

فلسفه

مدل کامپیوتری ذهن و هوش مصنوعی

دکتر محمود خاتمی

چکیده:

بحث درباره مدل کامپیوتری ذهن و هوش مصنوعی از مباحث بسیار مهم در فلسفه ذهن است. در این مقاله به صورت بسیار خلاصه و اجمالی به دو صورت از نظریه همسانی ذهن و کامپیوتر اشاره می‌شود تا مسئله مورد بحث روشن تر گردد سپس آزمون تورینگ ذکر می‌شود که نقش اساسی در پذیرش فرضیه هوش مصنوعی به معنی قوی دارد؛ آنگاه به انتقادات برجسته‌ای که از این آزمون و مدل کامپیوتری ذهن و هوش مصنوعی شده اشاره خواهد شد.

پرتال جامع علوم انسانی



پروہ شگاہ علوم انسانی و مطالعات فرہنگی
پرتال جامع علوم انسانی

طرح مسئله

پس از آن که پاتنم^۱ روایت اوّل خود از کارکردگرایی^۲ را با کمک ماشین محاسبه تورینگ^۳ صورتبندی کرد (۱۹۶۷)، اندیشه یکسان تلقی کردن ذهن با کامپیوتر نیز به طور جدی در فلسفه ذهن مطرح شد. در حالت کلی، این اندیشه حاوی دو دیدگاه است:

الف - دیدگاه کامپیوتری کردن^۴ ذهن: این دیدگاه اندیشه یکسانی ذهن و کامپیوتر را از این حیث بررسی می‌کند که می‌توان ذهن انسان را همچون کامپیوتر دانست. کامپیوتر ماشینی است که اطلاعات را دریافت می‌کند و به محاسبه، تفسیر، نگارش و پردازش می‌پردازد. این همه رفتارهایی هستند که ذهن هوشمند انسان در حالت عادی انجام می‌دهد. از این رو می‌توان به هوش انسان هم همچون پردازش اطلاعات کامپیوتری نظر کرد. این نظر بر دو فرض اصلی استوار است:

الف ۱- ذهن انسان دقیقاً شبیه کامپیوتر عمل می‌کند.

الف ۲- توانایی‌های روانی و شناختی ذهن انسان را می‌توان از موضع "شخص ثالث" و با طراحی سخت‌افزارهای پیشرفته عملی ساخت.^۵

این دو فرض، ابده اصلی کارکردگرایی

ماشینی را تشکیل می‌دهند. بنابراین دو فرض، حالات روانی و شناختی ذهن انسان همچون پردازش کامپیوتری است؛ و همان طور که هر پردازش کامپیوتری (مثلاً اجرای محاسبات عددی) توسط ماشین‌های محاسبه‌گر (که از لحاظ فیزیکی و ساختاری متفاوتند) قابل اجراست، حالات ذهنی هم به طرق گوناگون تحقق می‌یابند. این بدان معنی است که خواص روانی - شناختی ذهن از ساختار فیزیکی که آن را اجرا و محقق می‌نماید متمایزاند و بنابراین چنانکه ساختار زیستی انسان می‌تواند قابلیت‌های هوشی - روانی خاصی را به فعلیت درآورد، می‌توان تصور کرد که ساختار فیزیکی متفاوتی نیز بتواند همین قابلیت‌ها را به فعلیت درآورد. این دیدگاه زمینه لازم را برای طرح دیدگاه دوّم فراهم می‌آورد.

ب - دیدگاه هوش مصنوعی^۶: این دیدگاه اندیشه یکسانی ذهن و کامپیوتر را از حیث امکان ساخت کامپیوتری که بتواند حالاتی مشابه حالات ذهنی انسان داشته باشد بررسی می‌نماید. بررسی این مکان در سه سطح و براساس سه فرض انجام می‌گیرد:

ب ۱- کامپیوتر اعمالی هوشمند - نظیر اعمال ذهن انسان (مثل محاسبه و تفسیر) - انجام می‌دهد.

ب ۲- کامپیوتر این اعمال را به همان شیوه انجام می‌دهد که ذهن انسان.

ب ۳- فرض ب ۲ به این معنی است که کامپیوتر می‌تواند خواص روانی - شناختی (نظیر آگاهی و تفکر) داشته باشد.

