بسمه تعالی

بررسی امکان آگاهی در ماشین

(از رویکرد کلاسیک محاسباتی تا بدن‌مندی و پیوندگرایی)

زمستان ۱۳۹۹

آگاهی (Consciousness) و انواع آن

- آگاهی پدیداری (Phenomenal Consciousness): ند بلاک (Ned Block) ویژگی‌های تجربه‌ای و حسّ و حال درونی داشتن را اصطلاحاً آگاهی پدیداری می‌نامد و حالات ذهنی دارای این ویژگی‌ها (مثل غم و شادی و درد و ارگاسم و ادراکات حسّی و ...)، حالات پدیداری هستند.
- آگاهی دسترسی (Access Consciousness): صرفاً دارای کارکردهای شناختی (استدلال، تصمیم‌گیری، هدایت رفتار، ...) و فاقد جنبه پدیداری است. عنصر اصلی آگاهی دسترسی، گزارش‌پذیری است و نه تجربه. / البته نظریات جدید قائل به تعمیم پدیداری
- مقاله تاریخی What it is like to be a bat از تامس نیگل: آن‌چه که در جنبه پدیداری مهم می‌نماید، منظر اوّل شخص آنست که تئوری‌های فیزیکی و آبجکتیو، چنین نظرگاهی را وا می‌گذارند. استناد به حسّ درونی خفّاش‌ها در هنگام ردّیابی حشره‌ها از طریق امواج صوتی (تکانه‌های صادره از بازتاب‌های محیطی).

اهمیت و دشواری تبیین آگاهی پدیداری

- آگاهیِ پدیداری از موضوعات محوری فلسفه ذهن است که علیرغم همگانی و روزمرّه بودن تجربه آن، توضیحش در چارچوب تحلیل‌های فلسفی همچنان به نتیجه نرسیده و بلکه بر ابهامات آن افزوده شده است. دانشنامه آکسفورد در باب ذهن، آگاهی را واضح‌ترین و در عین حال رازآمیزترین خصیصه ذهن ما معرّفی می‌کند. چراکه «علوم تجربی توانسته اسرار بسیاری از پدیده‌های طبیعی – مانند مغناطیسم، فتوسنتز، هضم و گوارش، تولید نسل و ... – را آشکار سازد؛ ولی آگاهی اصلاً مشابه موارد مذکور به نظر نمی‌رسد».
- به تعبیر چالمرز، ما دلایل خوبی برای اعتقاد به این‌که آگاهی از یک سیستم فیزیکی – مانند مغز – ناشی می‌شود، در اختیار داریم. ولی ایده‌های چندان جالبی نداریم که چگونگی این فرآیند و حتّی چرایی آن را توضیح بدهد؛ یعنی این‌که چطور یک سیستم فیزیکی مانند مغز، می‌تواند یک "تجربه‌کننده" باشد؟ تا آن‌جا که دیوید چالمرز از این مسئله به «معضل دشوار» یاد می‌کند.)

تبیین کارکردگرایانه آگاهی

- تئوری‌های مختلفی در تبیین ذهن و حالات آگاهانه وجود دارد که به ترتیب تاریخی، با «رفتارگرایی» آغاز شد و بعد به نظریه این‌همانی مغز و ذهن رسید و سپس با طرح اشکالاتی (از جمله تحقق‌پذیری چندگانه: Multiple Realizibility) به طرح دیدگاه کارکردگرایی (Functionalism) توسط پاتنم و دیگران انجامید.
- اهمیت فرآیند پردازش اطلاعات در ساختار کارکردی ورودی‌ها و خروجی‌ها و عدم اهمیت محقق‌ساز (Realizer) که می‌تواند رشته‌های نورونی یا مغز سیلیکونی یا تراشه‌های الکتریکی یا ... باشد.
- پس از کارکردگرایی، دسته نظریات بازنمودگرایی برای تبیین آگاهی به میدان آمده‌اند، اما به هرحال، تلفیق کارکردگرایی با دیدگاه‌های محاسباتی (Computationalism) به ارائه تئوری‌هایی انجامید که تبیینی فلسفی برای امکان آگاهی در ماشین و هوش مصنوعی فراهم می‌کرد.

