غرب‌شناسی بنیادی، پژوهشگاه علوم انسانی و مطالعات فرهنگی
سال پنجم، شمارهٔ اول، بهار و تابستان ۱۳۹۳، صص ۱۰۳ – ۱۲۸

دریفوس و تاریخ فلسفی هوش مصنوعی

محمدحسین محمدعلی خلج*

چکیده

این مقاله برای ارائهٔ روایتی تاریخی از تحولاتِ هوش مصنوعی در قرن ۲۰ و ۲۱ از نظرگاهی فلسفی، سه دوره را از هم جدا کرده است: ۱. تسلطِ پارادایم نمادی؛ ۲. بازگشت به پیوندگرایی؛ ۳. ظهور حیات مصنوعی. مدعای اصلی مقاله آن است که هم بازگشت به پیوندگرایی و هم ظهور حیات مصنوعی، نقدهای دریفوس بر رویکردهای پیشین در هوش مصنوعی را تأیید می‌کند. به باور دریفوس یک نگرشِ اتمیستی ـ فورمالیستی به ذهن، دانش، و هستی، پشتوانهٔ اصلیِ پارادایم نمادی است. شکستِ هوش مصنوعی نمادی در صورت‌سازیِ فهمِ هرروزه و چرخش به سوی نگرشی کل‌گرایانه در پیوندگرایی مؤیدِ نقدهای دریفوس است. همچنین حیات مصنوعی با بدن‌مندانه و موقعیت‌مند دانستنِ شناخت موضع وی را تأیید می‌کند. با این وجود، حمایت واپسین دریفوس از هوش مصنوعی هیدگری و این ادعا که این جریان، هستی‌شناسیِ هیدگری را در آزمونی تجربی قرار می‌دهد قابل نقد است. همچنین همراه با ارزیابی نقد دریفوس نشان داده می‌شود که وی از نقد فلسفی هوش مصنوعی به نقدِ کلِ فرهنگ و اندیشهٔ غربی پل می‌زند و نشان می‌دهد که برای اصالت‌دادن به فهمِ تئوریک و فراموشیِ جهان انسانی، خطرِ پیشِ روی این فرهنگ ظهور انسان‌هایی با فهمِ ماشینی و نه ماشین‌هایی با توانایی‌های ورای انسانی است.

کلیدواژه‌ها: هوش مصنوعی، پارادایم نمادی، پیوندگرایی، حیات مصنوعی، هوش مصنوعی هیدگری.

---

* دانشجوی کارشناسی ارشد فلسفه، دانشگاه تهران mhmakh@gmail.com
تاریخ دریافت: ۱۳۹۲/۹/۱۰، تاریخ پذیرش: ۱۳۹۳/۳/۸

۱. مقدمه

هوش مصنوعی (artificial intelligence) حوزهٔ وسیعی از دانش را دربر می‌گیرد که در نقطهٔ تلاقی میان چند دانش بزرگ دیگر ازجمله علوم رایانه‌ای، الکترونیک، روان‌شناسی، زیست‌شناسی، زبان‌شناسی، منطق و فلسفه قرار گرفته است. می‌توان بیش از بیست زیرشاخه برای هوش مصنوعی برشمرد که یادگیری ماشین (machine learning)، شناخت الگو (pattern recognition)، پردازش زبان طبیعی (natural language processing)، روباتیک (robotic)، سامانه‌های خبره (expert system)، شبکه‌های عصبی (neural network)، منطق فازی (faze logic) و الگوریتم ژنتیک (genetic algorithm) از جملهٔ مهم‌ترین آن‌ها هستند. برای آشنایی مقدماتی با این دانش می‌توان سراغ تعاریف ارائه‌شده از آن رفت. استوآرت راسل تعاریف هوش مصنوعی را در چهار دسته جای می‌دهد: ۱. پژوهش در سامانه‌هایی که شبیه انسان می‌اندیشند؛ ۲. پژوهش در سامانه‌هایی که به گونه‌ای عقلانی می‌اندیشند؛ ۳. پژوهش در سامانه‌هایی که شبیه انسان رفتار می‌کنند؛ ٤. پژوهش در سامانه‌هایی که به گونه‌ای عقلانی رفتار می‌کنند (Russell and Norvig, 2003: 2). تعریف اول و دوم بر اندیشیدن تمرکز می‌کنند، اما تأکید تعریف سوم و چهارم بر رفتار است. از سوی دیگر تعریف اول و سوم شباهت به انسان را معیار هوشمندی قرار داده‌اند. اما تعریف دوم و چهارم عقلانی‌بودن را به‌گونه‌ای کلی معیار دانسته‌اند. هر یک از این تعاریف از نظرگاهی ویژه و از سوی یک جریان در هوش مصنوعی در مقطعی خاص از تاریخ آن ارائه شده است. اگر بخواهیم در نقدی جامع و از منظری فلسفی رشتهٔ هوش مصنوعی را ارزیابی کنیم نمی‌توانیم به هیچ‌یک از تعاریف ویژه بسنده کنیم بلکه باید به سراغ تاریخ تحولات این رشته برویم؛ تاریخی که این تعاریف مبتنی بر نگرش‌های گوناگون در هوش مصنوعی از آن سربرآوردند.

هوش مصنوعی در دههٔ چهلم میلادیِ قرن بیستم، با پژوهش‌های آلن تورینگ و کلود شانون متولد شد. از آن زمان تا به امروز این دانشِ جدید، دوره‌های گوناگونی را گذرانده است و در طیِ این مدت همواره مورد توجه فیلسوفان قرار داشته است. چرخش از نظرگاه هر دوره به دورهٔ بعد گرچه در درجهٔ نخست ریشه در کامیابی یا ناکامی پروژه‌های ساخت هوش مصنوعی داشته‌اند، اما از دلایل و دلالت‌های فلسفی ویژهٔ خود نیز برخوردار بوده‌اند؛ به بیان دیگر تاریخ هوش مصنوعی دارای جنبه‌ای فلسفی نیز است. مقالهٔ حاضر درصدد است این دلایل و دلالت‌های فلسفی را بازگو کند و منطقِ فلسفیِ حاکم بر چرخش از هر

دوره به دورهٔ بعد را آشکار سازد. مدعای اصلی این مقاله آن است که اگرچه نگرش واپسین دریفوس در هوش مصنوعی قابل نقد است، اما به طور کلی تحولات هوش مصنوعی تا حدود زیادی نقدهای دریفوس بر این رشته در دوره‌های گوناگون را تأیید می‌کند. در ادامه می‌کوشیم با روایت نقادانهٔ تاریخ تحولات هوش مصنوعی این مدعا را به اثبات برسانیم. اما پیش از آن نیاز است تا رویکرد کلی و روش دریفوس در فلسفهٔ ذهن به اختصار بیان شود تا بتوان ارتباط درستی با نقدهای وی برقرار کرد.

دریفوس در درجهٔ نخست مفسر اندیشه‌های هیدگر و مرلوپونتی است و دیدگاه کلی‌اش در فلسفه نیز متأثر از فلسفهٔ پدیدارشناسانهٔ ایشان شکل گرفته است. به باور وی «پدیدارشناس با وابستگی بستری آغاز می‌کند، همان چیزی که هیدگر و مرلوپونتی «بودن در جهان» می‌نامند. هیدگر در *هستی و زمان* توصیفی از این جهان که انسان در آن مسکن گزیده است ارائه می‌دهد. ... مطابق با نظر هیدگر اگر ما فرض کنیم که نخست واقعیت‌های ایزوله را تجربه می‌کنیم و سپس به آن واقعیت‌ها معنا می‌بخشیم آن‌گاه با مشکلات فلسفی مواجه خواهیم شد» (Dreyfus, 1971: 95). با این حساب می‌توان رویکرد کلی دریفوس به معنا را رویکردی عمل‌گرایانه، کل‌گرایانه، و بسترگرایانه دانست. یعنی تجربهٔ عملیِ ما از امر معنادار در بستر جهان هرروزه به مثابهٔ یک کل مقدم بر فهم صوریِ ما از واقعیت‌های اتمیک و ایزوله است. در ادامه، نگرش دریفوس به صورت کامل‌تری بیان می‌شود اما از همین شرح کوتاه هم پیداست که رویکرد وی در فلسفهٔ ذهن در برابر هر رویکرد بازنمودگرایانه، اتمیستی، و فورمالیستی قرار می‌گیرد. البته این نکته بسیار حائز اهمیت است که تفسیر دریفوس از فلسفهٔ پدیدارشناسانه به‌درستی درک شود؛ از نظر دریفوس شرح پدیدارشناسانه «نمی‌تواند در معنای سنتی شرحی علمی باشد» (ibid). بلکه پدیدارشناس فقط به توصیف جهان هرروزه می‌پردازد و امکان اثبات منطقیِ آموزه‌هایش منتفی است. دریفوس در تفسیرش بر هیدگر نیز به این امر اذعان دارد و بر آن است که «هیدگر انتظار ندارد که این مواضعش را اثبات کند» (Dreyfus, 1991: 60).[۱]

تنها با عطف نظر به این نگرش روش‌شناختی می‌توان به این پرسش پاسخ داد که چرا دریفوس علاقه‌مند است در اکثر موارد فقط نقدهایش را بر هوش مصنوعی بیان کند و کم‌تر به اثبات موضع ایجابی‌اش مبادرت می‌ورزد. در اصل مبنای روش‌شناختیِ او مستلزم به‌کارگیریِ چنین شیوه‌ای است. دریفوس معترف است که به صورت پیشینی نمی‌توان نگرش کل‌گرایانه ـ بسترگرایانهٔ پدیدارشناسان و همچنین شکست هوش مصنوعی در ارائهٔ

یک شرح اتمیستی و فرمال از جهان معنادار را اثبات کرد. از این رو «بار استدلال بـر دوش کسی است که انتظار دارد که این امر [یعنی ارائهٔ یک شرح اتمیک و فرمال از امـر معنـادار] با موفقیت انجام شـود» (ibid: 117) و پدیدارشـناس (در تفسـیر دریفوسـی از آن) تـرجیح می‌دهد فقط مشکلات ارائهٔ این شرح و دلایل ناکامی در ارائهٔ آن را برشمرد. با این حساب پدیدارشناسی «فقط نظرگاه واحدی را فراهم می‌آورد که از آن نظرگاه این مشـکلات ... بـه مثابهٔ امور گریزناپذیر دیده می‌شـود» (Dreyfus, 1971: 95). همـین نگـرش روش‌شـناختی سبب شده است که دریفوس شکست هوش مصنوع را دلیلی بـر حقانیـت هیـدگر بدانـد؛ چراکه گرچه هیدگر نمی‌تواند توصیف پدیدارشناسانه‌اش را به اثبات برساند، اما مـی‌توانـد ضرورت این شکست را تبیین کند (Dreyfus, 1991: 119). پس با این حساب نباید رویکرد نقادانه و سلبیِ دریفوس در هوش مصنوعی را تصادفی دانست؛ بلکه این شیوه ضـرورتاً از مبانی روش‌شناختی و تفسیرِ ویژهٔ وی از پدیدارشناسی برمی‌آید.