فرض اول (ب ۱) در حیطه مهندسی کامپیوتر بررسی می‌شود و به نظریه ضعیف هوش مصنوعی مشهور است. دو فرض بعدی (ب ۲ و ب ۳) در حیطه روان شناسی شناختی و فلسفه ذهن بررسی می‌شوند و معمولاً به نظریه قوی هوش مصنوعی (و گاه به ترتیب قوی و اقوی) شهرت دارند:^۷

آزمون تورینگ

در تأیید مدل کامپیوتری ذهن و دفاع از نظریه هوش مصنوعی به معنی قوی کلمه عمدتاً آزمون تورینگ مطرح می‌شود.^۸ این آزمون به قابلیت‌های انجام کار یک ماشین کامپیوتر در برابر یک انسان - با هوش متوسط - که از تمام جهات دارای ذهنیت متعارفی است می‌پردازد. طرح اصلی این است که ماشین کامپیوتری بتواند به همان خوبی که آن فرد اعمال و وظایف هوشمندانه خود را ایفا می‌کند اعمال مشابه را اجرا نماید. آزمون تورینگ را به صور مختلفی می‌توان تقریر کرد:

تقریر اول:

تورینگ وضعیتی را توصیف می‌کند که بتوان به کمک آن آزمایش را انجام داد: سه بازیکن در نظر می‌گیریم. یک پرسشگر، یک زن و یک مرد. پرسشگر در اتاقی است که از آن دو دیگر جداست. مرد و زن به عنوان فرد الف و ب معرفی شده‌اند (و جنسیت آنها برای پرسشگر نامعلوم است). هدف پرسشگر آن است که نشان دهد کدام یک زن است و کدام مرد. او این کار را براساس طرح پرسش‌هایی از طریق ترمینال کامپیوتری انجام می‌دهد. وظیفه مرد آن است که پرسشگر را حتی المقدور گمراه کند ولی وظیفه زن آن است که پاسخ درست به پرسشگر دهد.

در این مثال، تورینگ می‌گوید که می‌توان تصور کرد که به جای مرد از ماشین کامپیوتری استفاده کنیم و صورت مثال این شود که پرسشگر معین دارد کدام یک از الف یا ب ماشین و کدام انسان است. ماشین هم طوری برنامه‌ریزی شده که پرسشگر را گمراه نماید.

اکنون مسئله این است که آیا ماشین کامپیوتری می‌تواند به همان خوبی که آن مرد پرسشگر را گمراه می‌کند پرسشگر را به حدس غلط وا دارد. تورینگ نتیجه می‌گیرد که اگر

ماشین چنان کند که همان مرد می‌کرد پس همچون او هوشمند است و همان ذهنیتی را که بدو منسوب است بدین هم باید نسبت داد. تقریر دوم:

آیا می‌توان از طریق طرح سؤالات به وسیله یک ترمینال کامپیوتری دریافت که طرف مقابل یک انسان است یا ماشین؟ اگر طرف مقابل، در واقع، کامپیوتری باشد که همواره ما را به خطا وادارد چنانکه گمان بریم در طرف مقابل خود انسانی به ما پاسخ می‌دهد آنگاه باید نتیجه گرفت که آن کامپیوتر از ذهنیت منسوب به انسان برخوردار است.

بررسی آزمون تورینگ

آزمون تورینگ چنان طراحی شده که هوشمندی و ذهنیت را از مباحث مربوط به ظاهر کامپیوتر (نظیر این که آیا می‌تواند صحبت کند یا حرکت کند) متمایز می‌سازد. بحث تورینگ این است که حداقل شرایط داشتن ذهنیت انسان را در کامپیوتر نیز می‌توان اجرا کرد. بسیاری اشاره کرده‌اند که این آزمون از بسیاری جهات مخدوش است. نخست آن که داشتن یک زبان - که در این آزمون مفروض گرفته شده - نباید شرط لازم برای ذهنیت باشد. به علاوه هوش انسان حیطة گسترده‌ای دارد و نمی‌توان آن را به اعمالی که در این آزمون

