Biannual Journal of Metaphysical Investigations,
Vol.5, No. 9, Spring & Summer 2024, pp 188-211

Artificial Intelligence and Inference to the Best Explanation: Can a Machine Have the Same Level of Intelligence as Humans?

- *Javad Akbari Takhtameshlou
  jakbrit@gmail.com
Assistant Professor, Philosophy of Science Department. Sharif University of Technology

- Ali Farzan
  Alifarzaan110@gmail.com
MA. Philosophy of Science
Department. Sharif University of Technology

- Received:1/5/2024
- Confirmed: 22/6/2024

Abstract:

Artificial intelligence (AI) is now permeating most aspects of human life, making it necessary to study and assess it from various perspectives, including philosophical ones. A philosophical question in this regard is: Is it possible for AI to reach or even exceed the level of natural human intelligence? Are there any philosophical limitations or obstacles to AI that will basically prevent it from ever reaching such a level? This article aims to address such questions. Based on the article's discussions and results, AI, which has passed two waves so far in its development, in order to enter the third wave and become closer to natural human intelligence, requires the ability to perform a type of context-dependent inference known as "abduction" (also referred to as "inference based on the best explanation"). However, there is basically no clear horizon for the machine regarding this capacity because this type of inference is not formal and computational but rather a creative content-based inference. Indeed, the article comes to the conclusion that while machine intelligence can outperform humans in the purely formal-computational aspects of intelligence, it will never be able to match human intelligence in all aspects as long as it lacks the non-quantitative and non-formal capabilities that humans have, like 'imagination', 'creativity', 'understanding', 'thinking', 'good sense' and, above all, 'consciousness'.

Keywords:
Artificial (Machine) Intelligence, Natural (Human) Intelligence, Inference to the Best Explanation, Understanding, Consciousness, Mind.

Extended Abstract:

Introduction: An important philosophical question concerning artificial or machine intelligence is whether it is essentially possible to develop AI that achieves human-level intelligence, or maybe even beyond? In other words, is it possible for AI to attain every aspect of natural human intelligence (or even higher) during its development? Alan Turing, a pioneer of AI, in response to the question "Can machines think?" believes that if a machine can pass the test of answering a human interrogator's questions in such a way that the interrogator is deceived and confuses it with a human, it should be concluded that the machine is thinking (at a human level). But will passing the Turing test really mean that AI has reached the same level of intelligence as human intelligence? Will the realization of fully human-like responsiveness mean that the machine has (independently) acquired all the mental, epistemic, and inferential abilities that human intelligence has? Regardless of Turing's point and belief, is it essentially possible for a machine, *as a purely physical object*, to achieve true human-level intelligence? What conditions should artificial machine intelligence meet in order to be genuine intelligence? For this purpose, should it also be able to perform complex inferences like "abduction", commonly known as "Inference to the Best Explanation" (IBE)? If so, can a machine essentially afford to perform non-formal inferences such as IBE? In this article, we try to address such philosophical questions regarding artificial intelligence.

Discussion: One of the most innovative and, at the same time, essential views on our problem was recently expressed by computer scientist and philosopher Eric Larson, following his evaluation of the stages of AI development in recent decades. He says that in the first stage or wave of its development, AI employed deduction, and in the second wave, it employed induction. However, for the third wave and the achievement of "Artificial General Intelligence" (AGI), a third kind of inference called 'abduction', which is utilized by humans, is necessary. With this form of inference, we quickly and *creatively* conjecture plausible reasons for what we see. According to Larson, while the ability to perform this conjectural and creative inference is what forms the origin of true intelligence, the realization of AGI faces the fundamental challenge that no one (yet) knows how to program this kind of inference into a machine. The reason is that this type of inference is a 'creative leap' rather than something computational or syntactically derivable from data, and therefore it seems *magical* from an AI standpoint. Therefore, Larson concludes, we are not—at least, not yet—on a path to AGI. In this article, we try to support Larson's view by focusing on the nature and aspects of abductive inference and using more explicit and basic reasoning steps. This support even leads to a stronger conception of his claim being confirmed.

After introducing certain aspects of IBE, we will ask what basic abilities are needed to equip the machine with such an inference. Crucial in this regard is the fact that performing different stages of IBE requires a set of special abilities that are content-based, conceptual, semantic, or qualitative in nature rather than merely formal, syntactic, or quantitative. More explicitly, IBE is only possible for someone or something that possesses abilities such as 'imagination', 'creativity', 'understanding', 'thinking', 'intuition, 'consciousness', 'inspiration', 'good sense', and the like. IBE's reliance on these qualitative abilities leads us to investigate whether it is feasible to realize them in something like a machine as a purely physical entity. We advance this issue by examining their most important example, 'consciousness', which we argue serves as the basis and requirement for most of them. In addition to IBE's reliance on consciousness, this article's broader objective of "evaluating the possibility of AI reaching human-level intelligence" also compels us to examine the possibility of a machine having consciousness. Given the necessity of investigating the possibility of a machine having consciousness, the article advances the debate by presenting the opposing and supporting viewpoints of theorists in the area like John Searle,

David Chalmers, Hubert Dreyfuss, Antonio Damasio, Paul and Patricia Churchland. After presenting the arguments against the conscious machine, it is concluded that due to the special nature of consciousness, a (purely physical and material) machine cannot be given consciousness: giving such an ability to the machine requires special hardware, such as Cartesian mental essence. However, eliminative materialists like Paul and Patricia Churchland are exactly against this; therefore, supporting the above conclusion requires answering their arguments. Therefore, after presenting Churchlands' main line of arguments in defense of the possibility of a thinking and conscious machine, as well as their counterarguments, we critically evaluate their reasoning and conclude that their argumentations are insufficient, primarily reflecting the presumptions of their philosophical stance. Of course, we also agree that if their primary assumption, eliminative materialism, that there is no element or essence except the material essence in this world (or at least regarding the mind), is correct, then they will be right, and our mind is nothing but a physical brain. But we believe that the major issue is with this primary assumption: this assumption is too heavy and strong, lacking the necessary firm basis, to confidently affirm the possibility of a 'conscious machine', or, in Searle's words, 'strong AI'. The main reason that (people like) Churchlands usually provide for eliminative materialism is that, just as many entities in the history of science—crystalline spheres, phlogiston, caloric, ether, etc.—were eventually shown to not exist, so will the mind and mental states—beliefs, desire, pain, joy, intention and etc.—that we now think are definitive and signs of an immaterial essence or entity be shown to not exist. But in our opinion, such an analogical argument is too weak to be able to provide the necessary basis for its claim. We hold that mental internal phenomena such as pain, joy, and intention differ fundamentally from scientific outside-the-mind things such as phlogiston and ether; hence, analogizing them is nothing but a false analogy: our epistemological access to the former is direct, intuitive, and, in the words of the Islamic philosopher, via 'presence', in contrast to the latter, which is indirect, conjectural, and acquired. If these kinds of matters present in our cognitive faculties might turn out to be as illusory and untrue as the second ones, then nearly no claim, including the very eliminative materialism's that the mind is solely physical, will be assertible and affirmable by a human.

Conclusion: The article ultimately concludes that we cannot rule beforehand in favor of the feasibility of a conscious machine with human-level intelligence without presupposing philosophical positions such as '(eliminative) materialism' or at least that 'the human mind is purely physical'. However, these presuppositions are so strong that they're nearly impossible to verify or defend even with relatively firm reasons. Thus, belief in the possibility of strong AI will be more conditional and dependent on one's philosophical stances regarding the nature of the human mind. Of course, as weak AI is the outcome of a purely formal and computational function, these issues have no bearing on its occurrence. Indeed, this type of intelligence is currently in existence, continues to improve, and hits its peak every day.

Resources:

- Chalmers, D. J. (1997), *The Conscious Mind: In Search of a Fundamental Theory*, Oxford Paperbacks.
- Churchland, Paul (2013), *Matter and Consciousness*, 3rd ed., Cambridge (Massachusetts), London: Mit Press.
- Churchland, Patricia (1996), "The Hornswoggle Problem", *Journal of Consciousness Studies*, 3(5-6), 402-408.
- Gleiser, M. (ed.) (2022), *Great Minds Don't Think Alike: Debates on Consciousness, Reality, Intelligence, Faith, Time, AI, Immortality, and the Human*, Columbia University Press.
- Larson, E. J. (2021), *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do,* Harvard University Press.
- Searle, J. R. (1980), "Minds, Brains, and Programs", *Behavioral and Brain Sciences* 3(3), 417-424 and 450-457.
- Turing, A. M. (1950), "Computing Machinery and Intelligence", *Mind* 59, 433-460.

دوفصلنامهٔ علمی پژوهش‌های مابعدالطبیعی،
سال پنجم، شماره ۹، بهار و تابستان ۱۴۰۳،
صص۱۸۸-۲۱۱

هوش مصنوعی و استنتاج بر پایهٔ بهترین تبیین: آیا ماشین می‌تواند در حد ما انسان‌ها هوشمند باشد؟

جواد اکبری تختمشلو(نویسنده مسئول)
jakbarit@sharif.edu
عضو هیئت علمی گروه فلسفهٔ علم، دانشگاه صنعتی شریف.

علی فرزان
alifarzaan110@gmail.com
دانشجوی کارشناسی ارشد فلسفهٔ علم، دانشگاه صنعتی شریف.

- تاریخ دریافت: ۱۴۰۳/۲/۱۲
- تاریخ تایید علمی: ۱۴۰۳/۴/۲

چکیده

هوش مصنوعی یا ماشینی، که امروزه در حال گسترش به تمام ساحت‌های زندگی بشری است، بی‌تردید نیازمند بررسی و ارزیابی از ابعاد گوناگون از جمله بُعد فلسفی است. یکی از سوالات فلسفی در این زمینه این است که آیا هوش مصنوعی می‌تواند تا آنجا پیشرفت نماید که کاملاً به‌حد هوش طبیعی انسانی و یا حتّی فراتر از آن نائل آید؟ آیا در این خصوص برای هوش مصنوعی محدودیّت یا مانعی فلسفی قابل تشخیص است که اساساً آن را برای همیشه از نیل به‌چنین سطحی محروم سازد؟ این مقاله در صدد پرداختن به چنین سوالاتی است. طبق بررسی‌ها و نتایج مقاله، هوش مصنوعی، که تاکنون در مسیر توسعهٔ خود دو موج را پشت سرگذاشته، برای ورود به موج سوم و تقرّب به هوش طبیعیِ انسانی نیازمند توانایی انجام نوعی استنتاجِ وابسته‌به‌زمینه به‌نام "آبداکشن" ("استنتاج بر پایهٔ بهترین تبیین") است؛ درحالی‌که نسبت به چنین توانمندی، به‌دلیل خلّاقانه، محتوایی، غیرصوری و غیرمحاسباتی بودنِ این نوع استنتاج، اساساً افق روشنی برای ماشین متصور نیست. در واقع طبق استدلالات مقاله، هوش ماشینی، گرچه می‌تواند در ابعاد کارکردی خود (یعنی ابعاد صوری-محاسباتیِ محض) هوش انسانی را پشت سر گذاشته و از آن فراتر هم برود، ولی مادام که از توانمندی‌های غیرکمّی و غیرصوری نظیر "تخیّل"، "خلّاقیّت"، "فهم"، "تفکّر"، "شمّ خوب" و در رأس آنها "آگاهی" (که هوش طبیعی انسانی متنعّم از آنهاست) محروم باشد، امکان هوشمندی در تراز هوش انسانی را نخواهدداشت.

