بررسی انتقادی رویکردهای سرل، دنت و پاتنم به جایگاه فلسفی هوش مصنوعی

محمدحسین محمدعلی‌خلج*

چکیده

این مقاله به دو بخش کلی تقسیم شده است. در بخش نخست صورت‌بندی پاتنم، دنت و سرل به جایگاه فلسفی هوش مصنوعی به صورت کلی، طرح شده و نقاط ضعف هر یک نشان داده می‌شود. سپس صورت‌بندی برگزیده با تفکیک دو تراز فلسفه‌علمی و فلسفه‌ذهنی شرح داده می‌شود. در تراز فلسفه‌علمی مسئله آن است که آیا هوش مصنوعی می‌تواند دستگاهی بسازد که در آزمون نامحدود تورینگ پیروز شود؟ اما در تراز فلسفه‌ذهنی مسئله این است که اگر چنین دستگاهی ساخته شد، آیا ذهن دارد؟ در بخش دوم از نوشتار مهم‌ترین مسئله‌های طرح‌شده در این دو تراز ارزیابی می‌شود. مهم‌ترین مسئله در تراز فلسفهٔ علمی، مسئلهٔ چارچوب، یعنی مسئلهٔ چگونگی روبه‌رویی با رویدادهای تعریف‌ناشده و غیرمنتظره است. این مسئله بزرگ‌ترین مانع بر سر راه پیروزی هوش مصنوعی در آزمون تورینگ بوده است. در ادامه در تراز فلسفه‌ذهنی سه مسئله ارزیابی شده است: مغالطهٔ آدمک، استدلال اتاق چینی و نقد بر اساس تئوری گودل. مغالطهٔ آدمک نشان می‌دهد هوش مصنوعی امر ذهنی را به امری دلخواه تبدیل می‌کند، اما دو نقددیگر، اگرچه بر پارادایم کلاسیک هوش مصنوعی وارد است، بر شبکه‌های عصبی پیوندگرا وارد نیست.

واژگان کلیدی: *هوش مصنوعی، سرل، دنت، پاتنم.*

---

* دانشجوی دکتری فلسفه غرب معاصر دانشگاه شهید بهشتی. mhmakh@gmail.com

تاریخ دریافت: ۹۶/۶/۱۵ تاریخ تأیید: ۹۶/۱۱/۲۴

مقدمه

از زمانی که در سال ۱۹۵۶ میلادی در کنفرانس دارتموث، *جان مک‌کارتی* واژهٔ هوش مصنوعی را برای این رشته برگزید تا به امروز، همهٔ دانشمندان بزرگ هوش مصنوعی بر ماهیت میان‌رشته‌ای این دانش نوین تأکید داشته‌اند (Russell and Norvig, 2003, p.16). *استوآرت راسل* در کتاب مشهورش به نام **هوش مصنوعی رویکرد نوین** جایگاه هوش مصنوعی را در هفت دانش ریاضیات، مهندسی رایانه و رباتیک، تئوری کنترل و سایبرنتیک، نوروساینس، روان‌شناسی، زبان‌شناسی و فلسفه ارزیابی کرده است (Ibid, 2003, pp.5-16). می‌توان دانش‌های دیگری همچون منطق و زیست‌شناسی را نیز به این لیست افزود؛ اما در میان این رشته‌ها فلسفه در نسبت با هوش مصنوعی همواره جایگاه ویژه‌ای داشته است. در دههٔ پنجاه میلادی و با تولد هوش مصنوعی، پدران این رشته مدعی بودند که هوش مصنوعی به پرسش‌های ما دربارهٔ ذهن، یعنی پرسش‌هایی که فیلسوفان دو هزار سال است که پاسخ آنها را نیافته‌اند، پاسخ خواهد داد (Dreyfus, 1979, p.xxx). اما امروزه برخی از همان دانشمندان، موفقیت هوش مصنوعی را منوط به حل مسئله‌های فلسفی می‌دانند (McCarthy, 1995, p.1). در این مقاله جایگاه فلسفی هوش مصنوعی نه از منظر آن دانشمندان، بلکه از دیدگاه خود فیلسوفان روایت می‌شود و ارزیابی انتقادی نیز خواهد شد.

جایگاه فلسفی هوش مصنوعی در این نوشتار در دو سطح بررسی می‌شود؛ در سطح نخست صورت‌بندی کلی مسئلهٔ فلسفی در هوش مصنوعی مد نظر قرار می‌گیرد و در سطح دوم بر پایهٔ این صورت‌بندی، مهم‌ترین مسئله‌های این حوزه ازجمله مسئلهٔ چارچوب، مغاطلهٔ آدمک و آزمون اتاق چینی *سرل* مطرح می‌شود و پاسخ‌های داده‌شده به آنها به بحث گذاشته می‌شود. متناظر با این دو سطح این نوشتار نیز به دو بخش کلی- بخش اول صورت‌بندی کلی مسئلهٔ فلسفی هوش مصنوعی و بخش دوم مهم‌ترین مسئله‌های فلسفی هوش مصنوعی و پاسخ‌های داده‌شده به آنها- تقسیم می‌شود.

الف) صورت‌بندی کلی مسئلهٔ فلسفی در هوش مصنوعی

دیدگاه فیلسوفان دربارهٔ هوش مصنوعی طیف گسترده‌ای را در بر می‌گیرد. در یک سر این طیف *پاتنم* قرار دارد که مدعی است هوش مصنوعی چیزی بیش از یک رشتهٔ مهندسی نیست و ازاین‌رو شایستگی پرداختن به مسئله‌های فلسفی را ندارد؛ اما در سر دیگر این طیف *دنت* جای دارد که از اساس هوش مصنوعی را نوعی فلسفه می‌داند. صورت‌بندی فیلسوفان دیگری همچون *سرل*، *دریفوس* و *چالمرز* نیز در میانهٔ این طیف جای می‌گیرد. در این بخش از پژوهش نخست صورت‌بندی *پاتنم* و *دنت* به عنوان دو سر طیف نظرات فیلسوفان در مورد هوش مصنوعی روایت می‌شود و در ادامه رویکرد *سرل* بر پایهٔ تفکیک هوش مصنوعی قوی از هوش مصنوعی ضعیف به عنوان مشهورترین صورت‌بندی در این حوزه به بحث گذاشته می‌شود و در پایان با ارزیابی این نظرات، صورت‌بندی درست‌تر معرفی می‌شود.

۱. صورت‌بندی پاتنم

در آغاز دههٔ شصت میلادی، *پاتنم* با نوشتن مقالهٔ «اذهان و ماشین‌ها» (Putnam, 1960) رویکردی را در فلسفهٔ ذهن بنا نهاد که بعدها به کارکردگرایی ماشینی مشهور شد. کارکردگرایی ماشینی همواره از مهم‌ترین پشتوانه‌های فلسفی یک سری از جریان‌های هوش مصنوعی بوده است؛ اما *پاتنم* خود در اواخر دههٔ هشتاد هم کارکردگرایی را به‌طورکامل نقد کرده و هم در برابر هوش مصنوعی صریحاً موضع‌گیری کرد.* او به‌صراحت مدعی شد که هوش مصنوعی هیچ دستاورد فلسفی به همراه نداشته است. او گفت:

> آیا هوش مصنوعی برای ما چیز مهمی‌در مورد ذهن دارد؟ من فکرمی‌کنم که پاسخ منفی است... الگوی رایانه‌ای ذهن همان هوش مصنوعی نیست... اگر کسی الگوی رایانه‌ای ذهن را ساخته باشد، آن کس *آلن تورینگ* است. اما هوش مصنوعی... نه تئوری ماشین *تورینگ*، نه فلسفهٔ *آلن تورینگ*، بلکه چیزی بسیار

*برای آشنایی با کارکردگرایی ماشینی *پاتنم* و دلایل چرخش وی، ر.ک: محمدعلی‌خلج، ۱۳۹۲.

دم‌دستی‌تر است.... طراحی رایانه، شاخه‌ای از مهندسی است که ابتدا تنها در مورد طراحی نرم‌افزار بوده است. هوش مصنوعی نیز زیرشاخه‌ای از این شاخهٔ مهندسی است (Putnam(2), 1988, pp.269-270).

بدین ترتیب از نظر *پاتنم* حساب الگوی رایانه‌ای ذهن و هوش مصنوعی از هم جداست و هوش مصنوعی به‌مثابه رشته‌ای مهندسی، از نظر فلسفی اهمیت ندارد.

ارزیابی صورت‌بندی پاتنم

چنان‌که تعابیر خود *پاتنم* نشان می‌دهد، صورت‌بندی وی بر دو گزاره شکل گرفته است: ۱. الگوی رایانه‌ای ذهن چیزی متفاوت با هوش مصنوعی است؛ ۲. هوش مصنوعی اهمیت فلسفی ندارد. ما در جای دیگری به‌تفصیل گزارهٔ نخست و دلایل *پاتنم* را برای طرح آن نقد و ارزیابی کرده‌ایم و در اینجا به آن نخواهیم پرداخت.* اما ارزیابی گزارهٔ دوم: همهٔ مسائل فلسفی که در بخش دوم این مقاله ارزیابی شده‌اند، نادرستیِ این گزاره را نشان می‌دهند؛ چراکه همگی آنها به نوعی در نسبت با هوش مصنوعی قرار دارند و پیشاپیش، اهمیت فلسفی هوش مصنوعی را نشان می‌دهند. پس دست‌کم می‌توان گفت در پایانِ مقاله و با ارزیابی این مسائل مشخص خواهد شد که آیا ادعای *پاتنم* درست و هوش مصنوعی فاقد اهمیت فلسفی است یا خیر.

۲. صورت‌بندی دنت

دنت دقیقاً در نقطهٔ مقابل *پاتنم* جای دارد. از نگاه او هوش مصنوعی در معنایی عام نوعی فلسفه‌ورزی است. *دنت* به‌صراحت موضوع و مسائل هوش مصنوعی را موضوع و مسائل

* *دریفوس* در «پدیدارشناسی و مکانیزم» نشان می‌دهد که الگوی *تورینگ* از ذهن که از سوی *پاتنم* ارائه شده است، فاقد محتواست؛ یعنی ورودی- خروجی‌ها و قواعد حاکم بر ذهن به‌مثابه ماشین *تورینگ* مشخص نیست. پژوهشگران هوش مصنوعی می‌کوشند همین ورودی- خروجی‌ها و قواعد را کشف کنند و به تئوری *پاتنم* محتوا ببخشند؛ اما اگر ایشان شکست بخورند، تئوری *پاتنم* دربارهٔ ذهن- یعنی الگوی *تورینگ*- نیز شکست خورده است. پس سرنوشت هوش مصنوعی و الگوی *تورینگ* از هم جدا نیست (ر.ک: محمدعلی‌خلج، ۱۳۹۲).

فلسفی می‌داند. چنان‌که می‌گوید:

> هوش مصنوعی در معنایی گسترده، فلسفه است و به‌طورمستقیم با یک سری مسئله‌های فلسفی، ازجمله اینکه ذهن چیست؟ معنی چیست؟ عقلانیت چیست؟ چه شرایط ضروری برای شناخت وجود دارد؟ و چنین پرسش‌هایی سروکار دارد (Dennett, 1988, p.212).

البته از نظر *دنت* روش پاسخ به این مسئله‌های مشترک در فلسفه و هوش مصنوعی متفاوت است. به باور وی در پاسخ به مسئلهٔ یکسان، «پاسخ سنت فلسفی چنین است: تحلیلی پیشینی و برهان بیشتر و پاسخ محققان هوش مصنوعی چنین است: آن را بساز و ببین» (Ibid, p.273)؛ اما از نظر *دنت* این روش متفاوتِ هوش مصنوعی، یعنی محدودیتش به کار بر روی رایانه‌ها، مانعی برای پاسخ به مسئله‌های کلیِ فلسفی ایجاد نمی‌کند (Ibid, p.112). او برای نشان‌دادن این امر، هندسه را مثال می‌زند:*

> آیا رایانه به هوش مصنوعی نامربوط است؟ نامربوط است به همان معنا که کشیدن شکل بر روی صفحه در فهم هندسه نامربوط است؛ یعنی در عمل بدون آن ممکن نیست؛ اما در اصل نامربوط است (IbId, 1981, p.118).

به بیان دیگر هوش مصنوعی نیز همچون فلسفه در پی شرایط امکانِ برخورداری از ذهن و هوشمندی است و با انتزاع این شرایط از شرایط ویژهٔ هر آزمایش و سخت‌افزار آزمایشگاهی در پی یافتن قواعد کلی معرفت‌شناسی است:

> بنابراین فرد می‌تواند از مدل‌های معین- و البته بسیار غیرواقع‌گرایانهٔ- هوش مصنوعی چیزهای بسیاری را بیاموزد؛ همان‌گونه که از ذهن یک مریخی

* افزون بر مثال هندسه، *دنت* مثال دیگر چرخ را نیز برای رساندنِ منظورش می‌زند. *دنت* می‌گوید ما در طبیعت برای حرکت چرخ نداریم و چرخ ساختهٔ بشر است؛ اما می‌دانیم قواعد حرکت چه در چرخ چه در طبیعت، یکسان است و با کشف قواعد کلی حاکم بر حرکت چرخ می‌توان قواعد حاکم بر حرکت در طبیعت را نیز کشف کرد. به همین قیاس دربارهٔ شناخت نیز ما در طبیعت نرم‌افزار نداریم، اما قواعد کلیِ حاکم بر شناخت چه در طبیعت چه در نرم‌افزار یکی است؛ ازاین‌رو *دنت* نرم‌افزارهای هوش مصنوعی را چرخ‌های شناختی می‌نامد (Dennet 1984, p200).

می‌آموزد. یک روان‌شناسی خوب می‌تواند از یک مریخی، اگرچه شبیه ما نباشد، مبنایی برای قواعد کلی معرفت‌شناسی و روان‌شناسی به دست دهد که در موجود انسانی نیز به کار گرفته می‌شود (Ibid, p.113).