مورد نظر است محدود کرد. این آزمون حالات ذهنی را در سطح ادراکات حسی بررسی می‌کند و از این حیث از جامعیت کافی برخوردار نیست. دیگر آن که، آزمون تورینگ بر این فرضیه استوار است که اگر دو سیستم از نظر ورودی و خروجی معادل باشند موقعیت روان‌شناختی مشابهی دارند یعنی اگر یکی دارای ذهن باشد دیگری هم دارای ذهن خواهد بود. این فرضیه که آن را فرضیه تورینگ می‌نامند گذشته از این که مستلزم رفتارگرایی است، نمی‌تواند حالاتی از ذهنیت را توجیه کند که در اثر تفاوت در حالات درونی به وجود می‌آیند. بدین معنی که ممکن است بتوان تصور نمود که دو ماشین رفتار مشابهی (مثلاً محاسبه درست یک سلسله عملیات جبری برای هر ورودی به شکل $n+m$ و nm و n/m به ازای تمامی اعداد n و m در فاصله یک تا صد) داشته باشند ولی داشتن رفتار مشابه (که مفاد فرضیه تورینگ است) نمی‌تواند معیاری کافی برای تشخیص همسانی ذهنیت در دو سیستم جداگانه تلقی شود. یعنی برابر بودن ورودی / خروجی (که فرضیه تورینگ بیان می‌کند) لزوماً برابر بودن در ذهنیت را نتیجه نمی‌دهد.^۹ علاوه بر این انتقادات می‌توان از اشکالاتی که دریفوس و سرل بر مدل

دریغوس با یاری گرفتن از هوسرل و هیدگر امکان طراحی یک کامپیوتر را که چون انسان دارای ذهنیت باشد منتفی می‌داند.

دهد که ذهنیت انسان را نمی‌توان با محاسبات کامپیوتری - که ولو بسیار دقیق برنامه‌ای را اجرا می‌کند - برابر و یکی پنداشت.^{۱۱} به نظر او علوم شناختی به طور کلی و دیدگاه هوش مصنوعی قوی بویژه، که ذهن را با کامپیوتر یکسان می‌پندارد و نسبت ذهن و مغز را نسبت نرم‌افزار و سخت‌افزار تلقی می‌کند اساساً اشتباه است. زیرا برنامه‌های کامپیوتری از محتوای نحوی و گرامری برخوردار است. و با تعبیه نمادین و منطقی عددی (۰ و ۱) مشخص می‌شوند حال آن که ذهن از محتوای معنایی برخوردار است که از ترکیب نحوی ساده فراتر می‌رود. سرل برای تبیین موضع خود آزمون معروف اتاق چینی را طرح می‌نماید.^{۱۲} در این آزمون فرضی، شخصی که زبان چینی نمی‌داند در یک اتاق محبوس است و تعداد زیادی نماد چینی همراه با کتابچه راهنمای استفاده از آنها در اختیار او قرار دارد؛ فرض می‌شود که این شخص بدون آن که زبان چینی بداند در استفاده از این نمادها و تبدیل آنها برحسب قواعد موجود در آن کتابچه مهارت دارد. در این حال، اگر رشته‌ای از نمادهای چینی به درون اتاق

کامپیوتری ذهن و آزمون تورینگ وارد کرده‌اند نیز یاد کرد.

دریغوس با یاری گرفتن از هوسرل و هیدگر امکان طراحی یک کامپیوتر را که چون انسان دارای ذهنیت باشد منتفی می‌داند و استدلال می‌کند که در زندگی واقعی بر خلاف آزمون های فرضی (نظیر آزمون تورینگ) ذهنیت انسان قابلیت‌هایی را دارد که در واکنش به محیط انجام می‌گیرد. آنچه کامپیوتر می‌تواند انجام دهد کاملاً صریح و تعریف شده است و بنابراین می‌تواند به حل مسئله و پردازش اطلاعات و تفسیر آنها پردازد. این همه، براساس قواعد و مبانی تعریف شده‌ای انجام می‌گیرد که از قبل به کامپیوتر داده شده است. آنچه کامپیوتر نمی‌تواند انجام دهد واکنش‌های ذهنی (روانی - شناختی) انسان است که نه جنبه صوری دارند و نه براساس قواعد تعریف شده‌ای انجام می‌گیرند، بلکه در حیطه زیستی - فرهنگی خاصی صورت می‌گیرد که برای کامپیوتر امکان تجربه آن فراهم نیست.^{۱۰} سرل نیز آزمایشی فرضی را در تحدی نسبت به آزمون تورینگ طراحی کرد تا نشان

کامپیوتر براساس قواعد و علائمی که برایش تعریف شده است بدون آن که به درک معنایی عملکرد خود واقف باشد به ترکیب نحوی آنها می پردازد و پاسخ مناسب به ورودی های خود می دهد.