واژگان کلیدی: هوش مصنوعی (ماشینی)، هوش طبیعی (انسانی)، استنتاج بر پایهٔ بهترین تبیین، فهم، آگاهی، ذهن.

مقدمه

هوش مصنوعی امروزه یکی از ضروریاتِ زندگی انسان معاصر گشته و به‌سرعت نیز در حال گسترش است. گسترش روزافزون این فناوری در حول‌وحوش آدمیان، گذشته از بهبود زندگی و رفاهی که احیاناً برای بشر به ارمغان می‌آورد، به‌موجب تبعات گوناگونش نیازمند تأمل، تجزیه‌وتحلیل و بررسیِ ابعاد مختلف ازجمله بُعد نظری‌فلسفی است. یکی از سوالات فکری و فلسفی که توجه اندیشمندان را از همان اوان پیدایش ماشین‌های محاسبه‌گر به خود جلب نموده این است که آیا اساساً ایجاد هوش مصنوعی در حد هوش انسانی، یا چه‌بسا فراانسانی، امکان‌پذیر است؟ به‌عبارت دیگر، آیا از یک منظر فلسفی (و قطع نظر از مسائل فنی و عملی) این امکان برای هوش مصنوعی هست که در فرایند توسعهٔ خود بتواند از همه نظر به‌حد هوش حقیقی و طبیعی انسانی (یا حتّی فراتر) برسد؟ آلن تورینگ، از پیشگامان هوش مصنوعی، مقاله‌ای (1950) را حول این سوال تدوین نمود که «آیا ماشین‌ها می‌توانند فکر کنند؟» (Turing,1950:433). او در این مقاله طی یک اقدام مهم فلسفی، آزمون مشخصّی را برای آگاهی از وقوع هوش مصنوعی طراحی و معرفی نمود. طبق این آزمون که امروزه به «تست تورینگ» مشهور است، در اتاقی مجزا یک انسان و در اتاق دیگری یک ماشین (قادر به تولید نوشته‌هایی شبیه به نوشته‌های آدمیان) قرار می‌دهیم. بیرون از این دو اتاق، فرد سومی به‌عنوان بازجو یا داور شروع به ارتباطِ تایپی با حاضرین این اتاق‌ها می‌نماید تا با رد و بدل کردن سوال‌ها و پاسخ‌هایی هویّت هر کدام را حدس بزند. به‌اعتقاد تورینگ اگر ماشین بتواند پاسخ‌ها را به‌تقلید از انسان‌ها و چنان مشابه آن‌ها ارائه کند که بازجو فریب خورده و نتواند هویّت آن را با اطمینان از انسان حاضر در اتاق دیگر تمییز دهد، باید نتیجه بگیریم که ماشین آزمون را با موفقیت پاس نموده است: ماشین مزبور (در حدّ انسان) تفکّر دارد. از نظر عملی نیز تورینگ نه‌تنها هیچ مانعی برای این‌که ماشین‌ها واقعاً همچون انسان‌ها قادر به تفکر باشند نمی‌دید، بلکه اظهار امیدواری می‌کرد که آن‌ها نهایتاً (در آینده‌ای نه‌چندان دور) خواهند توانست در همهٔ زمینه‌های فکری با انسان‌ها به رقابت بپردازند (Ibid.:460).

علی‌رغم صراحت آزمون و باور تورینگ، این سوال جای طرح دارد که آیا پاس‌کردن این آزمون حقیقتاً به‌معنی تحقّق هوش مصنوعی در تراز هوش طبیعیِ انسانی خواهد بود؟ آیا برآورده شدنِ معیار پاسخ‌دهیِ سراسر شبیه به انسان، به‌معنی تحقّق امور ذهنی و فکری همچون درک، فهم، آگاهی و... نیز در ماشین خواهد بود؟ آیا بدین‌ترتیب ماشین همچون انسان (مستقلاً) برخوردار از توانایی ذهنی و معرفتی‌استنتاجیِ لازم مثلاً برای پیشبرد علم و تکنولوژی هم خواهد بود؟ در یک کلام، آیا به‌دنبال پاس شدنِ چنین آزمونی می‌توان ادعا نمود که هوش انسانی به‌طور تمام و کمال محقَّق شده است؟

فارغ از نکتهٔ تورینگ، ما با این سوال بنیادی و بسیار مهم فلسفی مواجهیم که آیا تحقّق هوشمندی حقیقی (در سطح انسانی) اساساً برای ماشین، به‌عنوان یک امر صرفاً فیزیکی میسر است؟ هوش مصنوعی‌ماشینی برای هوشمندی حقیقی باید چه شرایطی را برآورده سازد؟ آیا برخورداری از امور یا حالات ذهنی از قبیل تجربه، آگاهی، فهم، ادراک و... نیز جزء این شرایط است؟ در این صورت آیا بهره‌مندی از این امور برای هویتی صرفاً فیزیکی چون ماشین مقدور است؟ برای هوشمندی در حد حقیقی و طبیعی، قدرت تفکّر و استنتاج ماشین باید در چه حدّی باشد؟ آیا چنین ماشینی باید قادر به انجام استنتاجات پیچیده‌ای چون «اَبداکشن»(abduction) یا «استنتاج بر پایهٔ بهترین تبیین» (اختصاراً IBE) نیز باشد؟ در صورت مثبت بودنِ پاسخ، آیا انجام یک استنتاج غیرصوری چون IBE اساساً از عهدهٔ ماشین ساخته است؟ در این مقاله تلاش می‌کنیم به این گونه سوالات فلسفی در باب هوش مصنوعی بپردازیم.

هوش مصنوعی و استنتاج بر پایهٔ بهترین تبیین

یکی از جدیدترین و در عین حال مهمترین و جالب‌ترین یافته‌ها یا دیدگاه‌ها نسبت به مسئلهٔ مورد نظر این مقاله، دیدگاهی است که اِریک لارسون (2021)، فیلسوف و دانشمند علوم کامپیوتر، طی برّرسی خود از هوش مصنوعی و مراحل توسعهٔ آن در دهه‌های اخیر، به آن دست یافته‌است. چنان‌که ملاحظه خواهد شد، یافتهٔ لارسون که راجع به امکان یا شرایط وقوع

«هوش مصنوعی عمومی» (AGI) است، به‌جهت نو و بدیع بودگی‌اش و خصوصاً این‌که به‌نظر می‌رسد در پاسخ به سوالات پژوهشی فوق‌الذکرِ ما حائز اهمیتِ تعیین‌کننده‌ای است، در این‌جا سزاوار تمرکز و تبیین و حتّی تأیید و تقویت می‌باشد. منظور از هوش مصنوعی عمومی، یک هوش مصنوعی است که بتواند از هر جهت و در انجام هر وظیفه‌ای، خصوصاً حل مسائل جدیدی که نسبت به آن‌ها قبلاً آموزش ندیده، با ظرفیت‌های هوش انسانی برابری نموده یا حتّی فراتر برود. طبق بررسی‌ها و یافته‌های لارسون «حتمیّتِ تحقّق هوش مصنوعی حقیقی» که امروزه آن را با این ذهنیّت مسلَّم می‌گیرند که ما تاکنون در مسیری قدم گذاشته‌ایم که وقوع هوش مصنوعی در سطح انسانی و سپس اَبَرهوش نهایتاً اجتناب‌ناپذیر است، نه یک حقیقت مسلَّم بلکه افسانه است (Larson,2021:1). استدلال وی له این مدعا بر مفهوم «استنتاج» و این نکته مبتنی است که هر مرحله یا موج از توسعهٔ هوش مصنوعی به نوع استنتاج منطقی وابسته است که سامانهٔ هوشمند قادر می‌شود بهره‌برداری نماید.

قبل از بیان استدلال لارسون، لازم است بدانیم که تحولات و توسعهٔ هوش مصنوعی تا به امروز را اکنون به دو موج اصلی تقسیم می‌کنند (Launchbury,2017)[۶]. در موج اول (شامل بازهٔ ۱۹۵۶ الی دههٔ ۱۹۸۰)، رفتار ماشین هوشمند تماماً توسط انسان و به‌نحو قاعده‌محور (به‌شکل «اگر–آن‌گاه») طراحی و برنامه‌نویسی می‌شد. بدین‌ترتیب تمام تصمیمات و عملکردهای ماشین در این مرحله از هوش مصنوعی به‌نحو قطعی و ازپیش‌تعیین‌شده (متناسب با هر حالت یا وضعیت ممکنی که می‌توانست مواجه شود) تعریف می‌شدند. توانمندی ماشین برای انجام بازی شطرنج و رُبات شیکی(shakey) که قادر به انجام فرامین محدود همچون هل دادنِ اشیاء به اطراف، روشن و خاموش‌کردن کلید چراغ‌ها و بازوبسته‌کردن درها بود نمونه‌هایی از همین موج از توسعهٔ هوش مصنوعی بودند. روشن است که در چنین شرایطی هیچ موقعیّت و امکانی برای یادگیری ماشین نسبت به یک رفتار جدید و متفاوت فراهم نبود. ازاین‌رو موج دوّم هوش مصنوعی (از دههٔ ۱۹۸۰ تا به امروز) با رویکرد یادگیری ماشین از طریق روش‌های آماری شکل گرفت. طی این رویکرد، ماشین از روی نمونه‌ها و داده‌ها و ورودی‌های بیشتر یاد می‌گیرد که در موارد بعدی چگونه رفتار نماید. در این مرحله، دیگر لازم نیست همه‌چیز را انسان به‌طور دستی و با زبان قواعد به ماشین دیکته (برنامه‌نویسی) کند، بلکه صرفاً الگوریتم‌های یادگیری به ماشین داده می‌شود تا او خود رفتار و پیش‌بینی‌هایش را بهبود بخشد. در این مورد نیز می‌توان به‌عنوان نمونه به دستیارهای صوتی چون سیری(SIRI) (شرکت اَپل) و دستیار گوگل اشاره نمود. در موج سوّم (به‌عنوان رویداد آتی)، ماشین هوشمند می‌بایست قادر به «تطابق با زمینه» گردد. در این موج قرار است قدرت انتزاع و استنتاج ماشین افزایش یابد. به‌نحوی که قادر باشد تصمیم‌های خود را تبیین نماید و برای تصمیمات خود مدل‌های تبیین‌کنندهٔ مبتنی بر پدیده‌های دنیای واقعی ارائه کند.