ارزیابی صورت‌بندی دنت

صورت‌بندی *دنت* ابهام شدیدی دارد. این امر به‌خودی‌خود روشن نیست که روش آزمایشگاهی هوش مصنوعی چه کمکی به حل مسئله‌های انتزاعی فلسفه و معرفت‌شناسی می‌کند. ادعای *دنت* وقتی قابل اعتنا خواهد بود که وی نمونه‌هایی از کمک هوش مصنوعی به فلسفه را مثال بزند. در آن صورت است که می‌توان صورت‌بندی وی را ارزیابی کرد. در بخش دوم از این نگاشته دو نمونه از کمک‌های ادعایی *دنت* بیان می‌شود؛ یکی صورت‌بندی مسئلهٔ چارچوب و دیگری حل مسئلهٔ *هیوم*. دربارهٔ هر یک از این دو مسئله نشان داده می‌شود که ادعای *دنت* مبنی بر کمک هوش مصنوعی به فلسفه قابل دفاع نیست و درنتیجه صورت‌بندی کلی وی نیز مخدوش است.

۳. صورت‌بندی سرل

در سال ۱۹۸۰ میلادی *سرل* با تفکیک هوش مصنوعی ضعیف از هوش مصنوعی قوی، صورت‌بندی خود از مسئلهٔ فلسفی در هوش مصنوعی را ارائه کرد. بر پایهٔ این صورت‌بندی، هوش مصنوعی ضعیف به خودی خود یک شاخه از مهندسی است که می‌تواند به عنوان ابزار به شناخت ذهن کمک کند؛ اما هوش مصنوعی قوی ادعایی فلسفی دربارهٔ ذهن دارد. او می‌گوید:

> مطابق با هوش مصنوعی ضعیف، ارزش رایانه در آموزش در مورد ذهن، آن است که ابزاری بسیار قوی را به دست ما می‌دهد... اما مطابق با هوش مصنوعی قوی، رایانه تنها یک ابزار در مطالعهٔ ذهن نیست؛ بلکه رایانه‌های به‌درستی برنامه‌ریزی‌شده، همان ذهن هستند؛ بدین معنا می‌توان گفت که رایانه‌ای که برنامهٔ درستی را اجرا می‌کند، می‌فهمد و حالت‌های شناختی دارد (Searle, 1980, p.2).

به بیان دیگر هوش مصنوعی ضعیف رایانه را صرفاً ابزاری برای مطالعهٔ ذهن می‌داند

و خود رایانه را به‌خودی خود الگویی برای ذهن به شمار نمی‌آورد؛ اما هوش مصنوعی قوی دقیقاً در نقطهٔ مقابل قرار دارد و مدعی است رایانه الگو و نمونه‌ای برای ذهن است و به همان معنایی که انسان‌ها دارای ذهن و حالت‌های شناختی هستند، رایانه‌ها نیز ذهن و حالت‌های شناختی دارند.

ارزیابی صورت‌بندی سرل

دو نقد اساسی بر این صورت‌بندی *سرل* وارد است؛ نخستین نقد اینکه در بیان *سرل* جایگاه فلسفی هوش مصنوعی ضعیف چندان روشن نیست و گویا در این سطح مسئلهٔ فلسفی وجود ندارد. صورت‌بندی *سرل* در حالی بر اهمیت فلسفی هوش مصنوعی ضعیف سرپوش می‌نهد که بخش مهمی از مسئله‌های فلسفی هوش- همچون مسئلهٔ چارچوب یا خود آزمون *تورینگ*- در این سطح می‌گنجند. جدا از صورت‌بندی‌های گوناگون ایدهٔ اصلی آزمون *تورینگ* این است که آیا می‌توان کارکردهای هوشمندانهٔ انسانی را در رایانه پیاده‌سازی کرد؟ فارغ از اینکه رایانه حالت شناختی دارد یا ندارد، آیا می‌تواند به‌طورکامل همانند ما رفتار کند؟ این مسئله‌ای فلسفی است که در تراز هوش مصنوعی قوی قرار ندارد و *سرل* جای دیگری برای آن در نظر نگرفته است. نقد دوم بر صورت‌بندی *سرل* نیز آن است که بر پایهٔ آن، سخن بسیاری از مدافعان هوش مصنوعی نادرست فهمیده می‌شود. در اصل *سرل* موضع خود را مبنا قرار داده است و رویکرد مخالفان خود را به‌اشتباه، بر اساس آن صورت‌بندی کرده است. رویکرد *سرل* به ذهن در ادامه و در ذیل بحث از اتاق چینی شرح داده شده است؛ اما در اینجا کافی است به یکی از اصول اساسی آن اشاره کنیم؛ بنا بر این اصل، حالت‌های شناختی و سوبژکتیو وجود دارند (IbId, 1992, pp.18-19). از این لحاظ رویکرد *سرل* شبیه دوگانه‌انگاری خاصه‌ای است؛ زیرا ویژگی‌های ذهنی را نیز اموری موجود به شمار می‌آورد. پس ذهن‌مندی از نگاه *سرل* مستلزم وجود حالت‌های شناختی سوبژکتیو درون فرد است. با این حساب کسی که مدعی است رایانه‌ها دارای ذهن هستند- مدافعان هوش مصنوعی قوی- باور دارد که در رایانه‌ها جدا از مدارهای الکتریکی، چیز دیگری به عنوان حالت سوبژکتیو وجود دارد.

سرل چنین دیدگاهی را به مدافعان هوش مصنوعی قوی نسبت می‌دهد؛ درحالی‌که تقریباً هیچ یک از مدافعان هوش مصنوعی قوی این سخن را نمی‌پذیرد و بر این باور نیستند که در رایانه‌ها- جدا از مدارهای الکتریکی- چیز دیگری به منزلهٔ حالت سوبژکتیو وجود داشته باشد.* با این حساب صورت‌بندی *سرل* نیز از این جهت که بخش مهمی‌از مسئله‌های فلسفی در هوش مصنوعی را نادیده می‌گیرد و از جهتی که رویکرد مدافعان هوش مصنوعی را به اشتباه روایت می‌کند، صورت‌بندی مناسبی نیست.

سرل در سال ۱۹۹۰ میلادی و ده سال پس از نخستین صورت‌بندی‌اش، با قراردادن شناخت‌گرایی (Cognitivism) در کنار هوش مصنوعی قوی و ضعیف تا حدودی صورت‌بندی خود را تصحیح کرد. بر پایهٔ این صورت‌بندی جدید، وی میان این سه پرسش تفکیک قائل می‌شود: «۱. آیا مغز یک رایانهٔ دیجیتال است؟ ۲. آیا ذهن یک برنامهٔ رایانه‌ای است؟ ۳. آیا عمل مغز می‌تواند بر روی یک رایانهٔ دیجیتال شبیه‌سازی شود؟» (Searle, 1990, p.21) *سرل* در پاسخ به این سه پرسش، شناخت‌گرایی، هوش مصنوعی قوی و هوش مصنوعی ضعیف را این‌گونه تعریف می‌کند:

> من دیدگاهی که داشتن ذهن را همان داشتن یک برنامهٔ رایانه‌ای می‌داند، هوش مصنوعی قوی و دیدگاهی را که بر این باور است که پردازش‌های مغزی (پردازش‌های ذهنی) می‌تواند به صورت محاسبه‌گرایانه شبیه‌سازی شود، هوش مصنوعی ضعیف و دیدگاهی را که معتقد است مغز یک رایانهٔ دیجیتال است، شناخت‌گرایی می‌نامم (IbId, p.22)

این صورت‌بندی جدید از جهاتی صورت‌بندی بهتری است. بر اساس این صورت‌بندی رویکرد اصلی مدافعان هوش مصنوعی که مدعی‌اند تفکر در انسان چیزی جز پردازشگری دیجیتالی مغز نیست، در ذیل شناخت‌گرایی آورده شده است؛ اما در این

* البته *دیوید چالمرز* به صورتی عجیب، از یک سو مدافع ذهن‌مندی هوش مصنوعی است و از سوی دیگر حالت‌های شناختی را موجود می‌داند. *سرل* به‌خوبی تضاد این دو رویکرد را که *چالمرز* می‌کوشد با هم جمع کند، نشان می‌دهد (See: Searle, 1997).

صورت‌بندی نیز نسبت شناخت‌گرایی با هوش مصنوعی قوی تا حدودی مبهم است و همچنان اهمیت فلسفی هوش مصنوعی ضعیف نادیده گرفته شده است. از یک سو بسیاری از کسانی که مغز را یک سخت‌افزار دیجیتال می‌دانند، ذهن را برنامهٔ نرم‌افزاری آن به شمار می‌آورند و با توجه به نقل آورده‌شدهٔ پیشین، بر پایهٔ تقسیم‌بندی *سرل* همهٔ ایشان هم در شمار شناخت‌گرایان قرار می‌گیرند و هم در شمار مدافعان هوش مصنوعی قوی؛ اما- چنان‌که پیش از این هم دیدیم- *سرل* باور دیگری را نیز به مدافعان هوش مصنوعی قوی نسبت می‌دهد. به باور او «به‌طورکلی حالت‌های ذهنی یک هستی‌شناسی سوبژکتیو غیرقابل فروکاهی را دارد» (IbId, 1992, p.19)؛ ولی بیشتر مدافعان هوش مصنوعی نمی‌پذیرند که در رایانه‌ها افزون بر مدارهای الکتریکی، چیزی به منزلهٔ حالت‌های سوبژکتیو وجود داشته باشد. بر این اساس هیچ یک از شناخت‌گرایان مدافع هوش مصنوعی قوی نیستند. پس از یک زاویه، شناخت‌گرایان مدافع هوش مصنوعی قوی و از زاویهٔ دیگر مخالف آن هستند؛ ازاین‌رو نسبت شناخت‌گرایی و هوش مصنوعی قوی روشن نیست؛ به‌طورکلی صورت‌بندی *سرل* از مسئلهٔ فلسفی در هوش مصنوعی بسیار آمیخته به رویکرد ویژهٔ خودش در فلسفهٔ ذهن است و نمی‌تواند یک صورت‌بندی کلی مناسب برای معرفی جایگاه فلسفی هوش مصنوعی باشد. اما دو اصطلاح هوش مصنوعی قوی و ضعیف در ادبیات تئوریک هوش مصنوعی تعبیرهایی بسیار پرکاربرد هستند و باید تعبیر درستی از این دو ارائه شود. ما در ادامه کوشیده‌ایم در ذیل صورت‌بندی درست، تعابیر *سرلی* از هوش مصنوعی ضعیف و قوی را بهبود ببخشیم.

۴.صورت‌بندی درست

تا اینجا دیدیم که صورت‌بندی‌های *پاتنم*، *دنت* و *سرل* از مسائل فلسفی مرتبط با هوش مصنوعی، هر یک با مشکلاتی روبه‌روست. مدعای ما آن است که می‌توان بر اساس آزمون *تورینگ* صورت‌بندی ساده‌ای ارائه داد که به هیچ یک از این مشکلات دچار نباشد. در ادامه و همراه با شرح این صورت‌بندی، از این مدعا نیز دفاع خواهیم کرد. توضیح مفصل آزمون *تورینگ* برای صورت‌بندی مد نظر ما، مورد نیاز نیست؛ بلکه دانستن همین امر

کافی است که گذشته از محتوای خاص این آزمون، آزمون نامحدود *تورینگ* معیاری است از توانایی پیاده‌سازی رفتارهای هوشمندانهٔ انسانی؛ یعنی سخت‌افزاری که از این آزمون پیروز بیرون بیاید، می‌تواند کارکردهای هوشمندانهٔ انسانی را پیاده‌سازی نماید و مانند یک انسان رفتار کند.* با درنظرگرفتن این آزمون می‌توان دو پرسش کاملاً متفاوت را مطرح کرد. یک پرسش آن است که آیا ساختن چنین سخت‌افزاری ممکن است یا خیر؟ اگر آری، چگونه؟! این پرسش در درجهٔ نخست در تراز دانش هوش مصنوعی به منزلهٔ یکی از زیرشاخه‌های «علوم» رایانه‌ای مطرح می‌شود. دانشمندان و مهندسان هوش مصنوعی باید به این پرسش پاسخ دهند که چگونه سخت‌افزاری بسازند که توانایی پیاده‌سازی کارکردهای هوشمندانهٔ انسانی را داشته باشد. پس پرسش از چگونگی پیروزی

---

* *تورینگ* آزمونش را همچون یک بازی این‌گونه توصیف می‌کند: «این بازی با سه نفر انجام می‌شود: یک مرد (A)، یک زن (B) و یک بازپرس (C) که می‌تواند هر جنسیتی داشته باشد. بازپرس در یک اتاقی جدا از دو نفر دیگر قرار دارد. موضوع بازی برای بازپرس این است که باید تعیین کند که کدام یک از این دو، مرد و کدام یک، زن هستند. او آنها را با برچسب‌های X و Y می‌شناسد و در پایان بازی می‌گوید یا X، A است و Y، B است؛ یا X، B است و Y، A. بازپرس اجازه دارد تا از A و B بپرسد: C: ممکن است که X طولِ مویش را به من بگوید؟ حالا فرض کن که X واقعاً A است؛ درنتیجه باید جواب بدهد. نقش A در بازی این است که بکوشد و سبب شود که C تشخیص اشتباهی داشته باشد؛ بنابراین پاسخ او ممکن است این چنین باشد... بلندترین دستهٔ موی من بیست سانتی‌متر است... حالت ایدئال این است که ایشان به صورت تلگرافی ارتباط داشته باشند و صدای ایشان نباید به تشخیص C یاری برساند... موضوع بازی برای B این است که به تشخیص بازپرس کمک کند. بهترین راهبرد برای وی این است که پاسخ‌های حقیقی بدهد» (Turing, 1950, p.441). *تورینگ* خودش نام این بازی را که بعدها به آزمون *تورینگ* (Turing test) مشهور شد، بازی تقلید (Imitation Game) می‌گذارد. او برای روشن‌کردن نسبت این بازی و پرسش از امکان اندیشیدن ماشین، مقاله را این‌گونه ادامه می‌دهد: «حال ما می‌پرسیم هنگامی‌که یک ماشین به جای A در این بازی شرکت کند، چه اتفاقی خواهد افتاد؟ اگر بازی به این ترتیب- یعنی با حضور یک ماشین- انجام شود، آیا بازپرس اغلب دچار اشتباه خواهد شد؛ همان‌گونه که بازی میان یک مرد و یک زن انجام می‌شود؟ این پرسش‌ها جایگزین پرسش اصلی ما- یعنی آیا ماشین می‌تواند تفکر کند- می‌شود» (Ibid).