هدف‌گذاری آگاهی در هوش مصنوعی

- هوش مصنوعی قوی (Strong AI)، در پی خلق اشخاص هوشمند است؛ ماشین‌هایی که تمام قدرت‌های ذهنی ما ـ از جمله آگاهی پدیداری ـ 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 که با برنامه‌ریزی درست می‌تواند حقیقتاً آگاهی و حالات ذهنی داشته باشد. جان سرل – فیلسوف دانشگاه برکلی - این دیدگاه را هوش مصنوعی قوی می‌نامد و از دیدگاه محتاطانه‌تری که می‌گوید کامپیوتر یک ابزار مفید و کارآمد برای شبیه‌سازی ذهن و آزمایش فرضیات است (و آن را هوش مصنوعی ضعیف می‌نامد)، متمایز می‌کند.

نظریه محاسباتی ذهن: CTM

- ذهن ما واقعاً (و نه به وجه تمثیل)، یک کامپیوتر است و با دستکاری سمبل‌ها سروکار دارد. اساساً کلیه فرآیندهای ذهنی، حاصل دستکاری سمبل‌های "زبان فکر" است؛ چه این سمبل‌ها از جنس نورون باشد (در ذهن جاندار) و چه از جنس دیجیتال (در کامپیوتر).
- زبان فکر (LOT) – یا Mentalese – همانند زبان طبیعی سمبل‌ها (کلمات)ی دارد که با گرامر خاص خودشان، جملات مرکّب و معانی را می‌سازند.
- این اصل مهم، راهنمای هوش مصنوعی است. بدین ترتیب که ذهن می‌تواند به مثابه یک ماشین تفکرکننده لحاظ شود و یک ماشین می‌تواند فکر کند.
- البته قبلاً کامپیوترها با دستکاری داده‌ها (طبق دستورالعمل‌های برنامه‌نویسی) به حلّ مسئله می‌پرداختند؛ اما علیّت بازنمایی‌ها برای رفتار و تفکّر خردمندانه تبیین نشده بود. در این تئوری، بازنمایی ابژه‌ها توسّط نمادهای LOT و تولید معنا از طریق نحو را داریم.

ماشین تورینگ و برنامه GPS:

- آلن تورینگ پس از ساخت موفق اوّلین کامپیوتر عملگر جهان در جنگ جهانی اوّل و تحقیقات بعدی، در مقاله ۱۹۵۰ با عنوان Computing, Machinery and Intelligence، دستور کار پژوهش‌های نیم قرن بعد برای ساخت کامپیوترهای پیشرفته را ارائه کرد: تصمیم‌گیری، درک زبان، اثبات قضایا، یادگیری ماشین و ... که همگی از طریق کدگشایی سمبل‌ها صورت می‌گیرد. او آزمون مشهور تورینگ را هم همین‌جا ارائه کرد تا «فکرکردن» کامپیوتر را با «تمایزناپذیری زبانی» جایگزین کند.
- نول و سیمون در ۱۹۵۶، برنامه‌ای را به وجود آورد که نظریه منطق نامیده می‌شد و در ۱۹۵۷ نیز نسخه جدیدتر آن با نام حلّ مسئله عمومی (**GPS**)، عرضه گردید که تکنیک‌هایی را برای حلّ مسائل و اثبات قضایای دشوار ریاضیاتی به کار می‌گرفت.
- ماشین تورینگ و برنامه نول و سیمون، هرچند با دستکاری سمبول‌ها کار می‌کردند، اما هیچکدام در مورد علیّت برای رفتار خردمندانه و تفکّر خردمندانه توضیحی نداشتند.

ایده‌های مرکزی CTM:

- اوّلین ایده این بود که حالات ذهنی التفاتی را با رویکرد محاسباتی تبیین نماید. مثلا تفکّر درباره این‌که: "۷، یک عدد اوّل است"، یک گرایش ذهنی (مثل باور یا تفکّر) به یک گزاره را در بر دارد. این فرآیند بازنمایی، از طریق سمبل‌ها انجام می‌گیرد (دقیقا مثل زبان طبیعی) و جعبه باور شما، در بردارنده مصداقی از یک ساختار نمادین است که معنای آن اینست: ۷ عدد اوّل است.