۲. تاریخ تحولات هوش مصنوعی و نقدهای دریفوس

پس از آشنایی مقدماتی با نگرش کلی و روش دریفوس در فلسفهٔ ذهن اینک می‌توان سراغ تاریخ تحولات هوش مصنوعی و نقدهای دریفوس در دوره‌هـای گونـاگون بـر آن رفـت. اگرچه فقط شش دهه از تاریخ هوش مصنوعی گذشته است و هنوز مـرز میـان دوره‌هـای گوناگون آن کاملاً شفاف نیست اما در کلی‌ترین نگاه می‌توان دستِ کم سـه پـارادایم را در هوش مصنوعی از هم جدا کرد (Clark, 1996: 3) و تاریخ هوش مصنوعی را به سـه دورهٔ زیر تقسیم کرد:

۱. دورهٔ تسلط هوش مصنوعی نمادی (symbolic AI)؛ ۲. دورهٔ بازیابی هوش مصنوعی پیوندگرا (connectionist AI)؛ ۳. دورهٔ ظهور حیات مصنوعی (artificial life).

متناظر با این سه دوره می‌توان ورودهـای نقادانـهٔ دریفـوس بـه حـوزهٔ فلسـفهٔ هـوش مصنوعی را نیز به سه دسته تقسیم کرد.[۲] در ادامه پس معرفی هر دوره، نقدِ دریفوس بر هـر یک را مورد ارزیابی قرار می‌دهیم.

۱.۲ دورهٔ تسلط هوش مصنوعی نمادی

آیا ماشین می‌تواند بیندیشد؟ این پرسشی است که مقالهٔ «ماشین محاسـبه‌ای و هوشـمندی»

الن تورینگ با آن آغاز می‌شود (Turing, 1950). می‌توان این مقاله و ایـن پرسـش را نقطـهٔ آغازینِ فلسفهٔ هوش مصنوعی دانست. همهٔ ایده‌های پرداخته‌شده در فلسفهٔ هوش مصنوعی در پاسخ به این پرسش و یا بازسازی آن بـوده اسـت. در پاسـخ بـه ایـن پرسـش در دورهٔ نخستِ هوش مصنوعی، سامانه‌های نمادی، الگوی ماشین‌های اندیشنده به‌شـمار رفتنـد. از این رو نامِ هوش مصنوعیِ نمادی برای پارادایمِ حاکم بر آن برگزیده شده است. در تـاریخِ هوش مصنوعی، سه رویکرد از این پارادایم، دفاعی فلسفی انجام داده‌اند که عبـارت‌انـد از: دفاع محاسبه‌گرایانـه (computationalistic)؛ دفـاع کارکردگرایانـه (functionalistic)؛ دفـاع بازنمودگرایانه (representationalistic). در ادامه پخته‌ترین صورت‌بندی از هر یـک از ایـن دفاع‌ها را بیان خواهیم کرد.

الف) دفاع محاسبه‌گرایانه

نخستین صورت‌بندی از محاسبه‌گرایی به دستِ آلن تورینگ و با معرفی آزمون تورینگ و ماشین تورینگ ارائه شده است. اما در طول تاریخ هوش مصنوعی ایـن نگـرش شـاخ و برگ‌های زیادی یافت. دنت فصل مشترک همهٔ رویکردهای محاسبه‌گرایانه را در سه گزاره خلاصه می‌کند: «۱. تفکر، پردازش‌گری اطلاعـات اسـت. ... ۲. پـردازش‌گـری اطلاعـات، محاسبه است. ... ۳. معانی این نمادهـا [ی محاسـباتی] تفکـر را بـه جهـان خـارج ارتبـاط می‌دهد» (Dennett, 1998: 217) اگر این سه گزاره صادق باشد یک رایانـه بـه عنـوان یـک پردازش‌گر نمادی اطلاعات، یک سامانهٔ هوشمند و متفکر به‌شمار می‌آید.

ب) دفاع کارکردگرایانه

نخستین بار پاتنم (Putnam, 1960) با ایجاد قیاسی میان حالت‌های ذهنیِ انسان و حالت‌های محاسباتیِ ماشین تورینگ و حالت‌های فیزیکیِ انسان و حالت‌های ساختاری ماشین تورینگ، نگرشِ کارکردگرایانه به ذهن را تأسیس کرد (Dennett, 1969: 101).[۳] بر پایهٔ این قیاس حالت‌های ذهنی می‌توانند با حالت‌های محاسباتی ـ منطقیِ جدول ماشین تورینگ یا به بیان دیگر با توصیف نرم‌افزاری از رفتار رایانه یکی دانسته شوند (Block, 2007: 35-36). بر این اساس حالت‌های ذهنی (هم‌چون حالت‌های منطقیِ ماشین تورینگ)، کارکردهایی محاسباتی هستند که می‌توانند در سخت‌افزارهای گوناگونی ازجمله انسان یا رایانه پیاده شوند.[۴] از این رو باید رایانه را هم‌چون انسان هوشمند (ذهنمند) به‌شمار آورد.

ج) دفاع بازنمودگرایانه

جان مک‌کارتی به عنوان یکی از پدران هوش مصنوعی پس از نقد صورت‌بندی‌های محاسبه‌گرایانه و کارکردگرایانه از پارادایم نمادی، دفاعی بازنمودگرایانه از آن به‌دست می‌دهد (McCarthy and Hayes, 1969: 3). پخته‌ترین صورت‌بندی بازنمودگرایانه از پارادایم نمادی را دو تن از دیگر مهندسانِ پیش‌رو در هوش مصنوعی یعنی نیول و سایمون ارائه کرده‌اند. ایشان نظریهٔ کلی‌شان را در مورد هوشمندی چنین صورت‌بندی می‌کنند:

فرضیهٔ سامانه‌های نمادهای فیزیکی؛ یک سامانهٔ نمادهای فیزیکی ابزارهای لازم و کافی برای پیاده‌سازی هوشمندی عمومی را داراست (Newell and Simon, 1976: 116).

بر پایهٔ این فرض هر سامانهٔ هوشمند در درجهٔ نخست سامانه‌ای فیزیکی و در درجهٔ دوم یک سامانهٔ نمادی است. «یک سامانهٔ نمادی ماشینی است که در طی زمان ساختارهای نمادی تولید می‌کنند» (ibid). پس در جهان فیزیکی توانایی برای تولید بازنمودهایی نمادی، شرط لازم و کافی برای هوشمندی (ذهنمندی) است. بدین‌ترتیب هر سامانهٔ هوشمندی هم‌چون انسان یک سامانهٔ نمادهای فیزیکی است و هر سامانهٔ نمادهای فیزیکی هم‌چون رایانهٔ دیجیتال یک سامانهٔ هوشمند است.

به پشتوانهٔ سه دفاعِ روایت‌شده بر بسترِ محاسبه‌گرایی، کارکردگرایی، و بازنمودگرایی، هوش مصنوعیِ نمادی از دههٔ پنجاه به مدت سه دهه بر حوزه‌های پژوهشیِ هوش مصنوعی مسلط شد. بی‌شک یکی از علل افولِ پارادایمِ نمادی به چالش کشیده‌شدنِ پشتوانه‌های فلسفیِ آن بود. دریفوس که از اواسطِ دههٔ شصت نقدِ پارادایم نمادی را آغاز کرده بود پرچم‌دارِ این جریان در میان فیلسوفان به‌شمار می‌آمده است. در ادامه به نقشِ وی در نقد پارادایم نمادی خواهیم پرداخت.

۱.۱.۲ نقد هوش مصنوعی نمادی به دست دریفوس

نقدِ دریفوس بر پارادایم نمادی در درجهٔ نخست خوش‌باوریِ خام و پیش‌بینی‌های بلندپروازانهٔ مهندسان و پژوهش‌گران هوش مصنوعی را نشانه می‌رفت (Dreyfus, 1979: xxix, xxx)؛ دریفوس سه پیش‌فرضِ فلسفیِ توجیه‌ناشده را پشتوانهٔ این خوش‌باوری‌ها و به صورتِ کلی هم‌سان‌دانستنِ انسان و سامانه‌ای نمادی در پارادایم نمادی می‌داند؛[۵]

این فرض که کارکردهای انسانی شبیه یک وسیلهٔ نمادپرداز همه‌منظوره است مبتنی است بر ... [۱] یک فرض روان‌شناسانه که ذهن می‌تواند هم‌چون یک دستگاه نگریسته شود

که با بیت‌هایی از اطلاعات مبتنی بر یک سری قواعد صوری عمل می‌کند. ... [۲] یک فرض معرفت‌شناسانه که همهٔ دانش می‌تواند صوری‌سازی شود. ... [۳] درنهایت ... این فرض هستی‌شناختی است که هر آن‌چه هست مجموعه‌ای از واقعیت‌هایی است که هر یک منطقاً مستقل از یک‌دیگر هستند (ibid: 68).

این سه فرضِ فلسفی فرض‌هایی جدای از هم نیستند بلکه فرض معرفت‌شناختی پیش‌فرضِ فرضِ روان‌شناختی و فرضِ هستی‌شناختی پیش‌فرضِ فرضِ معرفت‌شناختی است؛ به بیان دیگر از نظرگاهِ نقادانهٔ دریفوس چون جهانِ هرروزهٔ ما ورای عالمِ واقعیت‌های منطقاً مستقل است و حقیقت قابل فروکاستن به تئوری نیست، نمی‌توان همهٔ دانش را صوری کرد و چون نمی‌توان همهٔ دانش را صوری کرد الگوی رایانه‌ای برای ذهن مناسب نیست. از این رو استدلال‌هایی که در نقدِ این سه فرض بیان می‌شوند نیز جدای از هم نخواهند بود. در ادامه هر یک از این سه فرض را از نظرگاه دریفوس ارزیابی خواهیم کرد.[۷]

نقدِ فرض روان‌شناسانه: این فرض بازگوکنندهٔ الگوی محاسبه‌گرایانه ـ کارکردگرایانهٔ ذهن است که بر مبنای آن می‌توان ذهن را نوعی رایانهٔ دیجیتال دانست. این الگو فرض می‌گیرد که چون هم ذهن و رایانه، اطلاعات را پردازش می‌کنند پس رایانه الگوی مناسبی برای فهمِ ذهن است. اما به باور دریفوس پشتوانهٔ این فرض، ابهام در معنای پردازش‌گریِ اطلاعات است. معنای پردازش‌گری اطلاعات در انسان و رایانه یکی نیست (ibid: 76-77). به طور کلی در آثار دریفوس دو نقد بر فرضِ روان‌شناسانه دیده می‌شود: ۱. پردازش‌گری رایانه با امر معنادار سروکار ندارد (ibid: 77)؛ ۲. پردازش‌گری ذهن با برنامهٔ رایانه‌ای سروکار ندارد (ibid: 78).