هوش مصنوعی در آزمون *تورینگ* به‌خودی‌خود پرسشی علمی‌است؛ ولی می‌دانیم که متناظر با هر علمی می‌توان از فلسفهٔ آن علم نیز سخن گفت. فلسفهٔ یک رشتهٔ علمی مبادی تصوری، تصدیقی و روش‌شناختی آن علم را از منظری فلسفی ارزیابی می‌کند؛ برای نمونه می‌توان متناظر با علوم گوناگون از فلسفهٔ علم فیزیک، فلسفهٔ علم زیست‌شناسی، فلسفهٔ ریاضیات و... نام برد. به همین قیاس می‌توان از فلسفهٔ علوم رایانه‌ای و مشخصاً فلسفهٔ هوش مصنوعی نیز سخن گفت. در فلسفهٔ هوش مصنوعی به این معنا، مبادی فلسفی دانشمندان هوش مصنوعی ارزیابی می‌شود. پس بر پایهٔ صورت‌بندی ما، پرسش نخست دربارهٔ آزمون *تورینگ* این است: چگونه می‌توان سخت‌افزاری ساخت که کارکردهای هوشمندانهٔ انسانی را پیاده‌سازی کند؟ چنان‌که گفتیم، این پرسش در درجهٔ نخست پرسش دانشمند هوش مصنوعی است؛ ولی فیلسوف نیز می‌تواند سراغ پاسخ‌های دانشمندان به این پرسش برود و پیش‌فرض‌های فلسفی آنان را در پاسخ به این پرسش نقد کند. بر پایهٔ صورت‌بندی ما، این نخستین تراز از مسئلهٔ فلسفی در هوش مصنوعی است که می‌توان آن را تراز فلسفه‌علمی نامید.

پرسش دوم دربارهٔ آزمون *تورینگ* آن است که به فرض اگر سخت‌افزاری ساخته شد که از آزمون *تورینگ* موفق بیرون آمد و توانست همهٔ رفتارهای هوشمندانهٔ انسانی را پیاده‌سازی کند، آیا این امر کافی است تا این سخت‌افزار را دارای ذهن بدانیم؟ این پرسش- بر عکس پرسش نخست- پرسشی علمی نیست؛ بلکه پرسشی فلسفی است و در تراز فلسفهٔ ذهن جای دارد. در این تراز پرسش اصلی آن است که ذهن چیست و چه موجودی دارای ذهن است؟ در پاسخ به این پرسش جریان‌هایی در فلسفهٔ ذهن بر این باورند هر موجودی که بتواند در آزمون *تورینگ* پیروز شود، دارای ذهن است. در اصل تراز نخست از پرسش فلسفی، نوعی نقادی درجهٔ دوم از دانش هوش مصنوعی است که بر اساس آن، مبادی معرفتی- روشی و امکانات و محدودیت‌های این دانش ارزیابی می‌شود؛ اما تراز دوم، یک تئوری فلسفه‌ذهنی است. پس می‌توان صورت‌بندی درست را با تفکیک این دو تراز و پرسش‌های متناظر با هر یک بیان کرد:

۱. تراز فلسفه علمی: آیا دانش هوش مصنوعی کنونی می‌تواند برنامه‌ای طراحی کند که در آزمون *تورینگ* پیروز شود؟

۲. تراز فلسفه ذهنی: آیا آزمون *تورینگ* معیار درستی برای ذهن‌مندی است؟

مدافع هوش مصنوعی در تراز فلسفه‌علمی ادعای کمینه و ضعیف‌تری را مطرح می‌کند. ادعای او آن است که رایانه می‌تواند مانند انسان رفتار کند؛ اما مدافع هوش مصنوعی در تراز فلسفه‌ذهنی ادعای بیشینه و قوی‌تری را مطرح می‌کند. او مدعی است نه تنها رایانه می‌تواند شبیه انسان رفتار کند، بلکه همین برای ذهن‌مند دانستن آن کافی است. بر این اساس می‌توان جریان مدافع هوش مصنوعی در تراز فلسفه‌علمی را هوش مصنوعی ضعیف و جریان مدافع هوش مصنوعی در تراز فلسفه‌ذهنی را هوش مصنوعی قوی دانست. در راستای نقد هوش مصنوعی ضعیف باید نشان داد که این رشته به سبب پذیرش پیش‌فرض‌های فلسفی نادرست، نمی‌تواند در آزمون *تورینگ* موفق باشد. اما در نقد هوش مصنوعی قوی باید نشان داد که آزمون *تورینگ* از اساس معیار مناسبی برای ذهن‌مندی نیست.

پیش از این گفتیم ادعای ما آن است که این صورت‌بندی با مشکلات صورت‌بندی‌های پیشین- به‌ویژه صورت‌بندی *سرل*- روبه‌رو نمی‌شود. از اشکال‌های صورت‌بندی *سرل* این بود که اهمیت فلسفی هوش مصنوعی ضعیف را نادیده می‌گرفت؛ اما از نظرگاه صورت‌بندی ما، هوش مصنوعی ضعیف نیز در تراز فلسفه‌علمی دارای جذابیت‌های فلسفی ویژهٔ خود خواهد بود. اگر *سرل* را نمونهٔ فیلسوفانی بدانیم که در تراز فلسفه‌ذهنی نقدهایش را بر هوش مصنوعی قوی وارد آورده است، *دریفوس* مشخصاً فیلسوفی بوده است که در تراز فلسفه‌علمی نقدهای فلسفی‌اش را بر هوش مصنوعی ضعیف وارد کرده است. گفتیم که در نقد هوش مصنوعی قوی- یعنی تراز فلسفه‌ذهنی- به آزمون *تورینگ* به عنوان معیاری برای ذهن‌مندی شک می‌شود؛ ولی در نقد هوش مصنوعی ضعیف- یعنی در تراز فلسفه‌علمی- چنین تشکیکی صورت نمی‌پذیرد. حال به این سخن *دریفوس* که در گفت‌وگو با *دنت* دربارهٔ برداشت اشتباه وی از سخنانش

اعتراض می‌کند، توجه کنید:

> به نظر می‌رسد تو [یعنی *دنت*] فکر کرده‌ای من [یعنی *دریفوس*] مثل *جان سرل* در استدلال اتاق چینی در پی این بوده‌ام که رایانه چیزی بیش از رفتاری را انجام دهد که گویا دارد زبان طبیعی را می‌فهمد... اما من هرگز چنین چیزی نخواسته‌ام. من همواره راضی بوده‌ام با قواعد آزمون نامحدود *تورینگ* بازی انجام شود (Dreyfus, 1997, p.267).

این تعبیر *دریفوس* به‌خوبی گواه آن است که وی نقدش را در ترازی متفاوت نسبت به *سرل* مطرح می‌کند و نقدهای *دریفوس* نه به هوش مصنوعی قوی- یعنی تراز فلسفه‌ذهنی- بلکه به هوش مصنوعی ضعیف- یعنی تراز فلسفه‌علمی— باز می‌گردد؛ به بیان دیگر چنان‌که خودش می‌گوید، او به معیار آزمون *تورینگ* شک نمی‌کند و مسئله‌اش این است که آیا هوش مصنوعی می‌تواند در آزمون *تورینگ* موفق باشد و کارکردهای هوشمندانهٔ انسانی را پیاده کند یا خیر.

روشن است که پرسش از امکان موفقیت هوش مصنوعی در پیاده‌سازی‌های کارکردهای انسانی در درجهٔ نخست پرسشی علمی است و در «علوم» کامپیوتر مطرح می‌شود؛ ولی همان‌گونه که متناظر با علم فیزیک و ریاضی می‌توان از فلسفهٔ علم فیزیک و فلسفهٔ ریاضی سخن گفت، می‌توان از فلسفهٔ علوم کامپیوتر و- به صورت خاص- فلسفهٔ هوش مصنوعی نیز سخن گفت. فلسفهٔ علم هوش مصنوعی- یعنی نقادی درجهٔ دوم این دانش- که مبادی تصدیقی، تصوری و روش‌شناسی این دانش را ارزیابی فلسفی می‌کند. برای این منظور باید به دنبال پژوهش‌های دانشمندان هوش مصنوعی رفت و پیش‌فرض‌های ایشان، نرم‌افزارها و سخت‌افزارهای طراحی‌شده از سوی آنان را نقد کرد. این دقیقاً همان کاری است که *دریفوس* انجام می‌دهد. *سرل* که در تراز فلسفه‌ذهنی موضع می‌گیرد، کاری به پژوهش‌های دانشمندان هوش مصنوعی ندارد؛ اما *دریفوس* مستقیم به دنبال برنامه‌های پژوهشی دانشمندان هوش مصنوعی رفته است و مبادی فلسفی آنان را نقد و ارزیابی کرده است. ماهیت فلسفه‌علمی نقدهای *دریفوس* را از تعابیر خود وی دربارهٔ الگووارهٔ کلاسیک هوش مصنوعی نیز می‌توان دریافت. چنان‌که می‌گوید: «آنچه

هوگلند هوش مصنوعی کلاسیک می‌نامد، نمونهٔ سرمشق‌گونه‌ای از چیزی است که فیلسوفان علم برنامهٔ پژوهشی رو به زوال می‌نامند» (IbId, 1992, p.xi).

از نگاه *لاکاتوش* برنامه‌های پژوهشی علمی به دو دسته تقسیم می‌شوند: برنامه‌های پژوهشی پیش‌رونده و برنامه‌های پژوهشی روبه‌زوال(Degenerating research program)؛ برنامه‌ای پیش‌رونده است که هر چه پیش می‌رود، مسئله‌های علمی بیشتری را حل کند؛ اما برنامهٔ علمی رو به زوال می‌کوشد با وصله‌زدن و به کمک‌آوردن تبصره‌های پی‌درپی، هستهٔ سخت خود را نجات دهد. جدا از پیش‌رونده یا روبه‌زوال بودن هوش مصنوعی کلاسیک، آنچه برای ما موضوعیت دارد، آن است که *دریفوس* با ارجاع به فیلسوفان علم و با وام‌گیری از *لاکاتوش* از اصطلاح فلسفه‌علمی برنامهٔ پژوهشی روبه‌زوال در وصف یک جریان هوش مصنوعی استفاده می‌کند و این امر به‌روشنی گواه آن است که او در تراز فلسفه‌علمی هوش مصنوعی را نقد کرده است. پس می‌بینیم که صورت‌بندی ما- بر عکس صورت‌بندی‌های پیشین- نه تنها سرشت متفاوت نقدهای *دریفوس* و *سرل* بر هوش مصنوعی را آشکار می‌کند، بر ماهیت فلسفی نقد هر یک از ایشان نیز پرتو می‌افکند.

ب) مهم‌ترین مسئله‌های فلسفی هوش مصنوعی و پاسخ‌های داده‌شده

در بخش دوم این نوشتار- بر پایهٔ صورت‌بندی برگزیده در بخش اول- مسئله‌های فلسفی در هوش مصنوعی را ارزیابی می‌کنیم؛ بر این اساس در دو بخش، مهم‌ترین مسئله‌های هر یک از دو تراز فلسفه‌علمی (هوش مصنوعی ضعیف) و فلسفه‌ذهنی (هوش مصنوعی قوی) را بیان می‌کنیم و پاسخ‌های داده‌شده به آنها را ارزیابی می‌کنیم.*

---

*شایان ذکر است، گسترهٔ بحث‌های صورت‌گرفته دربارهٔ هر یک از چهار مسئلهٔ طرح‌شده در این بخش بسیار زیاد است. در مشکل چارچوب افزون بر *دنت* و *دریفوس*، صاحب‌نظرانی همچون *فودور*، *مک‌درمت*، *پیلیشین*، *کلارک*، *مککارتی*، *ویلر* و... دیدگاه‌های مستقل خود را دارند. در استدلال اتاق چینی لیست افراد صاحب‌نظر بسیار بیش از اینهاست که به چند موردشان در متن اشاره شد. در استدلال بر پایهٔ فرضیهٔ *گودل* نیز *لوئیز*، *بارز*، *فدرمن*، *مککلوچ* و همچنین *مک‌درمت* و *مککارتی* صاحب‌نظر هستند. بدون

۱. مهم‌ترین مسئله‌های فلسفه‌علمی (هوش مصنوعی ضعیف)

پیش از این نشان دادیم نقدهای *دریفوس* بر هوش مصنوعی در این تراز جای دارد. وی در طی بیش از نیم قرن نقد جریان‌های گوناگون در هوش مصنوعی، بیان‌های متفاوتی از نقدش ارائه کرده است. می‌توان دست‌کم چهار بیان گوناگون از مسئلهٔ واحد *دریفوس* در هوش مصنوعی را از هم تفکیک کرد. در دورهٔ نخست مسئلهٔ پایگاه بزرگ و شرایط امکان هوشمندی به‌طورکلی (Dreyfus, 1979)، در دورهٔ دوم مسئلهٔ شرایط امکان خبرگی در فضای ایزوله (IbId, 1984) و در دورهٔ سوم مسئلهٔ فهم عرفی در مرکز نقد دریفوس قرار داشته‌اند (IbId, 1988)؛** اما در دورهٔ چهارم و پایانی *دریفوس* مسئله‌ای با نام مسئلهٔ چارچوب (Frame Problem) را معیار نقد نظریه‌های هوش مصنوعی می‌پذیرد؛ چنان‌که به‌صراحت می‌گوید:

> *ویلر* به‌درستی می‌اندیشد که ساده‌ترین آزمون برای تحقق هر برنامهٔ هوش مصنوعی شرح‌داده شده این است که آیا می‌تواند مسئلهٔ چارچوب را حل کند یا نه؟ (Dreyfus, 2007, p.26).