این شخص ارسال شود، او بسرعت می تواند پاسخ مناسب را (با استفاده از نمادها و کتابچه راهنما) به بیرون از اتاق ارسال دارد. از نظر شخصی که بیرون اتاق است روابط بین ورودی و خروجی دقیقاً همان روابطی است که اگر به جای آن شخص درون اتاق یک فرد چینی بود. با این وجود، شخص حاضر در اتاق، زبان چینی نمی داند و تنها براساس ترکیب نحوی عمل می کند و نه درک معنایی. این وضعیت به نظر سرل، دقیقاً همان وضعیتی است که در کامپیوتر اتفاق می افتد. کامپیوتر براساس قواعد و علائمی که برایش تعریف شده است بدون آن که به درک معنایی عملکرد خود واقف باشد به ترکیب نحوی آنها می پردازد و پاسخ مناسب به ورودی های خود می دهد. پس کامپیوتر فکر نمی کند و تصمیم نمی گیرد و به طور کلی از ذهنیت انسان - که درک معنایی ماورای ترکیب نحوی دارد - محروم است سرل نتیجه می گیرد که امور ذهنی و ذهنیت بیش از عملکرد پیچیده کامپیوتر است و از این رو آزمون تورینگ باطل است. استدلال سرل

مبتنی بر سه فرض است: اول آن که برنامه ها صوری (نحوی) اند. دوم آن که ذهن محتوا (معنا) دارد و سوم آن که نحو برای داشتن معنا کافی نیست. برهان اتاق چینی سرل صحت فرض سوم را تأیید می نماید. بنابراین ذهن یک برنامه صرف نیست و از این جا، سرل به نادرستی ادعای هوش مصنوعی به معنی قوی استدلال می نماید.

بسیاری بر آزمون سرل و استدلال او خدشه کرده اند. به طور عمده، عده ای می پندارند که لازمه این آزمون و استدلال این است که کامپیوترها نتوانند فکر کنند. سرل تاکید دارد که ذهن هم دارای قابلیت محاسبه کردن همچون کامپیوتر است ولی ذهن قابلیت دیگری که همان درک معنایی است دارد که کامپیوتر فاقد آن است. چنانچه محاسبه به معنی فکر کردن باشد بی گمان کامپیوتر فکر می کند و موجودات متفکر هم کامپیوترند: اما به نظر سرل، توجه به این نکته لازم است که فکر کردن همواره چیزی بیش از محاسبه کردن است، و به علاوه در محاسبه کردن گاه به

چنانچه محاسبه به معنی فکر کردن باشد بی‌گمان کامپیوتر فکر می‌کند و موجودات متفکر هم کامپیوترند.

معادل انگاشتن ذهن و کامپیوتر از نظر سرل باطل می‌شود.^{۱۳}

ملاحظه نهایی

ادله‌ای که علیه هوش مصنوعی قوی اقامه می‌شوند، به فرض تمام بودن فقط ناظر به حالتی است که ذهن همچون کامپیوتر ترسیم شود (یا نسبت ذهن و مغز مانند نرم‌افزار کامپیوتر به سخت‌افزار آن باشد)؛ لذا چنانچه بحث هوش مصنوعی با فعالیت‌هایی در حوزه ساخت و پرداخت بافت‌های زیستی - خواه در ارتباط با سخت‌افزار کامپیوتر و خواه مستقل از آن - گره خورد ورود این ادله محل مناقشه است. نکته دیگر این که، به فرض تمام بودن ادله مذکور و بطلان آزمون تورینگ و نهایتاً رد دیدگاه هوش مصنوعی منطقاً دیدگاه کامپیوتری ذهن (دیدگاه الف) که موضع کارکرد گرایی ماشینی است رد نمی‌شود. زیرا آنچه بر حسب دیدگاه کارکرد گرایی ماشینی نتیجه می‌شود این است که فقط سیستم‌هایی که ماشین تورینگ برای آنها توصیفی مناسب فراهم می‌آورد درجه مشابهی با ذهنیت دارند.

