمون تورینگ:

آلن تورینگ در سال ۱۹۵۰ آزمون معروفش را ارائه کرد تا یک تعریف عملیاتی رضایت‌بخش از هوش پیشنهاد داده باشد. برنامه‌نویسی برای یک کامپیوتر تا چنین آزمونی کاربردی و دقیق را پاس کند، کار بسیار زیادی طلب می‌کند. چنین کامپیوتری نیاز دارد که توانایی‌های ذیل را دارا باشد:

پردازش زبان طبیعی (تا کامپیوتر را قادر به ارتباط به زبان انگلیسی بگرداند)

بازنمایی معرفت (تا کامپیوتر را قادر به ذخیره‌سازی چیزهایی که می‌داند و می‌شنود، بگرداند)

استدلال خودگردان (تا اطلاعات ذخیره‌شده را برای پاسخ به سؤالات و نتیجه‌گیری‌های جدید به کار گیرد)

یادگیری ماشین (تا با شرایط جدید، سازگار شود و الگوها را تشخیص و استخراج نماید).

آزمون تورینگ عمداً از تعامل مستقیم فیزیکی بین قاضی و کامپیوتر، اجتناب می‌کند؛ زیرا در نظر تورینگ و عمده هم‌عصران وی، شبیه‌سازی فیزیکی یک شخص برای هوش، ضرورتی ندارد. هرچند آن‌چه آزمون تورینگ کامل[۱] نامیده می‌شود، شامل سیگنال‌هایی ویدئویی نیز هست تا قاضی بتواند توانایی ادراک حسّی سوبژکتیو را بیازماید. همچنین فرصتی برای عبور دادن برخی اشیاء فیزیکی توسّط قاضی از روزنه به داخل اتاق‌ها تدارک شده است. بنابراین کامپیوتر برای پاس کردن آزمون تورینگ کامل، نیازمند دو توانایی دیگر هم هست:

---

۱. Total Turing Test

بینایی کامپیوتری (برای ادراک حسّی اشیاء)
«روباتیک» (برای شبیه‌سازی اشیاء و حرکت به سمت آن‌ها)

این شش توانایی، مؤلّفه تشکیل دهنده اکثر موارد هوش مصنوعی است. با این‌حال محقّقان هوش مصنوعی، تلاش‌های کمی برای پاس‌کردن آزمون تورینگ به خرج می‌دهند؛ زیرا اعتقاد دارند مطالعه بر روی اصول زیرساختی هوش، اهمیت بیشتری از بازسازی یک نمونه دارد. جستجو به دنبال "پرواز مصنوعی"، زمانی به موفّقیت رسید که برادران رایت و دیگران، تقلید از پرندگان را کنار گذاشتند و شروع به یادگیری درباره آیرودینامیک و به‌کارگیری تونل‌های بادی کردند.

دوم. رویکرد تفکّر شبیه انسان یا رویکرد مدل‌سازی شناختی (Cognitive Modeling):

سه روش برای ورود به کارکرد واقعی ذهن‌های انسانی وجود دارد: از طریق درون‌نگری، از طریق آزمایشات روان‌شناختی (مشاهده یک شخص در حال فعالیت)، و از طریق تصویربرداری مغز (مشاهده مغز در حال فعالیت). وقتی که ما یک تئوری دقیق درباره ذهن داشته باشیم، ممکن است آن تئوری را در قالب یک برنامه کامپیوتری عرضه بداریم. آلن نول و هربرت سیمون – که برنامه حلّ مسئله عمومی (GPS) را نوشتند – تنها به این راضی نبودند که برنامه آن‌ها، مسائل را به درستی حل کند. بلکه آن‌ها بیشتر به دنبال مقایسه گام‌های استدلال در برنامه برای حلّ یک مسئله، با حلّ همان مسئله توسّط عامل انسانی بودند. این حوزه بین رشته‌ای از علوم شناختی، مدل‌های کامپیوتری را از هوش مصنوعی و تکنیک‌های تجربی را از روان‌شناسی را گرد می‌آورد تا تئوری‌هایی آزمون‌پذیر از ذهن انسان بسازد.

سوم. رویکرد تفکّر خردمندانه (ایده‌آل) یا رویکرد قوانین تفکّر:

ارسطو جزو اوّلین کسانی بود که تلاش کرد فکرکردن صحیح را به شکل قانون درآورد؛ یعنی فرآیند غیرقابل تکذیب استدلال را تدوین نماید. قیاس‌های او الگوهایی برای ساختار استدلال فراهم آورد که همیشه نتایج درستی به دنبال داشتند. قرار بود این قوانین تفکّر، عملیات ذهن را مدیریت کنند و مطالعات آن‌ها در این‌باره، حوزه‌ای به نام «منطق» را کلید زد.

تا ۱۹۶۵ برنامه‌هایی ارائه شده بود که علی‌الاصول می‌توانستند هر مسئله حل‌ناشده توصیف‌شده در قالب مفاهیم منطقی را حل کنند. هرچند اگر راه حلّی وجود نداشت، ممکن بود برنامه برای همیشه در یک چرخه گرفتار آید. آن‌چه سنّت منطقی خوانده می‌شود امیدوار است در هوش مصنوعی به برنامه‌هایی دست یابد که سیستم‌های هوشمند را خلق می‌کنند.

اما دو مانع اساسی در مقابل این رویکرد وجود دارد؛ اوّل آن‌که ممکن نیست معرفت بی‌قاعده[۱] را در قالب اصطلاحات صوری که لازمه مفاهیم منطقی است، بیان داریم. به ویژه وقتی معرفت دارای قطعیت کمتر از

[۱] Informal knowledge

۱۰۰٪ باشد. دوم آن‌که تفاوت بزرگی بین حل یک مسئله به‌صورت علی‌الاصول[۱]، با حل آن به‌صورت کاربردی[۲] وجود دارد.

چهارم. رویکرد رفتار خردمندانه (ایده‌آل) یا رویکرد عامل خردمند[۳]:

یک «عامل»، چیزی است که کاری را انجام می‌دهد و برنامه‌های کامپیوتری قطعاً کاری انجام می‌دهند. منتها از عامل‌های کامپیوتری، انتظاری فراتر می‌رود؛ عملیات خودگردان، ادراک محیط پیرامونی، پافشاری و استمرار در یک دوره زمانی طولانی، انطباق با تغییرات، خلق و پیگیری اهداف. یک عامل خردمند (عقلانی)، عاملی است که این کارها را برای دست‌یابی به بهترین نتیجه انجام می‌دهد.

در رویکرد پیشین (تفکّر خردمندانه)، تأکید بر روی استدلال صحیح بود. در حالی که استدلال صحیح آوردن، گاه «بخشی» از یک عامل خردمند بودن را تشکیل می‌دهد و نه همه آن را. در برخی موارد، هیچ چیز قابل اثباتی برای انجام دادن وجود ندارد. همچنین روش‌هایی برای انجام رفتار خردمندانه وجود دارد که با استدلال، سر و کار ندارد. مثلاً عقب‌کشیدن از یک اجاق گاز داغ، یک رفتار واکنشی است که معمولاً موفّقیت‌آمیزتر از رفتاری آهسته‌تر است که پس از بررسی و تأمّل صورت پذیرد.

رویکرد عامل خردمند، دو مزیّت نسبت به سایر رویکردها دارد. اوّل آن‌که جامع‌تر از رویکرد تفکّر خردمندانه است؛ زیرا استدلال صحیح فقط یکی از مکانیسم‌های متنوّع ممکن برای نیل به عقلانیت است. دوم آن‌که نسبت به رویکردهای مبتنی بر رفتار یا تفکّر انسانی، قابلیت بهتری برای توسعه علمی دارد. (See: Russell & Norvig, 2010, pp. 2-4.)

آن‌چه که پایه و اساس ساخت هوش مصنوعی (چه انسان‌گرا و چه ایده‌آل‌گرا) قرار گرفته، محاسبات است. این محاسبات در کامپیوترهای دیجیتال، به یک شیوه و در روبات‌ها و شبکه‌های مصنوعی به شیوه‌ای دیگر پیاده می‌شود. اما به هرحال مغزها و کامپیوترهای دیجیتال، ویژگی‌هایی تاحدّی متفاوت دارند. کامپیوترها یک چرخه زمانی دارند که میلیون‌ها بار سریع‌تر از مغز است. مغز، ذخیره‌سازی و اتّصالات بسیار بیشتری را در مقایسه با کامپیوترهای پیشرفته انجام می‌دهد؛ هرچند که بزرگترین سوپرکامپیوترها ظرفیتی مشابه ظرفیت مغز دارند. کارشناسان، آینده‌ای را پیش‌بینی می‌کنند که کامپیوترها به سطحی فرا انسانی از عملکردهای منحصر به فرد نائل خواهند شد. اما این مقایسه‌های خام، چندان عمیق به نظر نمی‌رسند. حتّی با یک کامپیوتر که دارای ظرفیتی نامحدود باشد، ما همچنان نمی‌دانیم چطور به سطح هوش انسانی یا هوش ایده‌آل – به معنای کامل آن - دست پیدا کنیم. مشابه همین مسئله در مورد روبات‌های بدن‌مند و شبکه‌های پیوندگرا هم وجود دارد.

---

[۱] In principle

[۲] In practice

[۳] Rational agent

۴. هوش مصنوعی قوی و ضعیف

می‌توان بین هوش مصنوعی قوی و ضعیف، با توجّه به اهداف مختلفی که این دو نسخه از هوش مصنوعی به دنبالش هستند، تمایز گذاشت. هوش مصنوعی قوی، در پی خلق اشخاص هوشمند است؛ یعنی ماشین‌هایی که تمام قدرت‌های ذهنی ما ــ از جمله آگاهی پدیداری ــ را داشته باشند. اما هوش مصنوعی ضعیف می‌خواهد نوعی از ماشین‌های پردازش اطلاعات بسازد که "علی‌الظاهر" تمام کارآمدی ذهنی اشخاص انسانی را دارا باشند.

هوش مصنوعی با تئوری‌های محاسباتی ذهن (CTM)، پیوند جدّی دارد. تندروترین نسخه‌های این تئوری‌ها می‌گویند ذهن فقط یک برنامه کامپیوتری است و هیچ چیز دیگری در کار نیست. جان سرل – فیلسوف دانشگاه برکلی - این دیدگاه را هوش مصنوعی قوی می‌نامد و از دیدگاهی که می‌گوید کامپیوتر یک ابزار مفید و کارآمد برای شبیه‌سازی ذهن است، متمایز می‌کند؛ دیدگاه محتاطانه‌تری که آن را هوش مصنوعی ضعیف نام‌گذاری می‌کند. (Searle, 1997, p. 9.) در واقع هوش مصنوعی قوی، تزی است که می‌گوید کامپیوترها با برنامه‌ریزی درست، واقعاً یک ذهن هستند. بدین معنا که کامپیوترها با برنامه‌ریزی درست، می‌توانند حقیقتاً "درک کنند" و دیگر حالات شناختی را نیز داشته باشند. در حالی که ادّعای هوش مصنوعی ضعیف فقط در این حدّ است که کامپیوترها ابزاری مفید برای فرمولیزه‌کردن و آزمایش فرضیات هستند.

بسیاری از دانشمندان و محقّقان علوم شناختی در سازگاری با رویکرد هوش مصنوعی قوی، اعتقاد دارند که رفتار هوشمندانه، با این واقعیت، ایجاد و تقویت می‌شود که ذهن یا مغز، یک ماشین اتوماتیک انتزاعی را پیاده‌سازی می‌کند؛ خواه ماشین تورینگ باشد یا یک برنامه یا یک شبکه نورونی انتزاعی یا ... . جاه‌طلبی هوش مصنوعی – که با دیدگاه هوش مصنوعی قوی (به بیان سرل)، تطابق دارد - بر ادّعایی راجع به کفایت محاسباتی، تکیه می‌کند؛ ادّعایی که می‌گوید یک کلاس خاص از ماشین‌های خودکار وجود دارند که هر پیاده‌سازی و اجرای یک برنامه خاص در آن کلاس از ماشین‌ها را دارای ذهن و آگاهی می‌داند. (Chalmers, 1996, p. 309.)

البته روشن نیست که تمامی محقّقان GOFAI به هوش مصنوعی قوی تعهّد داشته‌اند. ولی قطعاً برخی از آن‌ها چنین بودند. برای مثال آلن نول و هربرت سیمون صراحتاً شکلی از هوش مصنوعی قوی را در تئوری سیستم‌های نمادین فیزیکی مطرح کرده و ادّعا داشته‌اند که حیث التفاتی (یا دربارگی) به وسیله پیاده‌سازی انواع خاصّی از محاسبات صوری در این سیستم‌ها به دست آمده است. این‌ها به عنوان ابزار لازم و کافی برای عملکرد هوش عمومی لحاظ می‌شد. به عبارت دیگر، ذهن (یا ذهن/ مغز) یک سیستم نمادین فیزیکی است.

برنامه هوش مصنوعی نول و سیمون، سیستم‌های محاسبات صوری را همان‌گونه که آلن تورینگ تعریف کرده بود، مشخّص کردند. ولی آن‌ها – مانند روان‌شناسان – علاقمند به محاسباتی بودند که واکنش به محیط و قابل هدایت کردن رفتار به سوی محیط باشد. بنابراین آن‌ها – برخلاف تورینگ – محاسبات را در قالبی علّی تعریف می‌کردند. طبق دیدگاه ایشان، یک نماد مساوی با یک الگوی فیزیکی با تأثیراتی علّی بود. معنای یک نماد نیز یک‌سری تغییرات است که سیستم پردازش اطلاعات را برای پاسخ به برخی ابژه‌ها و فرآیندها (درون یا بیرون سیستم) توانمند می‌سازد. به طور مشابه، مفاهیمی مانند بازنمایی، تفسیر، تعیین، ارجاع، نام‌گذاری،

طرفداری، و دربارگی، به صورت علّی تعریف می‌یافتند. خلاصه آن‌که تئوری معناشناختی آن‌ها معنا را به نحو محاسباتی و محاسبات را به نحو التفاتی ارائه می‌داد. (Boden, 2014, p. 121.)

طبق دیدگاه تورینگ اما کامپیوتری که بتواند آزمون تورینگ را پاس کند، باید به عنوان یک ماشین متفکّر شناخته شود. ظاهراً رفتار زبانی مورد نیاز در آزمون تورینگ، به عنوان قلب شناخت انسانی، لحاظ شده است. ولی برخی مناقشات توسط فیلسوفان ذهن، پیش‌فرض رفتارگرایانه موجود در این آزمون را زیر سؤال برده‌اند. آزمون فکری بلاک (۱۹۸۱) با نام عمّه برتا[۱] می‌خواست با این پیش‌فرض درگیر شود و استدلال نماید که رفتار یک ارگانیسم، تنها چیزی نیست که هوشمندی آن را تعیین می‌کند. بلکه ما همچنین باید این را در نظر بگیریم که یک ارگانیسم، چگونه به هوش دست می‌یابد؛ یعنی سازمان کارکردی درونی سیستم هم باید در نظریه لحاظ شود. این نکته کلیدی کارکردگرایی بود که یکی دیگر از زیرساخت‌های فلسفی هوش مصنوعی را شکل می‌دهد.