مسئلهٔ چارچوب که نخستین بار از سوی *جان مک‌کارتی* و پس از آن *دنت* صورت‌بندی شده است، چنان در ادبیات تئوریک هوش مصنوعی پذیرفته شده است که *دریفوس* خود نیز آن را به رسمیت می‌شناسد؛ اما *دنت* بیان‌های متفاوتی در شرح این مسئله دارد و به‌صراحت بر پیچیده‌بودن آن صحه می‌گذارد. به‌طورکلی مسئلهٔ چارچوب یک نوع از مسئلهٔ یافتن امر مربوط در هر موقعیت است (Ibid)؛ یعنی این مسئله که در

شک روایت همهٔ این دیدگاه‌ها در یک مقاله ممکن نیست و ما در اینجا تنها نظریه‌های اصیل و قوی‌تر را به بحث گذاشته‌ایم. هر یک از این مسائل می‌تواند به خودیِ خود موضوع یک پایان‌نامهٔ فلسفی باشد. قصدِ ما در اینجا تنها این بود که یک دیدگاه جامع و کلی از ترازهای بحثِ فلسفی در هوش مصنوعی، جدی‌ترین مسائلِ هر یک از این ترازها و اصلی‌ترین رویکردهای هوش مصنوعی ارائه دهیم.

** ما در اینجا برای رعایت اختصار تنها صورت‌بندی پایانی *دریفوس* مبتنی بر مسئلهٔ چارچوب را ارزیابی کرده‌ایم (برای آشنایی کامل با رویکرد *دریفوس* به هوش مصنوعی در دوره‌های مختلف، ر.ک: محمدعلی‌خلج، ۱۳۹۳).

هر موقعیت کدام گزاره‌ها برای یافتن پاسخ مناسب، مربوط و کدام نامربوط است. *دنت* برای توضیح، مسئلهٔ چارچوب رباتی را مثال می‌زند که سه بار یا به دلیل روی‌دادن اتفاقات پیش‌بینی‌نشده یا عدم موفقیت در تفکیک امور مربوط و نامربوط، در موقعیت ناکام می‌ماند و می‌ترکد.* به باور *دنت* ریشهٔ مشکل در این است که ما باید در یک برنامهٔ رایانه‌ای بر اساس پیش‌بینی‌هایمان از هر موقعیت، آن را صوری نماییم و بر اساس یک سری تعاریف و پیش‌فرض‌های صوری صورت‌بندی آگزیوماتیک از جهان ارائه دهیم– دقیقاً همان کاری که در یک برنامهٔ نرم‌افزاری رایانه‌ای انجام می‌دهیم. اما مسئله از همین‌جا بر می‌خیزد که این آگزیوم‌های ایزوله، هرگز انعطاف طرح‌افکنی انسان بر پدیدارها را ندارد. به بیان دیگر انسان‌ها می‌توانند با احساسات سوبژکتیوشان چیزهای پیرامونشان را برای خودشان معنادار کنند؛ برای نمونه یک موجود ناشناخته می‌تواند برای انسان‌ها با حس سوبژکتیو ترس معنادار شود و انسان‌ها نیازی ندارند که همهٔ ابژه‌های پیرامونشان پیشاپیش به شیوه‌ای ابژکتیو تعریف شده باشد؛ اما ربات‌ها از این حس سوبژکتیو

* نخستین نمونه از رباتی که *دنت* مثال می‌زند R1 نام داشته است. به R1 آموزش داده شده بود که باتری ذخیره‌اش را که در یک واگن کوچک در یک اتاق است، پیدا کند؛ اما همزمان بمبی ساعتی نیز در واگن قرار داده شده بود. ربات در هنگام خارج‌کردن باتری از واگن بمب را هم حرکت می‌دهد و می‌ترکد. ربات می‌دانسته است که بمب در واگن است؛ ولی پیش‌بینی نمی‌کرده که بیرون‌کشیدن باتری از واگن به حرکتِ واگن و انفجار بمب بینجامد. درنتیجه سازندگان به گفتهٔ *دنت* نتیجه می‌گیرند که «ربات بعدی ما باید به‌گونه‌ای ساخته شود که دلالت‌های قصدشدهٔ اعمالش بلکه دلالت‌های مربوط به عوارض جانبی‌اش را هم بشناسد» (Dennett, 1981,p.183). آنها ربات دوم به نام R1D1 را به‌گونه‌ای ساختند که بتواند عوارض جانبی رفتارش را استنتاج کند؛ اما این ربات نیز به دلیل استنتاجی اشتباه که هل‌دادن واگن– نسبت به اینکه به داخل واگن برود و باتری‌اش را بردارد– می‌تواند سرعت حرکت چرخ‌هایش را افزایش دهد و این ربات نیز با همین استنباط اشتباه ترکید. بنا به گفتهٔ *دنت* سازندگان نتیجه گرفتند که «ما باید به او تفاوت دلالت‌های مربوط و نامربوط را بیاموزیم» (Ibid). اینها ربات R2D1 را برای این منظور ساختند؛ اما دیدند که این ربات هم به طرز عجیبی در جلوی در اتاقی که واگن در آن بود، نشست تا مهلت بمب ساعتی سر برسد و بترکد. *دنت* نتیجه‌گیری می‌کند که «همهٔ این ربات‌ها با مشکل چارچوب مواجه بوده‌اند» (Ibid).

برخوردار نیستند و ازاین‌رو هر تغییر در هر موقعیت باید پیشاپیش برایشان تعریف‌شده باشد و برای تغییر غیرقابل انتظار جایی وجود ندارد. «تلاش برای به‌چنگ‌آوردن این آگزیوم‌ها با این بینش شروع شد که هیچ تغییری در موقعیت رخ نمی‌دهد، جز آنچه توسط عمل تعریف‌شده به روشنی آمده است» (Dennett, 1988, p.196).

پس مشکل اصلی به امور تعریف‌ناشده و غیرمنتظره باز می‌گردد. اما انسان‌ها به پیرامون خود حسی دارند که به کمک این حس، امور غیرمنتظره و تعریف‌ناشده را برای خود معنادار می‌کنند؛ برای نمونه اگر انسان موجودی را ببیند که تاکنون برایش ناشناخته است، با حس ترس یا حس کنجکاوی، آن موجود را برای خود معنادار می‌کند و واکنش مناسب را نسبت به آن نشان می‌دهد؛ اما اگر یک ربات با موجودی روبه‌رو شود که برایش تعریف نشده است، قادر به بازنمایی آن نخواهد بود و بدین ترتیب واکنشی به آن نشان نخواهد داد. پس مشکل چارچوب با یک تعبیر مشکل امور غیرمنتظره و برنامه‌ریزی‌نشده است و با تعبیر دیگر مشکل یافتن چارچوبی است که بتوان به کمک آن در هر موقعیت امور مربوط را از امور نامربوط به گونه‌ای تفکیک کرد که ربات با امری برنامه‌ریزی‌نشده روبه‌رو نشود.

به جهت اینکه پروژه‌های هوش مصنوعی در حل مشکل چارچوب ناکام مانده‌اند، نمی‌توانند همهٔ کارکردهای هوشمندانه را در تراز انسانی انجام دهند و ازاین‌رو همهٔ این پروژه‌ها در آزمون نامحدود *تورینگ* ناکام می‌شوند. حتی *دنت* که خود همواره بزرگ‌ترین فیلسوف مدافع هوش مصنوعی بوده است، ناتوانی در پاسخ به این مسئله را می‌پذیرد؛ چنان‌که می‌گوید: «من در اینجا هیچ راه‌حلی برای مسئلهٔ چارچوب عرضه نخواهم کرد؛ حتی هیچ پیشنهادی ندارم که کجا می‌توان راه‌حلی یافت» (Ibid, p.183). این اعتراف *دنت* به‌خوبی گواه آن است که مسئلهٔ چارچوب از مهم‌ترین مسئله‌های تراز فلسفه‌علمی در هوش مصنوعی است که مانعی بزرگ بر سر راه پیروزی هوش مصنوعی ضعیف ایجاد کرده است.

اما *دنت* به گونه‌ای هوشمندانه می‌کوشد صورت‌بندی مسئلهٔ چارچوب در هوش

مصنوعی را گواهی بر ادعای همسانی فلسفه و هوش مصنوعی قلمداد کند. پیش از این- و در بخش نخست از مقاله- نشان دادیم از نظر *دنت* هوش مصنوعی همان فلسفه است و گفتیم در بخش دوم، دو نمونه از فرآورده‌های فلسفی هوش مصنوعی از نظر وی ارزیابی می‌شوند. مسئلهٔ چارچوب یکی از این دو نمونه است.

> تنها راه‌حل‌ها نیستند که برای مشکلات ارزشمند هستند؛ بلکه مسئله‌های جدید نیز به‌خودی‌خود ارزشمند هستند. بهترین مشارکت هوش مصنوعی در فلسفه، مسئلهٔ حل‌ناشدهٔ معرفت‌شناسانهٔ عمیق و جدیدی است که در تولیدهای فلاسفه نادیده گرفته می‌شد؛ یعنی مسئلهٔ چارچوب (Ibid, p.274).

اما دست‌کم این بخش از ادعای *دنت* مبنی بر اینکه فیلسوفان پیش از هوش مصنوعی، مسئلهٔ چارچوب را نادیده گرفته‌اند، درست نیست. خود *دنت* می‌پذیرد نقدهای *دریفوس* بر هوش مصنوعی ریشه در مسئلهٔ چارچوب دارد (Ibid, p.184) و روشن است که سرچشمهٔ اصلی نقدهای *دریفوس* بر هوش مصنوعی فلسفهٔ پدیدارشناسی است. پس مشکل چارچوب به گونه‌ای در فلسفهٔ پدیدارشناسی مد نظر بوده است که *دریفوس* توانسته است بینش‌هایش را از آن وام بگیرد؛ چنان‌که در یک تعبیر از این مسئله بیان شد مسئلهٔ چارچوب مسئلهٔ چگونگیِ معنادارکردن امور تعریف‌ناشده و غیرمنتظره است. *هوسرل*، *هیدگر* و *مرلوپونتی*، هر کدام به شیوهٔ خودشان به این مسئله پاسخ داده‌اند. به بیان *هوسرل* در چنین مواردی ما مادهٔ محض را تجربه می‌کنیم و افق درونی تازه‌ای در ما دربارهٔ امر رخ‌داده شکل خواهد گرفت (Dreyfus, 1979, p.242). به بیان *هیدگر* در این موارد رویارویی، تودستی ما با چیزها مختل می‌شود و آنها به صورت چیزهای فرادستی برای ما آشکار می‌شوند (IbId, 1991, pp.76-77) و به بیان *مرلوپونتی* در مواجهه با امر غیرمنتظره مطابقت و تعادل بدنی ما با پیرامون به هم می‌خورد و تشویش درونی برخاسته از آن، امر غیرمنتظره را معنادار می‌کند (IbId, 2007, p.16). پس محتوای مسئلهٔ چارچوب در فلسفهٔ پدیدارشناسی با تعبیرهای گوناگون بیان شده است و *دریفوس* نیز از آنها در نقدهایش بر هوش مصنوعی بهره گرفته است. پس فلسفه پیش از هوش مصنوعی این مسئله را مطرح کرده است و بر خلاف ادعای *دنت* از این جهت

وام‌دار هوش مصنوعی نیست.

۲. مهم‌ترین مسئله‌های فلسفه‌ذهنی (هوش مصنوعی قوی)

در قسمت اول، بحث بر سر این بود که آیا هوش مصنوعی می‌تواند رفتارهای هوشمندانهٔ انسانی را پیاده‌سازی کند؛ اما در این بخش بحث بر سر آن است که اگر به فرض توانست این رفتارها را پیاده کند، آیا این امر دلیلی بر ذهن‌مندی رایانه‌هاست یا خیر؟ به بیان دیگر آیا هوش مصنوعی می‌تواند به پرسش اصلی فلسفهٔ ذهن- یعنی پرسش ذهن چیست- پاسخ گوید یا خیر؟ بحث‌های بسیاری در این حیطه میان فیلسوفان در گرفته است؛ اما می‌توان مهم‌ترین گفت‌وگوها را در این تراز، در ذیل سه بحث کلی گنجاند:

۱. مغالطهٔ آدمک (Homunculus fallacy) و مسئلهٔ *هیوم*؛

۲. استدلال اتاق چینی *سرل*؛*

۳. نقد ذهن‌مندی رایانه‌ها بر پایهٔ تئوری *گودل*.

در ادامه، هر یک از این مباحث را به بحث خواهیم گذاشت؛ اما پیش از آن نیاز است دو الگوی اصلی در هوش مصنوعی را معرفی کنیم؛ زیرا نسبت این نقدها با این دو الگو یکسان نیست. هوش مصنوعی نمادی (Symbolic AI) یا محاسبه‌گرا، پارادایم کلاسیک در هوش مصنوعی است که تفکر را پردازشگری اطلاعات و محاسبه بر روی نمادهای اتمیک می‌داند (Dennett, 1998, p.217). اما از دههٔ هشتاد میلادی جریانی در هوش مصنوعی قوت گرفت که به شبکه‌های عصبی پیوندگرا (Connectionist Neural network) معروف شد. همهٔ ما درکی از شیوهٔ پردازشگری یک سامانهٔ محاسباتی نمادی همچون ماشین‌حساب داریم؛ ولی برای آشنایی با شبکه‌های پیوندگرا نیاز است شرحی کوتاه از آنها به دست دهیم. *اندی کلارک*- از اصلی‌ترین حامیان پیوندگرایی- دربارهٔ این پارادایم جایگزین می‌نویسد:

رویکرد پیوندگرای بسط‌یافته خودش را به منزلهٔ یک جایگزین حقیقی برای

* اگرچه مغالطهٔ آدمک در بیشتر موراد در سایهٔ استدلال اتاق چینی قرار می‌گیرد، اما *سرل* که طراح هر دو مسئله است، اصرار دارد اینها دو مسئلهٔ کاملا متفاوت هستند (Searle, 1990, pp.26-27).

سامانه‌های «قاعده و نماد» کلاسیک معرفی می‌کند. این رویکرد جدید بر ۱. یک صورت جایگزین بازنمایی معرفت؛ ۲. یک شیوهٔ جایگزین عمل بنیادین پردازشگری؛ ۳. یک مجموعه از الگوریتم‌های آموزشی قوی مبتنی شده است (Clark, 1996, p.7).