عملیات محاسبه پرداخته می‌شود و گاه به محتوای آن هم نظر می‌شود. کامپیوتر اولاً محاسبه‌ای دارد که در آن فکر نیست و ثانیاً در محاسبه کامپیوتری توجهی به محتوای نمی‌شود.

باز به سرل اشکال کرده‌اند که لازمه این استدلال آن است که مغز را به نحو ماتقدم واجد آگاهی و التفات بدانیم. به نظر او این ملازمه از استدلال او نتیجه نمی‌شود زیرا وی معتقد است که هر سیستم دیگری (غیر از مغز) نیز دارای ساختار علی لازم و کافی باشد می‌تواند واجد آگاهی و التفات هم باشند و لازم نیست که از سیستم عصبی نظیر مغز انسان برخوردار باشد. (همان طور که هواپیمای پرواز مثل پرندگان نمی‌پرد ولی از ساختار علی لازم و کافی برای غلبه بر موانع پرواز برخوردار است). به نظر سرل، مسئله این نیست که آگاهی و التفات را به نحو ماتقدم مخصوص مغز بدانیم؛ بلکه مشکل این است که آگاهی و التفات و به تعبیر درست‌تر ذهنیت فقط در سیستم‌های زیستی پیچیده به وجود می‌آیند و این دقیقاً همان چیز است که کامپیوتر فاقد آن است. بدین ترتیب ادعای هوش مصنوعی قوی برای

پی‌نوشت‌ها:

۱- هیلری پاتنم پایه‌گذار مکتب کارکردگرایی است. روایت او از این مکتب که به کارکردگرایی ماشینی معروف شد با انتشار مقاله زیر صورت گرفت.

Hilary Putnam, «Psychological predicates» (1967) reprinted as "The Nature of Mental states" in Rosenthal D., (ed), *The Nature of Mind*, (New York 1991) pp.197- 203; also see his *mind, language and reality: philosophical papers. V.2* (Cam. 1975).

۲- کارکردگرایی (Functionalism)، در سال ۱۹۶۷ با انتشار مقاله پاتنم ظهور کرد. این مقاله مختصر تمامی مباحثات در فلسفه ذهن را تحت تأثیر قرار داد و نخست زمینه انتقال از فیزیکیسم نوعی (Type physicalism) خصوصاً نظریه‌های همسانی مغز- ذهن (Identity theories) را به کارکردگرایی فراهم کرد. به علاوه، واکنشی تند علیه نظریه‌های تحویل‌گرا (Reductionism) شد. ایده اصلی نظریه کارکردگرایی در بحث از قابلیت تحقق چند گانه خواص ذهنی و مفهوم کارکردی از امور ذهنی نهفته است. روایت اولیه کارکردگرایی که به ماشینی معروف است بر اساس تشبیه ذهن به کامپیوتر انجام گرفت.

روایت دیگر کارکردگرایی که گاه غایت‌انگاران و گاه علی خوانده می‌شود به نقش علی (Causal Role) در تبیین امور ذهنی اهمیت می‌دهد (نگا):

Sydney shomaker" Some varieties of Functionalism" in his *Identity, Cause and Mind* (cambridge 1984).

هم چنین نگا: به مقاله Ned Block تحت عنوان Functionalism در:

A companion to philosophy of mind, ed.s. Guttenpland (oxford 1994).

برای انتقاداتی که از کارکردگرایی شده نگاه کنید به مقاله Ned Block تحت عنوان:

Troubles With Functionalism Rosenthal D. *The nature of mind*

ص ص ۲۱۱ تا ۲۲۹: همچنین نگا: به

(Hilary Putnam, *Representation and Reality* (cambridge 1988).