یکی دیگر از انتقادات فلسفی به آزمون تورینگ این بود که غیر واقع‌گرایانه است و شاید حتّی با نوع دیدگاه غیر بدن‌مند که به هوش دارد، در مسیر پیشرفت هوش مصنوعی مانع‌تراشی هم بکند. بسیاری از متفکّران اعتقاد دارند دستیابی به مصنوعات غیر بدن‌مند با سطح هوش انسانی، یک آرزوی واهی است که دست‌کم عملاً قابل تحقّق نخواهد بود. بر این اساس، هرناد[۲] تأکید دارد ظرفیت حسّی - حرکتی برای مصنوعاتی که می‌خواهند موفّقیت هوش مصنوعی را رقم بزنند، لازم است. او آزمون تامّ تورینگ[۳] را به عنوان ارتقاء و بهینه‌سازی آزمون تورینگ (TT) پیشنهاد می‌دهد. در شرایطی که یک برنامه کامپیوتری فاقد بدن می‌توانست – دست‌کم علی‌الاصول – آزمون تورینگ را پاس نماید، برای پاس‌کردن آزمون تورینگ تامّ به روبات‌هایی با توانایی عملکرد در محیط فیزیکی نیاز داریم؛ به گونه‌ای که عملکردشان از رفتارهای انسان‌های موجود در جهان فیزیکی، غیر قابل تمایز باشد.

هرچند تا وقتی هوش مصنوعی به صورت حوزه‌ای تخصّصی تعریف بشود که به مهندسی مصنوعات توانمند برای پاس‌کردن آزمون TT و TTT اختصاص دارد، می‌توان با اطمینان گفت که با هوش مصنوعی ضعیف سر و کار داریم. هوش مصنوعی ضعیف، ساخت ماشین‌هایی را هدف می‌گیرد که به نحوی هوشمند «عمل می‌کنند»؛ بدون این‌که موضعی درباره هوشمندی واقعی چنین ماشین‌هایی داشته باشد. (Arkoudas & Bringsjord, 2014, p. 51.)

در واقع کامپیوترهای مدرن به وسیله کدگذاری اطلاعات در قالب دوگانه‌های صفر و یک، ترجمه اطلاعات کدگذاری شده به پالس‌های الکتریکی، و سپس پردازش اطلاعات طبق قوانین برنامه عمل می‌کنند. این یکی

---

۱. Aunt Bertha

۲. Harnad

۳. Totally Turing Test: TTT

از مهم‌ترین دستاوردهای مهم قرن بیستم بود؛ ولی نکته مهم در نگاه فلسفی به موضوع، اینست که فرآیند فوق کاملاً در قالب دستکاری نمادها[۱] تعریف می‌شود که یک نوع عملیات صرفاً نحوی[۲] است. بدین معنا که تنها خصیصه‌های شکلی یا صوری نمادها برای پیاده‌سازی برنامه، موضوعیت دارند و جنبه‌های معنایی[۳] از همین طریق حاصل خواهد شد. (Searle, 1997, p. 10.) به عبارتی، یکی از مسائل مهم فلسفی این است که چطور می‌توان محتوا و معناداری – بلکه درک معنا – را صرفاً با برنامه نویسی براساس شکل نمادها به دست آورد؟

فرضیه کلیدی سرل این بود که سیستم‌های GOFAI در قالب محاسبات انتزاعی تورینگ تعریف شده‌اند. در نتیجه چنین برنامه‌های شامل صرفاً شکل‌هایی فاقد معنا (نمادهای صوری) بودند. این شکل‌ها می‌توانند علی‌الاصول هر تفسیری – بسته به فعّالیت‌های متنوّعی مانند ریاضیات، موسیقی، و ... – داشته باشند و با ملاحظه خود برنامه، انتخاب بین این معانی، دلبخواه خواهد بود. در واقع هر معنایی که بخواهیم به این نمادها بدهیم، تماماً از ناحیه ما اطلاق می‌شود.

حتّی در مورد هوش مصنوعی ضعیف هم سرل شک داشت که مغزها به طور کلّی، محاسباتی صوری را پیاده‌سازی می‌کنند. اگر مغزها این‌گونه کار نمی‌کنند، پس روان‌شناسی محاسباتی بر مبنای GOFAI نمی‌توانست حتّی حیات ذهنی ما را تبیین نماید؛ هرچند که شاید تئوری‌های هوش مصنوعی پیوندگرایانه بتوانند این تبیین را به انجام برسانند. ولی حتّی اگر مغز ما محاسبات صوری را پیاده‌سازی می‌کند، سرل در استدلال اتاق چینی سعی می‌کند نشان دهد که چیزی بیشتر برای حیث التفاتی نیاز است. (Boden, 2014, p. 122.)

البته بسیاری از فیلسوفان که با امکان هوش مصنوعی قوی مخالفت کرده و می‌کنند، در برابر اصل ایده بروز تفکّر و آگاهی به نحو مصنوعی، مقاومتی ندارند. آن‌چه ایشان در مقام ردّ و انکارش هستند، رسیدن به تفکّر و آگاهی ماشینی «از طریق صرفاً برنامه‌نویسی نمادمحور» است. همان‌گونه که خود سرل نیز تصریح دارد، بسیاری دچار اشتباه در فهم مسئله پیرامون هوش مصنوعی قوی شده‌اند و گمان کرده‌اند استدلال علیه هوش مصنوعی قوی، به معنای اینست که: «ماشین‌ها نمی‌توانند فکر کنند»، یا حتّی «کامپیوترها نمی‌توانند فکر کنند». هر دوی این‌ها به بیان سرل، یک فهم اشتباه از استدلال است. حتّی مغز هم یک ماشین بیولوژیکی است و می‌تواند تفکّر نماید. ممکن است مغزهایی مصنوعی ساخته شوند که بتوانند بیندیشند و چه بسا احساسات داشته باشند.

---

۱. Symbol Manipulation

۲. Syntactical

۳. Semantic

سوء تفاهم دیگر آنست که گمان کنیم استدلال علیه هوش مصنوعی قوی، امکان آگاهی (و تجربه پدیداری) به مثابه یک ویژگی نوظهوریافته[۱] را در کامپیوترها انکار می‌کند. در حالی که اگر مغز می‌تواند آگاهی را به عنوان یک ویژگی نوظهوریافته[۲] داشته باشد، چرا بقیه انواع ماشین‌ها نتوانند؟ چرا باید احتمال این‌که مثلاً یک شبکه نورونی مصنوعی – با مختصات و مؤلفه‌های خاص – دارای آگاهی باشد را از بنیان رد کرد؟ ادّعای هوش مصنوعی قوی، این نیست که گونه خاصّی از سخت‌افزارها یا ویژگی‌های سخت‌افزاری می‌تواند سیستم را به ایجاد حالات ذهنی آگاهانه برساند. بلکه هوش مصنوعی قوی ادّعا دارد صِرف پیاده‌سازی برنامه درست در هر سخت‌افزاری، دربردارنده حالات ذهنی خواهد بود. (See: Searle, 1997, pp. 12-13.)

به عبارت دیگر، تز هوش مصنوعی قوی این نیست که کامپیوتر می‌تواند حالت ذهنی تولید کند، یا حالات ذهنی را به مثابه ویژگی‌هایی نوظهوریافته داشته باشد. بلکه می‌گوید صرف اجرای برنامه، فی‌نفسه، حیات ذهنی را تضمین می‌کند و این، تئوری‌ای است که استدلال‌های فلسفی مانند اتاق چینی در صدد ردّ آنند. در واقع حامیان هوش مصنوعی قوی از این تشابه که نورون‌ها هم در یک وضعیت دوگانه (شلیک یا عدم شلیک) عمل می‌کنند، می‌خواهند به این نتیجه برسند که مغز نیز به‌گونه‌ای که گویا صفر و یک‌ها را دستکاری می‌کند، لحاظ شود. بدین ترتیب اگر مغز در قالب کدهای دوگانه عمل می‌کند، بنابراین مغز هم باید یک کامپیوتر دیجیتال باشد. حال آن‌که چند اشکال در این مقایسه وجود دارد، اما همان‌طور که سرل نیز تذکّر داده، مهم‌ترین تفاوت جدّی بین نورون‌ها و نمادهای کامپیوتر اینست که نورون‌ها به‌نحو علّی و از طریق مکانیسم‌های خاص بیولوژیک موجب آگاهی می‌شوند؛ حال آن‌که صفر و یک‌ها انتزاعی محض هستند و تنها قدرت علّی آن‌ها، قدرت اجرا و پیاده‌سازی در سخت‌افزار برای تولید مرحله بعدی برنامه است. در حقیقت نورن‌ها می‌توانند به وسیله یک برنامه کامپیوتری تقلید شوند؛ اما تقلید شلیک‌های نورونی، قدرت نورون‌ها برای ایجاد آگاهی را تضمین نمی‌کند. (Searle, 1997, pp. 59-60.)

نکته‌ای که در پس این اعتراض فلسفی نهفته است، اینست که محاسبات به عنوان خصیصه ذاتی جهان واقع محسوب نمی‌شود؛ بلکه وابسته به مشاهده‌گر (مفسّر) است. چرا که محاسبات در قالب دستکاری نمادها تعریف می‌شود و نماد هم مفهومی فیزیکی یا شیمیایی نیست. یک چیز فقط در صورتی نماد است که به مثابه نماد، «لحاظ یا تلقّی شود». (Searle, 2002, p. 17.) در واقع به بیان سرل، تا آن‌جا که فیزیک کامپیوتر مدّنظر باشد، یک‌سری جریانات بسیار پیچیده الکترونیکی را شاهد هستیم. اما چیزی که این پالس‌های الکتریکی را

---

۱. Emergent Property

۲. ویژگی نوظهوریافته یک سیستم، ویژگی‌ای است که به نحو علّی با رفتار عناصر سیستم، توضیح داده می‌شود؛ اما خود ویژگی هیچیک از عناصر سیستم به تنهایی، نیست و نمی‌تواند به صورت حاصل‌جمع مجموع ویژگی‌های عناصر آن سیستم، تبیین گردد. مثلاً مایع‌بودن (سیّالیت) برای آب؛ رفتار مولکول‌های H_2O، سیالیت را توضیح می‌دهد، ولی تک‌تک مولکول‌ها سیّال یا مایع نیستند.

نمادین می‌سازد، از همان دست واقعیتی است که نشانه‌های جوهر روی کاغذ کتاب را به نماد تبدیل می‌کند. ما این سیستم‌ها را طراحی، برنامه‌ریزی، و تولید کرده‌ایم؛ بنابراین می‌توانیم نشانه‌های مذکور را به مثابه نماد، تلقّی و استفاده کنیم. یک تکّه کاغذ فقط در صورتی پول خواهد بود که مردم، فکر و قبول کنند آن کاغذ، پول است. این واقعیت که کاغذ مذکور از فیبرهای سلولوزی تشکیل یافته، یک واقعیت مستقل از مشاهده‌کننده است و این واقعیت که معادل ۲۰ دلار سرمایه است، وابسته به مشاهده‌گر. به همین ترتیب، وقتی کاغذی روبه‌روی خودتان را مطالعه می‌کنید، نشانه‌های خاصّی از جوهر در برابرتان است. ترکیب شیمیایی این نشانه‌های جوهری، ویژگی ذاتی آن‌هاست، اما این واقعیت که آن‌ها حروف انگلیسی – یا گونه‌ای دیگر از نمادها - هستند، در ارتباط و وابسته به مشاهده‌کننده می‌باشد.

حاصل آن‌که هوش مصنوعی قوی – که به ماتریالیستی بودن خود و به این دیدگاه که مغز را یک ماشین می‌داند، مباهات می‌کند – به قدر کافی ماتریالیستی نیست. مغز در حقیقت یک ماشین است؛ یک ماشین ارگانیک. و فرآیندهای آن نیز – مانند شلیک‌های نورونی – فرآیندهای ماشین ارگانیکی هستند. اما محاسبات یک فرآیند مانند شلیک نورونی یا احتراق درونی نیست؛ بلکه محاسبات یک فرآیند ریاضیاتی انتزاعی است که فقط در نسبت با مشاهده‌گر و تفسیرگر آگاه وجود دارد. مشاهده‌کنندگانی مانند ما روش‌هایی را می‌یابیم که محاسبات را در ماشین‌های الکتریکی برپایه سیلیکون، پیاده‌سازی کنیم. اما این بدان معنا نیست که محاسبات را به شکل چیزی الکتریکی یا شیمیایی در می‌آوریم. (See: Searle, 1997, pp. 15-17.)

۵. به سمت شبیه‌سازی خاستگاه فیزیولوژیک

البته آگاهی در هوش طبیعی، از طریق نورون‌های بیولوژیکی زمینه‌سازی می‌شود. هرچند تشخیص چگونگی این زمینه‌سازی یا هم‌تغییری[۱] بین آگاهی و سطح نوروفیزیولوژیک، کار بسیار دشواری به نظر می‌رسد و هنوز - علی‌رغم تلاش‌ها و پروژه‌های بسیار - جزو معمّاهای حل ناشده علم به شمار می‌رود، ولی به هرحال این میزان تفاوت بین هوش طبیعی و هوش مصنوعی کلاسیک، مسلّم و واضح است که در یک سیستم بیولوژیک شبیه ذهن ما، آگاهی درون یک چارچوب متشکّل از آناتومی، شیمی و انرژی الکتریکی زمینه‌سازی می‌شود. حال آن‌که سیستم‌های کامپیوتری برای مشارکت و بازتولید فرآیند شناختی ما «برنامه‌نویسی» می‌شوند و داده‌های ورودی - براساس توانمندی ما در فهم معنای اطلاعات مختلف - سازمان‌دهی می‌گردند. ما این داده‌ها را با استفاده از الگوریتم‌ها و فرمول‌های منطقی و ریاضیاتی و علمی پردازش می‌کنیم؛ الگوریتم‌ها و فرمول‌هایی

---

۱. Correlation

که برای توصیف درک شناختی ما از جهان بیرون و درون، پیشنهاد شده‌اند. در نهایت، یکپارچه‌سازی حاصل از این برنامه‌ها، باید به شیوه‌ای که برای ما انسان‌ها قابل فهم باشد، نمایش داده شود.

در نتیجه، توانایی خلق سیستم‌هایی با آگاهی مصنوعی – که قابل مقایسه و سازگار با آگاهی بیولوژیکی ما باشد – به وسیله عوامل مختلفی محدود می‌شود:

الف) ما فهم بسیار محدودی از آگاهی خودمان داریم.