چنانکه *کلارک* توضیح می‌دهد این‌گونه نیست که بازنمایی در این الگوواره به وسیلهٔ محاسبه بر روی بازنمودهای اتمیک نمادی بر اساس یک سری قواعد صوری انجام شود؛ بلکه «به جای آن پیوندگرایی از الگوهای فعال‌ساز در میان شمار بسیار زیادی نورون‌های ایدئال‌شده (واحدهای پردازشگر کوچک) برای کدگذاری محتویات معین بهره می‌گیرد» (Ibid). در این نوع بازنمایی، چند محتوا و ارتباطاتشان به وسیلهٔ همهٔ شبکه- یعنی الگویی از ارتباطات میان نورون‌ها- بازنمایی می‌شود. پردازشگری نیز در آنها به شیوه‌ای متفاوت است و به وسیلهٔ بردارهای عددی تعریف می‌شوند. با دادن یک بردار ناقص به عنوان راهنما، بازیابی اطلاعات شامل فرایند تکمیل بردارها خواهد بود. تعمیم نیز به وسیلهٔ ذخیره‌سازی الگوهای فعال فراتر از موقعیت در یک مجموعهٔ منفرد از وزن‌دهی‌ها در درازمدت ممکن می‌شود (Ibid, p.8). این بردارها وزن‌دهی چگونگی ارتباطات میان نورون‌ها را مشخص می‌کند و معیاری از بازنمایی ارتباطات میان محتواهای گوناگون هستند. افزون بر اینها، اساس کار شبکه‌های پیوندگرا بر آموزش است.

چنین شبکه‌های برون‌داد الگوریتم‌های آموزشی قدرتمندی است. یک شبکه ابتدا با وزن‌دهی تصادفی شروع می‌کند و در ادامه می‌تواند به صورت اتوماتیک این وزن‌دهی‌ها را به شیوه‌ای تغییر دهد که یک الگوی ورودی- خروجی مطلوب را کدگذاری کند. این نوع از آموزش به‌طورمعمول به وسیلهٔ نمایش‌دادن یک مجموعه از ورودی‌ها به همراه یک مجموعه از خروجی‌های مطلوب هدایت می‌شود» (Ibid).

پس شبکه‌های پیوندگرا بر خلاف سامانه‌های نمادی یک معادله و فرایندهای محاسباتی مشخص ندارند؛ بلکه مجموعه‌ای ورودی- خروجی برای نمونه دارند که شبکهٔ خود را با آنها تطبیق می‌دهد و یاد می‌گیرد با تعمیم آنها، ورودی- خروجی‌های درست

دیگر را نیز تشخیص دهد. برای اینکه درکی شهودی و معنایی از شیوهٔ کار این شبکه‌ها بیابیم، شبکهٔ پیوندگرای فرضی «الف» را در نظر بگیرید. قرار است در تصاویری که به «الف» نشان داده می‌شود، این شبکه اسب‌ها را تشخیص دهد. با نشان‌دادن چند تصویر که در آن اسب‌هایی با رنگ‌های مختلف وجود دارد، شبکهٔ «الف» یاد می‌گیرد به رنگ‌ها حساس نباشد و می‌تواند با تعمیم این وضعیت، اسب‌هایی با رنگ‌های دیگر را نیز تشخیص دهد. پس شیوهٔ پردازش این شبکه‌ها بر پایهٔ معادله‌های محاسباتی از پیش‌داده‌شده نیست، بلکه بر پایهٔ آزمون و خطا به وسیلهٔ نمونه‌های آغازین و تعمیم آنها به موقعیت‌های بعدی است. همچنین این شبکه‌ها کل‌گرایانه هستند. مثلاً فرض کنید که در تصاویر شبکهٔ «الف» در کنار اسب‌ها انسانی نیز وجود داشته باشد. شبکه به وسیلهٔ نمونه‌های زیادی که به آن نشان داده شده است، آموخته است میان مفهوم اسب و مفهوم چهارپا ارتباط درونی قوی (وزن زیادی از پیوندهای نورونی) وجود دارد. حال فرض کنید که تصویری را به شبکه نشان می‌دهند که انسانی شبیه چهارپایان حرکت می‌کند. شبکه می‌آموزد که میان مفهوم انسان و مفهوم چهارپا نیز می‌تواند ارتباطی درونی وجود داشته باشد؛ چون این وضعیت تنها یک بار رخ داده است، وزن بسیار کمی از پیوندهای نورونی را به خود اختصاص می‌دهد. پس کل مفاهیمی که در یک شبکه بازنمایی می‌شوند- مانند مفهوم انسان و چهارپا- می‌توانند از درون با هم ارتباط داشته باشند، اما ارتباط نزدیک‌تر میان دو مفهوم- مانند دو مفهوم اسب و چهارپا- با وزن‌دهی بیشتر پیوندهای نورونی میان الگوهای مختص به آنها بازنمایی می‌شود. حال با درنظرگرفتن اختلاف‌های بنیادین میان این دو الگو در ادامه سراغ سه مسئلهٔ اصلی در تراز فلسفهٔ ذهن، در هوش مصنوعی می‌رویم.

۱-۲. مغالطهٔ آدمک و مسئلهٔ هیوم

مغالطهٔ آدمک در تاریخ فلسفه چالشی دامنه‌دار بوده است و بسیاری از فیلسوفان در چاه آن گرفتار شده‌اند. در اینجا و در بستر مسئلهٔ امکان ذهن‌مندی ماشین‌ها گفت‌وگوی نقادانهٔ *سرل* و *دنت* پیرامون این مغالطه را بازسازی می‌کنیم. دغدغهٔ *دنت* را می‌توان در ادامهٔ

تلاش *گیلبرت رایل* برای رسیدن به زبانی تماماً فیزیکالیستی دربارهٔ امر واقع به شمار آورد. او می‌گوید «از زمان *رایل* تا به امروز ما چیزی به‌مثابه ذهن درون مغز را به سخره گرفته‌ایم؛ اما تمسخر کافی نیست» (Dennett, 1969, p.99). به باور *دنت* قیاس انسان و رایانه‌ها (ماشین *تورینگ*) امکان کنارگذاشتن این ذهن درون مغز یا همان آدمک فهم‌کننده را فراهم می‌آورد. بنا بر مقایسهٔ انسان و رایانه در رویکرد کارکردگرایانهٔ *دنت*، بدن و مغز انسان متناظر با سخت‌افزار رایانه‌ای و ذهن انسان متناظر با نرم‌افزار رایانه‌ای است. بدین ترتیب همان‌گونه که در رایانه‌ها فقط سخت‌افزار وجود دارد و نرم‌افزارها اموری ارجاع‌پذیر نیستند، در انسان نیز تنها مغز و بدن وجود دارد و ذهن امری ارجاع‌پذیر نیست (Ibid, p.111).

سرل مدعی است که در این صورت ذهن‌مندی امری «در نگاه بیننده» خواهد بود؛ زیرا باید فردی باشد که از عملکرد مدارهای الکتریکی رایانه‌ها توصیفی انتزاعی به عنوان نرم‌افزار ارائه دهد. از نظر *سرل* «ذهنی و غیرذهنی نباید یک وصف در نگاه بیننده باشد؛ بلکه باید ذاتیِ سامانه باشد» (Searle, 1980, p.7). به باور *سرل* اگر ذهن‌مندی را امری در نگاه بیننده- و نه ذاتی – بدانیم، اگر بیننده بخواهد، می‌تواند هیچ انسانی را ذهن‌مند به شمار نیاورد؛ درحالی‌که سامانه‌های اکتشافی (Heuristic System) را ذهن‌مند بداند. بدین‌سان این تمایز- یعنی تمایز ذهنی و غیرذهنی- در نوشته‌های هوش مصنوعی، در راه‌های مصیبت‌باری که طی می‌کنند، برای اینکه نشان دهند، هوش مصنوعی یک تحقیق شناختی است، محو می‌شود (Ibid)؛

> آموزش دربارهٔ ذهن به‌طوردقیق از آنجایی آغاز می‌شود که ما بپذیریم که ذهن بر خلاف ترموستات و اتومبیل دارای باور است و اگر شما این را رد کنید، تئوری دربارهٔ ذهن را به‌کلی رد کرده‌اید. آنچه ما به دنبالش هستیم، این است که چه چیزی ذهن را از ترموستات و کبد جدا می‌کند (Ibid).

سرل با برشمردن این پیامدهای ناگوار، مشکل آدمک را نه حل، بلکه منحل می‌کند. در سال ۱۹۸۱ میلادی *دنت* کوشید در مقالهٔ «هوش مصنوعی» همچون فلسفه و روان‌شناسی با صورت‌بندی تازه‌ای از مشکل آدمک، به مسئله‌ای که *سرل* در سال ۱۹۸۰

میلادی پیش پایش گذاشته بود، پاسخ گوید. وی نام این صورت‌بندی تازه را مسئلهٔ *هیوم* گذاشت. به باور *دنت*، هوش مصنوعی این مشکل فلسفی را حل کرده است.

می‌توان استدلال *دنت* را در این باره چنین بازگویی کرد. فرض کنید در هر الگوریتم نرم‌افزاری، هر حالت منطقی را به شیوهٔ *دنت* با یک کارکرد منطقی – محاسباتی، مثلاً ضرب دو در دو، یعنی با یک ورودی و خروجی مشخص، درون یک جعبهٔ سیاه تعریف کنیم. گویی آدمکی در این جعبهٔ سیاه وجود دارد و فهمی از این عمل منطقی محاسباتی داشته است و بر پایهٔ آن فهم، آن کارکرد را به اجرا در می‌آورد. حال، کاری که هوش مصنوعی انجام می‌دهد، این است که می‌کوشد هر کارکرد منطقی را به فرایندی از چندین کارکرد منطقی ساده‌تر دیگر تحویل برد؛ به این ترتیب، هر جعبهٔ سیاه، در یک فرایند مشخص به چندین جعبهٔ سیاه تجزیه می‌شود و یک آدمک تبدیل به چندین آدمک می‌شود که این بار هر یک کارکردی ساده‌تر را به انجام می‌رسانند. سرانجام در ساده‌ترین تراز کارکردی، به جایی می‌رسیم که انجام کارکرد، آن‌قدر ساده است که به آدمک دارای فهم نیازی نیست و ما در اینجا ساختار فیزیکی را با آدمک جایگزین می‌کنیم. بدین ترتیب با رسیدن به پایین‌ترین لایهٔ کارکردی، آدمک‌ها محو می‌شوند (See: Dennett, 1981). اما *سرل*، در ادامهٔ گفت‌وگوی نقادانه‌اش با *دنت* بر سر مسئلهٔ آدمک، در سال ۱۹۹۰ میلادی در برابر این استدلال *دنت*، کاملاً موضع‌گیری می‌کند:

> بسیاری مثل *دنت* فکر می‌کنند مغالطهٔ آدمک اصلاً مشکلی اساسی نیست؛ چراکه آدمک می‌تواند نادید گرفته شود و کنار گذاشته بشود. ایدهٔ ایشان این است که اگر ما به پایین‌ترین تراز یک فیلیپ فلاپ ساده برسیم، دیگر یک آدمک واقعی وجود نخواهد داشت؛ یعنی تجزیهٔ بازگشتی (Recursive decomposition) آدمک‌ها را خواهد زدود (Searle, 1990, p.28).

سرل بر پایهٔ جداسازی میان امر سینتتیک– یعنی همان برنامهٔ نرم‌افزاری– و امر فیزیکی به *دنت* پاسخ می‌دهد. همان‌گونه که مدافعان هوش مصنوعی باور دارند یک نرم‌افزار و سینتکس رایانه‌ای می‌تواند بر روی ساختارهای فیزیکی گوناگون پیاده‌سازی شود؛ درنتیجه، تعریف یک نرم‌افزار برای رایانه‌ها، هیچ ربطی به تحقق فیزیکی آن ندارد. به

بیان *سرل* «امر سینتتیک برای امر فیزیکی ذاتی و اصلی نیست» (Ibid, p.26). حال مغالطه‌ای که *دنت* انجام می‌دهد، به‌طوردقیق در همین‌جا نهفته است. وی در پایین‌ترین درجهٔ تراز سینتتیک- نرم‌افزاری، ناگهان به تراز فیزیکی پرش می‌کند. درست است که در پایین‌ترین تراز تنها کاری ساده انجام می‌شود، اما در اینجا نیز همچنان نرم‌افزار- و به تعبیر دیگر ذهن برنامه‌نویس- را داریم که فرایند محاسبه را هدایت می‌کند. هویت فیزیکی این کار ساده‌ای که انجام می‌شود، به‌خودی‌خود قابل تعریف است؛ اما هویت نرم‌افزاری آن در ارتباط با کارهای دیگر و در یک الگوریتم کلی معنادار می‌شود. *دنت* با نادیده‌گرفتن ارتباطات این کار ساده با دیگر کارهای ساده- ارتباطی که دقیقاً آدمک آن را برقرار می‌کند- از تراز سینتتیک- نرم‌افزاری به تراز فیزیکی جهش کرده است. *سرل* همین نکته را از زاویه‌ای دیگر بیان می‌کند. گویا *دنت* خود در این فرایند بازگشتی، یک آدمک دیگر را وارد کرده است. *سرل* می‌گوید:

> بدون وجود یک آدمک که بیرون از این فرایند تجزیهٔ بازگشتی به ترازهای پایین‌تر ایستاده باشد، حتی یک سینتکس برای عمل نداریم؛ پس تلاش برای رفع مغالطهٔ آدمک به کمک تجزیه‌های بازگشتی شکست می‌خورد؛ زیرا تنها راهی که می‌توان امر سینتتیک را چیزی ذاتاً متعلق به امر فیزیکی دانست، آن است که یک آدمک را در فیزیک وارد کنیم (Ibid, p.29).