حال سخن لارسون در این میان این است که تکیه‌گاه منطقی و استنتاجی در مرحلهٔ نخست از توسعهٔ هوش مصنوعی استنتاج قیاسی و در مرحلهٔ دوم استنتاج استقرائی بوده است. اما به‌اعتقاد وی، برای دستیابی به هوش مصنوعی عمومی (به‌عنوان آرمان موج سوم یا به‌هرحال موج‌های بعدی) به نوع سوّمی از استنتاج که ما انسان‌ها از آن بهره‌مندیم، به‌نام «اَبداکشن (استنتاج ربایشی)» یا آن‌گونه که با مقالهٔ ۱۹۶۵ گیلبرت هارمن به «استنتاج بر پایهٔ بهترین تبیین» مشهور شده نیاز است. به‌کمک این نوع استنتاج، ما برای پدیده‌هایی که مواجه می‌شویم به‌سرعت و به‌نحوی خلّاقانه (البته با سرنخ قرار دادنِ فکت‌های مشاهده شدهٔ مرتبط) دلایل قابل قبولی را حدس می‌زنیم، و در صورتِ نیاز این حدس‌ها را به‌سرعت به‌روزرسانی می‌کنیم (Larson,2021:163). طبق دیدگاه لارسون، گرچه این حدس‌ها بعد از تولید توسط عامل هوشمند چه‌بسا به‌کمک قیاس (استخراج نتایج و استلزامات منطقی آن‌ها) و استقراء (مقایسهٔ نتایج و استلزامات مزبور با تجربه) اعتبارسنجی شوند، ولی به‌هرحال توانایی بر انجام این نوع استنتاجِ حدسی و خلّاقانه چیزی است که اساساً مبدأ و

---

۶. مرجع و منشاء اصلی این تقسیم‌بندی، یعنی همان (Launchbury,2017) سند یا درواقع ویدیویی است که از سوی «آژانس پروژه‌های تحقیقاتی پیشرفته دفاعی ایالات متحده» (اختصاراً مشهور به DARPA) منتشر شده است.

منشأ هوش به‌معنای واقعی را شکل می‌دهد. به‌عبارت دیگر، از نظر لارسون فرایند توسعهٔ هوش مصنوعی با این چالش اساسی و مبنایی مواجه است که ما مادامی که هوش مصنوعی را به استدلال ربایشی مجهز ننموده باشیم، اساساً در تشبّه و تقرّب به هوش انسانی قدمی برنداشته و همچنان با آرمان هوش مصنوعی (رسیدن به AGI) فاصله خواهیم داشت. این در حالی است که تاکنون قدم جدّی‌ای در این راستا برداشته نشده است:

> سه نوع استنتاج وجود دارد. هوش مصنوعی کلاسیک یکی از آن سه (قیاس) را مورد توجه قرار داد. توجه هوش مصنوعی مدرن به نوع دیگر آن (استقراء) است. این نوع سوم (اَبداکشن) است که به هوش عمومی منجر می‌شود و در کمال تعجب، هیچ‌کس روی آن اصلاً کار نمی‌کند. بالاخره می‌دانیم که چون هریک از این استنتاج‌ها نوع مجزّایی هستند ... شکست در ساخت سامانه‌های هوش مصنوعیِ استفاده‌کننده از این نوع استنتاجِ قوام‌بخش به هوش مصنوعی عمومی، به‌معنی شکست در نزدیک شدن به هوش مصنوعی عمومی یا AGI خواهد بود (Ibid.:4).

گرچه قیاس اثبات می‌کند که فلان چیز باید برقرار باشد و استقراء نشان می‌دهد که فلان چیز عملاً در جریان است و اَبداکشن صرفاً پیشنهاد می‌دهد که فلان چیز می‌تواند برقرار بوده باشد، اما درواقع «همین اَبداکشن است که موجب تفکّر در موقعیت‌های دنیای واقعی می‌شود» (ibid.:166).

بدین‌ترتیب، برای ایجاد هوش مصنوعی حقیقی (هم‌سطح هوش انسانی)، نه استنتاج قیاسی کفایت می‌کند و نه استنتاج استقرائی. درمورد قیاس، مشکل این است که این نوع استنتاج معمولاً در حال تولید یک‌سری حقایق و صدق‌های غیروابسته به زمینه و زمان است. به‌نحوی که قیاس با حقایقِ بی‌زمان (هرچند یقینی) خود چیز اندکی از جهان روزمرگی را می‌تواند به‌چنگ اندازد و خصوصاً ممکن است از ملاحظات ربط داشتگی غفلت ورزد. به‌عبارت دیگر، چون قیاس نوعی استنتاج صوری است، لزوماً به محتوا و ربط‌داشتگیِ محتواییِ نتیجه و مقدمات حساس نخواهد بود. به‌نحوی که «[نتیجه‌گیری] هوش مصنوعیِ تقویت‌شده با قیاس می‌تواند کاملاً مضحک و احمقانه باشد: مثلاً ممکن است نتیجه بگیرد که فلان شوهر چون قرص‌های ضدبارداری همسرش را مصرف کرده، باردار نخواهد شد» (ibid.:189). به‌طورکلی، قیاس چون معمولاً یک‌سری حقایق لازمان و خشک و نسبتاً کلی به بار می‌آورد و صرفاً به‌دنبال انتقال صدقِ موجود در مقدماتِ عیناً به نتیجه و یقینی بودنِ همه‌چیز است و لذا هیچ فضایی را برای جرح‌وتعدیل احیاناً مورد نیاز بعدی باقی نمی‌گذارد، نمی‌تواند برای استنتاج حقایق جزئی و وابسته به زمینه و شرایط خاص موقعیت‌های جهان روزمرگی مناسب باشد. این در حالی است که استنتاج این نوع حقایق، از ضروریاتِ هوشمندی واقعی است.

از نظر لارسون استنتاج استقرائی نیز بدین جهت برای ایجاد هوش عمومی نامناسب است که:

> استنتاج استقرایی به ما دانش موقّت می‌دهد؛ چرا که آینده می‌تواند شبیه گذشته نباشد (اغلب شبیه نیست). ... همچنین از این نیز رنج می‌برد که قادر نیست استنتاج‌های شناخت‌محوری را در بر بگیرد که لازمهٔ هوشمندی است؛ چرا که تکیه‌اش ناگزیر به [صرف] داده‌ها و فراوانیِ پدیده‌ها در داده‌هاست. ... نظام‌های استقرائی همچنین شکننده و فاقد استحکام هستند و به صرفِ داده‌ها به فهمی حقیقی نائل نمی‌آیند (Larson, 2021:189).

منظور لارسون این است که وقتی در استنتاج استقرائی ناگزیر بر صرف تواتر و فراوانی داده‌های مشابه تجربی تکیه می‌کنیم، همواره این خطر در میان خواهد بود که موارد آتی مطابق تواتر گذشته نباشند، کما این‌که بوقلمون راسل که روز

قبل از کریسمس بر پایهٔ فراوانی داده‌های گذشته و در اوج اعتماد به صاحب خود نتیجه‌گیری می‌کرد که همواره توسط وی تغذیه و مراقبت خواهد شد، سرش در روز کریسمس توسط همو بریده شد (ibid.:173)[۱۰]. از طرف دیگر، محدود بودن استقراء به داده‌های تجربی نیز خود محدودیّت بزرگ دیگری است که این نوع استنتاج را برای ایجاد هوش مصنوعی عمومی نامناسب می‌سازد. کما این‌که بوقلمون مورد بحث نیز چون تا قبل از کریسمس این دادهٔ تجربی را که «من در حال قطعه‌قطعه شدن هستم» نداشت، به‌عنوان استقراءگرای محض هرگز قادر به تصحیح دیدگاه خود نبود. او درواقع برای حدس زدنِ دیدگاه صحیح‌تر به استنتاجی متفاوت از استقراء که بدون چنین داده‌ای کار کند نیاز داشت. حال «در یادگیری ماشین این بدین معنی خواهد بود که تنها شناختِ قابل دستیابی برای یک سامانه [تقویت‌شده با استقراء]، شناختی است که به یک شکل صرفاً نحوی قابل بازیابی از داده‌ها باشد» (ibid.:173). ولی دراین‌صورت، ماشین دارای هیچ شناختی جز آنچه که مستقیماً در داده‌ها مشاهده می‌نماید نخواهد بود. درواقع طبق نظر لارسون درحالی‌که لازمه و نشانهٔ هوشمندی (حقیقی) برخورداری از شناختی است که اساساً در داده‌های تجربی، موجود و قابل‌مشاهده نیست، این نوع شناخت صرفاً از طریق استنتاج ربایشی قابل‌حصول است. قابلیتی که در ماشین ملاحظه نمی‌گردد:

> تفکّر هوشمند متضمّن شناختی است که از آنچه ما می‌توانیم صریحاً مشاهده کنیم فراتر می‌رود. لذا این یک راز است که ما [انسان‌ها] چگونه به چنین شناختی دست می‌یابیم و حتّی بالاتر از آن، چگونه این شناختِ صحیح را در زمانی مناسب و برای [حل] مشکلی به‌کار می‌بندیم. نه قیاس و نه استقراء، هیچ‌کدام قادر به گره‌گشایی از این راز محوری هوش انسانی نیستند. [تنها] استدلال ربایشی که پِرس مدت‌ها قبل آن را پیشنهاد کرده قادر است از عهدهٔ این کار برآید، اما ما نمی‌دانیم چگونه باید آن را برنامه‌نویسی کنیم (ibid.:189-190).