در دفاع از گزارهٔ نخست دریفوس به تئوری کلود شانون در سایبرنتیک (cybernetic) به عنوان پایهٔ پردازش‌گری رایانه‌ای اشاره می‌کند. این تئوری به‌روشنی معنا را در معنای معمولی‌اش نادیده می‌گیرد (ibid: 77). تعریفِ پردازش‌گری اطلاعات در رایانه به مثابهٔ تابعی ریاضیاتی هیچ نیازی به فهم و ادراکِ معنادار انسانی ندارد. اما هر تئوری در فلسفهٔ ذهن یا روان‌شناسی فلسفی به‌هیچ‌وجه نمی‌تواند فهم و ادراکِ معنادار انسانی را نادیده بگیرد. متغیرهای ریاضیاتیِ به‌کاررفته در یک عملیاتِ محاسباتی به خودیِ خود هیچ محتوای درونی و معناداری ندارند. آن‌چه اسبابِ خلطِ پردازش‌گریِ معنادار انسانی و پردازش‌گریِ بی‌معنای رایانه‌ای را فراهم می‌آورد نادیده‌گرفتنِ نقشِ برنامه‌نویس است؛

این دقیقاً نقش برنامه‌نویس است که گذر از جملاتی را که معنادار هستند (شامل اطلاعات در معنای معمولی) به رشته‌هایی از بیت‌های گسستهٔ بی‌معنا (اطلاعات در معنای فنی) که رایانه‌ها با آن‌ها کار می‌کنند ممکن می‌سازد (ibid: 78).

پس پردازش‌گری رایانه‌ای با متغیرهایی تهی و فاقدِ محتوای معنادار سروکار داشته است و این برنامه‌نویس است که به آن‌ها معنا می‌بخشد.

اما در دفاع از گزارهٔ دوم، یعنی فاقدِ برنامه‌بودنِ پردازش‌گری ذهن، می‌توان بهره‌ای را که دریفوس از آموزه‌های روان‌شناسان گشتالت‌گرا (gestaltist psychologist) و پدیدارشناسان می‌برد بازگو کرد؛ چنان‌که ایشان می‌گویند تجربهٔ ادراکی انسان دارای ساختار شکل ـ زمینه (figure-background) است. در هر تجربهٔ ادراکی یک شکل برجسته شده و باقیِ گسترهٔ آگاهی در زمینه یا به تعبیر پدیدارشناسان در افق بیرونی (outer horizon) جای می‌گیرد (ibid:152). آن‌چه در زمینه جای می‌گیرد مبهم و نامتعین است و بی‌شک آگاهیِ پردازش‌گرانه از آن نمی‌تواند از جنسِ پردازش‌گری رایانه‌ای باشد؛ چراکه پردازش‌گری رایانه‌ای صرفاً با اطلاعاتی متعین و روشن سروکار دارد. پس ادراکِ معنادار انسانی برخلافِ پردازش‌گری رایانه‌ای دارای زمینه و بسترمند است. افزون بر این پردازش‌گری معنادار در انسان از کل به جزء است. برای نمونه وقتی من از روبه‌رو به خانه‌ای می‌نگرم پشتِ آن را نمی‌بینم، اما با آن به‌گونه‌ای مواجه می‌شوم که گویی پشت هم دارد؛ یعنی آن را به مثابهٔ چیزی بیش از شکلِ ادراکیِ حاضر درک می‌کنم. پدیدارشناسان به این چیزی بیش از شکلِ ادراکی، افقِ درونی (inner horizon) می‌گویند؛

ما ابتدا با یک ابژه به صورت یک کل مواجه می‌شویم و اگر بخواهیم که آن را بهتر بشناسیم پشت و جزئیات آن را نیز ملاحظه می‌کنیم. اما یک ماشین [ازجمله هر رایانه] که چیزی معادل افق درونی ندارد این پردازش اطلاعات را به صورت برعکس انجام می‌دهد: یعنی از جزء به کل (ibid: 241).

یعنی ما نخست درکی از کلِ نامتعین پیدا می‌کنیم و سپس اجزای آن را در پرتوی آن کل خواهیم شناخت. پس چنان‌که در آغاز مقاله نیز اشاره شد از نظر دریفوس ادراکِ معنادار انسان افزون بر بسترمندبودن (contextual)، کل‌گرایانه (holistic) نیز است. اما امکانِ چنین فهمی در برنامهٔ رایانه‌ای فراهم نیست؛ چراکه چنین برنامه‌ای صرفاً با داده‌های متعین و روشن سروکار دارد.[۶]

نقدِ فرض معرفت‌شناسانه: حتی اگر توصیفِ ما از عملکردِ ذهن مغایر با پردازشِ

رایانه‌ای باشد، همچنان می‌توان از امکانِ هوشمندی رایانه‌ها دفاع کرد؛ چراکه فارغ از این‌که انسان‌ها به چه شیوه‌ای رفتار هوشمندانه انجام می‌دهند اگر بتوان همهٔ دانشِ انسانی را به‌گونه‌ای متعین و روشن صوری کرد می‌توان آن‌ها در رایانه نیز پیاده‌سازی کرد (ibid: 101). بنابراین فرضِ معرفت‌شناسانه این خواهد بود که همهٔ دانشِ انسانی می‌تواند صوری شود. به باور دریفوس این نگاه به دانش که وجود یک تئوری برای هر حوزه از حقیقت را پیش‌فرض می‌گیرد در عمقِ فرهنگِ غربی ریشه دارد (ibid: 103). پشتوانهٔ اصلیِ این فرض، موفقیت‌های دانش‌های کاملاً صوری در جهانِ جدید به‌ویژه فیزیک و زبان‌شناسی بوده است (ibid: 104).

چنان‌که دریفوس نشان می‌دهد صوری‌سازی در فیزیک هیچ ارتباطی با هوشمندی و جهانِ معنادار انسانی ندارد.[۸] اما زبان‌شناسی کاملاً به همین جهان مربوط است و از این جهت باید موردِ ارزیابی قرار گیرد (ibid: 106-108).[۱۲] دریفوس می‌پذیرد که زبان‌شناسی گشتاری (trancformental linguistic) توانسته است یک صورت‌بندی از فهمِ زبانی به‌دست دهد؛

چامسکی و زبان‌شناسان گشتاری کشف کرده‌اند که با انتزاع از اجرای (performance) انسانی، یعنی به‌کارگیری جملات جزئی در موقعیت‌های جزئی، آن‌ها می‌توانند آن‌چه را که باقی مانده است یعنی توانایی شناختِ دستور زبانیِ جملاتِ خوش‌ساخت و کنارگذاشتنِ مواردِ بدساخت، صوری‌سازی کنند؛ یعنی آن‌ها می‌توانند یک تئوری برای توانایی (competence) زبانی فراهم آورند (ibid: 110).

اما این تئوری به خودیِ خود برای اجرای رفتارِ زبانیِ هوشمندانه کافی نیست. فهمِ چگونگیِ به‌کارگیریِ جملاتِ جزئی در موقعیت‌های جزئی به دانشی جداگانه نیاز دارد. دریفوس در امکانِ صوری‌شدنِ این دانش اجرایی تشکیک می‌کند. دستِ کم موفقیتِ صوری‌سازی در زبان‌شناسی، دلیلی بر امکانِ صوری‌شدنِ دانش به‌کارگیری زبانی نیست و برای آن باید تئوری جداگانه‌ای یافت (ibid). از سوی دیگر دریفوس نشانه‌هایی را برمی‌شمرد که شاید این دانش مهارتی (know-how) از اساس قابلِ صورت‌بندی در دانشی گزاره‌ای (know-that) نباشد و به طور کلی نتوان برای آن یک تئوری یافت. برای نمونه به‌نظر می‌رسد که به‌کارگیریِ استعاری واژگان تن به صوری‌سازی نمی‌دهد (ibid: 111). نمونهٔ دیگری که در نوشته‌های دریفوس به چشم می‌خورد تعریف بر اساسِ شباهتِ خانوادگی (family resemblance) است. در این نوع از تعریف، هیچ ویژگیِ مشترکی میانِ

مصادیق وجود ندارد که بتوان بر اساسِ آن تعریفی صوری به دست داد (ibid: 39). اما بارزترین نوع از دانشِ مهارتیِ صوری‌ناشدنی را در کاربست خطای زبان می‌بینیم (ibid: 111)؛ در بسیاریِ موارد ما جملات را به‌گونه‌ای ناقص و بدساخت ادا می‌کنیم و همچنان فهمی کافی از آن خواهیم داشت. پس گویی فهم، به شیوه‌ای متفاوت روی می‌دهد. دریفوس بر آن است که این نمونه‌ها نشان می‌دهند که فهمِ مهارتی و عملیِ ما از زبان، فهمی بسترمند و غیر قابل صوری‌سازی است.

فراتر از این دریفوس مدعی می‌شود که نه‌تنها مهارتِ به‌کارگیری زبان بلکه هرگونه مهارت انسانی دانشی غیرتئوریک و صوری‌ناشدنی است. نگرشِ متافیزیکی به مهارت‌های انسانی که از سقراط تا لایب‌نیتس به‌روشنی بیان شده است این بوده است که در پسِ دانشِ مهارتی یک نظریه وجود دارد (Dreyfus, 1987: 18) اما دریفوس چه با مثال‌هایی که از دانش مهارتیِ به‌کارگیری زبان می‌آورد چه با پدیدارشناسی رفتار خبرگی نشان می‌دهد که وجود چنین نظریه‌ای کاملاً محل شک است. از نظر وی رسیدن به درجهٔ مهارت و خبرگی ضرورتاً نیازمند تجربهٔ عملی و درک شهودیِ هرچه بیش‌تر از یک سو و حس درونیِ پرورش‌یافته‌تر از سوی دیگر است و هیچ‌گونه دانش صوری و تئوریکی نمی‌تواند جای این دو را بگیرد. با این حساب اگر نتوان دانش مهارتیِ ما را به دانشی تئوریک و صوری فروکاست آن‌گاه فرض معرفت‌شناسانه نیز مخدوش است.[۹]

فرض هستی‌شناسانه: محتوای این فرض آن است که هر آن‌چه هست مجموعهٔ مستقلی از واقعیت‌های منطقاً جدا و به خودیِ خود مستقل است. دریفوس کلِ تاریخِ فلسفه چه سنتِ عقل‌گرا چه سنتِ تجربه‌گرا را پشتوانهٔ این فرض می‌داند:

مکاتب عقل‌گرا و تجربه‌گرا در اتمیسم منطقی راسل به یک نقطهٔ مشترک رسیدند و این ایده، کامل‌ترین بیانش را در تراکتاتوس ویتگنشتاین یافت؛ جایی که جهان به واسطهٔ مجموعه‌ای از واقعیت‌های اتمیک تعریف می‌شود که می‌توانند در گزاره‌های منطقاً مستقلی بیان شود. این خالص‌ترین صوری‌سازی از فرض هستی‌شناختی است (Dreyfus, 1979: 123).