پس به باور *سرل*، تجزیهٔ بازگشتی، خود پیشاپیش یک سینتکس تعریف‌شده‌ای را پیش‌فرض می‌گیرد و بر پایهٔ آن، نقش‌ها را ساده‌سازی می‌کند؛ یعنی همان نرم‌افزاری که باید در ذهن آدمکی (مثلاً برنامه‌نویس) تعریف شده باشد. پس به این روش هرگز نمی‌توان به مغالطهٔ آدمک پاسخ گفت. *سرل* به‌روشنی نشان می‌دهد که «بیشتر کسانی که در نظریهٔ محاسبه‌گرایانه عمل می‌کنند، به گونه‌های متفاوت مغالطهٔ آدمک دچار هستند» (Ibid, p.28). بدین ترتیب نتیجهٔ سخن همهٔ ایشان آن است که امر ذهنی صفتی در نگاه بیننده و نسبی است.

ارزیابی

به نظر می‌رسد استدلال *سرل* در نقد *دنت* کاملاً به‌جاست و به صورت یکسان بر هر دو

پارادایم هوش مصنوعی وارد است. حتی اگر این استدلال را نیز به کار نبریم، *دنت* به‌هیچ‌وجه توضیح نمی‌دهد که چگونه در پایین‌ترین تراز نرم‌افزاری، ناگهان به ترازی سخت‌افزاری می‌پرد. همین که در سال ۱۹۸۸ میلادی *دنت* به‌صراحت گفت: «هوش مصنوعی هنوز هیچ مسئله‌ای را دربارهٔ ذهن حل نکرده است» (Dennett, 1988, p.276)، نشان می‌دهد که خود نیز نسبت به ضعف استدلالش در پاسخ به مغالطهٔ آدمک واقف شده است. پس مغالطهٔ آدمک نخستین مسئله‌ای است که در تراز فلسفه‌ذهنی پیش پای هوش مصنوعی قوی قرار دارد و ایشان باید بازتعریفی از رویکردشان ارائه دهند که امر ذهنی را تبدیل به امری در نگاه بیننده و دلخواهانه نکند. همچنین باید به این نکته اشاره کرد که پاسخ به مغالطهٔ آدمک دومین مسئله‌ای بوده است که بنا بر ادعای *دنت*، فلسفه در آن وامدار هوش مصنوعی است؛ ولی دیدیم پاسخ وی به هیچ روی قانع‌کننده نیست و ازاین‌رو *دنت* نمی‌تواند مثالی از کمک هوش مصنوعی به فلسفه بیاورد و درنتیجه صورت‌بندی کلی‌اش از هوش مصنوعی به‌مثابه نوعی فلسفه‌ورزی نیز به کلی مخدوش است.

۲-۲. استدلال اتاق چینی سرل

«یک برنامهٔ رایانه‌ای به‌خودی‌خود هرگز شرط کافی برای قصدیت نیست. برهان اصلی مقاله، مربوط به این مدعاست» (Searle, 1980, p.1). این جملهٔ *سرل*، محتوای اصلی برهان اتاق چینی است. این برهان در صدد است تا مثال نقضی برای ادعای کلی همسانی برنامهٔ نرم‌افزاری و ذهن باشد. برای این کار، *سرل* وضعیتی خیالی را ترتیب می‌دهد که برنامهٔ نرم‌افزاری داریم، اما فهم و قصدیتی نداریم. صورت استدلالش چنین است: «صورت برهان به این صورت است که نشان دهد عامل انسانی می‌تواند یک برنامه را لحاظ کند، بی‌آنکه نیازی به فرض قصدیت باشد» (Ibid).

آزمایش ذهنی اتاق چینی را می‌توان چنین روایت کرد: فرض کنید داستانی به زبان چینی برای چند نفر- ازجمله خود من- که هر یک در اتاق‌های جداگانه‌ای قرار داریم، بیان شود. همهٔ افراد- به جز من- زبان چینی بلدند. سپس پرسش‌هایی را به زبان چینی

به صورت نوشته، به درون اتاق می‌فرستند و پاسخ‌هایی را به صورت نوشته دریافت می‌کنند. فرض کنید من متناظر با هر پرسش دستورالعمل‌هایی به زبان فارسی یا انگلیسی دارم که بر اساس آنها شکلک‌هایی را که همان الفبای چینی هستند، در کنار هم می‌گذارم و به بیرون اتاق می‌فرستم. دستورالعمل‌ها به‌گونه‌ای تنظیم شده است که وقتی حروف الفبای چینی در کنار هم قرار می‌گیرند، پاسخ پرسشی را بدهند که به زبان چینی به درون اتاق فرستاده شده است. بدین ترتیب من هیچ فهمی از داستان، پرسش‌ها و پاسخ‌های چینی نداشته‌ام؛ اما از بیرون تفاوتی میان پاسخ‌های من و افراد چینی نیست. «هیچ کس جز خود من نمی‌تواند به این امر پی ببرد که من چینی بلد نیستم» (Ibid, p.3).

حال فرض کنید داستان، پرسش‌ها و پاسخ‌ها به زبان انگلیسی و یا فارسی که من بنا به فرض بلد هستم، طرح شود. «فرض کنید من همان داستان‌ها را در انگلیسی بشنوم و به پرسش‌های آنها پاسخ دهم. از بیرون نمی‌توان تمایزی میان پاسخ انگلیسی و چینی من یافت؛ ولی من انگلیسی را می‌فهمم و چینی را تنها با تولید نمادهای صوری فهم‌ناشده پاسخ می‌دهم. تاآنجاکه چینی مد نظر باشد، شبیه یک رایانه عمل می‌کنم. من تنها اعمال محاسباتی را بر روی نمادهای معین، به صورت صوری انجام می‌دهم» (Ibid). پس آن مثال نقضی که *سرل* به دنبال آن بود، یافت می‌شود؛ چراکه ما در اینجا برنامهٔ رایانه‌ای داریم، ولی هیچ فهم، تفکر و حالت شناختی نداریم. عملکرد رایانه‌ها نیز همواره همانند عملکرد من دربارهٔ پرسش‌های چینی است؛ یعنی رایانه بدون اینکه هیچ فهمی از آنچه با آن سروکار دارد، داشته باشد، تنها یک سری دستورالعمل‌ها را اجرا می‌کند. پس انجام دستورالعمل‌ها، بدون داشتن درکی از محتوای دستورها، به‌هیچ‌روی برابر با فهم و تفکر انسانی نیست.

به بیان دیگری که *سرل* به کار می‌گیرد، رایانه‌های دیجیتال تنها سینتکس (دستورالعمل) و نه سمنتیک (وجه معنایی) دارند (Searle, 1980, p.12)؛ یعنی الگوریتم‌های برنامه‌ها را می‌توان همچون مجموعه قواعد (دستوری) در نظر آورد؛ اما اینها برای تولید معنا کافی نیست. این داوری به‌کلی به رویکرد معناشناسانهٔ *سرل* بسته

است. وی معناداری را همسان با برخورداری از محتوا و قصدمندی به سوی آن محتوا می‌داند؛ یعنی هر حالت ذهنی، قصدیتی به سوی یک محتواست و به‌هیچ‌رو- به‌گونه‌ای صرفاً صوری- نمی‌توان آنها را تعریف کرد. اما «برنامهٔ رایانه‌ای کاملاً صوری است؛ گو اینکه حالت‌های قصدی در این معنا صوری نیستند. آنها به واسطهٔ محتوایشان تعریف می‌شوند و نه به واسطه صورتشان» (Ibid). *سرل* برهان اتاق چینی را این‌گونه خلاصه می‌کند: «این برهان- یعنی برهان اتاق چینی- بر این حقیقت سادهٔ منطقی بنا شده است که وجه سینتتیک (دستوری)- نه همان وجه سمانتیک (معنایی)- است و نه شرط کافی برای وجه معنایی است» (Ibid, 1990, p.21).

ارزیابی

بی‌گمان در تاریخ تأمل فلسفی دربارهٔ هوش مصنوعی، هیچ موضوعی به اندازهٔ استدلال اتاق چینی *سرل* بحث و بررسی نشده است. چالش در استدلال اتاق چینی از خود مقالهٔ *سرل* آغاز شد. او کوشید شش نقدی را که ممکن است به این استدلال وارد شود، خود تقریر نماید و پاسخ گوید. *بادن*، *بلاک*، *تاگارد*، *هیوزر*، *بیل*، *دنت* و هر یک از فیلسوفان ذهن دیگر نیز پاسخ‌هایی درخور درنگ به این استدلال داده‌اند؛ ولی عمیق‌ترین پاسخ از آنِ *دیوید چالمرز* است. مشکل بنیادین در اینجاست که برای هر یک از نمادهای صوری در یک نرم‌افزار رایانه‌ای هیچ یک از ابژه‌های بیرونی تفاوتی نمی‌کند. نماد اسب که قرار است اسب را بازنمایی کند، می‌تواند بار دیگر برای نشان‌گری انسان به کار رود. نسبت نماد اسب با مفهوم چهارپا، با نسبت نماد انسان با این مفهوم تفاوتی ندارد؛ اما تصورات ما پندارهایی خنثی نیستند. تصور اسب برای ما، هرگز انسان را برای ما نشان‌گری نخواهد کرد. این‌گونه نیست که نماد اسب و نماد انسان نسبت به مفهوم چهارپا در یک جایگاه قرار داشته باشند. پس میان نمادهای درونی انسانی و مجموعهٔ مفاهیم و درون‌مایه‌هایی که این نمادها بازنمایی می‌کنند، ارتباطی ذاتی است. اما چنین ارتباطی در نمادهای رایانه‌ای و ابژه‌های بیرونی به چشم نمی‌خورد و برای این نمادها همهٔ ابژه‌ها در یک تراز یکسان قرار دارند. *چالمرز* خود با مثالی این نکته را بازگویی می‌کند:

یک نشانه‌ای که مفهوم فیل را بازنمایی می‌کند، می‌تواند با برچسب فیل نشانه‌گذاری شود؛ اما به همان میزان ممکن است با برچسب سیب نشانه‌گذاری شود... هیچ ویژگی درونی نشانهٔ فیل را به مفهوم فیل، نسبت به مفهوم سیب نزدیک‌تر نمی‌کند (Chalmers, 1992, p.13).

وقتی *سرل* می‌گوید نمادهای صوری بی‌محتوا هستند، دقیقاً چنین سخنی را مد نظر دارد؛ یعنی هیچ ویژگی درونی، نماد سیب را به معنای سیب و ارتباطات آن با دیگر معانی پیوند نداده است. پس حالت‌های ذهنی به گونه‌ای درونی، دارای معنا و محتوا هستند؛ درحالی‌که نمادهای برنامه‌های رایانه‌ای فاقد آنها هستند.

اما با وجود به‌دست‌دادن این بینش، *چالمرز* خود چگونه به استدلال اتاق چینی پاسخ می‌دهد؟ اگر بتوان به نوعی نمادها را از این حالت خنثی خارج کرد، آن‌گاه وجه سمانتیک نیز پیدا می‌کنند؛ یعنی اگر بتوان کاری کرد که نماد سیب به بازنمایی مفهوم سیب به همراه ارتباطات معنایی نسبت به هر مفهوم دیگری نزدیک‌تر باشد، آن‌گاه می‌توان نماد سیب را به‌گونه‌ای درونی دارای معنا و محتوای سیب دانست. *چالمرز* چارهٔ این مشکل را در ایجاد ساختاری درونی در لایه‌ای زیرین برای نمادها می‌داند. وی بر آن است که رویکرد پیوندگرا به‌طوردقیق در همین جهت گام بر می‌دارد. «مطابق با ادعای پیوندگرایان بازنمودهای توسعه‌یافته، دقیقاً به جهت ساختار درونی قوی‌شان، می‌توانند محتوایی را حمل کنند» (Ibid, p.15). پس ما در هر رویکرد پیوندگرا، دو تراز داریم؛ یکی تراز بازنمودهای معنادار است که قرار است هر یک با معنایشان و همچنین شبکه‌ای از معانی دیگر پیوندی درونی با قوت و ضعف گوناگون داشته باشند. دیگری ترازی زیرین که خود معنا ندارد، اما ساختاری را فراهم می‌کند که بازنمودها در تراز بالاتر را به سوی معنایشان جهت‌مند می‌کند. حال *چالمرز* بر آن است که در پیوندگرایی تراز بی‌معنای زیرین همان تراز محاسباتی است که با گره‌های خنثی سروکار دارد و تراز بازنمودی، همان تراز رویین بازنمایی‌های معنادار است.

اکنون باید سراغ مقدمهٔ اصلی استدلال *سرل* می‌رویم: «سینتکس برای سمانتیک نه لازم است و نه کافی». *چالمرز* می‌کوشد معنای دقیق‌تری از این مدعا به دست دهد. وی

نخست می‌پرسد: «سینتتیک‌بودن چه معنایی برای یک سامانه دارد؟ این دقیقاً به این معناست که سامانه به واسطهٔ دنبال‌کردن قواعد کار کند» (Ibid. p.16). برخورداری از وجه محتوای سمانتیک چطور؟ «همان‌گونه که دیدیم، اندیشهٔ محتوا در اینجا درون‌گرایانه است؛ یعنی محتوایی قصدی» (Ibid). حال می‌توان این مقدمهٔ اصلی را چنین ترجمه کرد: «هیچ سامانه‌ای که رفتاری بر اساس قاعده از خود نشان دهد، نمی‌تواند محتوای درونی داشته باشد» (Ibid).