لارسون می‌گوید مثلاً برای موقعیّت ساده‌ای چون سفارش غذا در دنیای واقعی نمی‌توان تمام دانش مرتبط و مورد نیاز را از قبل پیش‌بینی و به سامانه داد تا در تمام حالات ممکن بداند چه‌کار باید بکند. زیرا اگر هم درمورد مراحل عادی و روتینِ این ماجرا بتوان به‌صورت یک نمایشنامه به سامانهٔ هوشمند گفت که چه‌کار کند، مواردِ فراوانِ غیرمنتظره و پیش‌بینی‌نشده‌ای (از قبیل لزوم انتظار طولانی برای دریافت سرویس، برخورد به یک تابلویی که از مشتریان می‌خواهد امروز برای سفارش غذا از درب کناری استفاده کنند، یک گپ و گفت‌وگوی ساده ولی متفاوت با پیشخدمت و...) خواهند بود که موجب اختلال در برنامه و خروج از دانش نمایشنامه‌ایِ داده شده به سامانه و لذا ناتوانی سامانه در تشخیص اقدام مناسب در مقابل آن‌ها خواهند شد. برای انجام این خروج و رفع ناتوانی مزبور، سامانه «باید واجد مکانیسم استنتاجی باشد که بداند چه اتفاقی در حال رخ دادن است. باید کجا را و برای چه مورد توجه قرار دهد» (Larson,2021:182). ما انسان‌ها در چنین وضعیت‌هایی رفتار مناسب و شایستهٔ انجام را با تکیه بر دانش پیش‌زمینه‌ای و فهم متعارف و به کمکِ متد استنتاجی ربایشی و به یک نحو ابداعی و خلّاقانه استنباط می‌کنیم.

درواقع و در یک نگاه کلی و مبنایی‌تر، ما در هوش طبیعی با مشاهدهٔ معلول درصددِ استنباطِ علّت برمی‌آئیم (مثلاً خیس بودنِ خیابان را مشاهده می‌کنیم و سعی می‌کنیم چرایی و چگونگی آن را استنباط کنیم). این را می‌توان به‌تعبیر برخی نظریه‌پردازان به‌دست‌آوردنِ «تصویر بامعنی از جهان» نامید. در چنین تصویری دانش به این‌که چه چیز علّتِ چه چیزی

---

[۱۰]. به‌نظر می‌آید منظور لارسون از مسئله‌داریِ استقراء باید نه مسئلهٔ عمومی و قدیمی (مسئلهٔ هیوم مبنی‌بر این‌که همواره این امکان فلسفی در میان است که طبیعت مسیر خود را تغییر داده و طبق الگوهای گذشته عمل ننماید) بلکه بیشتر مسئلهٔ جدید استقراء (مسئلهٔ گودمن) بوده باشد مبنی‌بر این‌که تواترها یا فراوانی‌های *تصادفی* هیچ تفاوت صوری و نحوی با فراوانی‌های *قانون‌وار* ندارند و لذا آن‌ها به‌صرف صورت از هم قابل تفکیک نیستند. درواقع با در نظرگرفتنِ زمینهٔ بحث (هوش انسانی به‌عنوان هوش معیار) و هدفی که او از اشاره به کاستی‌های استقراء (*نسبت به آبداکشن*) دنبال می‌کند، صرفاً همین می‌تواند قابل قبول و موجه باشد.

است از اهمیت محوری برخوردار است. اما «به‌دست‌آوردن و استفاده از این دانش امر پیچیده‌ای است؛ زیرا اغلب رویدادهای دنیای واقعی عللِ ممکنِ بسیاری را مجاز می‌شمارند. مشکل گزینش در این است که علّت موثّر یا بهترین علّت یا علّت قابل قبول را از میان همهٔ ممکنات واقعی یا تخیّلی بیابیم» (ibid.:183). برای حل مشکل گزینش از میان علل یا عوامل رقیب، باید به‌نحوی درک و استنباط کنیم که در هر موقعیّتی چه چیزی ذی‌ربط است. اما مسئله این است که کسی نمی‌داند این کار را چگونه باید صورت داد. این درک و استنباط نوعی درک و استنباطِ محاسباتی و به‌شکل نحوی قابل اخذ از داده‌ها نیست تا قابل برنامه‌نویسی برای ماشین باشد، بلکه چیزی از جنس «جهش خلّاقانه»(creative leap) است. در حقیقت «استنتاج‌های واقعیِ ما اغلب نه قیاس یا استقراء بلکه حدس‌هایی هستند که ذی‌ربط یا قابل‌قبول در نظر گرفته می‌شوند. به‌همین خاطر است که آن‌ها از منظر هوش مصنوعی، به‌نظر سحرآمیز می‌آیند» (ibid.:183). با آن‌که استنتاج‌های اَسرارآمیز رایشی فرهنگ بشری را فراگرفته‌اند و «آن‌ها عمدتاً همان چیزی هستند که ما را انسان کرده‌اند» (ibid.:188)، ما فعلاً درمورد مکانیسم اخذ نتیجه در آن‌ها چیز زیادی نمی‌دانیم. ما فقط این را می‌دانیم که «به شگردی برای اجرایی نمودن استنتاج رایشی که مستلزم انبارهای عظیمی از دانش فهم متعارف است نیازمندیم. ما هنوز نمی‌دانیم چنین دانشی را چگونه برای ماشین‌ها فراهم آوریم و حتّی اگر روزی این را بفهمیم، نمی‌دانیم چگونه یک موتور استنتاج رایشی را برای استفادهٔ بلادرنگ از کل این دانش در دنیای واقعی راه بیاندازیم» (ibid.:186). لارسون نهایتاً نتیجه می‌گیرد که «پس، علی‌رغم جار زدن‌های اخیر و ادعای متفاوتی که می‌کنند، ما –حداقل فعلاً– در مسیر منتهی به هوش مصنوعی عمومی قرار نداریم. ما هنوز در جستجوی یک نظریهٔ بنیادی هستیم» (ibid.:190).

چنان‌که پیداست یافته یا ادعای لارسون دو مؤلّفه دارد: اوّل این‌که هوشمندی حقیقی (هوش عمومی) لازمه‌اش برخورداری از توانمندی استنتاج بر پایهٔ بهترین تبیین (بدان‌گونه که انسان‌ها قادر به آن هستند) است؛ دوّم این‌که برای تحقّق این توانمندی در ماشین اُفق روشن و امیدبخشی وجود ندارد. نکات لارسون درخصوص مؤلّفهٔ اول به‌نظر مکفی و بی‌نیاز از بحث و استدلال اضافی می‌آید. زیرا قابل انکار نیست که بخش قابل توجهی از تفکّر و چاره‌اندیشی و استنتاجاتِ (در یک کلام، هوشمندی) ما انسان‌ها، به‌عنوان هوش معیار مورد نظر در این مقاله، در مواجهه با موقعیت‌های گوناگون واقعی در زندگی روزمره و همچنین در علم معمولاً بر همین نوع استنتاج مبتنی است. اما مؤلّفهٔ دوم برخلاف اولی، ادعای بزرگ و تعیین‌کننده‌ای را حول موضوع این مقاله مطرح می‌کند. به‌نحوی که ارزیابی آن به بحث و بررسی مفصل‌تر و تصدیق آن به گام‌های استدلالی مشخص‌تر و مبنایی‌تری نیاز دارد. امر مهم و اساسی که برای این منظور نیاز است آشنایی عمقی با ماهیّت و ابعاد استنتاجِ رایشی است. ازاین‌رو ذیلاً سعی می‌کنیم ماهیّت این نوع استنتاج را هرچند به‌فراخورِ چنین مجالی، به‌نحو صرفاً اجمالی معرفی کنیم تا سپس به بررسی این موضوع بپردازیم که تجهیز ماشین به چنین استنتاجی مستلزم چه نوع توانمندی‌های مبنایی‌تری است. چنین بررسی‌ای، چنان‌که ملاحظه خواهد شد، گام‌های استدلالی مورد نیازِ فوق‌الذکر را برای ما فراهم آورده و در نتیجه ما را به تصدیق (حتی برداشت قوی‌تری از) ادعای مورد بحث دعوت خواهد نمود.

استنتاج رایشی و ملزومات آن

اصطلاح «اَبداکشن» (استنتاج رایشی) اولین‌بار توسط چارلز پِرس برای معرفی نوع مستقل و سومی از استنتاج، در کنار دو استدلال منطقی مشهور (قیاس و استقراء) استعمال گردید. طبق نظر پِرس «اَبداکشن عبارت است از مطالعهٔ فکت‌ها و طراحی نظریه‌ای برای تبیین آن‌ها» (Peirce,1998:205). به‌عبارت دیگر، طی اَبداکشن یا آن‌گونه که او گاهی آن را «فرضیه‌سازی» می‌نامد، برای پدیده‌های مشاهده‌شده یک فرضیهٔ تبیین‌کننده می‌سازیم. به گفتهٔ پِرس «تمام ایده‌های علم از طریق اَبداکشن است که وارد علم می‌شوند» (ibid.). درواقع «اَبداکشن تنها فرآیندی است که از طریق آن می‌توان عنصر جدیدی را وارد اندیشه ساخت» (ibid.:224). پِرس در معرفی این نوع استنتاج می‌گوید:

> فرض کنید من وارد اتاقی شده و در آن‌جا تعدادی کیسهٔ حاوی انواع مختلف لوبیا می‌یابم. روی میز یک مشت لوبیای سفید دیده می‌شود. پس از کمی برّرسی متوجه می‌شوم که یکی از کیسه‌ها فقط حاوی لوبیا سفید است. من بلافاصله به‌منزلهٔ یک احتمال یا یک حدسِ قابل‌قبول استنباط می‌کنم که این مشت از آن کیسه بیرون آورده شده است. به این نوع استنتاج، فرضیه‌سازی می‌گویند (Peirce,1992:188).