از آن جهت که مهم‌ترین پشتوانهٔ فرضِ هستی‌شناختی، نگرشِ سنتِ فلسفی به جهانِ هستی است؛ پس دریفوس نیز برای نقد این فرض باید به سراغِ آموزه‌های فیلسوفانی برود که نگرشِ سنتی به جهان هستی را به چالش کشیده‌اند. جریان‌هایی در فلسفه هم‌چون اگزیستانسیالیسم، هرمنوتیک، پدیدارشناسی، و نیز ویتگنشتاین متأخر امکانِ چنین نقدی را

برای دریفوس فراهم آورده‌اند. همهٔ ایشان با تعابیرِ گوناگون جهـانی انسـانی را ورای عـالم فیزیکی آشکار می‌سازند که نادیده‌گرفتنِ آن مهم‌ترین پشتوانهٔ فرضِ هستی‌شناختی است؛[1]

فرض هستی‌شناختی که به موجب آن جهان انسانی می‌تواند به واسطهٔ یک سری عناصر به نمایش درآید معقولیتش را هنگامی می‌یابد که فرد در تمایز میان جهان (world) و عـالم (universe) ناکام بماند، یا به بیان دیگر در جداسازی بین آن‌چه که از سـوی یـک چیـز در موقعیت انسانی و [آن‌چه از سوی یک حالت] فیزیکیِ سامانه به‌دست داده می‌شود نـاموفق شود (ibid: 125).

اگر نتوان جهانِ انسانی را به عالم فیزیکی فروکاست فـرضِ هسـتی‌شـناختی نیـز زیـر سؤال خواهد رفت. دریفوس به کمک اقتباسی کـه از فلسـفهٔ کیرکگـارد مـی‌کنـد امکـانِ فروکاهی فهم جزئیِ زمان‌مندِ ما در جهان انسانی به قواعد کلی و بی‌زمانِ (timeless) علـمَ را به چالش می‌کشد (ibid: 125). ویتگنشتاین متأخر نیـز بـا نقـدِ اتمیسـمی کـه خـود در تراکتاتوس فرض گرفته بود و با معرفیِ سبک‌های زندگی (forms of life) به مثابـهٔ بسـتر فهمِ انسانی، پایه‌ای را برای دریفوس فراهم می‌آورد تا بسترمند و غیر اتمیک‌بـودنِ جهـانِ انسان را آشکار سازد (ibid: 133). همچنین به کمک جهـان ادراکِ انسـانی کـه مرلوپـونتی معرفی می‌کند دریفوس نشان می‌دهد که تجربهٔ ادراکی انسان آمیختهٔ به معنا و گـران‌بـار از ارزش است (Dreyfus, 1967). افزون بر این دریفوس به یاریِ جهانِ هرمنوتیکی هیدگر نیز نشان می‌دهد که فهمِ انسانی در مواجههٔ هرروزه با ابزارها یک کلِ به‌هم‌پیوسته را مـی‌سـازد که نمی‌توان آن‌ها را به مجموعه‌ای از اشیای فرادستی (present-at-hand things) فروکاست (Dreyfus, 1979: 252). بنابراین چنان‌که دریفوس با یاری‌گرفتنِ از آموزه‌های این فیلسوفان نشان می‌دهد جهان انسانی (world) کاملاً متفاوت با عالم فیزیکی (universe) است. جهـانِ انسانی جهانی زمان‌مند، معنادار، آمیخته به ارزش، بسترمند، و کل‌گرایانه است که نمی‌تـوانَ به عالمِ فیزیکی که مجموعه‌ای از واقعیت‌های بی‌زمان، فاقد معنا، خنثـی، فـارغ از بسـتر، و اتمیک است فروکاست.

پس بدین‌سان دریفوس در نقدش بـر پـارادایم نمـادی هـر سـه فـرضِ روان‌شناسـانه، معرفت‌شناسانه، و هستی‌شناسانهٔ حامیِ این پارادایم را به چالش کشید.

۲.۲ دورهٔ بازیابی هوش مصنوعی پیوندگرا

در میانهٔ دههٔ ۱۹۸۰ شکست هوش مصنوعی نمادی هم در حوزهٔ عملی ـ مهندسی و هم در

حوزهٔ نظری ـ فلسفی بر همگان روشن شده بود. در حوزهٔ مهندسی، پروژه‌های هوش مصنوعی در پیاده‌سازی هیچ‌یک از وعده‌های داده‌شده کامیاب نبودند. به‌گونه‌ای که بودجهٔ بسیاری از این پروژه‌ها قطع و یا بسیار محدود شد (Russell and Norvig, 2003: 21). در حوزهٔ نظری نیز مهم‌ترین استدلال باقی‌مانده برای دفاع از پارادایم کلاسیک آن بود که چون هیچ بدیلی برای مطالعات علوم شناختی وجود ندارد چاره‌ای جز پذیرش محاسبه‌گرایی نمادی نیست (Dreyfus, 1992: xiii Dennett, 1998: 223). هوش مصنوعی پیوندگرا درصدد بود تا با فراهم‌آوردن چنین بدیلی هوش مصنوعی نمادی را کاملاً از رده خارج کند (Dreyfus, 1992: xv).

در دههٔ ۱۹۸۰ هم‌راستا با ناکامی هوش مصنوعی نمادی، گرایشی در بین پژوهش‌گران هوش مصنوعی برای بازگشت به الگوبرداری از مغز پیدا شد. این گرایش بر پایهٔ پژوهش‌های عصب‌شناسانهٔ دونالد هِب در دههٔ ۱۹٤۰ بر روی یادگیری شبکه‌های عصبی (Neural networks) متولد شده بود (Hebb, 1949) و از آغاز در میان برخی از پژوهش‌گران هوش مصنوعی ازجمله فرانک روزنبلات به‌گونه‌ای جدی پی‌گیری می‌شد (دریفوس، ۱۳۸٦: ۵)، با این حال تاریخ هوش مصنوعی به‌گونه‌ای پیش رفت که این پارادایم پژوهشی برای سه دهه در حاشیه و زیر سایهٔ پارادایم نمادی قرار گرفت.

پارادایم مبتنی بر آموزش شبکه‌های عصبی به پیوندگرایی مشهور شد و در برابر پارادایم نمادی بر سه پایهٔ متفاوت بنیان نهاده شده است: یک نوع جانشین برای بازنمایی، یک نوع جانشین برای پردازش، یک مجموعه الگوریتم آموزشی (Clark, 1996: 7).

بازنمود در این پارادایم برخلافِ پارادایم نمادی، متغیری خنثی و فاقد محتوا نیست؛ بلکه به واسطهٔ لایه‌ای زیرین‌تر دارای محتوایی درونی‌شده است (Chalmers, 1992: 15). لایهٔ زیرین‌تر، شبکهٔ پیوندهای عصبی است. تک‌تک عناصر این لایه به خودی خود فاقد معناست و هر نورون به‌تنهایی هیچ امری را بازنمایی نمی‌کند. اما نورون‌ها به صورت شبکه‌ای با هم در پیوند هستند. مجموعه‌ای از پیوندهای نورونی یک بازنمود را می‌سازند. بدین‌ترتیب هر بازنمود از طریقِ یک لایهٔ زیرین‌تر با یک سری بازنمودِ دیگر ارتباطی درونی دارد. پیوندهای نورونی به‌گونه‌ای ترتیب داده می‌شود که مثلاً بازنمودِ زینِ اسب از درون به بازنمودِ خودِ اسب نزدیک‌تر باشد تا به بازنمودی هم‌چون آدم فضایی. پس به جهت برخورداری از یک ساختار درونی، بازنمایی در شبکه‌های عصبی وجهی کل‌گرایانه دارد و این به‌روشنی با اتمیسمِ سامانه‌های نمادی در تضاد است.

شیوهٔ پردازشی که در پیوندگرایی به‌کار گرفته می‌شود مبتنی بر آموزش و به واسطهٔ یک سری الگوریتم‌های آموزشی است. آن‌چه به جهت فلسفی اهمیت دارد آن است که در رویکردِ نمادی و در رایانه‌های دیجیتال بار اولی که یک فرایند محاسباتی انجام می‌پذیرد با بار هزارم آن تفاوتی نمی‌کند. یعنی تجربهٔ عملی به مهارتِ بیش‌تر نمی‌انجامد. اما پیوندگرایی درصدد است این ویژگی زندگی انسانی را که به موجب آن تجربهٔ هرچه بیش‌تر به مهارت هرچه بیش‌تر منتهی می‌شود پیاده‌سازی کند. از این رو مبتنی بر یک سری الگوریتم‌های آموزشی هر بار به سوی بهبود نتایج پیش می‌رود. گویی شبکه هر بار در یافتنِ پاسخ ماهرتر شده است. در این رابطه در هر پردازشِ آموزشی، کلِ شبکهٔ عصبی درگیر است تا پیوندهای میانِ نورون‌ها به‌گونه‌ای اصلاح شود که شبکه، پاسخ مناسب‌تری بدهد (Clark, 1996: 8). بنابراین از نظرگاهی فلسفی پیوندگرایی از دو جهت با هوش مصنوعی نمادی متفاوت است. نخست به جهتِ بازنمایی کل‌گرایانه و دوم به جهتِ پیاده‌سازی آموزش یافتنِ مهارت در عمل.

۱.۲.۲ نقدِ هوش مصنوعیِ پیوندگرا به دست دریفوس

در نقدِ پارادایم پیوندگرا باید دو گام برداشت؛ نخست دلایل و دلالت‌های فلسفیِ شکستِ هوش مصنوعیِ نمادی و بازگشت به پیوندگرایی را برشمرد. سپس در گام دوم، نقد محتاطانهٔ دریفوس بر پیوندگرایی و محدودیت‌های این پارادایم را نیز بیان کرد.

ارزیابی بازگشت به پیوندگرایی: دریفوس پس از ظهور دوبارهٔ پیوندگرایی خود به ارزیابی این تحول اساسی مبادرت می‌ورزد و از این ارزیابی، فراتر از نقد یا تأیید یک جریانِ هوش مصنوعی، پایه‌ای برای نقدِ یکی از اساسی‌ترین بنیادهای اندیشه و فرهنگِ غربی می‌سازد. دریفوس پرسشی تأمل‌برانگیز را مطرح می‌کند؛ چرا با وجودِ این‌که هم پارادایم پیوندگرا و هم پارادایم نمادی در دههٔ ۱۹۵۰ با موفقیت‌های جزئی و مشکلاتِ کلی روبه‌رو شده بودند، به یک‌باره پارادایمِ پیوندگرا برای سه دهه به کلی به حاشیه رفت؟ چنان‌که شکستِ پارادایمِ نمادی نشان داد عامل به حاشیه‌رفتن پارادایم پیوندگرا دلایل علمی ـ فلسفی نبوده است؛ پس چه امری اسباب برتری نگرش نمادی به هوشمندی را فراهم آورد؟ به بیان دیگر سرچشمهٔ خوش‌باوری‌های پدرانِ هوش مصنوعی نمادی چه بود؟ پاسخ دریفوس روشن است؛ فرهنگ و اندیشهٔ غربی که از دل فلسفهٔ غرب بیرون آمده بود پشتوانهٔ هوش مصنوعی نمادی بوده است؛

فقط دکارت و اخلافِ وی نبوده‌اند که پشتِ پردازشِ اطلاعاتِ نمادی ایستاده بودند؛ کلِ فلسفهٔ غرب چنین بود (دریفوس، ۱۳۸٦: ۱۲۷).