اما مغز انسان دقیقاً مثال نقضی برای این سخن است؛ زیرا اگر سامانه‌ای دارای محتوای درونی باشد، آن سامانه مغز انسان است. همچنین مغز می‌تواند به‌مثابه سامانه‌ای در نظر گرفته شود که مجموعهٔ قواعد محکم (Iron-clad) را دنبال می‌کند (Ibid). قواعد فیزیکی دست‌کم از دیدگاه طبیعت‌گرایانهٔ بیشتر فیلسوفان ذهن- ازجمله خود *سرل*- مغز سامانه‌ای است که می‌تواند رفتار معنادار تولید کند و به این معنا دارای محتوای معنادار و وجه سمانتیک است. اما وجه سینتتیکی که *چالمرز* مد نظر دارد، پیروی از قواعد فیزیکی است. او پیش از این سینتتیک‌بودن را به پیروی از قواعد تعریف کرد. حال روشن است که مغز سامانه‌ای است که از قواعد فیزیکی پیروی می‌کند و درنتیجه دارای وجه سینتتیک است. نه تنها مغز، بلکه- همان‌گونه که در بالا توضیح دادیم- شبکه‌های پیوندگرا که در اصل از مغز الگوبرداری شده‌اند، نیز مثال نقضی بر این مدعایند؛ چراکه این شبکه‌ها نیز در تراز محاسباتی زیرین فاقد معنا و در تراز بازنمودی رویین دارای محتوای معنایی‌اند. پس با استناد به مغز و شبکه‌های عصبی پیوندگرا اگر مراد *سرل* از مقدمهٔ اصلی‌اش ترجمه‌ای باشد که *چالمرز* از آن به دست می‌دهد، آشکارا نادرست است؛ ولی به نظر می‌رسد که به گونه‌ای شهودی این مقدمه همچنان باورپذیر باشد.

چالمرز در پاسخ به این پرسش که قدرت شهودی ادعای اصلی *سرل* از کجا می‌آید، چنین پاسخ می‌دهد

> روشن است که این یقین از امر زبانی می‌آید: قواعد دستوری که واژه‌ها اطاعت می‌کنند، برای اینکه این واژه‌ها محتوای معنایی از خود نشان دهند، کافی نیست.

پس برای یک جمله، این درست است که بگوییم سینتکس برای سمانتیک کافی نیست» (Ibid).

در امر زبانی، سینتکس و سمانتیک- هر دو- در تراز جمله در نظر گرفته شده‌اند؛ درحالی‌که در مغز، سینتکس برای نورون‌ها و سلول‌های عصبی در نظر گرفته می‌شد و سمانتیک برای واژه‌ها و جمله‌ها به کار می‌رفت و هر یک در ترازی جداگانه قرار داشتند. پس جایی که سینتکس و سمانتیک در ترازی یکسان قرار داشته باشند (مانند امر زبانی)، مقدمهٔ اصلی *سرل* درست و جایی که این دو در دو تراز جدا باشند (مانند مغز)، مقدمهٔ *سرل* نادرست است. پس *چالمرز* به جای مقدمهٔ اصلی *سرل* این گزاره را قرار می‌دهد: «سینتکس در یک تراز معین، هرگز شرط لازم و کافی برای سمانتیک در همان تراز نیست» (Ibid, p.17).

رویکرد سرل دربارهٔ ذهن چیزی بیش از استدلال اتاق چینی اوست. وی باور دارد که ذهن مجموعهٔ حالت‌های شناختی وجودی است که تنها و تنها معلول مغز ارگانیسم‌های جاندار است. اگر این پیش‌فرض درست باشد، هیچ سخت‌افزار رایانه‌ای نمی‌تواند حالت‌های شناختی ایجاد کند و درنتیجه ذهن داشته باشد؛ ولی حکایت این پیش‌فرض با استدلال اتاق چینی متفاوت است. تا جایی که به این استدلال مربوط است، *چالمرز* صورت‌بندی درستی از آن ارائه می‌دهد و مغز انسان- و به تبع آن- شبکه‌های عصبی پیوندگرا از دیدگاهی طبیعت‌گرایانه مثال نقضی بر ادعای *سرل* است و بر این اساس *چالمرز* نشان می‌دهد که نقد *سرل* بر سامانه‌های پیوندگرا در هوش مصنوعی وارد نخواهند بود؛ زیرا در این سامانه‌ها- همان‌گونه که گفتیم- امر سمانتیک و امر سینتتیک در یک تراز نیستند؛ درنتیجه استدلال اتاق چینی به‌خودی‌خود امتناع ذاتی هوش مصنوعی قوی را به اثبات نمی‌رساند.*

* *چالمرز* در یک مبنای محاسبه‌گرایانه برای هوش مصنوعی، پاسخ دیگری را ذیل یکی از مشهورترین پاسخ‌های داده‌شده به استدلال اتاق چینی به نام پاسخ سامانه، به این استدلال می‌دهد که به‌هیچ‌روی خلاقیت موجود در پاسخی را که در اینجا از وی نقل کردیم، ندارد.

۳-۲. نقد هوش مصنوعی بر پایهٔ نظریهٔ گودل

از مباحث پیچیده و پردامنه در تاریخ هوش مصنوعی گفت‌وگوهایی است که پیرامون نظریهٔ *گودل* و نتایج آن دربارهٔ هوش مصنوعی شکل گرفته است. نخستین استدلال بر نادرستی یکی‌دانستن ذهن و هوش مصنوعی- که در این استدلال منظور از آن همان الگوی *تورینگ* است- در سال ۱۹۵۸ میلادی بر پایهٔ همین نظریه و از سوی *نیگل* و *نیومن* ارائه شد. *پاتنم* مهم‌ترین فیلسوفی بود که از هوش مصنوعی در برابر نقد این دو دفاع کرد. بعدها و در دههٔ ۹۰ میلادی *راجر پن‌رز* این استدلال را در یک صورت‌بندی متفاوت عرضه کرد و *چالمرز* و *سرل* مهم‌ترین ناقدان وی بودند. در ادامه به صورت فشرده مهم‌ترین بحث‌های مطرح‌شده در این گفت‌وگوها را طرح و ارزیابی خواهیم کرد. *پاتنم* استدلال *نیگل* و *نیومن* را به این شیوه بازسازی می‌کند:

> این امر (توسط *نیگل* و *نیومن* در کتاب **برهان** *گودل*) ادعا شده است که «این تئوری- یعنی تئوری *گودل* (Godel's theorem)- نشان داده است که ساختار و قدرت ذهن انسان، پیچیده و ماهرانه‌تر از آن است که از هر ماشین نازنده‌ای بتوان انتظار داشت»؛ بنابراین ماشین *تورینگ* نمی‌تواند همچون الگویی برای ذهن انسان به شمار آید؛ اما این یک اشتباه ساده است. در نظر بگیرید که T یک ماشین *تورینگ* باشد که برای من در این معنا «بازنمایی می‌شود» که T می‌تواند قضایای ریاضی را که من می‌توانم اثبات کنم، اثبات نماید. پس برهان (البته *نیگل* و *نیومن* برهانی نمی‌آورند؛ بلکه من فرض می‌کنم که چنین برهانی در ذهنشان بوده است.) این است که به واسطهٔ تکنیک *گودل*، من می‌توانم گزاره‌ای را کشف کنم که T نمی‌تواند اثبات کند (Putnam, 1975, p.366)

می‌توان این استدلال را در سه مقدمه بیان کرد:

۱. بر اساس تئوری *گودل* برای هر نظام منطقی سازواری، می‌توان قضیه‌ای (مثلاً U) را یافت که تصمیم‌ناپذیر باشد. بدین معنا که این نظام نمی‌تواند درستی آن را نشان دهد، اما ذهن من می‌تواند. مقصود از نظام منطقی سازوار نظامی است که در آن گزاره‌ای وجود ندارد که هم خودش و هم نقیضش بر پایهٔ آگزیوم‌ها قابل اثبات باشند.

۲. هر ماشین *تورینگ* در ذهن من به عنوان یک نظام منطقی سازوار (مثلاً T) تعریف شده است.

۳. من هم می‌توانم هر یک از (T)ها را تعریف کنم، هم می‌توانم U هر کدام از این (T)ها را اثبات کنم. درنتیجه من هیچ یک از این (T)ها نیستم؛ یعنی من یک ماشین *تورینگ* (یا یک نظام منطقی سازوار) نیستم.

با این صورت‌بندی، نقد *پاتنم* به *نیگل* و *نیومن* و استدلالشان نیز تا حدود زیادی روشن می‌شود. وی خطای این دو نفر را برآمده از کاربرد نادرست نظریهٔ *گودل* می‌داند. *پاتنم* میان تعریف یک ماشین *تورینگ* برای ذهن با برابری ذهن با یک ماشین *تورینگ* تفاوت قائل می‌شود. آنچه *نیگل* و *نیومن* فرض گرفته‌اند، تعریف یک ماشین *تورینگ* برای ذهن است؛ یعنی هم نظام منطقی T و هم گزارهٔ U- که برای ماشین تورینگ تصمیم‌ناپذیر است- درون نظام منطقی ذهن من جای می‌گیرند؛ اما گزارهٔ U وجود خواهد داشت که برای ذهن من تصمیم‌ناپذیر است. بر پایهٔ این نظریه کاملاً محتمل است گزاره‌ای وجود داشته باشد که برای ذهن تصمیم‌ناپذیر باشد. پس این حرف درست است که ذهن هیچ یک از ماشین‌های تورینگ تعریف‌شده برای خودش نیست؛ چراکه برای همهٔ این ماشین‌ها به‌مثابه نظام‌های بستهٔ منطقی، یک گزارهٔ (U)ای وجود دارد که برای آنها تصمیم‌ناپذیر و برای ذهن تصمیم‌پذیر است؛ ولی ممکن است گزارهٔ (U)ای وجود داشته باشد که برای ذهن تصمیم‌ناپذیر باشد؛ ازاین‌رو اگر وجود چنین گزاره‌ای ممکن است، به فرض ماشین *تورینگ*‌بودن ذهن نیز تناقضی وجود ندارد.

اما پاسخ *پاتنم* بدین معناست که ما نمی‌توانیم ذهن خود را به‌مثابه یک سیستم منطقی سازوار برای خود تعریف کنیم؛ یعنی ما نمی‌توانیم سازواری خود را نشان دهیم. پس پاسخ *پاتنم* به این مدعا ختم خواهد شد که ذهن ما می‌تواند ماشینی *تورینگ* باشد که قابل تعریف‌کردن نیست و سازواری‌اش را نمی‌توان نشان داد. *نیگل* و *نیومن* به‌شدت دربارهٔ این موضع *پاتنم* واکنش نشان دادند:

دکتر *پاتنم* این تأمل نتیجه‌بخش ما را که تئوری *گودل* نشان می‌دهد ذهن انسان

پیچیده‌تر از هر ماشین است که تحقق یافته است، به‌مثابه «خطایی ساده» و «کاربرد اشتباه از تئوری محض و سادهٔ *گودل*» رد کرده است... اما در نگاه ما موضع نقد دکتر *پاتنم*... یک جزمیت خالص است» (Nagel and Newman, 1961, pp.210-211)

از نظر این دو، هر ماشین *تورینگی* فرض سازواری را در خود دارد و اگر *پاتنم* مدعی است می‌توان ذهن انسان را یک ماشین *تورینگ* دانست، باید بتوان سازواری آن را نشان داد؛ یعنی بر خلاف نظر *پاتنم* باید بتوان ذهن را به عنوان یک ماشین *تورینگ* برای خودش تعریف کرد. این دقیقاً چالش‌برانگیزترین نقطهٔ این استدلال است. در ادامه خواهیم دید که گفت‌وگوهای بعدی دربارهٔ این استدلال نیز به همین‌جا ختم می‌شود.

پن‌رز صورت‌بندی متفاوتی از استدلال بر نقد همسانی ذهن و ماشین *تورینگ* بر پایهٔ نظریهٔ *گودل* عرضه می‌کند. استدلال او بر چندین مقدمه استوار است:

۱. فرض کنیم که ذهن من T- یعنی یک سیستم فورمال سازوار- است.

۲. پس گزارهٔ «من T هستم» نیز گزارهٔ درستی است.

۳. می‌دانیم بر پایهٔ تئوری *گودل* برای هر (T)ای یک قضیهٔ تصمیم‌ناپذیر (مثلاً U) وجود دارد که برای ذهن من تصمیم‌پذیر است.

۴. حال فرض کنید T' برابر باشد با T به اضافهٔ این گزاره که «من T هستم». ازآن‌رو که هر سیستم سازوار به اضافهٔ یک گزارهٔ درست، یک سیستم سازوار را نتیجه می‌دهد؛ درنتیجه T' نیز سازوار است و در قضیهٔ *گودل* جای می‌گیرد و در نتیجه U'اش برای من تصمیم‌ناپذیر خواهد بود؛ درنتیجه: بر اساس مقدمهٔ سوم هر (U')ای برای من تصمیم‌پذیر است و بر اساس مقدمهٔ چهارم، اما U' یک سیستم ویژه برای من تصمیم‌ناپذیر است. پس بر پایهٔ فرضیهٔ *گودل* این فرض که من یک سیستم فورمال سازوار- همچون یک ماشین *تورینگ*- هستم، به تناقض ختم می‌شود. پس این فرض نادرست است (Penrose, 1996, p.3).

اما *چالمرز* در نقد *پن‌رز* اگرچه وجود تناقض را می‌پذیرد، نتیجه‌گیری *پن‌رز* را به چالش می‌کشد:

> هنگامی‌که تناقض ایجاد می‌شود او [یعنی *پن‌رز*] این فرض را که نیروی استدلال ما می‌تواند با یک سیستم فورمال F [در مثال ما T] منطبق شود، مقصر می‌داند... [اما] مسئولیت تناقض در جایی به جز این فرض محاسبه‌گرا قرار دارد. فرضِ دانستن سازواری است که باید کنار گذاشته شود... شاید ما سازوار باشیم، اما ما نمی‌توانیم ضرورتاً بدانیم که سازوار هستیم» (Chalmers, 1995).

همان‌گونه که پیش از این دیدیم، *پاتنم* هم مشکل را دقیقاً در همین فرض دانسته بود و درنتیجه *چالمرز* نیز با او هم‌عقیده است که ما نمی‌توانیم به‌ضرورت سازواربودن خودمان را نشان دهیم. این امر نشان می‌دهد که گره اصلی بحث همین‌جاست و برای ارزیابی استدلال‌های مربوط به قضیهٔ *گودل* در هوش مصنوعی باید دربارهٔ این ادعا داوری کرد.