- ایده دوم: بوسیله دستکاری سمبل‌ها که منحصراً بر مبنای ویژگی‌های نحوی (Syntax) آن‌ها صورت می‌گیرد، امکان استخراج نمادهای دیگر که به نحوی منطقی از آن‌ها تبعیت می‌کنند، وجود دارد. بدین ترتیب نحو و ساختار می‌تواند معنا (Semantic) 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 در قالب ساختارهای نحوی اظهار می‌دارد. تأثیر هر بازنمایی ذهنی به وسیله همین ساختار آن مشخّص می‌شود؛ بسیار شبیه به شیوه‌ای که هندسه یک کلید، مشخّص می‌کند چه قفلی را باز خواهد کرد. (See: Fodor, 1987.)

سیستم سمبل‌های فیزیکی: PSS

- یک دیدگاه بسیار نزدیک به CTM، فرضیه سیستم نمادین فیزیکی (PSSH) بود که توسّط نول و سیمون در ۱۹۷۶ مطرح شد. طبق این فرضیه، یک سیستم نماد فیزیکی «تمام ابزار لازم و کافی برای عملکرد هوش عمومی را داراست» (Newell & Simon, 1976) سیستم سمبل‌های فیزیکی، ماشینی است که مجموعه‌ای از ساختارهای سمبلیک را تولید می‌کند. CTMو PSSH دو نظریه‌ای هستند که هوش مصنوعی کلاسیک یا آن‌چه که هاگلند، GOFAI می‌نامد را توصیف می‌کنند.

سیستماتیک بودن زبان فکر:

- تفکّرات ذاتاً و به روشی سیستماتیک با یکدیگر ارتباط دارند؛ یعنی با شناسایی فرمول ساختار جمله (مثلاً: فاعل و فعل و مفعول)، می‌توان جملات معنادار را ساخت. کسی که جمله «علی، دوست صمیمی رضا است» را درک کند، لوازم درک جمله «رضا، دوست صمیمی علی است» را هم در اختیار دارد. 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.)
- تبیین سیستماتیک بودن و بارور بودن، نیازمند روابط نحوی در میان بازنمایی‌های ذهنی، و نیز فرآیندی است که به چنین ساختار درونی حسّاس باشد. همانند آن‌چه که فرضیه زبان تفکّر فراهم می‌آورد.

(See: Fodor & Pylyshyn, 1988.)

ساختارمندی حامل‌های فیزیکی محتوا:

- CTM می‌گوید ساختار محتوا با ساختار جمله‌ای که آن را بیان می‌کند، تناظر یک‌به‌یک دارد. مثلاً اجزا و ساختار باور و فکر به "حسن، باهوش است"، با اجزاء جمله‌ای که بیان‌گر این باور و فکر است، تناظر یک‌به‌یک دارد.
- محتوا تحقّق‌بخش فیزیکی دارد؛ مثلاً جوهر ماژیکی که جمله خاصّی را محقّق می‌سازد، یا یک صوت خاص با طول موج خاص، یا علامات دیجیتالی در کامپیوتر، که حامل (Vehicle) و تحقّق‌بخش فیزیکی آن گزاره هستند. جملات ساختارمند، در حامل‌های فیزیکی ساختارمند محقّق (Realize) می‌شوند.
- CTM ادّعا می‌کند اگر بین ساختار محتوا و ساختار جمله‌ای که آن را بیان می‌کند، تناظر یک به یک داشته باشیم، و ساختار جمله نیز با ساختار حامل، تناظر یک به یک داشته باشد، پس ساختار محتوا، باید با ساختار حامل فیزیکی، تناظر داشته باشد.