ب) سیستم ادراک حسّی ما، اطلاعات را با استفاده از مودالیته‌های حسّی بیولوژیکی در انسان، یکپارچه می‌سازد. ما سیستم‌های هوش مصنوعی را مجبور می‌کنیم با ما درون محدودیت‌های این مودالیته‌های حسّی، مشارکت کنند.

ج) در حال حاضر تعامل با این سیستم‌ها، مقیّد به ظرفیت‌های حرکتی و صوتی ماست.

د) سیستم‌های کامپیوتری از طریق سوئیچ کردن‌های الکترونیکی، قدرت‌های الکتریکی، و بر پایه تراشه‌ها کار می‌کنند؛ یعنی کاملاً متفاوت از شیوه‌های چارچوب سه‌گانه بیولوژی آناتومی و الکتروفیزیولوژیک و شیمی. هرچند هر دو سیستم قابلیت سوئیچ از وضعیت دیجیتال به وضعیت آنالوگ را دارند.

ه) سیستم‌های کامپیوتری با استفاده از سطوح بسیار پیچیده و سنگین پردازش موازی اطلاعات (قابل مقایسه با انسان) کار می‌کنند.

و) سیستم‌های کامپیوتری با به‌کارگیری منطق‌های ریاضیاتی، برنامه‌نویسی می‌شوند. ولی مشخّص نیست که سیستم‌های پردازش بیولوژیک انسانی، بر هیچ شکلی از گونه‌های منطقی استوار باشد.

از این رو در پیشرفت‌های اخیر سیستم‌های هوش مصنوعی - هرچند همچنان از معادل‌های ریاضیاتی و الکترونیکی و فلسفی استفاده می‌شود و برنامه‌نویسی با منطق بشری ما صورت می‌گیرد – اما جایگزینی معدنی برای اجزای بیولوژیک آگاهی در انسان، شبکه‌های مصنوعی عصبی به‌جای نورون‌های سلولی در انسان، قدرت الکتریکی به‌جای نوروشیمیایی، و منطق فازی منعطف که جایگزین دستورالعمل‌های کارکردی براساس پروتئین و DNA در انسان باشند، به‌کار گرفته می‌شوند. (Pagel & Kirshtein, 2017, pp. 29-30.)

این در واقع یک نقطه عطف مهم بعد از بدن‌مندی و حتّی پیوندگرایی است؛ شبکه مصنوعی پیوندها در این رویکرد، هرچه بیشتر به سوی فیزیولوژی انسانی پیش رفته و شبیه‌سازی خاستگاه ظهور آگاهی – یعنی شبکه نورونی سلولی – مدّ نظر قرار گرفته است. روشن است که این تحوّلات به معنای نزدیک شدن در چند گام مهم به سیستم مولّد آگاهی طبیعی است و احتمال و امکان ظهور برخی کیفیات آگاهانه را به نحوی مصنوعی، یادآور می‌شود. به ویژه اگر تئوری حکمت متعالیه در باب حدوث جسمانی نفس و کیفیات نفسانی را بپذیریم که اساساً منشاء و خاستگاهی فیزیولوژیک برای بروز آگاهی ترسیم می‌کند و هیچ دلیلی وجود ندارد که اگر شبکه نورون طبیعی، علّت ایجاد کیفیات آگاهانه و نفس واجد آن‌هاست، این کیفیات از شبکه نورون مصنوعی – با مختصات و مؤلفه‌های خاص – امکان ظهور نداشته باشد.

۶. مسائل فلسفی

به موازات توسعه فناوری‌ها در حوزه هوش مصنوعی، تأملات فلسفی نیز شکل گرفته و به نحوی تعاملی و دو طرفه، تغییراتی در استراتژی‌های ساخت هوش مصنوعی گذاشته و از آن تأثیرتی پذیرفته است که در خلال بحث‌های گذشته به نمونه‌هایی از آن اشاره کردیم. انتقادات دریفوس، بلاک، سرل، فودور، لوکاس، پنروس و ... از جمله این مشارکت‌های فلسفی بود که البته چهار مورد اوّل، از شهرت و تأثیرگذاری بیشتری برخوردارند و ما به ترتیب تاریخی، آن‌ها را ذکر می‌کنیم و اشکال پنجمی را هم جداگانه و به عنوان مهم‌ترین اشکال شرح می‌دهیم:

۶-۱. مسئله حس عمومی یا متناسب بودن رفتار[۱]

انتقاد دریفوس، نخستین مورد از این دست تأملات فلسفی است که اوایل دهه هفتاد با ترکیبی از انتقادات فلسفی و تجربی، چشم‌انداز هوش مصنوعی را به چالش کشید. بار اصلی جنبه تجربی انتقاد او بر این نکته متمرکز بود که محقّقان هوش مصنوعی نمی‌توانند از عهده ارائه خروجی‌های مدّ نظر در موقعیت‌های متنوّع روزمرّه برآیند. بر خلاف پیش‌بینی‌های خوش‌بینانه - و غالباً اغراق‌گونه – اوّلیه، آن‌ها نمی‌توانستند سیستم‌های هوش با اهداف عمومی بسازند. این سیر از انتقادها البته با این عنوان که غیرمعتبر و غیرمنصفانه است، ردّ شد. چرا «غیر معتبر» است؟ چون در بهترین حالت نشان می‌دهند که هوش مصنوعی «هنوز» موفّق نبوده است؛ نه این‌که «هیچ‌گاه» نمی‌تواند موفّق باشد. و چرا «غیر منصفانه» است؟ چون هوش مصنوعی یک حوزه مطالعاتی بسیار جوان بود و پیشرفت‌های انقلابی تکنیکال را نمی‌توان - برخلاف اظهارات مشتاقانه پیشتازان این عرصه - از دوران آغازین آن توقّع داشت.

اما به لحاظ فلسفی دریفوس استدلال می‌کرد که توانایی ما در شناخت جهان و دیگر مردمان، یک نوع مهارت غیر توصیفی از سنخ دانستنِ چگونگی (علم به مهارت)[۲] است که قابل تقسیم به کدگذاری‌های گزاره‌ای در برنامه‌نویسی کلاسیک GOFAI نیست. این مهارت‌ها غیر قابل بیان و پیشا مفهومی‌اند و یک بُعد پدیدارشناسی ضروری دارند که نمی‌تواند به وسیله هیچ سیستم قانون‌محور به تصویر درآید. دریفوس همچنین بر اهمیت ظرفیت‌هایی مانند تخیّل، ایهام، و به‌کارگیری استعاره تأکید می‌کرد که همگی در مقابل تلقّی محاسباتی قرار داشت و دارد.

از این‌ها مهم‌تر، تأکیدی است که دریفوس برای اهمیت «مرتبط بودن» قائل بود؛ اهمیت توانایی انسان‌ها در تشخیص امر ذاتی از غیر ذاتی، و این‌که می‌توانند بدون زحمت براساس ملزومات تجربه و معرفت خویش

---

[۱]. گاه در ادبیات بحث، از این مسئله با عنوان مسئله چارچوب (Frame Problem) نیز یاد می‌شود. اما از آن‌جا که تقریرات مختلفی از مسئله چارچوب ارائه شده که بعضاً با مسئله حسّ عمومی متفاوت است، به عناوین «حسّ عمومی» و «متناسب بودن رفتار»، بسنده کردیم.

[۲]. know-how

در هر موقعیت مرتبط، پیش بروند. او به درستی احساس کرد که قائل‌شدن به چنین توانایی‌ای برای کامپیوترهای دیجیتال، لغزشگاهی در برابر هوش مصنوعی قرار می‌دهد که او معضل زمینه کل‌گرا[۱] نامید. معضل مرتبط و متناسب بودن، همچنان به مثابه یک چالش فنّی کلیدی در برابر هوش مصنوعی – قوی و ضعیف – و نیز علوم شناختی محاسباتی باقی مانده است.

مسئله متناسب و مرتبط بودن در رفتار اجتماعی انسان‌ها، به طور عادی و سریع و با تکیه بر حسّ عمومی، حلّ و فصل می‌شود و آن‌ها در موقعیت‌های متفاوت و متنوّع، رفتار و واکنشی معقول و متناسب انجام می‌دهند، بدون آن‌که نیاز داشته باشند در این فعّالیت‌های روزمرّه‌شان به دنبال پیروی از دستورالعمل‌ها بروند. این انتقاد – که در قالب‌های مختلف از سوی فیلسوفان متعدّد، از ویتکنشتاین و کواین تا دریفوس و سرل و دیگران، مطرح شده - بر تمایز بین توصیف و علیّت، و همچنین بین پیش‌بینی و تبیین دست می‌گذارد. یک دسته از قوانین (یا به نحو منطقی‌تر، یک برنامه کامپیوتری) باید یک پدیده شناختی را به اندازه کافی توصیف نماید؛ بدین ترتیب که قوانین مذکور، یک مدل معتبر از قوانین قابل مشاهده درشت‌بافت مرتبط با آن پدیده شکل بدهند. آن‌ها باید تمامی داده‌های تجربی در دسترس را متناسب‌سازی کرده و پیش‌بینی‌های درست را بسازند. اما این بدان معنا نیست که یک بازنمایی کدگذاری‌شده از قوانین (یا برنامه) درون سر ما وجود دارد که به نحوی علّی درگیر تولید پدیده است. برای مثال، گروهی از قوانین گرامری می‌توانند به نحوی درست، قیود خاص موجود در نحو زبان انگلیسی را توصیف نمایند. ولی این بدان معنا نیست که گویندگان زبان انگلیسی یک کدگذاری از این قوانین درون مغزهایشان دارند که باعث تولید سخن مطابق گ از سوی آن‌ها می‌شود. بنابراین با این‌که گ می‌تواند به درستی رفتار را پیش‌بینی نماید، اما ضرورتاً آن را تبیین نمی‌کند. (see: Arkoudas & Bringsjord, 2014, pp. 64-65.)

حدود یک دهه پس از طرح اشکال از سوی دریفوس، برخی مهندسان هوش مصنوعی هم کم‌کم به درستی زوایای این اشکال پی بردند و بر آن صحّه گذاشتند. به عنوان مثال، داگلاس لنت در سال ۱۹۸۴، می‌نویسد: «در اواسط دهه ۱۹۷۰ و پس از دو دهه پیشرفت با سرعت کم، محقّقان به یک نتیجه اساسی درباره رفتار هوشمندانه رسیدند: این‌گونه رفتار نیازمند حجم عظیمی از دانش است که مردم غالباً آن را به طور مفروض دارا هستند، اما باید به خورد کامپیوترها داده شود. ... حتّی درک آسان‌ترین عبارت‌ها در زبان انگلیسی رایج، مستلزم فهم کانتکست، گوینده و جهان است که بسیار فراتر از ظرفیت برنامه‌های کامپیوتری امروزه است.» (Lenat, 1984, p. 204.) و این ملاحظه او در سال ۱۹۹۱ حتّی شفّاف‌تر هم شده بود: «اکثر تحقیقات جاری هوش مصنوعی که ما در موردشان خوانده‌ایم، قاصر از حلّ این مسئله هستند.» (Lenat & Feigenbaum, 1991, p. 199.)

---

[۱] . Holistic Context

لنت در این مقاله اخیر توضیح می‌دهد که برای فهم جملات در زبان طبیعی، باید از معانی تک‌تک کلمات ابهام‌گشایی کرد، مرجع احتمالی ضمائر را به درستی یافت، کلمات از قلم‌افتاده را همراه با معانی‌شان تشخیص داد و ... . وی در ادامه، مثال جالبی را مطرح می‌کند که در عین سادگی برای فهم و عملکرد هوش انسانی، پیچیدگی‌های آن به منظور شبیه‌سازی در هوش مصنوعی را نشان می‌دهد: این جمله را در نظر بگیرید:

الف) مجسّمه آزادی را در حال پرواز بر روی نیویورک دیدم.

چه کسی در حال پرواز بوده، شما یا مجسّمه؟ ما هیچ سر نخی از زبان انگلیسی برای رفع این ابهام در اختیار نداریم و باید درباره مردم، مجسّمه‌ها، مسافرین هوایی، اندازه بارهایی که به وسیله هواپیما حمل می‌شود، اندازه و مکان مجسّمه آزادی، آسانی یا دشواری دیدن ابژه‌ها از فاصله دور، و بسیاری واقعیت‌ها و تخمین‌های دیگر بدانیم.

حال این دو جمله پشت سر هم را در نظر بگیرید:

ب) جعبه در آغل (pen) است. جوهر در خودکار (pen) است.

واژه pen در جمله اوّل به معنای آغل است و در جمله دوّم، به معنای خودکار. ولی شما چطور این را می‌فهمید؟ این با استفاده از حافظه شما درباره اشیاء جامد و مایع، اندازه اشیاء مختلف، توانمندی شما برای ملاحظه فوری و ناخودآگاه درباره این‌که چرا شخص باید یک جعبه را در هر نوع از pen بگذارد و چرا شخص باید جوهر را درون هر نوع از pen بریزد، و انتخاب تفاسیر مقبول از هرکدام از این‌ها ارتباط دارد. (Ibid, p. 200).

ممکن است برخی تصوّر کنند این قبیل مثال‌ها را می‌توان همچنان با فراهم آوردن یک بانک عظیم داده و ذخیره‌سازی تمامی گزاره‌ها و توضیحات درباره همه موقعیت‌ها و معانی ویژگی‌های تاریخی و جغرافیایی متفاوت و ... پاسخ گفت. اما – فارغ از میزان درستی این تصوّر – هرچه که به سمت کارکردهای تجربه‌ای‌تر برویم (مثلاً کارکرد تربیتی)، این راهکار بیشتر رنگ می‌بازد. مربّی در هر لحظه و در مواجهه با هر رفتار کودک، به فراخور تجربه‌ای که از تعامل با او و سایر کودکان اندوخته و برحسب ادراک درونی که از نوع واکنش کودک در آن لحظه دارد، تصمیم خاصّی می‌گیرد که می‌تواند کاملاً موردی باشد. به نظر می‌رسد بازسازی کردن چنین رفتارهایی صرفاً از طریق دستورالعمل‌های گزاره‌ای، قریب به محال باشد.

البته تلقّی مهندسان هوش مصنوعی در این زمینه، خود وام‌گیرنده از یک فهم سنّتی فلسفی بود. عقلانیت‌گرایان مانند دکارت و لایب‌نیتز، ذهن را با این تعریف در نظر می‌گرفتند که توانایی بازنمایی تمام قلمرو فعالیت‌ها را داراست. طبق این دیدگاه، تمام آن‌چه که ما درباره معرفت عمومی‌مان به چگونگی زندگی در جهان و مواجهه با اشیاء و مردم می‌دانیم، باید در قالب گزاره‌هایی در ذهن بازتاب داده شوند. بازنمودگرایی فرض می‌گیرد که اساس درک روزمرّه ما را سیستمی از باورهای ضمنی شکل می‌دهد. بعداً فیلسوفان التفات‌گرا مانند هوسرل و محاسبه‌گرایان مانند جری فودور و مهندسان هوش مصنوعی، در این فرضیه شریک شدند. مسئله خاص هوش مصنوعی برای بازنمایی تمام این معرفت‌ها در قالب خصیصه‌ها و قوانین «صوری»، پس از این فرضیه رخ نمود که حس عمومی انسان‌ها از «حجم وسیع داده‌هایی از معرفت گزاره‌ای» ناشی می‌شود.