پس آن‌چه هوش مصنوعیِ نمادی را در جایگاهِ برتر قرار داد فلسفهٔ غربی و نگرش اصالت تئوریک آن بود که در اعماق فرهنگ غربی ته‌نشین شده بود. به بیان دیگر می‌توان مدعی شد که نقد هوش مصنوعی این امر را آشکار می‌کند که اصالت تئوری و محاسبه‌گرایی در اعماقِ پس‌زمینهٔ فرهنگ غربی ریشه دوانده‌اند؛ به‌گونه‌ای که بی هیچ دلیل نظری و معرفتی یک پارادایم غیر محاسبه‌گرا و غیر تئوریک را برای سه دهه از رده خارج کرد. دریفوس اساساً مدعی است پارادایم نمادی در دل فلسفهٔ سنتیِ غربی پرورانده شده بود؛ وی به‌صراحت مدعی می‌شود که تاریخِ هوش مصنوعی با سقراط و افلاطون آغاز می‌شود و ایدهٔ فروکاستنِ همهٔ عقلانیت به محاسبه، نخستین صورت‌بندی‌هایش را با ایشان پیدا کرد (Dreyfus, 1979: xv). این ایده در عصرِ جدید با هابز صورت‌بندیِ شفاف‌تری یافت (Dreyfus, 2007: 1). در این میان دکارت با تحویلِ فهم به بازنمودهای درونی (ibid) لایب‌نیتس با تئوریک‌دانستن دانش عملی و نیز با تأسیسِ زبانِ باینری (Dreyfus, 1987: 18)، کانت با بازگرداندنِ مفاهیم به قواعد و فرگه با صوری‌سازیِ این قواعد (Dreyfus, 2007: 1) همگی در مسیری به پیش رفتند که در انتهای آن اتمیسم منطقیِ (logical atomism) راسل و ویتگنشتاینِ متقدم (Dreyfus, 1979: 123) و نیز پدیدارشناسی استعلاییِ (transcendental phenomenology) هوسرل قوی‌ترین پشتوانهٔ فلسفی را برای پارادایم نمادی فراهم آوردند (دریفوس، ۱۳۸٦: ۱۳۰). بی‌شک دریفوس نمی‌تواند ادعا کند که پدرانِ هوش مصنوعیِ نمادی به صورتِ مستقیم آثارِ این فیلسوفان را نخوانده بودند.[۱۱] اما با عطفِ توجه به این ادعای درست دریفوس که هیچ دلیلی برای برتری فاحش پارادایم نمادی نسبت به پارادایم پیوندگرا وجود نداشته است می‌توان هم‌سو با دریفوس مدعی شد که نگرشی به ذهن، معرفت، و هستی که در دل فلسفهٔ غرب پرورش یافته بود و در پس‌زمینهٔ فرهنگِ غربی نشست کرده بود عاملی را فراهم آورد که به پشتوانهٔ آن هوش مصنوعیِ نمادی درست پنداشته شد و پارادایمِ پیوندگرا نیز به جهتِ عدمِ برخورداری از چنین پشتوانه‌ای به حاشیه رفت.

اما شکست هوش مصنوعیِ نمادی و بازگشت به پارادایم پیوندگرا همهٔ پشتوانه‌های هوش مصنوعیِ نمادی را نیز به چالش کشید. چنان‌که گفتیم از جملهٔ مهم‌ترین پشتوانه‌ها، میراثی بود که از سنت فلسفهٔ غرب برای هوش مصنوعی نمادی به ارث رسیده بود و

گذر از این پارادایم به یک معنا بر آموزه‌های فیلسوفانی که این سنت فلسفی را نقد کـرده بودند نیز مهر تأییدی زد.[۱۲] به بیان دیگر همان مشکلی که سـنت فلسـفی را بـا بـن‌بسـت مواجه کـرده بـود هـوش مصـنوعی را نیـز زمـین زد. ایـن مشـکل در کلـی‌تـرین بیـان، به‌دست‌دادن یک تئوری صوری بر مبنای یک سری عناصر اتمی در مورد فهم ما در جهان هرروزه بود. به بیان دریفوس:

> همان بلایی که جهان روزمره بر سر فلسفهٔ سنتی آورد بر سـر هـوش مصـنوعی نیـز آمـد (همان: ۱۳۵).

بدین‌سان مهم‌ترین دلالت‌های فلسفیِ چرخش به سوی پیونـدگرایی نخسـت تأییـد کل‌گرایـی در برابـرِ اتـم‌گرایـی (همـان: ۱۲۹) و نیـز بسـترگرایی در برابـر فورمالیسـم (Dreyfus, 1992: xxxv- xxxvi) است و به صورت کلی‌تر نمایان‌گر آن است کـه فرهنـگ غربی با تقلیل دانشِ حقیقی به تئوری، و هستی به عالم فیزیکی، جهـان زنـدگی انسـانی و دانش بسترمندِ وی را به فراموشی سپرده است.[۱۳] با این حساب تـا آن‌جـا کـه بـه پـارادایم نمادی مربوط است ادعای اصلی مقاله مبنی بر این‌که تاریخ تحولات هوش مصنوعی مؤید نقدهای دریفوس بوده است به اثبات می‌رسد. این امری بود که حتی پدرانِ هوش مصنوعی نمادی ازجمله نیوول نیز به آن اذعان کردند (دریفوس، ۱۳۸٦: ۱۳۵).

محدودیت‌های پیوندگرایی: دریفوس گرچه از پیونـدگرایی در برابـر هـوش مصـنوعی نمادی پشتیبانی می‌کند اما در این امر بسیار محتاط است و می‌کوشد محدودیت‌های هوش مصنوعی پیوندگرا را نیز آشکار سازد. مهم‌ترین مشکلی که دریفوس پیشِ پای پـروژه‌هـای پیوندگرایانه برمی‌شمرد مشکلِ تعمیم (generalization) است. همهٔ پیوندگرایان «می‌پذیرنـد که شبکهٔ هوشمند باید بر تعمیم توانا باشد. به عنوان نمونه برای انجام یک طبقه‌بندی با دادن مثال‌های مناسب از ورودی‌هـای مـرتبط بـا یـک خروجـی مشـخص، شـبکه بایـد خـود ورودی‌های دیگر از همان نوع را به همان خروجی مرتبط سازد» (Dreyfus, 1992: xxxvi). اما دریفوس با آوردن مثال‌هایی عملی نشان می‌دهد که بسیاری از این شـبکه‌هـای عصـبیِ پیوندگرا در امر تعمیمِ درست با مشکل روبه‌رو شده‌اند و فقط با دخالت طراح شبکه، تعمیمِ درستی روی داده است. به باور دریفوس این مثال‌ها «این نکتهٔ کلی را روشن می‌سازند کـه یک شبکه باید در فهم کامن‌سنسِ (common sense) ما از جهان شریک باشد تا در درک ما از تعمیم مناسب شریک شود» (ibid). فهم کامن‌سنسِ ما بر بسـتر نیازهـای بـدنی، اهـداف شخصی و انتظارات فرهنگی شکل می‌گیرد. از این رو از نظر دریفوس پارادایمی مـی‌توانـد

هوش مصنوعی را یک گام از پیوندگرایی جلوتر ببرد که بتواند نیازها، اهداف و انتظارات انسانی را به مثابهٔ معیاری درونی برای تعمیم درست پیاده‌سازی کند. به‌نظر می‌رسد پارادایم حیات مصنوعی درصدد بوده است تا چنین نقشی را ایفا کند.

دریفوس در پایان، پیوندگراییِ مبتنی بر شبکه‌های عصبی را نیز به جهت ادعاهای گزاف در عین نبود معیار مناسبی برای تعمیم به نقد می‌کشد.

بنابراین پژوهش‌گران شبکه‌های عصبی در این موقعیت‌شان با پیروزی‌های کوچک بدون شیوه‌ای اصولی برای تعمیم به‌نظر می‌رسد در همان جایگاه پژوهش‌گران هوش مصنوعی کلاسیک باشند هنگامی که من در ۱۹۶۰ شروع به نوشتن در مورد آن‌ها کردم (ibid: xxxvii).

۳.۲ دورهٔ ظهور حیات مصنوعی

در دو دههٔ اخیر مجموعهٔ جریان‌هایی در هوش مصنوعی متولد شده‌اند که به حیات مصنوعی معروف شدند. در این پارادایم، هوش انسانی بریده از بستر محیطی‌اش درنظر گرفته نمی‌شود بلکه انسان و محیطش با هم به مثابهٔ یک کل یک سامانهٔ هوشمند را می‌سازند. سه گرایش اصلی را می‌توان در حیات مصنوعی تشخیص داد. گرایش نخست پژوهش برای پیاده‌سازی رفتار هوشمند را بر روی بافت‌های زندهٔ جان‌داران پی می‌گیرد. این گرایش بیش از هر دانشی به بیوشیمی وابسته است (Dennett, 1997)، اما گرایش دوم درصدد است تا روبات‌هایی با توانایی‌های زیستی بسازد. این گرایش در تراز سخت‌افزاری در پی تحقق جریان حیات در یک جان‌دار است. در این نگرش روبات و محیطش یک سامانهٔ واحد دانسته می‌شوند و روبات همواره در یک بستر محیطی درنظر گرفته می‌شود. اما گرایش سوم به مطالعه و شبیه‌سازی نرم‌افزاری رفتار جان‌دار در عرض کلِ محیط زیست و طول تکامل می‌پردازد و رفتار هوشمندانه را فرآوردهٔ درآمیختگی و داد و ستد جان‌داران با این محیطِ در حال تکامل می‌داند. این گرایش در محدودهٔ نرم‌افزاری به پژوهش می‌پردازد (Clark, 1997: 14-19). گرایش نخست رویکرد خیس (wet)، گرایش دوم رویکرد سخت (hard) (سخت‌افزاری) و گرایش سوم رویکرد نرم (soft) (نرم‌افزاری) نامیده می‌شود (Adamatzky and Komosinski, 2009).[۱۴]

مروری ساده بر گرایش‌های موجود در حیات مصنوعی ادعای اصلی مقاله را در مورد ظهور حیات مصنوعی نیز به اثبات می‌رساند. مدعای ما آن بود که تحولات روی‌داده در

هوش مصنوعی نقدهایی را که دریفوس پیش‌تر بر جریان‌های گوناگون این رشته وارد آورده بود تأیید می‌کند. به‌روشنی پیداست که میان گرایش‌های حیات مصنوعی و بینش‌های دریفوس در نقد دو پارادایم پیشین نزدیکی بسیاری وجود دارد و گویی این گرایش‌ها از آن بینش‌ها متأثر بوده‌اند. دریفوس در نقد هوش مصنوعی نمادی نشان داده بود که برای برخورداری از هوشمندی باید ضرورتاً بدن‌مند (embodiment) بود (رویکرد خیس)، همچنین وی همواره جدایی میان بستر زندگی هرروزه و مهارت‌های هوشمندانهٔ انسانی را انکار می‌کرده است (رویکرد سخت). افزون بر این دریفوس در نقد هوش مصنوعی پیوندگرا آورده بود که برای هوشمندی باید بتوان تعمیم‌هایی هم‌چون انسان انجام داد و برای برخورداری از این توانایی باید از نظرگاه انسان بدن‌مند درون یک فرهنگ به جهان نگریسته شود. نظرگاهی که تاریخ طبیعی و فرهنگی آن را فراهم می‌آورد (رویکرد نرم). پس هر سهٔ این رویکردها با بینش‌های کلی دریفوس در نقد دو پارادایم نخست هم‌سو است؛ بدین‌ترتیب مدعای اصلی مقاله مبنی بر این‌که تاریخ تحولات هوش مصنوعی مؤید نقدهای دریفوس بوده است در مورد دو تحول پارادایمی در تاریخ هوش مصنوعی یعنی بازگشت به پیوندگرایی و ظهور حیات مصنوعی به اثبات رسید. در ادامه باید نسبت نقدهای دریفوس با حیات مصنوعی را به صورت دقیق‌تر و با جزئیات بیان کنیم و همچنین موضع کلی وی در مورد حیات مصنوعی و آیندهٔ آن را نیز مورد نقد و ارزیابی قرار دهیم. دریفوس خود مدعی است اگر دلالت فلسفیِ حرکت از پارادایم نمادی به سوی حیات مصنوعی شکستِ ایجابی هستی‌شناسی دکارتی و تأیید سلبی هستی‌شناسی هیدگری (در تفسیر دریفوسی از آن) بوده است با حیات مصنوعی، هستی‌شناسی هیدگری (در تفسیر دریفوسی از آن) در آزمونی ایجابی مورد ارزیابی قرار خواهد گرفت (Dreyfus, 2007: 13).