زمستان GOFAI

- رویکرد کلاسیک به هوش مصنوعی (با اسم مستعار GOFAI)، از اواسط دهه ۵۰ تا اواسط دهه ۸۰ رویکرد غالب (و نه منحصر) در هوش مصنوعی بود. در آن روزها شبکه‌های نورونی مغز هم به مثابه دستگاه‌هایی لحاظ می‌شدند که قادر به حفظ و انتقال اطلاعات هستند.
- اما انتقادات گسترده‌ای وارد شد که ناتوانی آن‌ها برای انجام واکنش‌های هوشمندانه در محیط‌های متنوّع را هدف می‌گرفت. برخی منتقدان – که قابل توجّه‌ترین آن‌ها دریفوس بود – پیش‌بینی کردند که ماشین‌ها هرگز با سطح مهارت انسانی برابری نخواهند کرد.
- مشخّص شده بود مسائلی که ما فکر می‌کردیم مشکل باشند، از حلّ قضایای ریاضیاتی و بازی شطرنج گرفته تا راه‌حل‌های شیمی و داروسازی، آسان بودند و کامپیوترها با چندهزار فرمان در ثانیه، غالباً می‌توانستند نتایج رضایت‌بخشی را فراهم آورند. مسئله دشوار اما مهارت‌هایی بودند که هر بچّه پنج‌ساله هم داراست؛ مثل درک یک کارتون انیمیشینی.

اشکالات نظریه محاسباتی:

1. مسئله حس عمومی یا متناسب بودن رفتار:

گاه در ادبیات بحث، از مسئله حس عمومی (Common Sense) با عنوان مسئله چارچوب (Frame Problem) نیز یاد می‌شود. اما تقریرات مختلفی از مسئله چارچوب ارائه شده که بعضاً با مسئله حس ّعمومی متفاوت است که با «متناسب و مربوط بودن رفتار» (Relevance) سر و کار دارد.

دریفوس استدلال می‌کند که توانایی ما در شناخت جهان و دیگر مردمان، یک نوع مهارت غیر توصیفی از سنخ دانستنِ چگونگی (علم به مهارت: know how) است که قابل تقلیل به کدگذاری‌های گزاره‌ای در برنامه‌نویسی نیست. این مهارت‌ها غیرقابل بیان و پیشامفهومی‌اند و یک بُعد پدیدارشناسی دارند که نمی‌تواند به وسیله هیچ سیستم قانون‌محور به تصویر درآید.

- «مرتبط بودن» واکنش‌ها به هر موقعیت، اهمیت توانایی انسان‌ها در تشخیص امر دارای موضوعیت از غیر آن را نشان می‌دهد و این‌که می‌توانند بدون زحمت براساس ملزومات تجربه و معرفت خویش در هر موقعیت، عمل کنند. معضل مرتبط و متناسب بودن، همچنان به مثابه یک چالش فنّی کلیدی در برابر هوش مصنوعی و نیز علوم شناختی محاسباتی باقی مانده است.
- متناسب و مرتبط بودن در رفتار اجتماعی انسان‌ها، بسرعت و با تکیه بر حسّ عمومی، حلّ و فصل می‌شود. انسان‌ها در موقعیت‌های متنوّع، رفتار و واکنشی معقول و متناسب انجام می‌دهند، بدون آن‌که نیاز به مراجعه و پیروی از دستورالعمل‌ها داشته باشند. محقّقان هوش مصنوعی به این نتیجه اساسی رسیدند که رفتار هوشمندانه 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، مستلزم فهم کانتکست و جهان است که بسیار فراتر از ظرفیت برنامه‌های کامپیوتری است.» (Lenat, 1984, p. 204.)

- مثلاً 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؟ لاجرم باید درباره مردم، مجسّمه‌ها، مسافرین هوایی، اندازه بارهایی که به وسیله هواپیما حمل می‌شود، اندازه و مکان مجسّمه آزادی، چگونگی دیدن ابژه‌ها از فاصله دور، و بسیاری واقعیت‌ها و تخمین‌های دیگر بدانیم. (Lenat & Feigenbaum, 1991)

- 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.