۳- Turing Machine: ماشین تورینگ طرح یک ماشین محاسبه ساده است و دارای اجزاء زیر است: (۱) نواری که به خانه‌هایی تقسیم شده و از هر دو طرف نامحدود است. (۲) وسیله خواننده - نویسنده (هد) که در هر لحظه معین روی یکی از خانه‌های نوار قرار دارد. (۳) مجموعه‌ای محدود از حالات پیکربندی

داخلی نظیر $q_0, q_1, q_2, \dots, q_n$ (الف) الفبای محدودی که از علائمی نظیر $b_1, b_2, b_3, \dots, b_n$ تشکیل شده است. در هر خانه از نوار فقط یک علامت می تواند باشد. عملکرد کلی این ماشین از این قرار است که (الف) در هر لحظه ماشین در یک حالت (مثلاً q_1) است و هد خانه مشخصی را می خواند، (ب) آنچه ماشین در زمان مفروض T انجام می دهد توسط حالت (q_1) و علامتی که هد در لحظه T خوانده مشخص می شود؛ (ج) در این حالت ماشین یکی از سه کار زیر را انجام می دهد؛ (۱) هد علامت خوانده شده را به جای علامت دیگری از الفبا به کار می برد. (۲) هد روی نوار، یک خانه به سمت چپ یا راست می رود. (۳) ماشین وارد حالت جدیدی می شود. به وسیله ماشین تورینگ می توان هر محاسبه ای را که با کامپیوتر دیگری می توان محاسبه کرد انجام داد. این ماشین دارای یک ورودی و یک خروجی است. دلیل آن که ماشین تورینگ برای مکتب کارکرد گسرایمی پاشینی اهمیت دارد این است که شباهت زیادی وجود دارد میان مفهوم کارکردگرایانه از حالات ذهنی (از آن حیث که نقش علی خاصی بازی می کنند) و ایده یک حالت فیزیکی که حالت داخلی یک ماشین تورینگ را تحقق می بخشد. همان طور که در دیدگاه

کارکردگرایانه آنچه یک حالت ذهنی مفروض را نوعی حالت ذهنی می کند همان نقش علی آن است با توجه به ورودی های حسی، خروجی های رفتاری و دیگر حالات ذهنی، به همین ترتیب آنچه یک حالت فیزیکی را در ماشین تحقق می بخشد روابط علی آن با ورودی ها، خروجی ها و حالات فیزیکی دیگری است که دیگر حالات ماشین را محقق می کنند. از این روست که کارکردگرایان به ماشین تورینگ همچون مدلی از ذهن نظر می کنند. نگا:

Kim J., *Philosophy of Mind* فصل ۴ و نیز: (oxford 1996).

Martin Davis, *Computability and Unsolvability* (New York 1958).

4- Computation.

۵- نگاه کنید به صص ۵ و ۱۷۴ و Lycan

W.'philosophy of mind'

۶- هوش مصنوعی (Artificial intelligence) از موضوعات مورد بحث در علوم مختلف است از جمله مهندسی کامپیوتر، روانشناسی، شناخت شناسی، عصب شناسی، زیست شناسی و فلسفه ذهن. نگا به:

Charniak E. and Mc Dermott. D.V. *Artificial Intelligence* (NY 1985)

11. Searl J./ *Rediscovery of the mind* (amb, MIT 1992).

12. John Searle "Mind, Brains and Programs" in Rosenthal D./ *The Nature of Mind* pp. 509 - 20.

۱۳- برای اشکالاتی که به سرل شده و پاسخ او به آنها نگاه:

Leporo E. and Van Gulick. R. (eds) *John Searl and His Critis* (oxford 1991).

۷- لیکان، پیشین

8. Turing Test نگا:

A. M. Turing' Computing Machiner and Intelligence' in Anderson A.R. (ed) *Minds and Machines* (NJ 1964) PP. 4-30.

9. Daniel Dennett. *Consciousness Explained* (Boston 1991) pp. 435- 440.

10. Dreyfus H.L. *What Computers Can't Do* (NY 1979) also see: Dreyfus H.L and Dreyfus S. *Mind over Machine* (NY 1986).



پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



پرو، شگاه علوم انسانی و مطالعات فرهنگی
رتال جامع علوم انسانی