۱.۳.۲ نقد حیات مصنوعی به دست دریفوس

دریفوس هیچ‌گاه تعبیر حیات مصنوعی را به‌کار نبرده است اما چند جریان درون این پارادایم را مورد ارزیابی قرار داده است و بر تأثیر همهٔ این جریانات از هیدگر و مرلوپونتی اذعان دارد (Dreyfus, 2005). مهم‌ترین جریانی از پارادایم حیات مصنوعی که وی مورد پژوهش قرار می‌دهد هوش مصنوعی هیدگری (Heideggerian AI) نامیده می‌شود که می‌توان آن را ذیل حیات مصنوعی سخت‌افزاری گنجاند.[۱۵] این رویکرد با عنوان توجه‌برانگیزش برای دریفوس چیزی بیش از یک جریان در میان دیگر جریان‌های هوش مصنوعی است. پژوهش‌گران هوش مصنوعی که در آغاز نیمهٔ دوم قرن بیستم کار خود را با

دعوی کنارگذاردن فلسفه و شناخت و بازسازی عقلانیت و هوش انسانی آغاز کرده بودند پس از نیم قرن تلاش ناکام اینک برای بازیابی امید به کشف راز هوشمندی انسانی باید پای درس هیدگر می‌نشستند. فیلسوفی که دریفوس در همهٔ نقدهایش بیش‌ترین الهام را از وی گرفته بود. از میان پژوهش‌گران هوش مصنوعی تِری وینوگراد نخستین و بزرگ‌ترین پژوهش‌گر هوش مصنوعی بود که تحت تأثیر دریفوس هیدگر خواند و ایمانش را نسبت به هوش مصنوعی کلاسیک (GOFAI) از دست داد (Dreyfus, 2007: 4). وینوگراد همچنین خود نخستین کسی بود که تعبیر هوش مصنوعی هیدگری را به‌کار برد.

این جزو طنزهای تاریخ است که آزمایشگاه M.I.T مهد هوش مصنوعی هیدگری شد. پس از بیست سال از صورت‌بندی نخست نقد دریفوس در آن‌جا و پس از این‌که ایشان به تأمل در آن‌چه او می‌گفت پرداختند بدون شک بعضی از کارهایی که الان در این آزمایشگاه انجام می‌شود تحت تأثیر هیدگر و دریفوس است (وینوگراد به نقل از Dreyfus, 2007: 5).

چنان‌که وینوگراد می‌گوید یک سری از جریان‌های هوش مصنوعی به صورت مستقیم از دریفوس تأثیر پذیرفتند. دریفوس دو ویژگی کلی برای هیدگری‌بودن یک برنامهٔ پژوهشی در هوش مصنوعی برمی‌شمرد؛

[یکی این‌که] نقد هیدگر بر بازنمودگرایی درونی دکارتی را پذیرفته‌اند و ... [دیگری این‌که] شعار جان هوگلند را مبنی بر این‌که شناخت، بدن‌مندانه و موقعیت‌مند (embedded) است سرلوحهٔ عمل قرار می‌دهند (ibid).

پذیرشِ بدن‌مندی و بسترمندیِ شناخت وجهِ ایجابی و ردِ بازنمودگرایی وجهِ سلبیِ هر جریانِ هیدگری در هوش مصنوعی است.[۱۶ ۱۷] این جریان درصدد است نگرشی به هوشمندی را پیاده‌سازی کند که برخلاف جریان‌های پیشین در هوش مصنوعی و نیز برخلاف کلِ تاریخ فلسفه تا پیش از هیدگر و مرلوپونتی، پیوند درونی انسان و جهانش را نادیده نگیرد و نقش ضروریِ بدن در فرایند شناخت را تأیید کند.

بنابر شرح پدیدارشناسانهٔ دریفوس از رفتار هوشمندانه، ما همواره در تعادلی درونی و بدنی با جهانمان قرار داریم. هرگاه این وضعیتِ تعادلی و تسلط ما بر پیرامونمان برهم بخورد بی‌آن‌که نیاز به تحلیلِ بازنمودیِ امر بیرونی باشد حسی درونی ما را به سوی حداکثر تعادل با محیط و حداکثر تسلط بر آن سوق خواهد داد. نمونه‌ای که دریفوس از مرلوپونتی وام می‌گیرد نگریستن به یک تابلوی نقاشی است (ibid: 16). اگر ما زیاد به تابلو نزدیک شده باشیم خودبه‌خود کمی عقب می‌رویم و اگر زیاد دور باشیم بی‌نیاز به تأمل کمی جلو

می‌رویم تا تابلو را تار ادراک نکنیم. حس درونی تشویشِ حاصل از ادراک تار، ما را به سوی حداکثر تعادل هدایت خواهد کرد. به باور دریفوس این تنظیم تعادل از درون، ویژگی کلی نسبت انسان و جهان است (ibid).

همین پیوند درونی بنیاد رفتار هوشمندانهٔ ما را فراهم می‌آورد. روبات‌ها به علت این‌که بهره‌ای از چنین پیوندی با جهان ندارند در پیاده‌سازی رفتار هوشمندانه ناکام مانده‌اند. علت اصلی شکست هوش مصنوعی آن بوده است که روبات‌ها نمی‌توانند در برابر تغییرات غیر منتظره و برنامه‌ریزی‌نشده پاسخ مناسبی بدهند (Dennett, 1984). آن‌چه سبب می‌شود تا تغییرات غیر منتظره و تعریف‌نشده برای انسان‌ها معنادار شود همین پیوند درونیِ انسان و جهان و حس ما از مطابقت با محیط است. تشویش برآمده از یک تغییر غیرمنتظره، ما را از درون به سوی کوشش در جهت بازیابی دوبارهٔ تعادل‌مان با محیط سوق می‌دهد (Dreyfus, 2007: 27) هنگامی که پاسخ مناسب یافته شد و تغییر مورد نظر معنادار شد تشویش درونی ما برطرف می‌شود و بار دیگر در مطابقت حداکثری با محیط قرار خواهیم گرفت.

حال هوش مصنوعی هیدگری می‌کوشد تا همین پیوند درونی انسان و جهان و حس درونی ما از تعادل با محیط را پیاده سازد. به باور دریفوس فقط جریانی در هوش مصنوعی هیدگری که بر پایهٔ الگوی نورونی فریمن پیاده شده باشد از پس این امر برخواهد آمد.[۱۸] متناظر با حسِ درونی ما از تعادل با محیط در شرح مرلوپونتی و هیدگر، در الگوی فریمن حداقل انرژی در هر غشای مغزی قرار دارد. متناسب با هر محرک بیرونی، شبکه‌هایی نورونی در بخش مربوطه در مغز شکل می‌گیرند که پاسخ به محرک را هدایت می‌کنند. زمانی که پاسخ ما به محرک، فعالیت مغز را در حداقل انرژی قرار نمی‌دهد همان زمانی است که در ساحت پدیداری حس درونی ما حداقل تشویش را ندارد و تسلط ما بر محیط و تعادلمان با آن برهم خورده است. در این زمان کل بخش مورد نظر وارد یک بی‌نظمی نورونی می‌شود و تا ایجاد شبکه‌ای جدید که به محرک پاسخی مناسب بدهد و وضعیت فعالیت را به حداقل انرژی برساند بی‌نظمی ادامه می‌یابد. این کنش و واکنش‌هایی که به جهت رسیدن به حداقل انرژی در تراز نورونی صورت می‌پذیرد به باور دریفوس در تراز پدیداری با کوشش در راستای رسیدن به کم‌ترین تشویش و بیش‌ترین تعادل با محیط متناظر است (ibid: 19-25).[۱۹] بدین‌سان الگوی فریمن در تناظر کامل با شرح پدیداری‌شناختی قرار دارد[۲۰] و می‌تواند پیوند درونیِ انسان و جهانش را تبیین کند و پروژهٔ

ساخته‌شده بر پایهٔ آن به مشکلاتی که همهٔ برنامه‌های هوش مصنوعی تا به امروز را از پا در آورده است دچار نمی‌شود. بنابراین دریفوس پس از نیم قرن نقد هوش مصنوعی، سرانجام امکان آن را پذیرفت.

اما به‌نظر نمی‌رسد این رویکردِ واپسینِ دریفوس چندان قابل دفاع باشد. دستِ کم این رویکرد با بخشی از نقدهای پیشینِ خودِ دریفوس ناسازگار است. دریفوس همواره عدم برخورداری از دانش نامتعین و سوبژکتیو را دلیلی بر ناکامی پروژه‌های هوش مصنوعی می‌دانسته است. اما وی روشن نمی‌سازد که یک پروژهٔ رایانه‌ای مبتنی بر الگوی فریمن چگونه می‌تواند این دانش غیرفورمال و سوبژکتیو را تولید کند؟ چطور متغیرهای ابژکتیو می‌توانند حس درونی ما را بازسازی کنند؟ دریفوس پاسخ روشنی به این پرسش‌ها نمی‌دهد. به همین ترتیب، سخن‌گفتن از آزمون تجربی هستی‌شناسیِ هیدگری برمبنای علوم شناختی و هوش مصنوعی هیدگری دستِ کم با تفسیر پیشینِ خود دریفوس از هیدگر ناهم‌خوان است. دریفوس شاید تحت تأثیر گفت‌وگوهای مداومش با فیلسوفان تحلیلی هم‌چون سرل و دنت درنهایت طبیعت‌گرایی را به صورت کامل می‌پذیرد و بر پایهٔ آن مدعی می‌شود که شرح فریمن از عملکرد مغز برای توضیح امر معنادار کافی است (ibid: 19). این در حالی است که دریفوس پیش از این بیان داشته بود که یک شرح علّی از طبیعت به‌هیچ‌وجه نمی‌تواند برای توضیح امر معنادار کافی باشد (Dreyfus, 1991: 121). از سوی دیگر می‌دانیم که هوسرل پدیدارشناسی را در مقابل روان‌شناسی طبیعت‌گرا تعریف کرده بود و از این جهت تا جایی که هستی‌شناسی هیدگری را یک رویکرد پدیدارشناسانه بدانیم سخن‌گفتن از آزمون تجربی آن به واسطهٔ یک دانش طبیعی هم‌چون عصب‌شناسی قابل پذیرش نیست.