- به بیان دریفوس، شاید هیچ دسته واقعیت‌های رها از کانتکست (Context-free) وجود نداشته باشد و ما باید فقط از تجربیات وسیع خود بیاموزیم که چطور به هر مورد پاسخ بدهیم.
- در دیدگاه پیوندگرایی، یک شبکه نورونی شامل تعدادی از واحدهاست که مانند نورن‌های مغز هستند. هر واحد، تعدادی از سیگنال‌های ورودی را دریافت و یک سیگنال خروجی تحویل می‌دهد. واحدها با یکدیگر اتّصال دارند و خروجی یک واحد، ورودی واحد دیگر است.
- مرتبط بودن واکنش‌ها در انسان‌ها بر تجربیات پدیداری آن‌ها استوار می‌شود. اما شبکه‌های نورونی مصنوعی، چنین توانایی‌ای را بر پایه دسته‌بندی رفتارها برنامه‌ریزی می‌کنند. دستگاه‌های پیوندگرا قادر به تعمیم هستند و هرگاه یک الگوی ورودی جدید را – که شبیه الگوی تمرین‌شده قبلی است – تمرین می‌کنند، خروجی‌ای مشابه تولید خواهند کرد.
- دریفوس و حامیانش، اشکال مرتبط بودن را در رویکرد پیوندگرایی هم دارند. «از یک نوع بودن» را در دسته بندی و تعمیم رفتارها، چه چیزی مشخّص می‌کند؟ مدل‌سازی نمی‌تواند جلوی تعمیم‌های ناجور و بیجا را بگیرد.

2. مسئله اتاق چینی:

- جان سرل استدلال خود 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.
- او خودش را در اتاقی مجسّم کرده که یک کتاب راهنمای سمبل‌های زبان چینی در آن وجود دارد، و یک برنامه به زبان انگلیسی که کار با این کتاب را فقط از روی شکل سمبل‌های آن میسّر می‌سازد. (جایگزین برنامه 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 راهنما (جایگزین بانک اطلاعات در کامپیوترها) نگاه می‌کند و سمبل‌های متناسب با کاراکترهای دریافتی را گزینش و از اتاق به بیرون می‌فرستد که در واقع، همان پاسخ به سؤالات هستند. احتمالاً پاسخ‌های مذکور، آن‌قدر خوب باشند که شخص چینی زبان بیرون اتاق را راضی سازد. ولی در واقع سرل چیزی از پرسش‌ها درک نکرده است. پس رویکرد محاسباتی نمی‌تواند آگاهی را تبیین کند. (See: Searle, 1984, p. 39.)

- سرل ۱۷ سال بعد و در کتاب *راز آگاهی*، با اشاره به بیش از ۱۰۰ مقاله‌ای که در پاسخ به استدلال وی به چاپ رسیده، توضیح می‌دهد که غالب آن‌ها هسته مرکزی اشکال اتاق چینی را درنیافته‌اند. لذا استدلال را در سه مقدّمه صورت‌بندی می‌کند:

1) برنامه‌ها کاملاً نحوی هستند.
2) ذهن‌ها دارای معناشناسی هستند.
3) نحو و ساختار، نه معادل با و نه فی‌نفسه کافی برای جنبه‌های معناشناسی است.

بنابراین ذهن‌ها همان برنامه‌ها نیستند.

- 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.)

- 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 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.

۳. مسئله خصیصه سابجکتیو تجربه پدیداری:

- از مهم‌ترین اشکالات بر استراتژی محاسباتی اینست که اساساً جنبه پدیداری تجربیات و احساسات را نادیده می‌گیرد. بدین ترتیب می‌توان گفت رویکرد محاسباتی عمدتا به آگاهی دسترسی و معرفت گزاره‌ای نظر دارد.
- این نکته البته قلب اشکال دریفوس درباره حسّ عمومی را شکل می‌داد؛ اما فقدان پدیدارشناسی در استدلال دریفوس، تنها ناظر به ارائه پاسخ‌های متناسب بود و نه درباره کلیّت حضور و اهمیت آگاهی پدیداری.

- شکاف تبیینی (Explanatory Gap) همین‌جا بروز می‌کند که هرچه درباره وقایع فیزیکی و کارکردی دیدن رنگ بگویید، همچنان خواهم پرسید: چرا این وقایع فیزیکی و کارکردی، *آن* خصیصه پدیداری خاص را تولید می‌کنند و نه خصیصه پدیداری دیگری را؟ یا "مسئله دشوار" (Hard Problem) – همسایه معضل شکاف تبیینی – که 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)
- مادام که یک تبیین درست برای چگونگی و چرایی بروز آگاهی از فرآیند خاص پردازش اطلاعات یا فرآیند محاسباتی خاص وجود نداشته باشد، این شکاف وجود دارد.
- از رهگذر همین چالش بوده که پروژه آگاهی ماشین، با استفاده از محاسبات و ساخت ابزاری مصنوعی، معمولاً به مدل کردن «رفتارها و عملکردها»ی آگاهانه ختم می‌شود.