همچنان‌که قبلاً نیز اشاره شد، بعدها هارمن برای این نوع استنتاج عنوان گویاتر «استنتاج بر پایهٔ بهترین تبیین» را برگزیر. هارمن در توضیح ماهیّت آن می‌گوید اگر بتوانیم از میان تبیین‌های ارائه‌شده بهترین تبیین را برای توضیح یک پدیده برگزینیم، این استنتاج تضمین‌کنندهٔ صدق آن خواهد بود:

> در این استنتاج از این حقیقت که فرضیه‌ای خاص شواهد را توضیح می‌دهد، صدق آن فرضیه استنتاج می‌گردد. در کل چندین فرضیه وجود خواهند داشت که ممکن است شواهد را توضیح دهند. لذا فرد قبل از مجاز شدن به اقامهٔ این استنتاج باید بتواند همهٔ فرضیه‌های بدیل را رد کند. بنابراین، از این مقدمه که فلان فرضیهٔ خاص نسبت به هر فرضیهٔ دیگری تبیین «بهتری» را برای شواهد ارائه می‌کند، نتیجه گرفته می‌شود که فرضیهٔ مزبور صادق است (Harman,1965:89).

هارمن در ادامه به این نکتهٔ مهم اشاره می‌کند که IBE را نمی‌توان به استقراء شمارشی فروکاست؛ زیرا برخی از استنتاج-های مربوط ما چنان هستند که آن‌ها را نمی‌توان ذاتاً جز امری بی‌سابقه و یک حدس ابداعی به‌شمار آورد: «[مثلاً] استنتاج از داده‌های تجربی به‌نظریهٔ ذرات زیراَتمی قطعاً قابل‌توصیف به‌عنوان نمونه‌ای از استقراء شمارشی به‌نظر نمی‌آید» (ibid.:90). بنابراین این نوع استنتاج را که اساساً بر پایهٔ ملاحظات تبیینی (و نه مثلاً ملاحظات تواتری و آماری) قوام می‌یابد، می‌بایست نوعی استنتاج مستقل در نظر گرفت.
نهایتاً لازم به ذکر و توجه است که گذشته از مرحلهٔ تولید حدس‌های تبیین‌گر ممکن، تشخیص و گزینشِ «بهترین» هم یک مرحلهٔ چالشی در IBE است: این‌که مراد از بهترین تبیین چیست و طبق چه معیاری می‌توان بهترین را از میان تبیین-های گوناگون تشخیص داد. صاحب‌نظران معمولاً معیارهایی را برای این منظور ارائه می‌دهند. مثلاً خود هارمن در این باره می‌گوید: «... البته این مشکل وجود دارد که چگونه باید قضاوت کرد که یک فرضیه به‌اَندازهٔ کافی بهتر از فرضیهٔ دیگر است. احتمالاً چنین قضاوتی بر ملاحظاتی از این دست مبتنی خواهد بود که کدام فرضیه ساده‌تر است، کدام قابل‌قبول‌تر است، کدام بیشتر توضیح می‌دهد، کدام کمتر ارتجالی است و امثالهم» (ibid.:89). یا چاکراوارتی معتقد است که تبیین-های مختلف را می‌توان با ملاک‌های نظری از قبیل سادگی، سازگاری درونی، هماهنگی با دیگر دانسته‌ها و تبیین‌ها، میزان وحدت‌بخشی به پدیده‌های گوناگون و... امتیازدهی و رتبه‌بندی کرد و درنهایت بهترین آن‌ها را انتخاب نمود (Chakravartty,2017:35). علی‌رغم چنین اظهاراتی، به‌نظر می‌رسد با توجه به ماهیّتِ عمدتاً غیرکمّی و غیرالگوریتمی اغلبِ چنین معیارهایی، باید حداقل از منظر موضوع و مباحثِ این مقاله در نظر داشت که خلاقیّت و اَسرارآمیزبودگی که لارسون درخصوص اَبداکشن تأکید می‌کند، شامل همین مرحلهٔ انتخاب بهترین تبیین نیز می‌شود: درمورد این مرحله نیز صریحاً و دقیقاً پیدا نیست که ما انسان‌ها آن را چگونه طی می‌کنیم.
به‌دنبال آشنایی با IBE، اینک باید به این سوال مهم بپردازیم که «آیا هوش ماشینی اساساً قادر به انجام این نوع استنتاج است؟» به‌اعتقاد ما، در پاسخ به این سوال نکتهٔ بسیار مهمی که در کنار تأمّل در تمام نکات و مطالب فوق تعیین‌کننده و روشنی‌بخش می‌باشد، این حقیقت است که اجرای مراحل مختلف IBE از جمله ارائهٔ تبیین‌های بدیلِ ممکن خصوصاً از نوع نوآورانه و خلّاقانه و تشخیص و انتخاب بهترینِ آن‌ها به‌نحوی که از بالاترین احتمالِ صدق برخوردار باشد، به یک‌سری

توانایی‌های ویژه‌ای متّکی است که ماهیتاً نه امور صوری و نحوی و کمّیِ صرف، بلکه یک‌سری توانایی‌های محتوایی، مفهومی، معنایی و در یک کلام کیفی هستند. به‌عبارت صریح‌تر، انجام IBE صرفاً برای کسی یا چیزی میسّر است که از توانایی‌هایی چون «تخیّل»، «خلّاقیّت»، «تفکّر»، «فهمیدن»، «شعور»، «آگاهی»، «حدس‌زنی صائب»، «الهام‌گیری»، «جهش خلّاق» یا به‌تعبیری «شمّ خوب» و امثالهم برخوردار باشد. ریشهٔ نیازمندی این استنتاج به امور کیفی مزبور هم درواقع به ماهیّت و ساختار خود آن، خصوصاً این جنبه از آن برمی‌گردد که تبیین پدیده‌ها را ناگزیر با به میان کشیدنِ امور کاملاً جدید و ابتکاری و فراتر از مشاهدات و تجارب گذشته صورت می‌بخشد. به‌نحوی که مثلاً در فرایندهای علمی، دانشمندان برای توضیح پدیده‌های مشاهده‌شدهٔ خود به هویات و مکانیسم‌های بی‌سابقهٔ اساساً مشاهده‌ناپذیر و نظری متوسل می‌گردند. به‌موجب چنین ابعادی استنتاج بر پایهٔ بهترین تبیین، برخلاف قیاس و استقراء، واجد یک ماهیتِ عمدتاً محتوایی و مفهومی و کیفی است تا صوری و نحوی و محاسباتی. از همین رو مراحل مختلف IBE، از قبیل تولید تبیین‌های ممکن، سنجش بهتریِ آن‌ها (حتی قبل از آن، لحاظ نمودن معیارهای بهتربودگی) و انتقال از مقدمات به نتیجه، قابل‌اجرا در قالب یک الگوی صوری و مشخص نبوده و در هر مورد عملی نوآورانه و خلاقانه و البته وابسته به شرایط مفهومی و محتواییِ عناصر ذی‌ربط از قبیل پدیدهٔ تبیین‌خواه و زمینهٔ آن، طیف تبیین‌های ممکن، بهترین تبیین قابل‌انتخاب، دانش پیش‌زمینه و... می‌باشد. در استنتاج تبیینی، اگر از سوی اقامه‌کننده نسبت به مثلاً محتوای مقدمات و نتیجه و سپس اتّصال مفهومی آن‌ها درک و فهم و شهود در میان نباشد و به تعبیر لورنس بونژور پیوندِ میان نتیجه و مقدماتِ استدلال و صدق احتمالیِ نتیجه در پرتوی صدق مقدمات به‌نحو فکری و شهودی دیده و دریافته نشود (Bonjour,1998:203)، اساساً استنتاجی شکل نمی‌گیرد. یا اگر فهم پدیدهٔ تبیین‌خواه، فهم دانش پیش‌زمینه و سپس تخیّل و خلّاقیّت (معطوف به‌صدق) در میان نباشد، چه‌بسا تبیین مناسب (احتمالاً‌صادق) هرگز وارد گزینه‌های پیشنهادی تحت برّرسی نگردد. مثلاً، به‌عنوان یک مورد واقعی‌تاریخی، معروف است که اگوست ککوله که مدّت‌ها به‌دنبال ساختار ملکول بَنزن بود، ساختار حلقوی بنزن را خلاقانه از رؤیایی الهام گرفت که طی آن ماری را دید که دُم خود را گاز گرفته‌است. تاریخ علم نشان می‌دهد که کشفیّات و پیشرفت علمی معمولاً با چنین فهم و خلاقیّتِ متعاقب یا به‌هرحال با توانمندی محتوایی و غیرصوریِ مشابه رقم می‌خورد. اکتشافات معمولاً نه با برقراری ارتباط با دانش و تجربیات گذشته، بلکه عمدتاً به‌واسطهٔ ارائهٔ تبیین‌هایی رقم می‌خورند که به‌نحو بی‌سابقه و برای اولین بار به فکرها خطور می‌کنند و حتّی نحوهٔ خطورشان به‌ذهن نیز بعضاً قابلیّت بازگویی زبانی ندارد.

اما در این صورت آیا چنین عمل بدیع و خلاقانه‌ای[۱۵] از ماشین‌هایی که لااقل فعلاً مجهز به توانمندی‌های کیفی فوق‌الذکر نیستند قابل صدور است؟ برای تجسّم وضعیّت، مثلاً به این بیندیشید که وقتی آن‌ها قرار باشد تبیین یا تبیین‌هایی را پیش بکشند که در انبار داده‌هایشان سابقه ندارد، با تکیه بر چه چیزی و با چه مکانیسمی باید عمل نمایند؟ می‌دانیم که اکنون مثلاً ربات هوش مصنوعی ChatGPT برای پاسخ‌دهی به سوالات صرفاً به‌دنبال داده‌هایی می‌گردد که از قبل در آن لحاظ شده‌اند. روشن است که این امر برای دست به ابتکار عمل زدن و ارائهٔ تبیینی ابداعی که برای پدیده یا موقعیت تبیین‌خواه بهترین نیز بوده باشد کافی نخواهد بود: اگر قرار باشد هوش مصنوعی صرفاً براساس داده‌هایی که بشر از قبل به او خورانده به پاسخ برسد، جز این نخواهد بود که در محدودهٔ بسته‌ای از پاسخ‌ها به دام خواهد اُفتاد. ولی ماشین با تکیه بر چه چیزی از نزد خود می‌تواند تبیین‌های به‌کلی جدیدِ احتمالاً صادق ارائه نماید؟ آیا ماشین می‌تواند ابتدا تمام یا اغلب حالات و تبیین‌های بدیلِ ممکن را با تکیه‌بر قوهٔ تخیّل یا خلّاقیّتِ خود حدس بزند، احتمالاً کاذب‌ها را شناسایی و حذف نماید و سپس در میان باقی‌مانده‌ها بهترین، یعنی محتمل‌ترین از نظر صدق را با تکیه‌بر مثلاً شمّ خوب (یا هرچیزی که در هوش طبیعیِ انسانی مسئول این مرحله است) تشخیص داده و نهایتاً آن را گزینش و ارائه نماید؟

---

[۱۵]. لازم به تأکید بر این نکتهٔ مهم نیست که عمل بدیع و خلاقانهٔ مورد بحث در IBE به‌هیچ وجه یک عمل آزادانه و دلبخواهانه نیست که از هر عاملی ساخته باشد، بلکه اساساً عملی مقیّد و معطوف و رهنمون به «صدق» است.