ولی دریفوس در این اشکال، بر این نکته انگشت می‌گذارد که معرفت ما به چگونگی (علم به مهارت) برای مواجهه با اشیاء و حوادث روزمرّه، نمی‌تواند در یک پایگاه داده جامع که ابژه‌ها و ویژگی‌های مختلف را لحاظ می‌کند، به تفصیل درآید. حتّی شاید هیچ دسته واقعیت‌های رها از کانتکست[۱] وجود نداشته باشد که شیوه‌های رفتاری را ترسیم نماید و ما باید فقط از تجربیات وسیع خود بیاموزیم که چطور به هزاران مورد معمول و رایج، پاسخ بدهیم. دریفوس نتیجه می‌گیرد که حسّ عمومی حتّی برای مواجهه با انواع طبیعی (مثل آب) هم مانند انواع اجتماعی (مثل هدیه)، بر پایه معرفت به چگونگی و نه معرفت گزاره‌ای بنا شده است. و این معرفت به چگونگی می‌تواند شیوه‌ای برای ذخیره کردن تجربیات‌مان از جهان باشد که با بازنمایی جهان به شیوه هوش مصنوعی نمادین سر و کار ندارد. (Dreyfus, 1999, p. xxvii.)

آن‌هایی که انتقاد دریفوس به GOFAI را جدّی می‌گیرند، می‌توانند همانند خود وی، برخی تردیدهای مشابه را درباره رویکرد پیوندگرایی هم داشته باشند. کار در حوزه تقویت یادگیری شبکه‌های پیوندگرا بر مسائل کوچکی تمرکز دارد و معلوم نیست که از نتایج آن بتوان برای مدیریت پیچیدگی‌های جهان واقعی سود برد. رفتارها می‌توانند به خوبی تعریف شوند و به آسانی تمایز یابند، ولی هوش باید با رفتارهایی سر و کار داشته باشد که دسته‌بندی مناسب آن‌ها وابسته به شرایط و زمینه[۲] است. برای مثال، رفتار دویدن می‌تواند «فرار» یا «تعقیب» باشد (بسته به این‌که مثلاً یک دونده پشت سر شخص می‌دود یا یک صید پیش روی اوست). چه بسا روال رایج در شبکه‌های نورونی مصنوعی که مستلزم دسته‌بندی‌های ساده رفتارهاست، یک عنصر اساسی هوش را دور می‌زند و نادیده می‌گیرد؛ یعنی یادگیری این که چطور شرایط و زمینه، بر اهمیت و معنای رفتار تأثیر می‌گذارد. (Robinson, 2014, pp. 101-102.)

۶-۲. مسئله آدمک چینی

انتقاد تأثیرگذار دوم، به طور مشخّص کارکردگرایی ماشینی را هدف گرفته بود و توسّط بلاک (۱۹۷۸) در قالب آزمایشی فکری مطرح شد. او از ما می‌خواهد تصوّر کنیم تمام جمعیت چین، یک ذهن انسان را شبیه‌سازی می‌کنند. او چین را از این جهت مثال می‌زند که یک میلیارد جمعیت دارد و تا حدّی می‌تواند کمیت نورون‌های درگیر در مغز ما را شبیه‌سازی نماید. به تمام شهروندان چینی بیسیم‌های دو موجی داده شده که آن‌ها را به روشی درست، با یکدیگر مرتبط می‌سازد. می‌توانیم هر تک شهروند چینی را به عنوان یک نورون یا هر عنصر پایه‌ای مغز در نظر بگیریم. مردم همچنین از طریق بیسیم به یک بدن مصنوعی مرتبط هستند که از طریق آن می‌توانند محرّک‌های حسّی را دریافت کنند و سیگنال‌های خروجی را برای تولید رفتار فیزیکی (مانند بالا بردن دست) به آن برسانند.

---

۱. Context-free

2. Context

طبق کارکردگرایی ماشینی باید نتیجه گرفت که چنانچه انتقال‌های درست با یکدیگر و با ورودی‌ها و خروجی‌ها شبیه‌سازی شود، آن‌ها در مجموع یک ذهن آگاه به شمار خواهند آمد. ولی این نتیجه، اگر نگوییم یاوه‌گویی، خلاف شهود به نظر می‌رسد. سیستم نهایی ارتباطات متناظر خوبی با مغز دارد، اما به نظر نمی‌رسد که حامل هیچ‌گونه احساسات بدنی (درد، خارش، ...) یا باورها و تمایلات باشد. به دلایل مشابه نتیجه می‌شود که در مورد هیچ سیستم هوش مصنوعی محاسباتی خالص هم نمی‌توان گفت که یک ذهن حقیقی دارد.

برخی کارکردگرایان حاضر به پذیرش نتیجه شده‌اند که مغز چینی (یا یک روبات که به نحو مناسب، برنامه‌نویسی شده)، واقعاً دارای محتوای ذهنی حقیقی است؛ برخلاف شهود تعصّب‌آمیز ما درباره مغز که فقط شبکه نورونی را دارای قابلیت حیات ذهنی می‌دانیم. ولی هضم این مطلب، دشوار است و آزمون فکری فوق، خیلی‌ها را متقاعد کرد که کارکردگرایی، به افراط رفته و باید یا ترک و یا به طور قابل ملاحظه‌ای محدود شود. (Arkoudas & Bringsjord, 2014, pp. 65-66.)

شبیه این اشکال را برخی فیلسوفان دیگر – مانند ایلیت سوبر - هم مطرح کرده‌اند. کارکردگرایی ماشینی که اساس شبیه‌سازی هوش مصنوعی از هوش طبیعی قرار گرفته، بر خروجی‌های رفتاری یکسان در شرایط ورودی‌های یکسان تأکید دارد. اما استدلال بلاک و دیگر فیلسوفان، نشان می‌داد که امکان دارد شرایط کارکردی و محاسباتی یکسانی وجود داشته باشد، اما حالت روان‌شناختی و ذهنی، کاملاً متفاوت باشد. سوبر در مثال خویش می‌گوید فرض کنید به دسته‌ای از مردم پول داده شده تا ساختار شبیه ساختار مغز من را وقتی دارم فکر می‌کنم که «من بستنی می‌خواهم»، شبیه‌سازی کنند. هنگامی که یک سلول عصبی، سیگنالی را به سلول دیگر می‌فرستد، یکی از این اشخاص اجیرشده نیز با چکش به سر فرد کناری می‌زند. این گروه دارند ساختار و عملکرد مغز مرا شبیه‌سازی می‌کنند، حال آن‌که من در این هنگام دارای این فکر هستم که «من بستنی می‌خواهم»، ولی آن‌ها چنین باوری ندارند. (See: Sober, 1996.)

در واقع این آزمون‌های فکری می‌خواهند بگویند روبات‌ها با رویکرد کارکردگرایی طراحی شده و هرچقدر هم که در حیطه رفتار و واکنش‌ها به درستی عمل کنند، همچنان به لحاظ ذهنی و روان‌شناختی، قابل مقایسه با انسان نیستند و این مشکلی نیست که با پیشرفت و توسعه فناوری‌ها – مادام که در چارچوب استراتژی کارکردگرایی طراحی می‌شوند - برطرف شود.

۶-۳. مسئله اتاق چینی

حمله عمده سوم به هوش مصنوعی قوی توسّط سرل در ۱۹۸۰ راه‌اندازی شد، با استدلالی که امروزه به استدلال اتاق چینی (cra) مشهور است و حجم عظیمی از مباحث و مناقشات را پدید آورده است. در واقع، اتاق چینی یک صنعت فلسفی پر رونق را پایه گذاری کرد که چرخ کارخانه‌های آن هنوز به خوبی در گردش هستند! بسیاری از محقّقان اعتقاد دارند که استدلال نمادهای تهی سرل، دقیقاً چیزی که او گفته را اثبات می‌کند، در حالی که بسیاری دیگر استدلال وی را از اساس به اشتباه رفته می‌دانند. سرل نیز استدلال خویش را بر یک آزمون فکری استوار می‌کند که البته - لااقل به نحو علی‌الاصول - امکان تحقّق در عالم خارجی هم دارد و در حدّ یک مثال خلاف واقع نیست.

سرل خودش را در اتاقی مجسّم کرده که یادداشت‌ها و گزارش‌هایی در قالب یک کتاب راهنما وجود دارد، و یک برنامه به زبان انگلیسی هم هست که انجام عملیات بر روی کاراکترهای چینی را فقط از روی شکل آن‌ها میسّر می‌سازد. بنابراین نه هیچ ترجمه‌ای و نه هیچ ابزاری برای ترجمه در اختیار او قرار ندارد. او زبان چینی نمی‌داند و فقط در اتاقی رها شده که تعداد زیادی جعبه از نمادهای زبان چینی دارد (جایگزین بانک اطلاعاتی[۱] در کامپیوترها). سرل دسته کوچکی از نمادهای زبان چینی را از پنجره دریافت می‌کند که در واقع، سؤالاتی به زبان چینی است. سپس به کتاب قواعد (جایگزین برنامه[۲] در کامپیوترها) نگاه می‌کند و نمادهای مخصوص و متناسب با نمادهای دریافتی را گزینش و از اتاق به بیرون می‌فرستد که در واقع، همان پاسخ به سؤالات هستند. گاه برنامه، فرمان کپی‌کردن یک شکل بر روی تکّه‌ای از کاغذ را می‌دهد و در نهایت یک‌سری از این پرسش و پاسخ‌ها به یک سخنگوی بومی زبان چینی که بیرون اتاق است، تحویل داده خواهد شد. احتمالاً پاسخ‌های مذکور در قیاس با سؤالات، به آن میزان خوب باشد که شخص چینی زبان بیرون اتاق را راضی سازد. چه بسا مردمی که بیرون اتاق هستند، تصوّر کنند کسی در اتاق است که پرسش‌ها را می‌فهمد. ولی در واقع سرل زبان چینی نمی‌داند و چیزی هم از پرسش‌ها درک نکرده است. او در واقع چیزی بیش از شناسایی یک‌سری اشکال در وضعیت و ترتیب‌های مختلف انجام نداده و در نهایت، شکل‌هایی متناسب را (از روی کتاب راهنمای قواعد) بر کاغذ خروجی کپی کرده است. (See: Searle, 1984, p. 39.)

او هفده سال بعد و در کتاب *راز آگاهی*، با اشاره به بیش از یکصد مقاله‌ای که در پاسخ به استدلال وی به چاپ رسیده، توضیح می‌دهد که غالب آن‌ها هسته مرکزی اشکال اتاق چینی را درنیافته‌اند. لذا به تبیینی از این نکته محوری استدلال اتاق چینی می‌پردازد و آن را در سه مقدّمه صورت‌بندی می‌کند:

برنامه‌ها کاملاً نحوی[۳] هستند.

ذهن‌ها دارای معناشناسی هستند.

نحو و ساختار، نه معادل با و نه فی‌نفسه کافی برای جنبه‌های معناشناسی است.

بنابراین ذهن‌ها همان برنامه‌ها نیستند.

مقدّمه اوّل، خصیصه ذاتی تعاریف تورینگ را بیان می‌دارد: برنامه کامپیوتری نوشته شده، تماماً متشکّل از قوانینی در مورد هویّات نحوی است؛ یعنی قوانینی برای دستکاری نمادها. ساختار فیزیکی اجراکننده (یا پیاده‌ساز برنامه) – یعنی ویژگی‌های شیمیایی، الکتریکی و فیزیکی کامپیوتر – ربطی به ماهیت عملیات محاسباتی ندارد. مثلاً اگر ما از تراشه‌های سیلیکونی برای ساخت کامپیوترها استفاده می‌کنیم، هیچ ارتباط ذاتی بین فیزیک و شیمی سیلیکون با ویژگی‌های صوری و انتزاعی برنامه‌های کامپیوتری وجود ندارد.

---

۱. Database

۲. Program

۳. Syntactical

مقدّمه دوم نیز چیزی را بیان می‌دارد که همه ما درباره تفکّر انسانی می‌دانیم: وقتی که ما در قالب کلمات یا نمادهای دیگر فکر می‌کنیم، باید بدانیم که آن کلمات و نمادها چه معنایی می‌دهند. این دلیل آنست که مثلاً من می‌توانم به فارسی فکر کنم، اما به چینی نه. در واقع، در ذهن من، چیزی بیش از نمادهای صوری تفسیرنشده می‌گذرد که همان محتوای ذهنی یا محتوای معناشناختی است.

مقدّمه سوم هم بیان‌گر این اصل کلّی است که آزمون فکری اتاق چینی شرح می‌داد: صرف دستکاری نمادهای صوری به خودی خود، متشکّل از محتواهای معناشناختی نیست و فی‌نفسه برای تضمین وجود محتواهای معناشناختی کفایت نمی‌کند. مهم نیست که سیستم چطور رفتار کسی که واقعاً زبان چینی را می‌فهمد، تقلید می‌کند. همچنین مهم نیست که چقدر عملیات دستکاری نمادها پیچیده است. به هرحال شما نمی‌توانید معنا را صرفاً از یک فرآیند نحوی بیرون بکشید. برای ردّ کردن استدلال، باید نشان داد که یکی از این سه مقدّمه، اشتباه است. چشم‌اندازی که بنا به دیدگاه سرل، محتمل به نظر نمی‌رسد. (Searle, 1997, p. 13)

سرل در واقع، درستی واکنش‌ها و خروجی‌های رفتاری و کلامی ماشین را انکار نمی‌کند و بلکه می‌پذیرد. هوش مصنوعی طبق دیدگاه او، انعطاف‌پذیر است و می‌تواند به طیف وسیعی از شرایط جدید، پاسخی متناسب بدهد. اما آن‌چه که سرل به طور مشخّص انکار می‌کند، وجود درک[۱] توسط هوش مصنوعی است. بدین ترتیب برای کامپیوترها حیث التفاتی[۲] وجود ندارد؛ یعنی نمادها برای آن‌ها فقط یک‌سری شکل هستند و نه «درباره» هیچ چیزی. او با اشاره به یک نکته کلیدی، اساساً فهم هر نمادی را وابسته به مشاهده‌گر (مفسّر) می‌داند؛ نه این‌که معنای نمادها ذاتی و مستقل از قرارداد مفسّرین باشد. در نتیجه نه تنها آگاهی، بلکه هیچ چیز «ذاتاً» محاسباتی نیست و به تعریف و تفسیر و فهم از سوی ناظر بستگی دارد. در این صورت، فهم نمادها توسّط ماشین، موکول به فهم او از تفسیر و تعریف نمادهاست و این فهم دوم نیز موکول به فهم دیگری از تفسیر و تعریف از تفسیر و تعریف نمادهاست و این مسیر تسلسل‌وار ادامه خواهد یافت.