- مهم‌تر از خصیصه پدیداری شاید «اراده آزاد» باشد که معلوم نیست چطور می‌تواند در رویکرد محاسباتی تبیین شود. در منابع کلاسیک فلسفه اسلامی اساساً دلیل بر وجود نفس را از آثاری که «لا علی وتیرۀ واحدۀ» (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. اما روبات‌ها – مهم نیست چقدر پیچیدگی داشته باشند – به وسیله قوانین متعیّن (Deterministic Laws) یا لااقل با تمرین و کارآموزی مبتنی بر قواعد خاص مدیریت می‌شوند و در نتیجه ممکن است حتّی زمینه و قابلیت اراده آزاد را نیز نداشته باشند.

- البته نظریه محاسباتی برای تبیین حالات التفاتی ارائه، اما بعدها به منظور نیل به چشم‌انداز هوش مصنوعی قوی به کار گرفته شد و بلکه تمام حالات ذهن (از جمله حالات پدیداری) نیز طبق دیدگاه پرطرفدار بازنمودگرایی، به حالاتی التفاتی تقلیل داده شد. در نتیجه اشکالات مربوط به آگاهی پدیداری و اراده آزاد مطرح می‌شوند.

هوش مصنوعی بدن‌مند

- امروزه دیگر اهمیّت بدن‌مندی (EMBODIMENT) و موقعیت‌مندی تقریباً به نحوی جهانی پذیرفته شده و مکرّراً اظهاراتی از این دست که «هیچ مغزی/ذهنی بدون یک بدن وجود ندارد» رواج یافته است. این البته مرهون سرخوردگی‌هایی بود که نسبت به هوش کلاسیک (GOFAL) ظهور کرد و نگرش جدید را در چند دهه اخیر به وجود آورد که مفهومی‌سازی هوش بمثابه یک فرآیند صرفاً کامپیوتری را کافی نمی‌داند؛ آگاهی به یک ارگانیسم کامل فیزیکی – به یک بدن – برای تعامل با جهان خارج احتیاج دارند.
- در واقع 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. ایده‌هایی شبیه به این، اصول عقاید رویکرد رفتارگرا (در برابر رویکردهای تفکّرگرا) به هوش مصنوعی را تشکیل می‌داد که در سرتاسر آزمایشگاه‌های مهم و مطرح هوش مصنوعی در اوایل دهه ۹۰ (بعد از زمستان GOFAI)، به چشم می‌خورد.

- 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.
- 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. امروزه شناخت و عمل، نتیجه یک ظهور جهشی و رو به رشد لحاظ می‌شود و نه چیزی که بتواند مستقیماً در یک روبات برنامه‌ریزی شود. چنین رویکردی عمیقاً بر بصیرت‌های دریافتی از سیر رشد کودک، تکیه دارد. (Lungarella & Iida & Bongard & Pfeifer, 2007.)
- محققان هوش مصنوعی عطف توجّه از وظایف نمادین مرتبه بالاتر (مانند استدلال قیاسی) به سمت وظایف حسّی حرکتی – که ظاهراً ساده‌تر بودند – را خواستار شدند؛ وظایفی مانند احساس‌کردن، حرکت کردن، دور زدن، به دست گرفتن، پرهیز از موانع و ... آن‌ها می‌گفتند مطالعه توانایی‌های مذکور می‌تواند روشن کند که چطور شناخت مرتبه بالاتر می‌تواند از چنین فعالیت‌هایی ظهور نماید.

بازگشت اشکالات

- با وجود اهمیت بدن‌مندی، همه پژوهشگران این حوزه قبول ندارند که بدن‌مندی در آگاهی ماشین، محوریت دارد. 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» (For-me-ness) را به همراه دارد.
- هرچند توافقی نسبی در مورد محوریت تصوّر و تخیّل برای پروژه آگاهی ماشین در بین بخشی از مهندسین هوش مصنوعی وجود دارد. اما پرسش‌هایی مطرح شده – و البته همچنان باقی مانده – که تخیّل و تصوّر از چه مؤلفه‌هایی تشکیل شده و چطور با آگاهی ارتباط دارد؟ و چطور می‌توان درک اول‌شخص آن را بازسازی کرد؟