حال اگر مراحل مختلف استنتاج تبیینی (حدس‌زنی گزینه‌های متنوع ممکن، سبک‌وسنگین‌کردن آن‌ها از جهت توان تبیین‌گری‌شان نسبت به پدیدهٔ نیازمندِ تبیین، تشخیص و انتخاب بهترین آن‌ها با نظر به صدق و...) بدون توانایی‌های کیفی‌محتوایی چون تخیّل، خلاقیّت، الهام‌گیری، درک و فهم، شعور، آگاهی، تفکّر، شهود و امثالهم قابل اجرا نباشد، پس پاسخ سوالات مطروحه‌مان درخصوص هوش مصنوعی در گرو این خواهد بود که آیا تحقق چنین توانمندی‌هایی در چیزی چون ماشین به‌عنوان یک هویّت فیزیکی محض امکان‌پذیر است یا خیر؟ روشن است که ارزیابیِ خود این امر مستلزم تحلیل دو طرف ماجراست. ولی باید اعتراف نمود که تحلیل طرف اول، یعنی توانمندی‌های متنوّع فوق به‌عنوان یک‌سری امور کیفی‌مفهومی، کار ساده و آسانی، خصوصاً در این مجال نیست. بااین‌حال، به‌نظر می‌رسد این موضوع را بتوان بدون پرداختن به همگی این توانمندی‌ها و صرفاً با بررسی مهمترین نمونهٔ آن‌ها یعنی «آگاهی» پی گرفت. زیرا آگاهی گذشته از این‌که خود یکی از این توانمندی‌هاست، در میان بقیه از جایگاه مهم‌تر و تعیین‌کننده‌تری برخوردار است. به‌نحوی که گوئی بقیه شأنی از شئونات آگاهی یا حداقل آگاهی مبنا و از جمله ملزومات همگی است. در پشتیبانی از این ادعا و همچنین اشاره به ریشهٔ نیازمندی IBE به عنصر آگاهی، بایستی توجه خواننده را به این نکتهٔ کلیدی جلب نمائیم که «آگاهی» در این بستر و در معنای عمیق خود، چنان‌که طی مباحث آتی خواهیم دید، نه صرف اطلاع داشتن از محیط پیرامون، بلکه «یک تجربه و شهود خاص درونی نسبت به آنچه در ذهن و ساحت مفاهیم می‌گذرد» می‌باشد. گرچه مراد از این عبارت بیشتر در بخش‌های بعدی مقاله روشن خواهد شد، ولی به‌هرحال با دریافت و لحاظ نمودن این معنای خاصِ آگاهی، پی بردن به گره‌خوردگی تک‌تک توانمندی‌های کیفیِ مورد بحث به مقولهٔ آگاهی و بالتّبع توقف IBE بر آگاهی چندان دشوار نخواهد بود؛ زیرا با نظر به معنای مزبور، روشن است که مثلاً «درک و فهم ارتباطات تبیینی» که از عناصر بنیادیِ IBE است، بدون جریان آگاهی (تجربه و شهود از نوع مزبور) ابتدا نسبت به خود طرفین رابطه و سپس پیوند میان آن‌ها به‌دست نخواهدآمد. یا «دیدن و شهود اتّصالِ مفهومی‌استنتاجیِ نتیجه به مقدمات» که گفتیم در IBE از ضروریّات است، در واقع چیزی جز آگاهی به‌معنای مزبور نخواهد بود.

گذشته از توقف IBE بر آگاهی، جنبهٔ دیگری نیز از مسئلهٔ مورد پیگیریِ این مقاله ما را به برّرسی «امکان بهره‌مندی ماشین از آگاهی» وامی‌دارد؛ زیرا فارغ از بحث توقف یا عدم‌توقف استنتاج بر پایهٔ بهترین تبیین بر آگاهی، هدف عام‌تر و مهم‌تر این پژوهش نیز یعنی «ارزیابی امکان رسیدن هوش مصنوعی به‌حدّ هوش انسانی» برّرسی مزبور را به این سبب که طرف دوم (انسان) از آگاهی که در تأثیرگذاری‌اش بر سطح و نوع هوشمندی تردید نیست بهره‌مند می‌باشد، مستقلاً ایجاب می‌نماید. نتیجتاً لازم است جستجویمان را ذیلاً در این راستا پی بگیریم که آیا ماشین به‌منظور توانایی بر تفکّر در غالب IBE یا به‌هرحال به‌منظور نیل به هوشمندی در سطح انسانی امکان برخورداری از درک و آگاهی را دارد یا خیر؟

ماشین و آگاهی

هوش مصنوعی ضعیف و قوی

جان سِرل (1980) در مقایسهٔ هوش مصنوعی با ذهن آدمی، دو نوع نگاه را به ماهیّت هوش مصنوعی از هم تفکیک می‌نماید که برای روشنی و دقّت مباحث در این‌جا بسیار مفید است: «هوش مصنوعی ضعیف» و «هوش مصنوعی قوی». سِرل این دو را به‌نحو زیر تعریف می‌کند:

> طبق [نظریهٔ] هوش مصنوعی ضعیف، ارزش اصلی کامپیوتر از منظر مطالعات ذهن این است که یک ابزار بسیار قدرتمندی را در اختیار ما می‌گذارد. ... اما طبق [نظریهٔ] هوش مصنوعی قوی، کامپیوترها از منظر مطالعاتِ ذهن صرفاً یک ابزار نیستند، بلکه کامپیوترِ به‌نحو مناسب برنامه‌ریزی‌شده واقعاً یک ذهن می‌باشد. به

این معنا که کامپیوترهایِ دارایِ برنامه‌هایِ مناسب را حقیقتاً می‌توان گفت که می‌فهمند و دارای دیگر حالات شناختی هستند (Searle,1980:417).

به‌اعتقاد سِرل، هوش مصنوعیِ ضعیف درواقع چیزی جز یک «شبیه‌سازی» ذهن نیست و لذا با توجه به این‌که «در شبیه‌سازی کلِ چیزی که نیاز است، ورودی و خروجی صحیح و یک برنامهٔ واسطی است که اولی را به دومی تبدیل کند» (ibid.:423). در هوش مصنوعی ضعیف حالات واقعی ذهنی (از قبیل فهمیدن، احساس، آگاهی و...) در میان نخواهد بود. کما این‌که شبیه‌سازیِ کامپیوتریِ یک آتش‌سوزیِ مهیب جایی را نمی‌سوزاند یا شبیه‌سازیِ کامپیوتریِ یک بارانِ سیل‌آسا کسی را خیس یا غرق نمی‌کند. بنابراین درخصوص حالت ذهنی، نظیر فهمیدن نیز باید گفت که در یک شبیه‌سازیِ کامپیوتری از فهمیدن، چیزی به‌نام فهمِ حقیقی در میان نخواهد بود (ibid.:423). اما درمورد هوش مصنوعی قوی که تقریباً می‌توان آن را معادل با همان هوش مصنوعی عمومی تلقی نمود چه باید گفت؟ خُب، اقتضاء فرضیِ خود این برنامه که طبق آن قرار است برای یک کامپیوتر چنان برنامهٔ خوب و مناسبی ساخته شود که فراتر از صرف شبیه‌سازی ذهن عیناً یک ذهن واقعی بوده و همانند انسان‌ها درک و فهم و تفکّر داشته باشد، همین است که هوش مصنوعی قوی می‌تواند واجدِ همهٔ حالات ذهنی این‌چنینی باشد. اما سوال اساسی این است که آیا چنین امر فرضی‌ای واقعاً و عملاً هم می‌تواند در هویّاتی چون ماشین‌ها و کامپیوترها تحقّق و عینیّت یابد؟ از نظر سِرل پاسخ این سوال منفی است: هوش مصنوعی قوی امکان وقوع ندارد (نظریهٔ هوش مصنوعی قوی کاذب است).

سرل برای به چالش‌کشیدنِ ادعای مرکزی فرضیهٔ هوش مصنوعی قوی، به آزمایش فکری معروف خود تحت عنوان «اتاق چینی» متوسّل می‌شود. طبق این آزمایش فکری فردی را در اتاقی بسته و محصور تصوّر کنید که هیچ آشنایی با زبان چینی ندارد. به‌نحوی که نمادهای چینی برایش جز خطوط بی‌معنی نیستند. همچنین تصوّر کنید که در این اتاق سبدهایی باشد مملو از حروف یا نمادهای چینی با یک کتاب راهنما که به زبانی آشنا برای فرد بیان می‌کند که او چگونه با نمادهای چینی براساس صرفاً شکل ظاهری‌شان سر و کار داشته باشد (این کتاب درواقع حاوی قواعد و دستورالعمل‌هایی از این دست است که وقتی فرد با فلان دسته از نمادهای صوری مواجه می‌شود، متقابلاً کافی است فلان دسته از نمادهای درون سبدها را با شکل و ترتیب مشخص گزینش و ارائه نماید). حال بیرون از این اتاق، افراد مسلّط به زبان چینی را در نظر بگیرید که سوالاتی را به چینی نوشته و از شکاف درب به داخل اتاق می‌اندازند. آن فرد نیز متقابلاً به‌پیروی از کتاب، قواعدِ دستهٔ معینی از نمادها را از درون سبدها جمع‌آوری و به بیرون می‌اندازد. اکنون که فرد پاسخ مناسب سوالات اشخاص بیرونی را به‌کمک دستورالعمل‌ها به‌خوبی آماده و تحویل می‌دهد، گویش‌ورانی که زبان مادری‌شان چینی است خیال می‌کنند که فرد محصور در اتاق کاملاً به‌زبان چینی مسلّط بوده و آن را کاملاً می‌فهمد.