ارزیابی

به نظر می‌رسد دست‌کم تا جایی که به ذهنیت پژوهشگران هوش مصنوعی و تعریف خود *تورینگ* از ماشین تورینگ باز می‌گردد، *چالمرز* و *پاتنم* در اشتباه باشند و به فرض ماشین تورینگ سازواری نیز نهفته است. مدعای اصلی روش *تورینگ* آن است که تفکر و پردازشگری ذهنی چیزی جز همین محاسبهٔ منطقی- ریاضیاتی نیست؛ یعنی یک سری آگزیوم و قواعد محاسباتی وجود دارد که می‌توان همهٔ کارکردهای ذهنی را به وسیلهٔ الگوریتم‌های نوشته‌شده بر اساس آنها در ماشین *تورینگ* پیاده‌سازی کرد؛ اما اگر این نظام منطقی- محاسباتی با تناقض درونی همراه باشد، ایدهٔ *تورینگ* از اساس به چالش کشیده خواهد شد (Turing, 1950, p.449). اما *چالمرز* بار دیگر راه چاره را در استناد به شبکه‌های عصبی پیوندگرا می‌داند. او در نقد *پن‌رز* می‌گوید: «*پن‌رز* چگونه استدلال می‌کند که ما می‌توانیم ببینیم که F [در مثال ما T] سازوار است؟ او در بخش ۳-۳ استدلال می‌کند که ما می‌توانیم F را به‌مثابه یک سیستمی از آگزیوم‌ها و قواعد استنباط در نظر آوریم» (Chalmers, 1995).

یک سیستم ساخته‌شده از آگزیوم‌های سازوار به اضافهٔ قواعد استنباط سازوار خود نیز سازوار است؛ ولی *چالمرز* در ادامه معتقد است که همهٔ پروژه‌های هوش مصنوعی را

نمی‌توان به این آگزیوم‌ها و قواعد تجزیه کرد. «حتی در حوزهٔ موجود تحقیق هوش مصنوعی فرایندهای محاسباتی بسیاری وجود دارند، برای نمونه شبکه‌های پیوندگرا که به آگزیوم‌ها و قواعد استنباط تجزیه نمی‌شود» (Ibid). این شبکه‌ها به گونه‌ای کل‌گرایانه به سوی پاسخ حرکت می‌کنند و هیچ الگوریتم برنامه‌نویسی‌شدهٔ جامعی بر آنها حاکم نیست که سازواری آنها را تضمین کند؛ پس نمی‌توان سازواری این شبکه‌ها را نشان داد و ازاین‌رو نقد *پن‌رز* بر آنها وارد نیست. بدین ترتیب دفاع *پاتنم* از هوش مصنوعی در برابر نقد *نیگل* و *نیومن* نادرست است؛ چراکه وی تنها ماشین‌های تورینگ کلاسیک را مد نظر داشته است؛ اما *چالمرز* تا جایی که سخنش را به شبکه‌های عصبی پیوندگرا محدود نماید، دفاع درستی از هوش مصنوعی در برابر نقدِ *پن‌رز* و هر نوع نقد دیگری که بر پایهٔ فرضیهٔ *گودل* صورت‌بندی شده باشد، عرضه کرده است.

نتیجه‌گیری

نخستین هدف مقالهٔ پیش رو این بوده است که یک صورت‌بندی کلی از جایگاه فلسفی هوش مصنوعی ارائه کند. برای این منظور مهم‌ترین صورت‌بندی‌های عرضه‌شده از این جایگاه از سوی *پاتنم*، *دنت* و *سرل* طرح و ارزیابی شد و نقاط ضعف هر یک از اینها نشان داده شد. سپس صورت‌بندی برگزیده به کمک آزمون *تورینگ* و با تفکیک دو تراز فلسفه‌علمی و فلسفه‌ذهنی ارائه شد. بر این اساس پرسش اصلی در تراز فلسفه‌علمی‌این بود که آیا دانش هوش مصنوعی می‌تواند ماشینی بسازد که در آزمون نامحدود *تورینگ* پیروز شود؛ یعنی همهٔ مهارت‌های هوشمندانهٔ انسانی را پیاده‌سازی کند یا خیر؟ به بیان دیگر و با استفاده از تعابیر *سرل*، آیا هوش مصنوعی ضعیف پیروز خواهد شد؟ بدین ترتیب، بر خلاف باور عمومی که هوش مصنوعی ضعیف را فاقد اهمیت فلسفی می‌داند، در این پژوهش نشان داده شد که بسیاری از مهم‌ترین گفت‌وگوهای فلسفی در هوش مصنوعی، ازجمله بیشتر نقدهای *دریفوس* و دفاع‌های *دنت*، دربارهٔ هوش مصنوعی ضعیف و در تراز فلسفه‌علمی بوده است؛ اما پرسش اصلی در تراز فلسفه‌ذهنی چنین بود: اگر ماشینی ساخته شود که همهٔ کارکردهای هوشمندانه را پیاده‌سازی کند و در آزمون

تورینگ پیروز شود، آیا بدین معناست که ذهن دارد؟ به بیان دیگر آیا هوش مصنوعی الگوی مناسبی برای ذهن فراهم می‌آورد؟

با داشتن این صورت‌بندی کلی، فرد می‌تواند ترازهای گوناگون بحث فلسفی در هوش مصنوعی را از هم تفکیک کند و لابه‌لای گفت‌وگوهای پیچیدهٔ فیلسوفان در این حوزه سردرگم نخواهد شد. اما در ادامهٔ مقاله بر پایهٔ این صورت‌بندی مهم‌ترین مسئله‌های هر یک از دو تراز فلسفه‌علمی و فلسفه‌ذهنی را طرح و ارزیابی کردیم. در تراز فلسفه‌علمی دیدیم که *دریفوس*، پس از آنکه در طی پنجاه سال، صورت‌بندی‌های گوناگونی از چالش اصلی هوش مصنوعی ارائه می‌کند، سرانجام صورت‌بندی *دنت* ذیل مسئلهٔ چارچوب را مهم‌ترین معیار برای موفقیت هوش مصنوعی در آزمون *تورینگ* می‌پذیرد. همچنین مشاهده کردیم *دنت* که خود همواره بزرگ‌ترین مدافع هوش مصنوعی بوده است، این نقد را می‌پذیرد که تابه‌حال هیچ راه‌حل روشنی برای مسئلهٔ چارچوب ارائه نشده است. این بهترین گواه بر این مدعاست که هوش مصنوعی ضعیف با سدی بلند به نام مسئلهٔ چارچوب در راه موفقیت روبه‌روست. از سوی دیگر دیدیم *دنت* کاملاً در اشتباه است که فکر می‌کند پیش از هوش مصنوعی مسئلهٔ چارچوب در فلسفه صورت‌بندی نشده است و مسئلهٔ چارچوب به عنوان مسئلهٔ چگونگی روبه‌رویی با امر غیرمنتظره و پیش‌بینی‌ناشده در فلسفه‌های پدیدارشناختی با تعابیر گوناگون صورت‌بندی شده است.

در تراز فلسفه‌ذهنی نیز سه مسئلهٔ اصلی را بررسی کرده‌ایم. در مسئلهٔ نخست- یعنی مغالطهٔ آدمک به کمک نقد *سرل*- نشان دادیم که ادعای *دنت* مبنی بر اینکه هوش مصنوعی با تجزیهٔ بازگشتی، مغالطهٔ آدمک- یا به تعبیر *دنت*، مسئلهٔ *هیوم*- را حل کرده است، از اساس اشتباه است. *دنت* در مغالطه‌ای آشکار از تراز معنادار ذهنی به تراز فاقد معنای فیزیکی می‌پرد. پس هوش مصنوعی با ارتکاب مغالطهٔ آدمک امر ذهنی را به چیزی در نگاه بیننده و درنتیجه نسبی و دلخواهانه تبدیل کرده است. اما دربارهٔ دو مسئلهٔ دیگر- یعنی استدلال اتاق چینی و نقد بر پایهٔ فرضیهٔ *گودل*- *چالمرز* دفاع نسبتاً مناسبی از شبکه‌های عصبی پیوندگرا در هوش مصنوعی ارائه می‌دهد. هستهٔ اصلی استدلال اتاق

چینی آن است که سینتکس برای سمانتیک کافی نیست؛ ولی *چالمرز* نشان داد که این گزارهٔ زمانی درست است که امر سینتتیک و امر سمانتیک در یک تراز باشند؛ مانند جمله‌های زبانی که در آنها سینتکس زبانی- یعنی دستور زبان- برای تولید معنا کافی نیست؛ اما اگر در یک تراز نباشند، این گزاره درست نیست؛ برای نمونه رفتار مغز در تراز فیزیکی از سینتکس فیزیکی پیروی می‌کند و در ترازی دیگر، همین مغز معنا را تولید می‌کند. با این حساب شبکه‌های عصبی پیوندگرا که در آنها تراز سینتتیک متفاوت با تراز سمانتیک است، از نقد *سرل* مصون خواهند بود.

در مسئلهٔ سوم نیز *چالمرز* نشان می‌دهد شبکه‌های عصبی پیوندگرا از گزند نقد بر پایهٔ فرضیهٔ *گودل* در امان هستند؛ زیرا این شبکه‌ها از اساس یک سیستم فورمال سازوار با الگوریتم‌های منطقی- محاسباتی نیستند که تئوری *گودل* دربارهٔ آنها صدق کند. پس این دو نقد اخیر- یعنی استدلال اتاق چینی و نقد بر پایهٔ تئوری *گودل*- بر شبکه‌های عصبی پیوندگرا وارد نیست؛ اما اولاً نقد نخست، یعنی مغالطهٔ آدمک- و هر نوع صورت‌بندی دیگری از مسئلهٔ امر سوبژکتیو- بر جریان پیوندگرا نیز وارد است؛ ثانیاً نکتهٔ مهم آن است که این دو نقد در صدد بوده‌اند که نشان دهند هوش مصنوعی به عنوان یک تئوری برای ذهن، غلط است و از اینکه این دو نقد وارد نیست، هرگز نمی‌توان نتیجه گرفت که هوش مصنوعی الگوی درستی برای ذهن است. داوری در این مورد به حل مسئله‌های کلان فلسفهٔ ذهن منوط است و ازاین‌رو هوش مصنوعی در همهٔ اشکالش راه درازی تا موفقیت در تراز فلسفهٔ ذهن دارد.

منابع و مآخذ

۱. محمدعلی خلج، محمدحسین و محمدتقی طباطبائی؛ «از کارکردگرایی تا بسترگرایی: بازسازی گفت‌وگوی دریفوس و پاتنم»؛ **مجلهٔ فلسفه**، ش۲، پاییز و زمستان ۱۳۹۲.

۲. محمدعلی خلج، محمدحسین؛ «دریفوس و تاریخ فلسفی هوش مصنوعی»؛ **مجلهٔ غرب‌شناسی بنیادی**، ش۱، بهار و تابستان ۱۳۹۳.

3. Chalmers, D; "Subsymbolic Computation and the Chinese Room The Symbolic and Connectionist Paradigms: Closing the Gap"; **Lawrence Erlbaum**, J Dinsmore Addressing from this site: http://consc.net/papers/subsymbolic.pdf, 1992.

4. _____; "Minds, Machines, and Mathematics. Published in a PSYCHE symposium on Roger Penrose's book Shadows of the Mind in 1995"; http://consc.net/papers/penrose.html, 1995.

5. _____; A Computational Foundation for the Study of Cognition; http://moodle.tau.ac.il/mod/resource/view.php?id=341764&redirect=1, 2011.

6. Clark A ;"Philosophical Foundations" in **Artificial intelligence,** edit. Margaret Boden, Academic Press, Ing, 1996.

7. Dennett, D.C; **Content and Consciousness**; Routledge & Kegan Paul, 1969.

8. _____; "Cognitive Wheels: The Frame Problem of AI"; **Minds, Machines and Evolution;** Cambridge: Cambridge University Press, 1984, pp.129–151.

9. _____; "Artificial intelligence as Philosophy and Psychology"; in **Brainchildren Essays on Designing Minds;** Cambridge, Mass.: MIT Press, 1981.

10. _____; "When philosophers encounter Artificial intelligence"; **Brainchildren Essays on Designing Minds;**

Cambridge, Mass.: MIT Press, 1981.

11. Dreyfus, Hubert L; "Phenomenology and Mechanism"; **Noûs**; Vol.5, Feb., 1971, pp.9-81.
12. _____; **What Computers Can't Do: A Critique of Artificial Reason**; Harper and Row, 1979.
13. _____; **Being-in-the-world: A Commentary on Heidegger's Being and Time**; División I, MIT Press, 1991.
14. _____; **What Computers Still Can't Do**; New York: MIT Press, 1992.
15. _____; "Why Heideggerian AI Failed and how Fixing it would Require making it more Heideggerian"; Addressing from this site: http://leidlmair.at/doc/WhyHeideggerianAIFailed.pdf,2007.
16. Dreyfus, H. and Dennett; **Mechanical Bodies**; Computational Minds A Bradford Book, 1997.
17. Dreyfus, H and Stuart Dreyfus; **Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer**; New York: Free Press, 1986.
18. McCarthy, J; "What has AI in common with philosophy"; Addressing from this site: http://www-formal.stanford.edu/jmc/, 1995.
19. Nagel and Newman; "Communication"; **The Journal of Philosophy**, Vol.58, N.8, 1961.
20. Penrose, Roger; **Beyond the Doubting of a Shadow: A Reply to Commentaries on Shadows of the Mind**, Cambridge: Cambridge University Press, 1996.
21. Putnam, H; "Minds and machines; Dimensions of Mind, ed. Sidney Hook"; New York: Published in Mind, Language and Reality; **Philosophical Papers**, Vol.2, Cambridge University Press 1975.
22. Russell, S and p. Norvig; "Modern Approach"; **Pearson Education International**, Second edition, 2003.
23. Searle, J.R; "Minds, brains, and programs"; **Behavioral and Brain Sciences**, Vol.3, 1980, pp.417-457.

24. _____; **Rediscovery of the mind**; Cambridge, Mass.: MIT Press, 1992.
25. _____; "Is the brain a digital computer?"; **Proceeding and addresses of the American philosophical association**, Vol.64, N.3, 1990.
26. _____; **The Mystery Of Consciousness**; New York: The New York Review of Books, 1997.
27. Turing, A.M. "Computing machinery and intelligence"; **Mind**, Vol.59, 1950, pp.433-460.