- یکی از تلاش‌های مفصّل برای استدلال راجع به چگونگی عالَم درونی ماشین، در کار هِسلو و جیرِنهِد دیده می‌شود. رویکرد آن‌ها بر فرضیه شبیه‌سازی هسلو تکیه دارد که به توانایی شبیه‌سازی تعاملات با جهان خارج وابسته است.
- کریسلی و پارث‌مور چشم‌انداز دیگری در مورد نقش محوری تصوّر در آگاهی ماشین، ارائه می‌دهند. آن‌ها با اتّخاذ یک رویکرد حسّی حرکتی، آگاهی را در بردارنده توانایی تصوّر ورودی‌های حسّی فرض می‌کنند که باید برای حرکت به هر سمتی، دریافت بشود.
- حداقل سابجکتیویتی در نظریه کیورستین با عنوان "دیدگاه حسّی حرکتی دینامیک": DSM اما فقط نیازمند تمرین مهارت‌های حسّی حرکتی ماشین است. او اعتقاد دارد آگاهی ماشین از تمرین غلبه نظم‌های حسّی حرکتی ناشی می‌شود.
- با تمام این تلاش‌ها، تردیدهایی جدّی درباره خروجی این پروژه وجود دارد؛ به عنوان مثال تورنس اکثر محقّقین آگاهی مصنوعی را متهم می‌کند که در محوریت پدیدارشناسی (و منظر اوّل شخص) برای آگاهی، قصور می‌ورزند. (See: Clowes & Torrance & Chrisley, 2007.)

تفکرگرایی در رویکرد محاسباتی و رفتارگرایی در رویکرد بدن‌مندی

- راسل و نورویگ در AIMA *(رویکردی مدرن به هوش مصنوعی)* – که یکی از مهم‌ترین و جامع‌ترین منابع آکادمیک در این حوزه محسوب می‌شود – عمده رویکردها در هوش مصنوعی را در چهار تعریف خلاصه کرده‌اند که براساس یک جدول دو بُعدی استخراج شده است. یک بُعد را دو رویکرد تفکّر و استدلال در برابر معیارهای رفتاری شکل می‌دهد. بُعد دیگر، دوگانه وفاداری به کارآیی انسانی را در مقابل کارآیی ایده‌آل مدّ نظر دارد. در نتیجه چهار رویکرد تفکر انسانی، رفتار انسانی، تفکر ایده‌آل، رفتار ایده‌آل شکل می‌گیرد که در هر چهار رویکرد در ادبیات بحث، طرفدارانی داشته و دارد.
- بدن‌مندی در واقع احیاء رویکرد رفتاری در برابر رویکرد تفکری محاسباتی بود و تمام اشکالات رفتارگرایی در برابر بدن‌مندی نیز قابل بازسازی است.

بازسازی اشکال سرل علیه روبات‌ها

- سرل خود تعدادی اشکال به استدلال خویش مطرح کرده که یکی از آن‌ها «جواب روبات» است که می‌گوید اگر خروجی با هدایت‌کردن یک روبات مورد استفاده قرار گیرد، به گونه‌ای که رفتار روبات متناسب با کلمات باشد، کلیّت روبات درک خواهد داشت. مثلاً یک کامپیوتر دیجیتال به پرسش «چه می‌کنید اگر بوی دود سیگار استشمام نمایید؟» پاسخ می‌داد: «ترک اتاق»؛ بدون آن‌که درکی از این پاسخ داشته باشد. اما فرض کنید که یک روبات، حس‌ّگر بو داشته باشد و بتواند پاسخ مناسبش را به مکانیسم‌هایی منتقل نماید که آن را برای ترک اتاق حرکت بدهد.
- پاسخ سرل به جواب روبات اینست: تصوّر کنید که او در اتاق چینی در واقع درون یک روبات قرار دارد. مانند قبل از روی دفترچه راهنما ولی این‌بار نه فقط سخنان روبات بلکه مثلاً با تکان‌دادن اهرم‌هایی، رفتارهای متناسبش را شکل خواهد داد. سرل توجّه می‌دهد که او در این سناریو نیز همچنان هیچ چیزی از معانی را نمی‌فهمد.