سرل خود تعدادی اشکال به استدلال خویش، مطرح کرده است که یکی از آن‌ها اشکال «جواب سیستم‌ها»[۳] است. می‌توان ادّعای پاسخ سیستم‌ها را این‌گونه در نظر گرفت که کجا بهترین حدّ و مرز برای یک درک‌کننده را شکل می‌دهد؟ ممکن است دور سیستم – شامل فرد داخل اتاق و کتاب راهنمای قواعد و تناظر شکل‌ها – یک خط بکشیم و بگوییم کلّ سیستم است که ماجراها و پرسش‌ها را درک می‌کند؛ همان‌طور که کلیّت شخص، زبان را درک می‌کند و نه صرفاً نواحی زبانی یا لوب‌های فرونتال مغز او. در واقع، پاسخ سیستم‌ها می‌خواهد

---

۱. Understanding

۲. Intentionality

۳. Systems Reply

بگوید فقط یک بخش خاص از سیستم (مانند مرد داخل اتاق چینی) زبان را نمی‌فهمد؛ بلکه این کلیّت سیستم اتاق چینی است که درک زبان بدو منتسب می‌شود.

پاسخ سرل به جواب سیستم‌ها اینست که فرض می‌کند برنامه و یادداشت‌ها را کاملاً به خاطر سپرده باشد و برنامه را با مراجعه به حافظه خویش، اجرا نماید. نتیجه درونی‌سازی برنامه و یادداشت‌ها در حافظه اینست که مرز سیستم دیگر با مرز بدن سرل، یکسان است. ولی همان‌طور که سرل استدلال می‌کند، او هنوز یک کلمه چینی نفهمیده است، حتّی اگر نتایج اجرای برنامه توسّط او، مکتوبی باشد که گوینده بومی زبان چینی، آن‌ها را کاملاً صحیح بیابد.

اشکال مهم دیگری که خود سرل متوجّه استدلال اتاق چینی کرده، «جواب روبات»[۱] است. این جواب می‌پذیرد که یک کامپیوتر محض، درک ندارد و مهم نیست که تا چه میزان پاسخ‌های زبانی او درست باشد. ولی اگر خروجی کامپیوترها با هدایت‌کردن یک روبات مورد استفاده قرار گیرد، به گونه‌ای که رفتار روبات متناسب با کلمات باشد، کلیّت روبات درک خواهد داشت. زیرا در این صورت، تناظرهایی بین ادراکات و رفتارها وجود خواهد داشت که کلمات را به ابژه‌ها و موقعیت‌های جهان خارج متّصل می‌سازد. مثلاً وقتی یک کامپیوتر به پرسش «چه می‌کنید اگر بوی دود سیگار استشمام نمایید؟» پاسخ خواهد داد: «ترک اتاق»؛ بدون آن‌که درکی از این پاسخ داشته باشد. اما فرض کنید که یک روبات، حسّ‌گر بو داشته باشد و فرض کنید می‌تواند پاسخ مناسبش را به مکانیسم‌هایی منتقل نماید که آن را – وقتی حسّ‌گر بو تحریک می‌شود – برای ترک اتاق حرکت بدهد. ممکن است گفته شود چنین روباتی فقط پاسخ مناسب را بر زبان نمی‌آورد؛ بلکه می‌فهمد چه می‌گوید. وقتی طیف وسیعی از پاسخ‌های بدیع با رفتارهای متناسب همراه گردد، مقبولیت این پیشنهاد بیشتر می‌شود.

پاسخ سرل به جواب روبات اینست: سناریویی را تصوّر کنید که او درون اتاق چینی است که درون یک روبات قرار دارد. او دقیقاً مانند قبل عمل خواهد کرد، اما اکنون – به‌طور نادانسته – پیام‌هایی که به بیرون اتاق چینی می‌فرستد، نه فقط سخنان روبات بلکه رفتارهای متناسبش را هم شکل خواهد داد. سرل توجّه می‌دهد که او در این سناریو نیز همچنان هیچ کلمه‌ای از چینی را نمی‌فهمد و نتیجه می‌گیرد روباتی که – نادانسته – در حال هدایت و کنترل آنست، حیث التفاتی ندارد. (See: Searle, 1980, pp. 419-420.)

دانیل دنت به سرل انتقاد می‌کند که علی‌الظاهر بین دو ادّعا خلط کرده است: ادّعای اشتقاق ناپذیری معنا از نحو، و ادّعای اشتقاق ناپذیری آگاهی معنا از نحو، و او امکان معناشناسی غیرآگاهانه را مدّنظر قرار نمی‌دهد. دنت در واقع، مقدّمه دوم استدلال سرل (نحو فی‌نفسه برای معناشناسی کافی نیست) را زیر سؤال می‌برد. از نگاه دنت، این مقدّمه در صورتی می‌تواند درست انگاشته شود که مرتکب یک اشتباه ساده درباره نحو بشویم؛ یعنی آن را به صورت برنامه اجرانشده (و اصطلاحاً روی طاقچه) در نظر بگیریم. اما نحوِ در حال اجرای بدن‌مند

---

[۱]. Robot Reply

در یک ماشین متناسب، از نظر دنت برای اشتقاق حیث التفاتی کافی است و بلکه این تنها نوع معناشناسی است که وجود دارد. (See: Dennett, 1998, pp. 335-336.)

هرچند دنت مشخّص نمی‌کند که چطور اجرای قواعد کتاب راهنما در قالب سیستم اتاق چینی، باعث ایجاد معناشناسی می‌گردد؛ به ویژه اگر به این نکته سرل توجّه کنیم که هر نوع معناشناسی از نمادها، در گرو تفسیر و فهم شاهد و مفسّری است که آن نمادها را به مثابه «نماد» لحاظ و تفسیر نماید. ولی به هرحال، نکته اصلی سرل در استدلال اتاق چینی، ناظر به آگاهی ماشین و داعیه هوش مصنوعی قوی است. روشن است که او «آگاهی از معناشناسی» و نه صرف معناشناسی را هدف گرفته و بنابراین حتّی بنا به نظر دنت نیز همچنان اتاق چینی می‌تواند در برابر هوش مصنوعی قد علم کند.

۶-۴. مسئله سیستماتیک بودن

فودور و پایلیشاین در اواخر دهه هشتاد، یک تحلیل انتقادی منتشر کردند (Fodor & Pylyshyn, 1988.) که انتقادی مؤثر از ارتباط تبیینی مدل‌های شبکه نورونی شناخت بود. در آن زمان پیوندگرایی در حال ظهور به مثابه یک حوزه جدید و جذّاب مطالعاتی بود. ولی این دو فیلسوف معاصر اعتقاد داشتند رویکرد فوق بر اساس مدلی ناقص از ذهن انسانی شکل گرفته است که به یک خصیصه ذاتی تفکّر بی‌توجّه است: ماهیت سیستماتیک بودن. کامل‌ترین آزمون استدلال‌های مرتبط با چالش سیستماتیک بودن را آیزاوا[۱] در کتاب *استدلال‌های سیستماتیک بودن*[۲] در سال ۲۰۰۳ آورده است. طبق دیدگاه فودور و پایلیشاین اما اگر انکار کنیم که بازنمایی‌ها با ترکیبی کلاسیک و به شیوه‌ای نحوی سازماندهی شده‌اند، این واقعیت روان‌شناسی را نمی‌توانیم توضیح بدهیم که تفکّرات ذاتاً و به روشی سیستماتیک با یکدیگر ارتباط دارند.

پیوندگرایان – به نظر این دو – ناگزیر نمی‌توانند تبیینی معنادار از شناخت فراهم آورند. این اشکال البته به سرعت منجر به ارائه پاسخ از سوی دیگر فیلسوفان و دانشمندان علوم شناختی شد و بحث‌های به دنبال آن، ادبیاتی عظیم را از منظرهای مختلف پدید آورد. فودور قبلاً در دهه ۷۰ استدلال کرده بود که فرضیه زبان تفکّر، خصیصه سیستماتیک بودن فکر را تبیین می‌نماید. طبق دیدگاه وی تمام تفکّراتی که محتوای گزاره‌ای دارند، ماهیّتاً بازنمودی[۳] هستند و این بازنمودها دارای خصیصه‌های نحوی و معناشناختی‌اند که به شیوه‌ای شبیه قوانین زبان طبیعی، سازمان یافته‌اند. سه توضیح اساسی که فودور درباره شاخصه‌های زبان تفکّر برجسته می‌سازد، عبارتند از:

---

۱. Aizawa

۲. The Systematicity Arguments

۳. Representational

الف. توانایی ما برای اندیشیدن و درک تفکّرات جدید و جملات ناشنیده: باروری تفکّر.[۱]

ب. درک واقعی یک تفکّر، مستلزم درک دیگر تفکّرات مرتبط با آنست: سیستماتیک بودن تفکّر.

ج. معنای یک جمله از معنای واژگانش حاصل می‌شود: اصل ترکیبی بودن.[۲]

در واقع انتقاد فودور و پایلیشاین از پیوندگرایی، قبلاً توسّط فودور در هنگام تبیین فرضیه زبان تفکّر، زمینه‌سازی شده بود. به هرحال مقاله این‌دو با بیش از ۲۶۰۰ ارجاع، نقطه‌ای کانونی از فعّال‌ترین مباحث فلسفه علوم شناختی در ۲۵ سال گذشته بوده است. ملاحظه اساسی فودور و پایلیشاین در این انتقاد، ماهیت عینی شناخت انسانی بود و نه جزئیات این‌که چطور شناخت در سیستم عصبی پیاده‌سازی می‌شود. واضح به نظر می‌رسد که ظرفیت یک سخنگوی زبان برای فهم جملات بومی زبانش، در ارتباط با فهم او از تعدادی دیگر از جملات معنادار مرتبط است. مشکل است تصوّر کنیم کسی بتواند جمله «جان، عاشق مری است» را درک کند، بدون آن‌که منابع درک جمله «مری، عاشق جان است» را در اختیار داشته باشد. بنابراین همان‌طور که مک‌لافلین توضیح می‌دهد، ادّعای سیستماتیک بودن اینست: یک نفر توانایی بازنمایی ذهنی جمله اوّل را در فکر خود دارد، اگر و تنها اگر توانایی بازنمایی جمله دوم را هم در فکر خود داشته باشد. توانایی داشتن تفکّرات، هیچ‌گاه به صورت جزیره‌وار یا نقطه نقطه نیست؛ بلکه به صورت خوشه‌ای شکل می‌گیرد. (See: McLaghlin, 2013.)

مثلاً تلاش برای یادگیری یک زبان، صرفاً از طریق کتاب لغت راهنمای سفر را در نظر بگیرید. دسترسی انحصاری شخص به این کتاب، یادگیری جزیره‌وار را به دنبال دارد؛ یعنی چیزی که شخص یاد می‌گیرد، وابسته به واژگان یا جملاتی است که بنا به اتّفاق، در کتاب جستجو می‌کند. این کاملاً با روشی که سخنگویان بومی آن زبان، جملات را می‌فهمند، متفاوت است. طبق دیدگاه فودور و پایلیشاین، استدلال سیستماتیک بودن تفکّر علیه پیوندگرایی، به قرار زیر است:

این واقعیتی روان‌شناختی است که تفکّرات، سیستماتیک هستند. به نحوی که تفکّرات ما ذاتاً با یکدیگر در ارتباطند؛ یعنی داشتن یک تفکّر، به معنای داشتن ظرفیت دسترسی به مجموعه بزرگی از دیگر تفکّرات است. بنا به توضیح متعارف، کسی که فکر می‌کند «جان، عاشق مری است»، باید ظرفیت فکر کردن به این را هم داشته باشد که «مری، عاشق جان است».

تبیین سیستماتیک بودن، نیازمند روابط نحوی و معنایی در میان بازنمایی‌های ذهنی، و نیز فرآیندی است که به چنین ساختار درونی حسّاس باشد. همانند آن‌چه که فرضیه زبان تفکّر فراهم می‌آورد.

فرضیه پیوندگرایی نه روابط نحوی و معنایی در میان بازنمایی‌های ذهنی را نشان می‌دهد و نه فرآیندی که به چنین ساختار درونی، حسّاس باشد.

---

۱. The Productivity of Thought

۲. Compositionality

بنابراین پیوندگرایی نمی‌تواند برای سیستماتیک بودن تفکّر، به کار آید. (See: Symons, J. & Calvo, P. 2013, pp. 10-11.)

البته چالمرز اشاره می‌کند که استدلال فودور و پایلیشاین، نتیجه‌ای افراطی را اثبات می‌کند؛ این‌که هیچ شبکه نورونی، نمی‌تواند سیستماتیک‌بودن را نمایش بدهد. با در نظر گرفتن این نکته که مغز یک شبکه نورونی است، نتیجه می‌شود که سیستماتیک بودن در مغز انسانی هم غیر ممکن باشد. (See: Chalmers, 1993.)

از نتایج استدلال سیستماتیک بودن - که غالباً مورد اشاره قرار گرفته - اینست که ساختار هوش کلاسیک هم در تبیین سیستماتیک بودن، بهتر عمل نمی‌کند.[1] مدل‌های کلاسیکی وجود دارند که می‌توانند برای تشخیص جمله «جان، عاشق مری است» برنامه نویسی شوند، بدون آن‌که بتوانند جمله «مری، عاشق جان است» را تشخیص بدهند. زیرا مسئله دقیقاً به این نکته بستگی دارد که کدام قانون نمادین، پردازش‌های کلاسیک را مدیریت می‌کند. نکته این‌جاست که نه استفاده از ساختار پیوندگرا و نه استفاده از ساختار کلاسیک، به تنهایی قدرت کافی برای تبیین سیستماتیک‌بودن فراگیر را ندارد. در هر دو ساختار، باید فرضیه‌های دیگری راجع به ماهیت پردازش نیز اضافه شود.

استدلال‌های سیستماتیک بودن، به‌طور مستمرّ مباحثی را درباره ساختار شناخت شکل داده‌اند؛ از دوران اوج پیوندگرایی در دهه‌های ۸۰ و ۹۰ تا ظهور یک دوره پسا پیوندگرایی در اوایل قرن ۲۱. پس از گذشت ۲۵ سال از مقاله پایه‌ای فودور و پایلیشاین، در پرتو تنوّع وسیعی از رویکردها در کنار تئوری پیوندگرایی، تلاش‌هایی جمعی برای بازاندیشی در مورد پرسش سیستماتیک بودن در جریان است. هرچند فودور و پایلیشاین قبول می‌کنند که شبکه‌های نورونی، پیچیدگی‌های زیادی یافته‌اند؛ اما اصول مبنایی همچنان همان است که بود. در نتیجه اگر آن‌ها مقاله‌شان را امروز هم بنویسند، استدلال‌شان شاید این‌گونه خواهد شد که ساختار ذهن در سطح شناختی، یک تجزیه‌کننده پیوندگرای غیرکلاسیک نیست (See: Symons, J. & Calvo, P. 2013. pp. 14-15.). و پیوندگرایی غیر کلاسیک، در بهترین حالت یک فرضیه پیاده‌سازی فراهم می‌آورد. در واقع محور اصلی استدلال فودور و پایلیشاین همچنان و علی‌رغم توسعه‌های علمی چشمگیر، کار می‌کند.