۳. نتیجه‌گیری

سه مدعا در این مقاله مورد نظر بوده است؛ بر پایهٔ مدعای اصلی مقاله نشان دادیم که منطق فلسفی حاکم بر تاریخ هوش مصنوعی، نقدهای دریفوس بر این رشته در دوره‌های گوناگون را تأیید می‌کند. در آغاز تاریخ هوش مصنوعی در دههٔ ۱۹۵۰ پارادایم نمادی خیلی زود پارادایم پیوندگرا را به حاشیه راند. پشتوانهٔ پارادایم نمادی نگرشی بود که الگوی رایانه‌ای را برای ذهن مناسب می‌دانست و دانش را به فورمالیسم و هستی را به عناصرِ اتمیک تحویل می‌برد. این نگرش که به وسیلهٔ سنت فلسفی تا رویکرد اتم‌گرای ویتگنشتاین

متقدم و پدیدارشناسی استعلایی هوسرل پشتیبانی می‌شد در عمق اندیشه و فرهنگ غربی نشست کرده است. اما شکست پارادایم نمادی و بازگشت به پارادایم پیوندگرا نظر منتقدان سنت فلسفی ازجمله هیدگر، ویتگنشتاین متأخر و مرلوپونتی را در نقد این سنت تأیید کرد. بنابر نظر ایشان، نمی‌توان فهم کل‌گرایانه، بدن‌مندانه، مهارتی، و بسترمندِ ما در جهان هرروزه را به واسطهٔ یک سری عناصر اتمیک صوری کرد و بدین‌سان پژوهش‌گران هوش مصنوعی نمادی نیز در این امر ناکام بودند. پس از این شکست در دههٔ ۱۹۸۰ بازگشتی به پارادایم پیوندگرا در میان پژوهش‌گران هوش مصنوعی روی داد. پیوندگرایی به جهت پذیرش فهم کل‌گرایانه و بسترمند انسانی یک گام به سوی تأیید نقدهای دریفوس به‌شمار می‌آمد، با این حال به باورِ خودِ دریفوس این پارادایم نیز در راستای بازسازیِ تعمیم‌های انسانی با مشکل روبه‌رو است. چنان‌که دریفوس مدعی است فقط با برخورداری از فهمی تاریخ‌مند (چه تاریخ فرهنگی، چه تاریخ طبیعی) و بدن‌مند هم‌چون انسان‌ها می‌توان تعمیم‌هایی از نظرگاهی انسانی انجام داد. در راستای رسیدنِ به چنین نظرگاهی پارادایم حیات مصنوعی در دو دههٔ اخیر سر برآورده است.

اما مدعای دوم آن بود که نگرش اصالت تئوری و محاسبه‌گرا در اعماق فرهنگ غربی نشست کرده است. در حاشیهٔ روایتِ تاریخِ فلسفیِ هوش مصنوعی نشان دادیم که چطور دریفوس از نقدِ هوش مصنوعی به مثابهٔ یک رشتهٔ دانشگاهی به نقد کلیت فرهنگ و اندیشهٔ غربی پل می‌زند. او متأثر از هیدگر بر آن است که این فرهنگ که از دل سنت فلسفی سر برآورده است با جدایی‌انداختن میان انسان و جهانش و اصالت‌دادن به دانش تئوریک و نیز با فراموشی دانش زیسته، مهارتی، و بدن‌مندِ ما در جهان هرروزه، انسان را به رایانه‌ای دیجیتال فرو می‌کاهد. از این رو برخلاف نگرانی عمومی مبنی بر ظهور روبات‌هایی با هوش و توانایی‌های فراانسانی، هم‌سو با دریفوس باید نگرانِ ظهور انسان‌هایی با هوش ماشینی بود. مدعای سوم نیز این بود که رویکرد واپسین دریفوس در تأیید هوش مصنوعی هیدگری و این ادعا که هوش مصنوعی هیدگری، هستی‌شناسی هیدگری را به آزمونی تجربی می‌گذارد دستِ کم با رویکردهای پیشین دریفوس هم‌خوان نیست.

پی‌نوشت

۱. از نظر دریفوس چنین نگرش محتاطانه‌ای به مدعیات پدیدارشناسانه حتی در کلامِ خود هیدگر نیز قابل ردیابی است. وی پس از بیان نقل قولی از هیدگر می‌گوید: «جملهٔ هیدگر محتاطانه

است. او می‌داند که نمی‌تواند *اثبات کند* که مدل‌های صوری فهم هرروزه باید ... شکست بخورند» (Dreyfus, 1991: 120).

۲. دریفوس در اواسط دههٔ هشتاد و پس از نقدِ نخستش بر هوش مصنوعی صورت‌بندیِ تازه‌ای از آن را این‌بار به صورت ویژه در رابطه با سامانه‌های خبره (expert system) به‌دست داد (Dreyfus, 1987; Dreyfus, 1986). اگرچه می‌توان سامانه‌های خبره را نسلِ جدیدی از سامانه‌های هوشمند به‌شمار آورد اما به صورت کلی در ذیلِ پارادایمِ هوش مصنوعی نمادی می‌گنجد.

۳. کارکردگرایی نگرشی است که در فلسفهٔ ذهنِ معاصر چالش‌های بسیاری را برانگیخته است. دستِ کم دو صورت‌بندی از کارکردگراییَ عرضه شده است: کارکردگرایی ماشینی و کارکردگرایی علّی (متافیزیکی). کارکردگرایی ماشینی ذهن را نوعی ماشین تورینگ به‌شمار می‌آورد و از این جهت با هوش مصنوعی در ارتباط تنگاتنگی است. در سوی مقابل «کارکردگرایان متافیزیکی [علّی] حالت‌های ذهنی را به واسطهٔ نقش‌های علّی‌شان صورت‌بندی می‌کنند؛ به صورت مشخص به واسطهٔ ارتباطات علّی میانِ انگیختارهای حسی، خروجی‌های رفتاری و دیگر حالت‌های ذهنی» (Block, 2007: 28).

٤. از همین تعریف پیداست که مشابهت‌های بسیاری میان کارکردگرایی و محاسبه‌گرایی به مثابهٔ دو پشتوانهٔ هوش مصنوعی به چشم می‌خورد به‌گونه‌ای که بسیاری این دو را هم‌معنا می‌دانند (Chalmers, 1992).

۵. دریفوس یک فرض تماماً علمی به نام فرض زیست‌شناسانه نیز معرفی می‌کند که به موجب آن مغز یک رایانهٔ دیجیتال دانسته می‌شود. پیشرفت علوم شناختی نشان داد که این نگرش به‌هیچ‌وجه با شواهد تجربی سازگار نیست (Clark, 1996: 6; Dennett, 1998: 217).

٦. سرل نیز دو دهه پس از نقد دریفوس بر فرض روان‌شناسانهٔ این دو گزاره در استدلال اطاق چینی نقد وارد می‌کند (Searl, 1980).

۷. دریفوس به‌روشنی این سه فرض را از هم جدا می‌کند و در نقد هر یک نیز استدلال‌هایی می‌آورد. اما زبانِ وی در نقد گاهی آشفته می‌شود و به جهتِ ارتباطِ درونی فرض‌های فلسفی سخنش مغشوشَ می‌نماید. ما در این‌جا کوشیده‌ایم سخنِ دریفوس را در شفاف و سازوارترین صورت‌بندی بازسازی کنیم.

۸. این‌که رایانه و بدنِ انسان هر دو به مثابهٔ چیزهایی فیزیکی از قواعد فیزیک پیروی می‌کنند ممکن است این پندار را تقویت کند که رایانه می‌تواند هم‌چون انسان هوشمند باشد. اما دریفوس باور دارد که صوری‌سازی در فیزیک هیچ ربطی به جهانِ فهمِ هوشمندانهٔ انسانی ندارد. همهٔ ابژه‌های فیزیکی به جهتِ پیروی از قواعدِ فیزیک در شرایط یکسانی قرار دارند و اگر به صرفِ این پیروی بتوان یک ابژه را هوشمند دانست آن‌گاه همهٔ ابژه‌های فیزیکی ازجمله هواپیما و میز نیز باید هوشمند دانسته شوند (Dreyfus, 1979: 107-108).

۹. دلیلِ دیگری که فرضِ معرفت‌شناسانه را با تردید مواجه می‌کند مشکلاتی است که هوش مصنوعی خود در مسیرِ صوری‌سازیِ فهمِ انسانی با آن مواجه شده است؛ یکی از مشکلاتی که دریفوس در این مورد به آن اشاره می‌کند مشکل دیتابیسِ بزرگ (large data base) است. مشکل آن است که رایانه چگونه می‌تواند از میانِ بی‌نهایت واقعیتی که می‌تواند به هر موقعیت مربوط باشد، امرِ واقعاً مربوط را از آن‌چه نامربوط است تفکیک کند. حتی اگر بپذیریم که می‌توان همهٔ دانشِ انسانی را به صورتِ گزاره‌ای صورت‌بندی کرد مشکلِ یافتن مکانیزمی برای بازیابیِ اطلاعاتِ مربوط به هر موقعیت در زمانی معقول مشابه با رفتارِ انسانی کاری بی‌نهایت دشوار به‌نظر می‌رسد. بر مبنایِ پاسخِ دریفوس چنان‌که دیدیم آن‌چه از میانِ بی‌نهایت اطلاعات، اطلاعات مربوط را گزینش می‌کند و معنای آن را معین می‌سازد بستری است که آن اطلاعات در آن داده شده است. «برای فهم یک سخن، ساختن یک مسئله یا شناخت یک الگو، رایانه باید اطلاعاتش را به وسیلهٔ یک بستر، گزینش و تفسیر کند» (ibid: 120). موقعیتِ انسانی این بسترِ گزینش‌گر و متعین‌کنندهٔ واقعیات را برای ما فراهم می‌آورد. «به طور کلی ما فهمی ضمنی از موقعیت انسانی داریم که ما را از واقعیت‌های معین برخوردار می‌کند و آن‌ها را شفاف می‌سازد» (ibid)، اما از آن جهت که رایانه فقط با اطلاعات سروکار دارد درکی از بستر و موقعیتِ انسانی ندارد و از این جهت است که با مشکلِ دیتابیس بزرگ یا به بیان دیگر مشکل فهمِ امرِ مربوط و نامربوط مواجه می‌شود.

۱۰. متأسفانه دریفوس در وام‌گیری از اندیشه‌های فیلسوفانِ گوناگون تساهلِ بسیاری به خرج می‌دهد و به صورتی دقیق و جزئی در امکانِ کنارِ هم نهادنِ این اندیشه‌ها تأمل نمی‌کند. برای نمونه او نگرشِ پدیدارشناسیِ ذات‌انگارِ مرلوپونتی و هیدگر را به‌راحتی در کنارِ ضدِ واقع‌گرایی ویتگنشتاین می‌نهد.