از نظر سِرل ماجرای هوش مصنوعی و کامپیوترها نیز (هر قدر هم که پیچیده باشند) دقیقاً از همین نوع است که در آن نقش کتاب قواعد را برنامه‌های داده‌شده به کامپیوترها و نقش فرد محصور در اتاق را خود کامپیوترها ایفاء می‌نمایند. وقتی کامپیوتر سوالی به زبان چینی دریافت می‌کند، آن را به‌کمک برنامه‌اش به‌یک‌نحوِ تعبیرنشده و کاملاً صوری با بانک اطلاعاتی خود مقایسه کرده و پاسخی را چه‌بسا به‌همان خوبیِ گویش‌ورانِ چینی (اگر برنامه‌اش کاملاً مناسب بوده باشد) تولید می‌کند. بااین‌حال، آیا می‌توان گفت که این کامپیوتر زبان چینی می‌فهمد، دقیقاً به‌همان معنایی که گویش‌وران چینی آن را می‌فهمند؟ به‌اعتقاد سِرل روشن است که پاسخ منفی است؛ زیرا برنامه‌های کامپیوتری امری صرفاً صوری و سینتکتیک هستند، در حالی که فهمیدن اساساً مستلزم ابعاد محتوایی و سمانتیکی است.

به‌اعتقاد سِرل ایراد اصلی هوش مصنوعی قوی درواقع این است که توجه ندارد که صرف داشتنِ یک نرم‌افزار خوب برای تضمینِ داشتن هوش مصنوعی قوی کافی نیست؛ بلکه برای این منظور به سخت‌افزار ویژه نیاز است که واجد قوای علّی معادل با مغز بوده باشد (Searle,1980:421-422-452). به‌عبارت دیگر، از نظر سِرل برای ایجاد هوش مصنوعی

قوی، (گذشته از نرم‌افزار) سخت‌افزار نیز مهم است: (در حالی که قوای علّی یک کامپیوتر با اجزاء صرفاً فیزیکی محدود و منحصر به تعامل صوری با نمادها و تولید خروجی متناسب با ورودی است) برای تحقّق حالات شناختی چون ادراک معنا، تجربهٔ کیفی، احساس، آگاهی و... به سخت‌افزاری با قوای علّی منحصربه‌فرد مغز نیاز است.

پیداست که آزمایش فکری اتاق چینی به‌معنی نقض پروژهٔ تورینگ است؛ زیرا در حالی که طبق پروژهٔ مزبور برنامهٔ کامپیوتری که بتواند اعتماد انسان‌ها را چنان جلب نماید که خیال کنند در حال مراوده با انسان دیگری هستند حقیقتاً واجد درک و تفکّر است. آزمایش اتاق چینی نشان می‌دهد که یک برنامهٔ کامپیوتری را می‌توان چنان ساخت که آزمون تورینگ را بدون داشتن فکر یا درکی از معنای نمادهای به‌کارگرفته‌اش پشت سر بگذارد: صرف تولید خروجی مناسب، به‌معنی حصول هوش مصنوعی قوی نیست (Searle,1980:419).

درخصوص رفتار هوشمندانه‌ای که پیروی کارکردِ صوری ماشین (اتاق چینی) رقم می‌خورد، لازم به تذکّر این نکتهٔ مهم است که به‌نظر ما حتی در این گونه موارد نیز هوشمندی با حضور و به‌موجب آگاهی است که پدید می‌آید؛ زیرا در این موارد نیز درواقع مجموعهٔ «ماشین و طراح (طراحِ کتاب راهنما)» از آگاهی (شعور و درک و تفکّر) بهره‌مند می‌باشد. به‌همین سبب است که افراد بیرون از اتاق احساس می‌کنند با کسی طرفند که کاملاً چینی می‌فهمد. بنابراین باید دقّت داشت که آزمایش فکری اتاق چینی حامل این دلالت نیست که پس لااقل هوشمندی صوری و نحوی بی‌مددِ آگاهی و درک و فهم قابل حصول است. درواقع به‌نظر می‌رسد درمورد هیچ رفتار هوشمندانه (غیرجسته‌وگریختهٔ تصادفی) و حساب‌شده‌ای نمی‌توان مدعی عدم دخالت و حضور اموری چون آگاهی و درک و فهم گردید. مثلاً هوشمندیِ حتّی ماشین ساده‌ای چون «ترموستات خودرو» نیز به‌موجب آگاهی و فهمی است که در ذهنِ طراح آن مستقر می‌باشد. شاید تنها تفاوت هوشمندی برآمده از کارکرد صحیح صوری و غیرصوری در این باشد که اولی قابل تفویض به ماشینِ فاقدِ آگاهی است، اما همین اندازه هم نسبت به هوشمندی محتوایی و غیرصوری مقدور نیست. نکتهٔ مندرج در این بند در صورت صحّت بدین معنی است که اگر ماشین می‌خواهد در هوشمندی استقلال داشته باشد، گریزی از داشتنِ مستقیم و مستقل توانمندیِ اموری چون آگاهی و درک و فهم ندارد.

آگاهی: ماهیّت و مسئلهٔ آن

دیوید چالمرز، از صاحب‌نظران برجستهٔ حوزهٔ آگاهی، در تمهید مقدمه برای معرفی «مسئلهٔ آگاهی» می‌پرسد که «ربط آگاهی به مغز چگونه است؟ و چگونه می‌توانیم آگاهی را برحسب مغز و شاید بدن توضیح دهیم؟» (Chalmers,2022:9). به گفتهٔ وی، آگاهی علی‌رغم این‌که آشناترین امر برای ماست، درعین‌حال مبهم و مسئله‌دار نیز است: «آگاهی سرزنده‌ترین پدیده است؛ به‌نحوی‌که در نزد ما چیزی از آن واقعی‌تر نیست. بااین‌حال می‌تواند به‌طور ناامیدکننده‌ای مبهم نیز باشد» (Chalmers,1997:3). به‌اعتقاد چالمرز آگاهی به پدیده‌های گوناگونی اشاره و دلالت دارد که همگی نیازمند تبیین هستند. اما چون دشواری تبیین این پدیده‌ها در یک سطح نیست، مسائل آگاهی را باید به دو سطح «آسان» و «دشوار» تقسیم نمود (Chalmers,2017:32). مسائل آسان آگاهی آن‌هایی هستند که به‌نظر می‌رسد مستقیماً به روش‌های استانداردِ علوم شناختی و توضیحات مبتنی بر مکانیسم‌های عصبی یا محاسباتی تن می‌دهند. مسائل دشوار نیز آن‌هایی هستند که دربرابر این روش‌ها مقاومت می‌کنند، به‌نحوی که به‌نظر می‌رسد برای آن‌ها هیچ‌گونه توضیح فیزیکی نمی‌توان فراهم ساخت.

چالمرز از میان پدیده‌های متصل به آگاهی، تبیین پدیده‌هایی چون «توانایی تشخیص و طبقه‌بندی تحریکات دریافتی از محیط و واکنش به آن‌ها»، «توانایی گزارش‌دهی از حالات ذهنی»، «توانایی دسترسی به حالات درونی خود»، «تمرکز توجه»، «کنترل ارادی رفتار» و «بازشناسی بیداری از خواب» را در زمرهٔ مسائل آسان ذکر می‌کند. منشأ این آسانی هم به این برمی‌گردد که «این پدیده‌ها را می‌توان بدون مشکل واقعی، به‌نحو علمی توضیح داد» (ibid.:32)، چرا که این‌ها

عمدتاً توانایی‌های شناختی و درواقع کارکردهایی از سامانه‌اند و لذا تبیین‌شان صرفاً شامل معرفی مکانیسمی عصبی یا محاسباتی خواهد بود که به تولید آن‌ها منجر می‌شود: «مثلاً تبیین توانایی گزارش‌دهی صرفاً شامل توضیح این می‌شود که یک سامانه کارکرد تولید گزارش‌هایی از حالات درونی خود را چگونه می‌تواند اجرا نماید» (ibid.:32).

اما مسئلهٔ دشوار آگاهی از نظر چالمرز، «عبارتست از مسئلهٔ تجربه [درونی]» (ibid.:33). گفتنی است که به هنگام ادراک و تفکّر ما زنجیره‌ای از پردازش اطلاعات را داریم. اما گذشته از این، در این هنگام ما یک جنبهٔ ذهنی (نوعی نگاه یا حس و تجربهٔ درونی) نیز داریم: «همان‌طور که نیگل (۱۹۷۴) گفته است، [در این هنگام] حسّی درونی از به‌چه‌مانندیِ فلان اُرگانیسمِ آگاه بودن(something it’s like to be a conscious organism) وجود دارد. این جنبهٔ ذهنی تجربه است. مثلاً به‌هنگام دیدن، احساس‌های بصری را تجربه می‌کنیم: کیفیّت حس‌شدهٔ قرمز بودن، تجربهٔ تاریک و روشن بودن، کیفیت عُمق حوزهٔ دید» (ibid.:33). از همین قبیل است تجربه و ادراک صدای کلارینت، بوی نفتالین، یا احساس‌های بدنی نظیر درد، لذّت و عشق و عاطفه. در همهٔ این تجارب آگاهانه در نزد فاعل شناسا «حسّی درونی از به‌چه‌مانندیِ [اُرگانیسم خاص مثلاً «خفّاش»] بودن[۲۲]» (ibid.:33) حضور دارد. درصورتی‌که در نزد جسم فیزیکی محض همچون این بطری پپسی، (احتمالاً) هرگز چنین چیزی وجود ندارد. بدین‌ترتیب طبق تعریف چالمرز، «یک موجودْ آگاه است، اگر واجد حسّی درونی از به‌چه‌مانندیِ آن موجود بودن باشد. ... به بیان دیگر، یک حالت ذهنی در صورتی می‌توان گفت آگاهانه است که توأم با یک احساس کیفی –کیفیتی ناشی از تجربه– بوده باشد. این احساس‌های کیفی را کیفیات پدیداری یا اختصاراً کوالیا نیز می‌گویند» (Chalmers,1997:4). حال از نظر چالمرز مسئلهٔ دشوار آگاهی دقیقاً عبارت است از تبیین همین تجربهٔ درونی و این‌که چرا اصلاً چنین تجربهٔ آگاه شکل می‌گیرد. چگونه است که فرایندهای فیزیکی و عصبی می‌توانند به ظهور آگاهی منتهی گردند؟ با توجه به این‌که جهانی از مخلوقات (زامبی‌ها) را می‌توان با همین فرایندهای مغزی تصور نمود که به‌کلی فاقد تجربهٔ آگاه باشند، پس درمورد موجوداتِ آگاهِ این جهان چگونه است که این فرایندهای مغزی می‌توانند به تجربهٔ آگاه منجر گردند؟ به‌اعتقاد چالمرز در این‌جا، بر خلاف مسئلهٔ آسان، دیگر کاری از تبیین‌های کارکردی تقلیلی (فروکاستن ماجرا به مکانیسم‌ها و مدل‌های عصبی‌فیزیولوژیکی) ساخته نیست: «حتی اگر نحوهٔ به اجرا درآمدنِ همهٔ کارکردهای شناختی و رفتاری پیرامون تجربه را ... تبیین کنیم، باز هم یک سوال اضافی بی‌پاسخ می‌ماند: چرا به اجرا درآمدن این کارکردها توأم با تجربه است؟» (Chalmers,2017:35).