۶-۵. مسئله خصیصه سابجکتیو تجربه پدیداری

از مهم‌ترین – و چه بسا مهم‌ترین – اشکالات بر تعاریف و استراتژی‌های ساخت هوش مصنوعی، اینست که اساساً جنبه پدیداری تجربیات و احساسات را نادیده می‌گیرد. حال آن‌که رفتار هوشمندانه، هدفمند است و دشوار است عملی هدفمند را تصوّر کنیم، مگر آن‌که برخی احساسات خوشایند و ناخوشایند، هیجانات، نشاط یا افسردگی، یا ... مشروط به این عمل باشد. این قبیل احساسات و هیجانات، به اصطلاح ند بلاک – فیلسوف

---

[1]. For Example: Aizawa, 1997, "Explaining Systematicity," *Mind and Language*, 12: 115–136. / Also: Matthews, R., 1997, "Can Connectionists Explain Systematicity?" *Mind and Language*, 12: 154–177. / Also: Hadley, R., 1997, "Cognition, Systematicity and Nomic Necessity," *Mind and Language*, 12: 137–153.

ذهن معاصر - حالات آگاهانه پدیداری‌اند. بدین ترتیب می‌توان گفت هوش مصنوعی، مستلزم آگاهی پدیداری مصنوعی است. اما از آن‌جا که عمده رویکردها به هوش مصنوعی، معطوف به تولید آگاهی پدیداری نیست، می‌توان در موفقیت رویکردهای رایج، تردید داشت.

هرچند اکثر محقّقین هوش مصنوعی اگر بتوانند فقط واکنش‌هایی متناسب برای رسیدن به اهداف کارکردی فراهم آورند، راضی خواهند شد. در حالی که اهداف کارکردی، تمام اهداف هوش مصنوعی و شبیه‌سازی حیوان یا انسان نیستند. هدف کارکردی یک ماشین شطرنج‌باز، اینست که برنده بازی بشود و موفقیت با این استاندارد، احتمالاً موفقیتی برای هوش مصنوعی تلقّی خواهد شد؛ حتّی اگر هیچ احساس سرافرازی در هنگام برد یا احساس سرشکستگی در هنگام باخت برای چنین ماشینی در نظر گرفته نشده باشد. یا مثلاً روباتی که بتواند خرید از سوپر مارکت (شامل جداکردن میوه‌های رسیده از خام، انتخاب بهترین گونه میوه با تنظیم درست بین کیفیت و قیمت آن‌ها، هماهنگی با تغییرات برنامه حرکت اتوبوس‌ها جهت رفت و آمد به میوه‌فروشی، و ...) را انجام بدهد، دارای هوش مصنوعی به حساب خواهد آمد؛ بدون آن‌که هیچ‌گونه حالت آگاهانه یا احساس پدیداری را در هنگام انجام موفق کارها برای او مفروض گرفته باشیم. (Robinson, 2014, pp. 88-89.)

این نکته البته قلب اشکال دریفوس درباره حسّ عمومی را شکل می‌داد؛ بنا به نظر دریفوس، این‌که روبات‌ها نمی‌توانند در موقعیت‌های متنوّع روزمرّه، رفتاری متناسب و مرتبط را به نمایش بگذارند، ناشی از این واقعیت بود که انسان‌ها از تجربیات درونی خویش برای ارائه واکنش‌های سریع و درخور کمک می‌گیرند؛ حال آن‌که روبات‌ها فاقد چنین حسّی درونی از تجربیات پیشین هستند. اما فقدان پدیدارشناسی در استدلال دریفوس، تنها ناظر به ارائه پاسخ‌های متناسب بود و نه درباره کلیّت حضور و اهمیت تجربه پدیداری.

آن‌چه در تجربه پدیداری مهم و دشوار جلوه می‌نماید، منظر اوّل شخص و سابجکتیو آنست. به تعبیر تامس نیگل – فیلسوف ذهن دانشگاه نیویورک - «اساساً داشتن تجربه‌های پدیداری در یک ارگانیسم، بدین معناست که چیزهایی وجود دارد که عبارت است از حسّ و حال آن ارگانیسم بودن ... ما می‌توانیم آن‌ها را خصلت‌های سابجکتیو تجربه بنامیم که در قالب هیچ‌یک از تحلیل‌های تقلیل‌گرایانه اخیر از پدیده‌های ذهنی، قابل تبیین نیست. چرا که تمامی این تحلیل‌ها منطقاً با فقدان این تجربه‌های پدیداری نیز سازگار هستند.» (Nagel, 1974, p. 435.) او برای این‌که تبیین کند چرا تجربه پدیداری از دسترس نظریه‌پردازی سوّم شخص و آبجکتیو خارج است، به حسّ درونی خفّاش‌ها استناد می‌کند؛ چیزی که حسّ درونی خفّاش‌ها در هنگام ردّیابی حشره‌ها از طریق امواج صوتی است. می‌دانیم مغز خفّاش‌ها به گونه‌ای است که با تکانه‌های صادره از بازتاب‌های محیطی، همبسته است و بدین ترتیب آن‌ها را قادر می‌سازد تا فرق دقیقی بین مسافت‌ها، اندازه‌ها، شکل‌ها و ... بگذارند. اما ردیاب صوتی خفّاش در نوع عملکرد خویش – با وجود این‌که یک نوع از ادراک را سامان می‌دهد – شبیه نحوه ادراک در ما نیست و دلیلی هم وجود ندارد که به لحاظ سابجکتیو، شبیه همان تجربه‌ای باشد که ما در ادراکات‌مان داریم. در نتیجه مشکل است تصوّر کنیم که خفّاش بودن، چه حسّ و تجربه‌ای دارد؟ او این اشاره و استناد را برای رسیدن به این نتیجه‌گیری دنبال می‌کند: «قبول واقعیّتی ورای دسترسی مفهومی». (Ibid, p. 441.)

این‌چنین است که برای یک فیلسوف فیزیکالیست، تبیین کیف نفسانی با تکیه بر اساس و پایه فیزیکی و جسمانی، کاری بسیار دشوار جلوه می‌نماید. تا آن‌جا که دیوید چالمرز از این مسئله به «معضل دشوار»[۱] یاد می‌کند. (Chalmers, 1995, p. ix.) و ند بلاک از فاصله بزرگ میان اکتشافات نورولوژیک و بیولوژیک با تبیین حالات پدیداری ذهن، به «شکاف تبیینی»[۲] یاد می‌کند. (Block, 2007, p. 399.) البته مایکل تای بین معضل شکاف تبیینی و مسئله دشوار، فرق می‌گذارد؛ تعریف او از شکاف تبیینی، همین است که شما هرچه درباره وقایع فیزیکی و کارکردی دیدن رنگ به من بگویید، من همچنان خواهم پرسید: چرا این وقایع فیزیکی و کارکردی، *آن* خصیصه پدیداری خاص را تولید می‌کنند و نه خصیصه پدیداری دیگری را؟ اما آن‌چه که در نظر تای، مسئله دشوار آگاهی نامیده می‌شود، همسایه[۳] این معضل شکاف تبیینی است. «پرسش در این‌جا این نیست که چرا فرایندهای فیزیکی حاضر به این خصیصه پدیداری منجر گردیده و نه به آن خصیصه پدیداری؛ بلکه اصلاً چرا به‌طور کلّی، فرآیندهای فیزیکی به خصیصه پدیداری منجر می‌شوند؟ چرا همین فرآیندهای فیزیکی که الان در من وجود دارند، نمی‌توانند بدون هیچ‌گونه تجربه درونی در من وجود داشته باشند؟ چرا من یک زامبی نیستم؟» (See: Tye, 2009, pp. 31-32)

از رهگذر همین چالش بزرگ بوده که پروژه آگاهی ماشین، دو مسیر متفاوت را دنبال کرده است؛ یکی استفاده از محاسبات و ساخت ابزاری مصنوعی برای مدل کردن آگاهی (MMC)[۴] که معمولاً به مدل کردن «رفتارها و عملکردها»ی آگاهانه ختم می‌شود و دیگری، تلاش جهت خلق مصنوعی آگاهی. این رویکرد دوم را شاید بتوان به تبعیت از سرل در تفکیک هوش مصنوعی قوی و ضعیف، آگاهی قوی ماشین[۵] نامید. تا کنون مقالات بسیاری به منظور پیشبرد پروژه MMC نوشته شده‌اند و حتّی برخی از این محقّقان - مانند هسلو و جیرنهد - هیچ مؤلفه کشف نشده‌ای که برای تحقّق آگاهی کامل و اصیل ماشین لازم باشد، نمی‌بینند و اعتقاد دارند فقط باید مقداری پیچیدگی بیشتر به روش‌های موجود طراحی سیستم‌ها اضافه شود. هرچند اکثر کسانی که به طور خاص بر پروژه آگاهی قوی ماشین تمرکز کرده‌اند، موافق با چشم‌اندازی تئوریک‌تر هستند. از میان ایشان، استدلال کیوراستین[۶] دلالت بر یک خوشبینی شدید درباره محقّق‌سازی آگاهی قوی ماشین دارد؛ البته به نحو علی الاصول و شاید در یک دوره زمانی معقول. (Clowes & Torrance & Chrisley, 2007, p. 13.)

---

[۱] Hard Problem

[2]. Explanatory gap

[۳]. تعبیر تای از مسئله دشوار، این است: Cousin of explanatory gap (پسر خاله شکاف تبیینی)

[۴] . Machine Modeling of Consciousness

[۵] . Srtong MC

[۶]. Kiverstein

به هرحال با ملاحظه تمام پیشرفت‌ها در حوزه هوش مصنوعی، همچنان به نظر می‌رسد در مسیر ساخت مصنوعی آگاهی – به معنای تولید تجربه پدیداری – حتّی به گام‌های نخستین هم وارد نشده‌ایم و معلوم نیست چنین مسیری – در صورت وجود – با چه استراتژی و اصولی باید پیش برود.

البته طراحی‌های مهندسین هوش مصنوعی، امروزه کاربردهای فراوانی دارند و از جمله استفاده‌های جدید و رایج آن، در بازی‌های ویدئویی و انیمیشن‌های هالیودی است. سیستم‌های کارشناس، بازی‌های سه بُعدی (واقعیت مجازی)، و موتورهای جستجوگر اینترنتی مانند گوگل، همگی این واقعیت را توضیح می‌دهند که طیفی از کاربردهای نامرئی هوش مصنوعی وجود دارد که عموم مردم حتّی از آن اطلاع هم ندارند. بنابراین اتهام به GOFAI به عنوان یک تکنولوژی شکست خورده، درست نیست و نمی‌تواند مورد حمایت متخصّصین و مهندسین قرار گیرد. اما درباره GOFAI - و به طور کلّی، هوش مصنوعی - به عنوان یک پروژه روان‌شناختی (و نه فنّی) چطور؟ به یاد بیاوریم که در هوش مصنوعی ضعیف، هیچ پرسش یا ادّعای روان‌شناختی وجود نداشت. اما هوش مصنوعی قوی، داعیه روان‌شناختی مهمّی درباره ماهیت آگاهی و قابلیت بازسازی آن به نحو مصنوعی را مطرح می‌کند. در این‌جا حداقل دو دلیل وجود دارد که چرا کسی نمی‌تواند ادّعای موفقیت هوش مصنوعی قوی – به عنوان یک طرح روان‌شناختی - را داشته باشد: از یک‌سو، این تصوّر که محاسبات می‌تواند تمام جنبه‌های هوش را تبیین کند/ پیاده سازد، اشتباه است. از سوی دیگر، معلوم نیست که هوش مصنوعی قوی، از اساس صحیح باشد.

در حقیقت این ادّعا همواره یک مسئله مناقشه برانگیز بوده؛ زیرا با یک تبیین علمی و طبیعت‌گرایانه از حیث التفاتی سر و کار دارد. فیلسوفانی که متعهّد به یافتن چنین تبیینی هستند، درباره چگونگی انجام آن توافق ندارند. برخی‌ها امیدوارکننده‌ترین تلاش را رویکرد تکاملی می‌دانند، ولی حتّی طرفداران این رویکرد هم می‌پذیرند که موضع آن‌ها از جهات مختلف، غیر شهودی است. علاوه بر این، بسیاری فیلسوفان – شامل اکثر کسانی که حامی دیدگاه پدیدارشناسی هستند – استدلال می‌کنند که ارائه یک نظریه طبیعت‌گرایانه درباره حیث التفاتی، علی‌الاصول ممکن نیست. حتّی هیلاری پاتنم که فلسفه کارکردگرایانه او، سوخت لازم برای اوج‌گیری هوش مصنوعی قوی بر مبنای GOFAI را تأمین کرد، اکنون نسخه‌ای از همین دیدگاه (عدم امکان علی‌الاصول ارائه نظریه‌ای طبیعت‌گرایانه درباره حیث التفاتی) را قبول دارد و بر همین اساس، از علوم شناختی به عنوان تخیّل علمی[۱] یاد می‌کند.

هسته ادّعای ضد طبیعت‌گرایانه اینست که زبان انسان، معنا و آگاهی، منبع تمام مفاهیم و دانش ما (از جمله علوم) هستند و بنابراین خود این پدیده‌ها نمی‌توانند به وسیله علوم تبیین شوند. طبق این دیدگاه، هوش مصنوعی قوی، مبتنی بر یک دور باطل در تأمین معانی و مفاهیم است و در نتیجه، دیدگاهی غیر معتبر به نظر می‌رسد. هرچند شقاق طبیعت‌گرایی/ ضد طبیعت‌گرایی، عمیق‌ترین تقسیم در فلسفه معاصر است که به

---

[۱] . Science Fiction

نزاع‌های علمی تندی نیز انجامیده؛ نزاع‌هایی که در آن علم به طور کلّی مورد حمله قرار گرفته یا به شدّت حمایت شده است. اما به هرحال، با توجّه به ادّعای تبیین پدیده ذهنی، هوش مصنوعی روان‌شناختی (چه بر مبنای GOFAI یا مبانی جدیدتر) در خط مقدّم این نزاع قرار دارد. (See: Boden, 2014, pp. 125-127.)