- این استدلال را نه فقط برای درک زبان، بلکه برای هیجانات و احساسات هم می‌توان بازسازی کرد. مواجهه با صحنه‌ای که شادی‌آور/ هیجان‌انگیز است، انسان داخل روبات را به حرکت دادن اهرم‌های متّصل به بازوها و اندام مرتبط می‌اندازد و نهایتاً شاهد مثلاً رفتار دست زدن و بالا پایین پریدن ناشی از خوشحالی باشیم. اما در واقع نه شخص داخل روبات و نه کلّ سیستم مشتمل بر شخص داخل روبات و کلیّت روبات و ورودی‌ها و خروجی‌ها، هیچ حسّ پدیداری از شادی و هیجان ندارند.

- پروژه آگاهی ماشین، دو مسیر متفاوت را دنبال کرده است؛ یکی استفاده از محاسبات و ساخت ابزاری مصنوعی برای مدل کردن آگاهی (Machine Modeling of Consciousness: MMC) که معمولاً به مدل کردن «رفتارها و عملکردها»ی آگاهانه ختم می‌شود و دیگری، تلاش جهت خلق مصنوعی آگاهی. رویکرد دوم را شاید بتوان آگاهی قوی ماشین نامید. تا کنون مقالات بسیاری به منظور پیشبرد پروژه MMC نوشته شده، اما اکثر کسانی که به طور خاص بر پروژه آگاهی قوی ماشین تمرکز کرده‌اند، متمرکز بر چشم‌اندازی عمیق‌تر هستند.

امکان ظهور آگاهی مصنوعی برپایه حکمت متعالیه

- صدرالمتألهین در پی نظریه حدوث جسمانی نفس، بر این نتیجه تأکید دارد که نفس را نمی‌تواند از ابتدا و در حین پیدایش، مجرّد محض دانست؛ بلکه از آن‌جا که صورت جسم طبیعی و منوّع یک نوع مادّی است که در مرحله جنینی و به صورت تشدید شونده از سطح نوروفیزیولوژیک پدید می‌آید.
- بر این اساس، علی‌الاصول چه دلیلی بر محال بودن پیدایش نفس 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 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 آگاهی بشوند.

- به ویژه آن‌که ساخت شبکه‌های نورونی در رویکردهای نوین، به سوی فیزیولوژی انسانی پیش رفته و شبیه‌سازی خاستگاه ظهور آگاهی – یعنی شبکه نورونی سلولی – مدّ نظر قرار گرفته است. (See: Pagel & Kirshtein, 2017)
- 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 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 کیفیتی که در انسان شاهد هستیم، تولید نمایند. اما – مطابق نظریه صدرا – این احتمال و امکان را نمی‌توان نفی کرد که سطوحی از جوهر نفسانی و آگاهی مصنوعی در این مسیر به ظهور برسد.

منابع:

- Dreyfus, H., 1999, *What Computers Still Can't Do*, MIT Press.
- Fodor, J., 1987, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, MA: MIT Press.
- Fodor, J., & Pylyshyn, Z., 1988, "Connectionism and Cognitive Architecture: A Critical Analysis," *Cognition*, No. 28, pp. 3–71.
- Haugeland, J., 1985, *Artificial Intelligence: The Very Idea,* MA:The MIT Press.
- Lenat, D., 1984, "Computer Software for Intelligent Systems," *Scientific American*.
- Lenat D. & Feigenbaum, E., 1991, "On the Thresholds of Knowledge," *Artificial Intelligence*, Vol. 47, Nos. 1-3.
- 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.

- Nagel, T., 1974, "What is it like to be a bat?", *The Philosophical Review*, Vol.83, No. 4., pp. 435-450.
- Newell, A. & Simon, h., 1976, "Computer science as empirical inquiry: symbols and search", *Communications of the ACM*, Vol. 19., No. 3., pp. 113-126.
- Searl, J., 1984, *Minds, Brains, and Science,* Cambridge, MA: Harvard University Press.
- ----------, 1997, *The Mystery of Consciousness*, New York, NY: New York Review of Books.
- Sun, R., 2014, Connectionism and neural networks, In: Franklin, Keith & Ramsy, William (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press.
- Tye, M., 2009, *Consciousness Revisited; Materialism Without Phenomenal Concepts,* The MIT Press.