۱۱. اگرچه مک‌کارتی در ۱۹۶۹ به اهمیت اندیشیدنِ فلسفی در هوش مصنوعی پی برده بود اما وی در آن‌جا به‌گونه‌ای ساده‌انگارانه و بدونِ ارجاعِ مستقیم به اندیشه‌های فیلسوفان به پرسش‌های فلسفیِ پیرامونِ هوش مصنوعی پاسخ گفت. همین امر نمایان‌گرِ آن است که پدرانِ هوش مصنوعی درس‌هایشان از فلسفهٔ سنتی را بیش‌تر از طریقِ پس‌زمینهٔ فرهنگی‌شان و نه به صورت مستقیم از خودِ فیلسوفان اخذ کرده بودند.

۱۲. این پرسشی قابلِ تأمل است که آیا شکستِ پروژه‌های هوش مصنوعی در تجربهٔ ساخت، دلیلی برای زیرسؤال‌بردنِ پشتوانهٔ فلسفیِ این پروژه‌ها به‌شمار می‌آید؟ دستِ کم دریفوس چنین نگاهِ تجربه‌گرایانه‌ای دارد و حتی تفسیرش از هیدگر نیز با این نگاه هم‌سو است. به باور وی هیدگر به جهت نبودِ دلیلِ پیشینی در پشتیبانی از آموزه‌هایش پیرامون انسان، با گونه‌ای احتیاط آن‌ها را بازگو می‌کند و می‌کوشد بار استدلال را بر دوشِ منتقد قرار دهد (Dreyfus, 1991: 116-117). از

این رو در نبودِ دلیلِ پیشینی، شکستِ بازسازیِ صوریِ فهمِ انسان در رایانه‌ها در عمل می‌تواند دلیلی در حمایت از آموزه‌های پدیدارشناسانهٔ هیدگر فراهم آورد (ibid: 117).

۱۳. دریفوس در نقدش بر هوش مصنوعی بیش از هر فیلسوفی از هیدگر متأثر بوده است. جایی که دریفوس از نقدِ هوش مصنوعی به نقدِ کلِ فلسفه و اندیشهٔ غربی پل می‌زند نیز به‌نظر می‌رسد همچنان بیش از هر فیلسوفی متأثر از هیدگر و نقدِ تاریخِ متافیزیک است. دریفوس حتی صراحتاً به سخنی از هیدگر اشاره می‌کند که در آن سایبرنتیک (به مثابهٔ پایهٔ نظری هوش مصنوعی) را قلهٔ متافیزیک می‌داند (Dreyfus, 1979: 72).

۱٤. البته نامِ حیات مصنوعی بیش‌تر برای گرایشِ نرم به‌کار می‌رود (Johnston, 2008).

۱۵. دریفوس افزون بر هوش مصنوعیِ هیدگری، جریانِ یادگیریِ تقویتی (reinforcement learning) را نیز مورد ارزیابی قرار می‌دهد. می‌توان این جریان را نیز در ذیلِ پارادایمِ حیات مصنوعی قرار داد. دریفوس به‌صراحت بر این باور است که یادگیری تقویتی از آموزه‌های مرلوپونتی متأثر است (Dreyfus, 2005).

۱٦. دریفوس جریانِ دیگری را نیز که خود را هوش مصنوعیِ هیدگری می‌داند موردِ ارزیابی قرار می‌دهد. جریانی که مایکل ویلر بر پایهٔ خارجیت‌گراییِ (externalism) دیوید چالمرز آن را صورت‌بندی کرده است. اما دریفوس به جهتِ این‌که رویکردِ ویلر و چالمرز نوعی از بازنمودگرایی را می‌پذیرد آن را جریان هوش مصنوعی هیدگری دروغین می‌نامد (Dreyfus, 2007: 12-15).

۱۷. البته ضرورت در بدن‌مندیِ شناخت امری است که مرلوپونتی در پدیدارشناسی از آن دفاع می‌کند. در حالی که هیدگر با تردید از ضرورتِ بدن‌مندیِ دازاین سخن می‌گوید (Dreyfus, 1991: 44-45). با این وجود به جهتِ آن‌که هیدگر الهام‌بخش‌ترینِ فیلسوف برای دریفوس بوده است و مرلوپونتی نیز بسیار وام‌دارِ هیدگر است دریفوس نامِ این جریان را هوش مصنوعی هیدگری می‌نامد.

۱۸. دریفوس دو جریان هوش مصنوعی هیدگری راستینِ دیگر نیز برمی‌شمرد که عبارت‌اند از رفتارگرایی بروک و اینترکشنیسم آگره و چاپمن. رویکرد نخست این ایده را از هیدگر و دریفوس پذیرفته بود که بازنمودِ ما از جهان خودِ جهان ماست (Dreyfus, 2007: 5) و رویکرد دوم نیز درصدد بود مواجههٔ تودستی با اشیا را پیاده‌سازی کند (ibid: 10). به باور دریفوس هر دو رویکرد به جهتِ ابژکتیوسازیِ فهمِ پیشاابژکتیوِ ما از جهان ناکام ماندند (ibid: 12).

۱۹. دریفوس مدعی است از هر شبکهٔ عصبیِ شکل‌گرفته به شبکهٔ عصبیِ بعدی ما با حالاتی گسسته روبه‌رو هستیم. به همین جهت است که منطقاً امکانِ پیاده‌سازیِ آن‌ها در حالاتِ گسستهٔ رایانه‌ای فراهم است (Dreyfus, 2007: 25).

۲۰. دریفوس اگرچه نشانه هایی در سخنِ مرلوپونتی مبنی بر پذیرشِ الگویی عصبی همچون الگوی

فریمن می‌یابد (Dreyfus, 2007) اما به‌صراحت اعلام می‌دارد که پذیرش یک شرح تماماً فیزیولوژیک از سوی مرلوپونتی بسیار بعید به‌نظر می‌رسد (Dreyfus, 2005). اساساً پدیدارشناسی با نقدِ رویکردهای طبیعت‌گرایانه متولد می‌شود و از این رو نه‌تنها مرلوپونتی که هر پدیدارشناسِ دیگری نیز با یک شرحِ تماماً طبیعت‌گرایانه کنار نخواهد آمد. دریفوس اگرچه بسیار گذرا به این مسئله می‌پردازد. اما به‌نظر می‌رسد که با اذعان به این امر که می‌توان چگونگیِ تولیدِ یک فعالیتِ معنادار را فقط در شرحی فیزیولوژیک توضیح داد از پدیدارشناسی به سوی یک نگرشِ طبیعت‌گرایانه گذر می‌کند. او به‌صراحت اعلام می‌دارد که اگر نتوانیم فعالیتِ معنادار را در شرحی طبیعت‌گرایانه توضیح دهیم آن‌گاه باید چنان‌که دنت می‌گوید مغز را بافتی عالی به‌شمار آوریم (Dreyfus, 2007: 19).

منابع

دریفوس هیوبرت و دریفوس استوآرت (۱۳۸٦). «ساختن یک ذهن به جای مدل‌سازی از مغز»، ترجمهٔ محمدرضا طهماسبی، ذهن، ش ۳۲.

Block, N. J (2007). *Consciousness, Function, and Representation*, Collected Papers, Vol. I, the MIT Press.

Chalmers, D. (1992). 'Subsymbolic Computation and the Chinese Room' The *Symbolic and Connectionist Paradigms: Closing the Gap*, J Dinsmore (ed.), Lawrence Erlbaum (available in: http://consc.net/papers/subsymbolic.pdf).

Clark, A. (1996). *Philosophical Foundations in Artificial intelligence*, Margaret A. Boden (ed), Academic Press, INC.

Dennett, D. C. (1969). *Content and Consciousness (International Library of Philosophy & Scientific Method)*, Routledge & Kegan Paul.

Dennett, D. C. (1984). 'Cognitive Wheels: The Frame Problem of AI' In C. Hookway (ed.), *Minds, Machines and Evolution*, Cambridge: Cambridge University Press.

Dennett, D. C. (1986). 'The Logical Geography of Computational Approaches: A View from the East Pole Brand', In M. Harnish, M. Brand (eds.), *The Representation of Knowledge and Belief*, Tucson: University of Arizona Press.

Dennett, D. C. (1998). *Brainchildren Essays on Designing Minds*, The MIT Press.

Dreyfus, H (1967). 'Why computers must have bodies in order to be intelligent', The Review of Metaphysics, Vol. 21, No. 1.

Dreyfus, H (1979). *What Computers Can't Do: A Critique of Artificial Reason*, Harper and Row.

Dreyfus, H (1986). *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*, New York: Free Press.

Dreyfus, H (1987). ‘Socrates to Expert Systems: The Limits of Calculative Rationality’, *Bulletin of the American Academy of Arts and Sciences*, Vol. 40, No. 4.

Dreyfus, H (1991). *Being-in-the-world: A Commentary on Heidegger's Being and Time, División I*, MIT Press.

Dreyfus, H (1992). *What Computers Still Can't Do*, New York: MIT Press.

Dreyfus, H and Dennett, D (1997). Mechanical Bodies, Computational Minds, *A Bradford Book*, The M.I.T. Press.

Dreyfus, H (2005). ‘Merleau-Ponty and Recent Cognitive Science’, in Taylor Carman and Mark B. N (eds.), *The Cambridge Companion to Merleau-Ponty*, Hansen Cambridge University Press.

Dreyfus H (2007). ‘Why Heideggerian AI Failed and how Fixing it would Require making it more Heideggerian’, available in: http://leidlmair.at/doc/WhyHeideggerianAIFailed.pdf.

Hebb, D. O. (1949). *The organization of behavior: a neuropsychological theory*, John Wiley & Sons.

Johnston J, (2008). *The Allure of Machinic Life: Cybernetics, Artificial Life, and the New AI*, MIT Press.

Komosinski, M and Adamatzky, A (2009). *Artificial Life Models in Software*, Springer-Verlag London Limited.

Komosinski, M and Adamatzky, A (2009). *Artificial Life Models in Hardware*, Springer-Verlag London Limited.

McCarthy, J and Hayes P (1969). ‘Some Philosophical Problems From The Standpoint Of Artificial Intelligence’, available in: http://www-formal.stanford.edu/jmc/.

Nagel, E, and James Newman (1957). *Godel’s Proof*, New York University Press (2001).

Newell, A, and Simon, H (1976). ‘Computer Science as Empirical Inquiry: Symbols and Search’, Common ACM, Vol. 19, No. 3.

Putnam, H (1960). ‘Minds and machines’, In Sidney Hook (ed.), *Dimensions of Mind*, New York University Press.

Putnam, H (1975). *Mind, Language and Reality, Philosophical Papers*, Volume 2, Cambridge University Press.

Russell, S and Norvig. P (2003). *Artificial Intelligence: A Modern Approach*, Pearson Education International.

Searle, J. R. (1980). ‘Minds, brains, and programs’, *Behavioral and Brain Sciences*, Vol. 3, No. 3.

Turing, A. M. (1950). ‘Computing machinery and intelligence’, *Mind*, 59.