این نکته مهمّی است که هیچ‌یک از عمده رویکردهای چهارگانه – که از راسل و نوریگ نقل کردیم – به دنبال ساخت آگاهی مصنوعی – به معنای پدیداری آن – نیستند. شبیه‌سازی رفتار یا حتّی تفکّر (چه عقلانی و چه انسانی)، فاصله بسیار و معناداری با شبیه‌سازی تجربه پدیداری دارد. حال آن‌که چه بسا مهم‌ترین و رازآمیزترین جنبه هوش، همین آگاهی پدیداری است و بدون آن، در بهترین حالت، با یک زامبی درست کردار مواجه هستیم. شاید اگر رویکرد پدیداری به نحوی جدّی در دستور کار ساخت هوش مصنوعی قرار گیرد، تئوری‌های فلسفی درباره ماهیت آگاهی و چگونگی ظهور آن بتواند مسیرهای فنّی جدیدی را پیش روی محقّقان بگذارد.

حتّی مهم‌تر از خصیصه سابجکتیو تجربه آگاهانه شاید «اراده آزاد» باشد که معلوم نیست چطور می‌تواند در روبات‌ها شبیه‌سازی شود. اراده آزاد از عناصر مهم مرتبط با هوش و عقلانیت دانسته شده (See: Aquinas, 1265–72/1945, I, Q83, a. 1) و در منابع کلاسیک فلسفه اسلامی هم اساساً دلیل بر وجود نفس را از آثاری که «لا علی وتیرة واحدة» (نه بر یک منوال مشخص و ثابت) هستند، استنتاج می‌کنند؛ یعنی رفتار و واکنش‌های انسانی همواره در شرایط یکسان، به نحو یکسان بروز نمی‌نمایند و موکول به آیتم انتخاب‌گری و اراده «عامل» است. این نکته می‌تواند چالشی جدّی فرا روی هوش مصنوعی باشد. زیرا روبات‌ها – مهم نیست چقدر و به چه نحو پیچیدگی داشته باشند – به وسیله قوانین متعیّن[۱] مدیریت می‌شوند و در نتیجه ممکن است حتّی زمینه و قابلیت عقلانیت را نیز نداشته باشند؛ زیرا بنا به سنّت پرطرفدار فکری، سیستم‌های متعیّن، اراده آزاد ندارند. هرچند ممکن است گفته شود چنین موضعی، با در نظر گرفتن هوش انسانی به گونه‌ای کاملاً قابل تبیین براساس فعالیت مغز مادّی، سازگاری ندارد.

۷. امکان ظهور آگاهی مصنوعی برپایه فلسفه اسلامی

نفس در نگاه فیلسوفان مسلمان تا پیش از زمان ملاصدرا، یک جوهر مجرّد و روحانی بود که به بیان ابن‌سینا، حدوث و بقای روحانی دارد. نفس در این دیدگاه، در عین تجرّد، حادث است؛ یعنی آن‌گاه که - بنا به نظر بوعلی - ترکیب عناصر طبیعی بدن (جنین) به حدّی از اعتدال رسید، نفس هم بدان تعلّق می‌گیرد. چه این‌که عناصر مادّی (که تشکیل‌دهنده جنین هستند)، در غایت تضادّ با هم‌اند. کیفیت غالب در هریک از عناصر مادّی چهارگانه (آب و هوا و آتش و خاک) - از دید طبیعیات قدیم – با هم متنافرند و بالطّبع از یکدیگر

[۱]. Deterministic Laws

می‌گریزند و قبول حیات نمی‌کنند. مگر آن‌که به سبب اختلاط، سردی آب و گرمی آتش، از شدّت همدیگر بکاهند و رطوبت هوا و خشکی خاک نیز موجب انکسار متقابل شوند. تا آن‌جا که در ترکیب حاصل، درجه گرمی و سردی، و نیز درجه رطوبت و خشکی به حدّی از اعتدال برسد و همه اجزاء به‌نحوی یکسان‌تر در محصول ممتزَج که «مزاج» نام دارد، ترکیب شوند. مزاج، برخلاف هریک از عناصر اربعه‌ی تشکیل‌دهنده آن، ضدّی ندارد و بنا به میزان اعتدالش، می‌تواند استعداد پذیرش نفس نباتی یا حیوانی را داشته باشد، و حتّی تا بدان حدّ از اعتدال رسد که قابل تعلّق به نفس انسانی گردد. (ابن‌سینا، ۱۳۸۳، صص۶۳-۶۱.)

نفس در این نگاه، جوهری کاملاً متفاوت از جوهر بدن است که به‌طور کلّی، دو دسته فعل انجام می‌دهد: یکی در قیاس با بدن؛ که سیاست و تدبیر بدن باشد. و دیگری در قیاس با ذات خود؛ که ادراک معقولات است. این دو دسته فعل از نگاه شیخ‌الرّئیس، متنافر از هم و مانع یکدیگر هستند؛ به‌گونه‌ای که هرگاه نفس به یکی مشغول شود، از دیگری منصرف گردد و جمع بین این دو برای نفس، بسیار صعب است. (ابن‌سینا، ۱۴۰۴، ج۲، ص۱۹۵)

تعلّق و ارتباط نفس به بدن، از طریق قوای مختلف نفس (غاذیه و مُنمیه و مولده و محرّکه و حاسّه و متخیّله و متوهّمه و حافظه و عالمه و عامله) صورت می‌پذیرد. بین بدن و قوا، جرم لطیف گرمی وجود دارد که حامل اوّلیه تمامی این قواست و روح بخاری نامیده می‌شود. روح لطیف یا روح بخاری، به‌دلیل شدّت لطافتی که دارد، در درون اعصاب، نفوذ می‌یابد و قوای مختلف نفس را حمل می‌کند. روح بخاری مادام که در قلب قرار دارد، «روح حیوانی» نامیده می‌شود و آن‌گاه که به مغز منتقل و در خُلل و فُرج آن واقع گردید و انفعال پذیرفت، «روح نفسانی» خوانده می‌شود. (ابن‌سینا، ۱۳۶۳، صص۹۵-۹۴) دیدگاه شیخ‌الرّئیس و دکارت در این‌باره، گونه‌ای تبعیت از ایده ارسطویی درباره حادث بودن نفس است که در مقابل نظریه افلاطونی پیرامون ازلی بودن نفس قرار دارد.

البته قول به حدوث روحانی نفس از اشکالاتی ناشی می‌شود که قِدم نفس روحانی، دچارش بود. هرچند که حدوث روحانی نفس نیز بعدها توسّط منتقدین بوعلی در این‌باره و به‌ویژه خواجه طوسی و صدرالمتألّهین، با اشکالات مهمّی مواجه گشت.

صدرالمتألّهین در استدلال مهمّی که مستقیماً به تعریف و تصویر نفس در حکمت متعالیه اشاره دارد، نفس را در یک ترکیب حقیقی واحد، «صورت» جسم طبیعی می‌بیند: صورت در اندیشه ارسطویی – آن‌گونه که توسّط فیلسوفان اسلامی، پذیرفته و تقریر شده است – کمال مادّه و در سنخیت با آن است؛ به گونه‌ای که از ترکیب آن دو، یک نوع مادّی شکل می‌گیرد. یعنی حقیقتاً در خارج، یک وجود واحد شکل گرفته و در غیر این‌صورت، ترکیب این‌دو نمی‌تواند یک نوع مادّی را درست کند.

صدرالمتألهین در پی استدلال‌های خویش، بر این نتیجه تأکید دارد که نفس را نمی‌تواند از ابتدا و در حین حدوث، مجرّد محض دانست؛ بلکه از آن‌جا که صورت جسم طبیعی و منوّع یک نوع مادّی است، حدوثی مادّی و وجودی مادّی دارد. نحوه وجود نفس در اندیشه صدرا، نحوه اتّصال تدبیری برای بدن است که از مجموع نفس و بدن، یک وجود واحد و یک نوع واحد درست شده است. بدین ترتیب، نحوه وجود نفس، نه مانند نحوه وجود مجرّدات محض، مفارق تامّ است و نه مانند نحوه وجود مادّه محض، منغمر در مکان. بلکه طبق نظر او،

نحوه وجود نفس، نحوه‌ای خاص خودش می‌باشد و در نتیجه، تعلّق و تصرّف در بدن، امری ذاتی برای نفس است (چرا که در نحوه وجود آن اخذ شده است) و نه امری عارضی برای آن. حال آن‌که طبق نظر بوعلی، تعلّق به بدن و تصرّف در آن، امری عارض بر ذات مفارق نفس بود.

این‌جاست که یکی از مهم‌ترین و اساسی‌ترین اشکالات حکمت متعالیه بر دوئالیسم سینوی (و به طریق اولی بر دوئالیسم دکارتی) آشکار می‌گردد: «اتّصال نفس و بدن یک همراهی اتّفاقی نیست که بین‌شان ارتباطی ذاتی نباشد. چگونه [حامیان مکتب مشّاء] به این اندیشه متمایل شده‌اند، در حالی‌که خودشان تصریح دارند که نفس، صورت کمالی بدن و کمال اوّل جسم طبیعی آلی است و به حصول نوع طبیعی از آن‌دو حکم کرده‌اند. امکان ندارد چنین ترکیبی از دو امر بدون ارتباط علّی و معلولی، حاصل شده باشد.» (صدرالمتألهین، ج۸، ص۳۸۲)

در نتیجه وی در مقابل اندیشه شیخ‌الرّئیس، بدن را از سنخ علّت مادّی برای نفس دانسته و بدین‌ترتیب، تعلّقی ذاتی بین آن‌دو برقرار می‌بیند. حال آن‌که به تعبیر او، نظر بوعلی به این می‌انجامد که نسبت نفس مجرّد مفارق با بدن مادّی، همانند نسبت سنگ در کنار انسان (بدون هیچ‌گونه ارتباط ذاتی و الزامی) باشد. ملاصدرا تأکید دارد که ابزاریّت بدن برای نفس، مانند ابزاریّت ارّه برای نجّار و کشتی برای ناخدا نیست که اگر بخواهد، آن را به‌کار گیرد و اگر نخواهد، واگذارد. (ن.ک: همان، صص۳۸۴-۳۸۳)

در این دیدگاه، نفس دارای مقامات و درجاتی است که ابتدا از پایین‌ترین و مادّی‌ترین مرتبه شروع می‌شود و تکوّن می‌یابد. به بیان دیگر حدوث نفس، صرفاً مادّی و جسمانی است که تدریجاً و در طول زمان با حرکت جوهری خویش، کمالات بالاتر را می‌پذیرد و واجد جنبه مجرّد و کمالات ذومراتب آن می‌شود. بدین ترتیب، نفس در حکمت متعالیه، نه تنها قدیم و ازلی نیست؛ بلکه در ابتدا حدوث روحانی هم ندارد و در یک زمان خاص، از وجود مادّی جسم (بدن) حادث می‌شود. (صدرالمتألهین، ۱۳۸۲، ص۱۷۰) در این دیدگاه، اوّلین نشئه‌ای که از نفس تکوّن می‌یابد، قوّه جسمانی است و سپس صورت طبیعی و سپس نفس حسّاس در مراتب مختلف آن، سپس نفس متفکّر و سپس نفس ناطقه و در نهایت، مرتبه عقل در درجات مختلفش تا حدّ عقل فعّال.» (صدرالمتألهین، ۱۴۲۰، ص۲۳۵) «نیاز نفس به بدن در ابتدای وجودش، نیازی جوهری و طبیعی است؛ چه این‌که نفس بدون بدن، استقلال در وجود ندارد و دانستیم که نفسیّت نفس، نحوه وجود خاص خودش را دارد که از قبیل نسبت و اضافه عارض بر ذات نیست. البته نفس در مراتب کمالش، می‌تواند عقل بالفعل و غیر محتاج به بدن گردد.» (صدرالمتألهین، ۱۳۶۸، ج۹، ص۸۵)

بنابراین نفوس انسانی در نظر صدرا، در زمره صور مادّی هستند که حدوث شان هم مادّی است. البته در طول زمان و پس از نیل به استعداد خاصّی در مادّه، از آن ترفّع یافته و حالت تجرّد می یابند. ضمن اینکه همین حالت تجرّد، یک وضعیت تشکیک پذیر و ذومراتب بوده و می تواند در نتیجه حرکت جوهری، شدّت و ضعف پذیرد. نکته دیگر آنکه روح مجرّد و بدن مادّی را باید دو جنبه از یک جوهر واحد دانست که همان نفس آدمی است.

این یک تحوّل اساسی در نوع نگرش عرفی و سنّتی به «روح» بود که آن را (لااقل در بدو امر) به موجودی مادّی با منشاء پیدایش کاملاً طبیعت‌گرایانه تبدیل کرد. کما این‌که در ادامه وجود خویش نیز – تا زمانی که

به بدن اتّصال دارد – موجودی مابین مادّی و مجرّد باقی می‌ماند. نکته مهم اما در بحث مدّ نظر ما اینست که اگر نفس و حالات آگاهانه آن نه به صورت روح مجرّد فرافیزیکی مستقل از بدن، بلکه در مرحله جنینی و به صورت تشدید شونده از سطح نوروفیزیولوژیک پدید می‌آید، علی‌الاصول چه دلیلی بر محال بودن پیدایش آن از شبکه‌های مصنوعی پیچیده و مشابه وجود دارد؟ قبلاً اشاره کردیم که اشکالات فلسفی – مانند استدلال دریفوس درباره حسّ عمومی و استدلال سرل و بلاک – ناظر به این بود که صِرف اجرای برنامه‌های محاسباتی – چه به نحو کلاسیک (GOFAI) و چه در قالب شبکه‌های نورون مصنوعی – نمی‌تواند معادل یا در بردارنده آگاهی باشد. اما - همان‌طور که خود سرل نیز تصریح می‌کرد – منافاتی ندارد با این که شبکه‌های مصنوعی، به نحو متافیزیکی منشاء پیدایش و ظهور[۱] آگاهی بشوند.

به ویژه آن‌که ساخت شبکه‌های نورونی از سطح جریان شلیک‌های الکتریکی در آکسون‌ها و سیناپس‌های عصبی فراتر رفته و شبکه مصنوعی پیوندها در رویکردهای نوین، هرچه بیشتر به سوی فیزیولوژی انسانی پیش رفته و شبیه‌سازی خاستگاه ظهور آگاهی – یعنی شبکه نورونی سلولی – مدّ نظر قرار گرفته است. (See: Pagel & Kirshtein, 2017, p. 30.) هرچند که با شبیه‌سازی و ساخت جایگزین برای شبکه اتّصالات و جریان انرژی الکتریکی و نوروشیمی و بدن‌مندی، هنوز و همچنان کارکرد پیشینه زیست‌شناختی و DNA، بی‌بدیل و مختص هوش طبیعی به نظر می‌رسد و معلوم نیست نقش هریک در پیدایش آگاهی – اگر دیدگاه حکمت متعالیه درست باشد – چقدر و چگونه است. با این اوصاف شاید پروژه شبکه‌های نورونی مصنوعی در اوج کمال خویش نیز نتوانند حالات آگاهانه و کیفیات نفسانی را به کمال و کیفیتی که در انسان شاهد هستیم، تولید نمایند. اما – مطابق نظریه حدوث جسمانی نفس - این احتمال و امکان را نمی‌توان نفی کرد که چه بسا سطوحی از جوهر نفسانی و آگاهی مصنوعی در این مسیر به منصه ظهور برسد. خصوصاً آن‌که نفس در نظر صدرا، دارای مراتبی از حرکت تشکیکی و استکمالی است و ممکن است لااقل برخی از مراتب اوّلیه آن، به نحوی مصنوعی و از طریق شبیه‌سازی‌های جدید، ایجاد گردد.

طبیعی است که بسیاری از دانشمندان مادّی‌انگار از حامیان سرسخت چشم‌انداز ساخت آگاهی مصنوعی و شبیه‌سازی انسان باشند. در این شرایط، شکل‌دهی یک دوقطبی از حامیان و منکران این چشم‌انداز که عمده گروه دوم را نیز طرفداران نگرش دینی و الهیاتی شکل بدهد، می‌تواند با تحقّق درصدی از مانیفست هوش مصنوعی قوی، به ضرر اردوگاه دینداران و دینداری تمام خواهد شد. حال آن‌که هیچ دلیل قاطعی از متون دینی – لااقل متون شیعی – برای مقاومت و انکار جزم‌انگارانه در برابر این احتمال وجود ندارد و بلکه – همان‌طور که توضیح دادیم – حتّی نگرش فلسفی غالب شیعی در چهارصد سال اخیر، چشم‌انداز مذکور را دست‌کم در حدّ یک احتمال، روا می‌دارد.

---

[۱]. Emergence

آن‌چه در این گزارش بدان پرداختیم، فلسفه هوش مصنوعی از زاویه فلسفه ذهن بود که بیشتر به جنبه‌های متافیزیکی هوش، مربوط می‌شود. اما پس از عبور از این مرحله و پذیرش احتمال و امکان تحقّق هوش مصنوعی قوی، آن‌گاه نوبت به قانون‌گذاری و ترسیم خطور اخلاقی و حقوقی برای روبات‌ها و ماشین‌های هوشمند می‌رسد که آینده نزدیک زندگی بشر را تسخیر خواهند کرد. این جنبه از مباحث فلسفه هوش مصنوعی البته ذیل فلسفه اخلاق قرار می‌گیرد که آن هم حوزه وسیعی از آثار و تأمّلات را در بر گرفته و مسائل مهمّی از «حقوق روبات‌ها» و «انتساب "عاملیت" به روبات‌ها» گرفته تا «مسئولیت اخلاقی در روبات‌ها» و «تئوری‌هایی برای ساخت روبات‌های اخلاقی» تا «تهدیدهای به کارگیری روبات‌ها در پست‌های حساس (مانند قضاوت و نظامی‌گری و ...)» و ... را به بحث گذاشته است که علی‌القاعده در آینده‌ای نه چندان دور، مبنای طراحی حقوق و قوانین مربوط به ساخت و استفاده از روبات‌ها در حوزه‌های مختلف زندگی اقتصادی و اجتماعی بشر را سامان خواهد داد.

جمع‌بندی

بنا بر آن‌چه در این گزارش آمد، توجّه به جنبه‌های فلسفی در حوزه هوش مصنوعی بسیار اهمیت داشته و پرداخت صرفاً فنّی و مهندسی به این حوزه، می‌تواند به تقلیدهای فاحش یا غفلت‌ها و اشتباهات سنگین بینجامد. اساساً دو پارگی مطالعات فلسفی و مبنایی از مطالعات فنّی و تکنولوژیک، در شرایطی که فی الواقع ارتباطی جدّی بین این‌دو برقرار است، نوعی نابینایی (آگاهانه یا ناخودآگاه) در هر دو حوزه پدید می‌آورد که تأثیرات این عدم ارتباط، به طور خاص در حوزه فنّی و تکنیکال، گسترده‌تر و جدّی‌تر است.

به طور مقابل، توجّه به جنبه‌های مبنایی و تأمّلات فلسفی می‌تواند رویکردهای تولید و ساخت هوش مصنوعی و اهداف و چشم‌اندازهای تعریف شده را متحوّل ساخته و چه بسا بصیرت‌هایی جدید برای اصلاح روندهای فنّی و تکنیکی فراهم آورد. کما این‌که می‌تواند مباحث انسان شناختی در فلسفه را نیز با تحوّلات و پیشرفت‌هایی قابل توجّه مواجه سازد.

در متن گزارش به برخی از تعاملات دو سویه بین فیلسوفان و مهندسان غربی اشاره شده که چطور مثلاً اشکال دریفوس درباره فقدان حسّ عمومی و معرفت پدیداری، مهندسان هوش مصنوعی را به تکاپو انداخته و بعضاً به اعتراف در مورد اشتباه بودن رویکرد برنامه‌نویسی واداشته است. هرچند که بازسازی همین اشکال در برابر رویکرد جایگزین (رویکرد پیوندگرا) نیز تأمّلاتی دوباره را برانگیخته است. همچنین اشکال فودور در مورد سیستماتیک بودن، سایر فیلسوفان و نیز برخی مهندسان را به تلاش برای تبیینی بهتر از رویکرد جدید یا انجام برخی اصلاحات در آن کشانده است. این تأثیرگذاری البته یک‌طرفه نبوده و یافته‌های مهندسی و پیشرفت‌ها یا شکست‌های تکنیکال در حوزه هوش مصنوعی، بصیرت‌ها و افق‌های جدیدی را در نظریه‌های فلسفی گشوده و مثلاً رویکرد بازنمودگرایی درباره تمام حالات ذهنی – به عنوان رویکرد غالب در دهه‌های اخیر – را با چالشی مهم روبه رو کرده یا مثلاً از جمله دلایل کنارگذاشتن دیدگاه‌های کارکردگرایانه توسط برخی بوده است.

آن‌چه از مجموع اشکالات فلسفی مطرح در برابر رویکردهای رایج هوش مصنوعی ذکر شد و نیز آن‌چه از مبانی فلسفه اسلامی (لااقل در پارادایم حکمت متعالیه) بر می‌آید، چه بسا بتوان تعاملاتی دو سویه بین حوزه

مهندسی و فلسفی در این‌باره در محافل علمی و تصمیم‌گیری داخل کشور برقرار ساخت و افق‌هایی جدید را در هر دو عرصه فنّی و نظری گشود. به ویژه آن‌که هوش مصنوعی امروزه به یکی از معدود حوزه‌های دانشی مرتبط با الحاد و انکار فراطبیعت تبدیل گشته و نمودهای آن در مراکز دانشگاهی و پژوهشی داخل کشور نیز قابل مشاهده است.

در نهایت و علاوه بر آن‌چه در مورد دو حوزه دانشی مرتبط (فلسفه و هوش مصنوعی) ذکر شد، بر این نکته نیز باید تأکید کرد که توجّه درست به مباحث فلسفه ذهن و فلسفه اخلاق در مورد هوش مصنوعی، اثراتی مهم در تصمیم‌گیری و مدیریت این حوزه رو به رشد و استراتژیک دارد.

منابع

[۱] ابن سینا، حسین، ۱۴۰۴ه.ق، *الشّفاء*، ج۲، کتاب النّفس، قم، مکتبه آیت‌الله المرعشی.

[۲] ----------، ۱۳۶۳، *المبدأ و المعاد*، النّص، تهران، مؤسّسه مطالعات اسلامی.

[۳] ----------، ۱۳۸۳، *رساله نفس*، متن، همدان، دانشگاه بوعلی.

[۴] صدرالمتألهین، محمد، ۱۳۶۸، *الحکمه المتعالیه فی الأسفار العقلیه الأربعه*، قم، مکتبة المصطفوی.

[۵] ----------، ۱۳۸۲، *الحاشیة علی الهیات شفا*، تصحیح نجفقلی حبیبی، تهران، بنیاد حکمت اسلامی صدرا.

[۶] ----------، ۱۴۲۰ه.ق، *الحکمة العرشیة*، بیروت، مؤسّسه التاریخ العربی.

[7] Aquinas, T., 1265–72/1945, *Summa Theologica*, in A. C. Pegis (tr.), *Basic Writings of St. Thomas Aquinas*. New York: Random House.

[8] Arkoudas, K., & Bringsjord, S., 2014, Philosophical foundations, In: Franklin, Keith & Ramsy, William (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press.

[9] Bellman, R. E. (1978). *An Introduction to Artificial Intelligence: Can Computers Think?* Boyd & Fraser Publishing Company.

[10] Block, N., 2007, *Consciousness, Function, and Representation*, The MIT Press.

[11] Boden, M., 2014, GOFAI, In: Franklin, Keith & Ramsy, William (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press.

[12] Burge, Tyler, 1982, "Other Bodies", In: *Thought and Object: Essays on Intentionality*, A. Woodfield, (ed.), Oxford: Oxford University Press, pp. 97–120.

[13] Charniak, E. & McDermott, D., 1985, *Introduction to Artificial Intelligence*, Reading, MA: Addison Wesley.

[14] Chalmers, D., 1993, "Why Fodor and Pylyshyn Were Wrong: The Simplest Refutation," *Philosophical Psychology*, 6(3): 305–319.

[15] ----------, 1995, *The Conscious Mind*, University of California, Santa Cruz.

[16] ----------, 1996, Does A Rock Implement Every Finite-State Automaton?, Synthese, 108, pp.309-333.

[17] Clowes, R., & Torrance, S., & Chrisley, R., 2007, "Machine Consciousness, Embodiment and Imagination", *Journal of Consciousness Studies*, 14, No. 7, pp. 9–13.

[18] Dennett, D., 1998, *The Intentional Stance*, A Bradford Books, The MIT Press.

[19] Dreyfus, H., 1999, *What Computers Still Can't Do*, MIT Press.
[20] Fodor, J., 1987, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.
[21] Fodor, J., & Pylyshyn, Z., 1988, "Connectionism and Cognitive Architecture: A Critical Analysis," *Cognition*, No. 28, pp. 3–71.
[22] Harnad, S. 1990, "The symbol grounding problem", *Physica* D 42: 335–46.
[23] Haugeland, J. 1985, *Artificial Intelligence: The Very Idea*, Cambridge, MA: MIT Press.
[24] Johnson-Laird, P. N., 1977, "Procedural semantics", *Cognition* 5: 189–214.
[25] Kurtzweil, R., 1999, *The Age of Spiritual Machines; When Computers Exceed Human Intelligence*, Viking Adult.
[26] Lenat, D., 1984, "Computer Software for Intelligent Systems," *Scientific American*.
[27] Lenat D. & Feigenbaum, E., 1991, "On the Thresholds of Knowledge," *Artificial Intelligence*, Vol. 47, Nos. 1-3.
[28] Lungarella M. & Iida F. & Bongard J. & Pfeifer R., 2007, *50 Years of Artificial Intelligence*, Springer.
[29] Maynard-Smith, J., 2000, "The Concept of Information in Biology", *Philosophy of Science*, 67, pp.177-194.
[30] McLaghlin, B., 2013 "Can an ICS Architecture Meet the Systematicity and Productivity Challenges?", in: Paco Calvo & John Symons, *The Architecture of Cognition Rethinking Fodor and Pylyshyn 's Systematicity Challenge*, The MIT Press.
[31] Nagel, T., 1974, "What is it like to be a bat?", *The Philosophical Review*, Vol.83, No. 4., pp. 435-450.
[32] Nilsson, N. J. 1998, *Artificial Intelligence: A New Synthesis*, Morgan Kaufmann.
[33] Pagel, J. F. & Kirshtein, Philip, 2017, *Machine Dreaming and Consciousness*, Academic Press.
[34] Poole, D., & Mackworth, K., & Goebel, R., 1998, *Computational intelligence: A logical approach*, Oxford University Press.
[35] Putnam, Hillary, 1975. "The Meaning of 'Meaning'", *Language, Mind and Knowledge*, *Minnesota Studies in the Philosophy of Science*, vol. 7, University of Minnesota Press, pp. 131-193.
[36] Rich, E. & Knight, K. 1991, *Artificial Intelligence* (2th ed.), McGraw-Hill.
[37] Robinson, W., 2014, Philosophical challenges, In: Franklin, Keith & Ramsy, William (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press.
[38] Rumelhart, D. E., McClelland, J. L., and the PDP Research Group. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. I: Foundations. Cambridge, MA: MIT Press.
[39] Russell, S., & Norvig, P., 2010, *Artificial Intelligence, A Modern Approach* (3rd ed.), Prentice Hall.
[40] Schmidhuber, J., 2007, "Celebrating 75 Years of AI; History and Outlook: The Next 25 Years", in: Lungarella M. & Iida F. & Bongard J. & Pfeifer R. (eds.), *50 Years of Artificial Intelligence*, Springer.
[41] Searle, J., 1980, "Minds, Brains, and Programs", *The Behavior and Brain Science*, No. 3.

[42] ----------, 1984, *Minds, Brains, and Science,* Cambridge, MA: Harvard University Press.
[43] ----------, 1997, *The Mystery of Consciousness*, New York, NY: New York Review of Books.
[44] ----------, 2002, *Consciousness and Language,* Cambridge university press.
[45] Sober, Elliott, 1996, "Putting the Function Back into Functionalism", in: *Mind and Cognition*, Lycan (ed.), Basil Blackwell.
[46] Steels, L., 2007, "Fifty Years of AI; From Symbols to Embodiment - and Back", in: Lungarella M. & Iida F. & Bongard J. & Pfeifer R. (eds.), *50 Years of Artificial Intelligence*, Springer.
[47] Sun, R., 2014, Connectionism and neural networks, In: Franklin, Keith & Ramsy, William (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press.
[48] Sun, R., & Slusarz, P., & Terry, C. (2005). "The interaction of the explicit and the implicit in skill learning: A dual-process approach", *Psychological Review* 112: 159–92.
[49] Symons, J. & Calvo, P. 2013. Systematicity: An Overview, In: *Paco Calvo and John Symons (eds.), The Architecture of Cognition; Rethinking Fodor and Pylyshyn 's Systematicity Challenge*, The MIT Press,
[50] Tye, M., 2009, *Consciousness Revisited; Materialism Without Phenomenal Concepts,* The MIT Press.
[51] Winston, P. H. 1992, *Artificial Intelligence* (3th ed.), Addison-Wesley.

حـوزه فضـای مجـازی بـه انـدازه انقـلاب اسـلامی اهمیـت دارد. ایـن فضـا مثـل یـک رودخانـه پـر از آب و خروشـان اسـت کـه مـی آیـد و دائمـاً هـم بـر آب آن افـزوده و خروشـان تـر مـی شـود. اگـر مـا بـر ایـن رودخانـه تدبیـر کنیـم و برنامـه داشـته باشـیم، زه‌کشـی کنیـم و هدایـت کنیـم ایـن رودخانـه را تـا بـه سـد بریـزد، می‌شـود فرصـت. اگـر رهایـش کنیـم و برنامـه‌ای بـرای آن نداشـته باشـیم می‌شـود یـک تهدید.

csri.majazi.ir