آگاهی و تکامل

برخی نظریه‌پردازان در تلاش خود برای توضیح آگاهی به تکامل زیستی توسّل جسته‌اند. مثلاً آنتونیو داماسیو، دانشمند معروف علوم اعصاب، طی یک گفتگو با چالمرز سعی می‌کند (Damasio,2022) موضوع (پیدایش) ذهن و آگاهی را در یک بستر علمی و به‌عنوان نتیجهٔ یک مسیر طولانی (در حد میلیاردها سال) از فرایند تکامل موجودات زنده ببیند که انسان اکنون جزء حلقه‌های پایانی آن به‌شمار می‌آید:

> این واقعیت که ما درد یا خوشی یا میل جنسی را احساس می‌کنیم تصادفی نیست. این‌ها درواقع نتیجهٔ یک فرآیند تکاملی طولانی که درد، به‌عنوان مثال در آن نقشی داشته هستند. ... این نقش عبارت بوده از محافظتِ موجود زنده در برابر خطرات. ... من درمورد این نیز که چرا ما صاحب تجربهٔ ذهنی هستیم از تفکر تکاملی استفاده می‌کنم. از افراد می‌شنوم که می‌پرسند [ولی] «فایده و نقشِ داشتنِ تجربهٔ ذهنی چه بوده است؟»

---

۲۲. برای معنای دقیق‌تر اصطلاح فنی «something it’s like to be» در نزد صاحب‌نظرانی چون توماس نیگل مثلاً نگاه کنید به Nagel 1974:440, note 6.

> خُب، من فکر می‌کنم که تجربه‌های ذهنی در فرایند تکامل حائز اهمیت‌اند. مثلاً در این‌جا اگر هیچ‌یک از ما هیچ استفاده‌ای از احساس نداشتیم و اگر هیچ‌یک از ما محتوای ذهنی‌مان را که معطوف به سوژه‌ای است که خودِ او نیز دارای احساسات است نداشتیم، دیگر چرا خوب یا بدیِ رفتارمان یا با شور و حرارت بودن آن برای ما مهم می‌بود؟ (ibid.:16)

از نظر داماسیو فرایند تکامل نهایتاً منجر به شکل‌گیری بدن، سیستم عصبی و مغز در ما شده است و این سه (خصوصاً بدن) و همکاری‌شان برای تنظیم حیات ما چیزی است که می‌بایست در تبیین ذهن و آگاهی روی آن‌ها به‌جدّ حساب نمود (ibid.:7). او در این راستا تمایل دارد که تفاوت هوش انسانی با هوش مصنوعی یعنی بهره‌مندی اولی از آگاهی را با استناد به زنده بودن انسان (برخورداری‌اش از گوشت و بدن زنده) قابل‌تبیین معرفی نماید (ibid.:23).

اما برخلاف داماسیو، چالمرز معتقد است که همهٔ جوانب مسئلهٔ آگاهی به‌صورت علمیِ صرف پاسخ‌دادنی نیست و مسائل عمیقِ فلسفی همچنان پابرجا می‌مانند (Chalmers,2022:9). چالمرز می‌گوید آگاهی را اساساً نمی‌توان به اموراتی چون «توانایی پاسخ‌دهی به محرّک‌ها» فروکاست: «توانایی پاسخ‌گویی و توانایی ابراز رفتار، آگاهی نیست. آگاهی تجربهٔ خاص ذهنی است» (ibid.:11). این تجربهٔ ذهنیِ خاص را همچنین نمی‌توان منتج از محاسبات و پردازش مغز به‌شمار آورد؛ چرا که اگر چنین است، پس چرا یک ربات یا یک زامبی نیز با داشتن همهٔ این پردازش‌ها و محاسبات دارای تجربهٔ ذهنی مورد بحث نمی‌شود[۱]؟ تجربهٔ ذهنی مورد بحث نوعی داده، نه از منظر سوم شخص (بیرونی) بلکه از منظر «اول شخص» است و همین است دلیل این‌که چرا علم نمی‌تواند همهٔ ابعاد آگاهی، خصوصاً تجربهٔ ذهنی مورد بحث را توضیح دهد. به‌اعتقاد چالمرز، نظریه‌ای که بتواند تبیین‌گر آگاهی و رابطه‌اش با عملکرد مغز باشد، اساساً «... هرگز یک نظریهٔ کاملاً تقلیل‌گرایانه نخواهد بود. ما هرگز نخواهیم گفت که آگاهی صرفاً یک فرآیند در مغز است. ما هرگز آگاهی را به‌طورکامل بر حسب فرآیندهای فیزیکی تبیین نخواهیم کرد» (ibid:14).

به‌دنبال نکات این‌چنینی از سوی چالمرز، نهایتاً داماسیو نیز در این گفتگو مخالفت خود را با تقلیل‌گرایی افراطی اعلام می‌دارد: «من [نیز] فکر نمی‌کنم ذهن ما و آنچه که به‌لحاظ ذهنی هستیم، آنچه که شخصیت ما را می‌سازد، قابل تقلیل به نورون‌ها باشد» (Damasio,2022:15). به‌هرحال، نکتهٔ اصلی چالمرز این است که آگاهی را نه می‌توان همچون نوخواسته‌گرایان[۲] نتیجهٔ پیچیدگیِ امور فیزیکی دانست و نه آن را همچون فروکاست‌گرایان به امور فیزیکی فروکاست. از نظر چالمرز برای تعریف تجربهٔ آگاه، ظاهراً راهی جز این نیست که آن را مفهومی بنیادین به‌شمار آوریم: «دیدگاه من این است که ما باید آگاهی را به‌عنوان عنصر بنیادی طبیعت درنظر بگیریم» (Chalmers,2022:22).

درمورد رویکرد تکاملی فوق، گذشته از نکات چالمرز دو نکته شایان توجه است. اولاً در این رویکرد نیز نسبت به اصلِ آگاهی و این‌که انسان به‌عنوان «هوش طبیعی معیار» جهت تنظیم رفتارهای هوشمندانهٔ خود نیازمند آگاهی است هیچ تردیدی نشده، بلکه این امر مورد تأیید کامل قرار می‌گیرد. ثانیاً توضیح تکاملی تدارک دیده شده برای نحوهٔ بروز آگاهی (حداقل در شکل کنونی‌اش) به‌گونه‌ای نیست که آن را بتوان یک توضیح حقیقی در این زمینه به‌حساب آورد؛ زیرا تمرکز این نوع توضیح اصلاً به اصل بحث معطوف نیست، این‌که چگونه است که از ماده و فیزیکِ محض هویّتی چون آگاهی که به‌نظر یک امر فیزیکی نمی‌آید برخاسته است: این توضیح اساساً به مسئله و ابهام مرکزی نمی‌پردازد.

با توجه به‌مباحث و نکات فوق پیرامون ماهیّت آگاهی (از جمله این‌که آگاهی یک امر سابجکتیو و از جنس تجربهٔ درونی است)، به‌نظر دشوار یا حتّی ناممکن می‌آید که بتوان برای شیء فیزیکیِ محض خصوصاً از قبیل ماشین (کامپیوتر)،

---

[۲۴]. چالمرز با استدلال زامبی و صورت‌بندی منطقی عدم ضرورت وجود آگاهی در جهان ممکن زامبی‌ها (Chalmers,1997:94-99) درواقع در حال رد کردن فیزیکالیسم می‌باشد.

[۲۵]. طبق نوخاسته‌گرایی (Emergentism) ذهن و ویژگی‌های ذهنی، برآمده از فرآیند تکامل بدن فیزیکی و در عین حال غیرقابل‌تقلیل به آن است. به عبارت دیگر، بدنِ فیزیکی سیستمی سیّال و پویا است که در بالاترین سطح پیچیدگی سیستمی، موجب ظهور ذهن و آگاهی می‌شود.

به‌عنوان هویّت دست‌ساز بشر، قابلیّت برخورداری از چنین امر ذهنی در نظر گرفت (جز این‌که برای اعطاء این امر بنیادی به آن راه قابل‌قبولی قابل‌تصور باشد). بااین‌حال هستند صاحب‌نظرانی که برای تصاحب آگاهی توسط ماشین هیچ مانعی نمی‌بینند. نظر به چنین دیدگاهی لازم است نگاه و ادلهٔ این گروه نیز ذیلاً مورد بررسی قرار گیرد تا امکان قضاوت درخصوص موضوع مقاله با جامعیّت کافی فراهم آید.

امکان ماشین آگاه

از جمله فلاسفه‌ای که تمایل دارند برای ماشین قابلیّت برخورداری از آگاهی در نظر بگیرند، زوج فیلسوف پاول چرچلند و پاتریشا چرچلند می‌باشند: کسانی که درصددند کلیه پدیده‌های ذهنی و شناختی را حتّی‌الامکان به‌نحو کاملاً فیزیکی و بر پایهٔ صرف قوانین علمی و علوم اعصاب توصیف نمایند. پاول چرچلند درمورد مهم‌ترین مناقشهٔ مرتبط با هوش مصنوعی می‌گوید:

> پرسش پیش روی برنامهٔ پژوهشی هوش مصنوعی این نیست که آیا کامپیوترهای دارای برنامه‌ریزیِ مناسب قادرند رفتارهای ورودی/خروجی تولیدشده توسط روش‌های محاسباتی موجود در حیوانات طبیعی، از جمله آن‌هایی که در انسان یافت می‌شوند را شبیه‌سازی کنند یا خیر؟ این سوال معمولاً پاسخ‌داده‌شده تلقی می‌شود. کامپیوترها حداقل اصولاً باید قادر به انجام این باشند. سوال مهم این است که آیا فعالیت‌هایی که هوشِ آگاه را شکل می‌دهند همگی نوعی رویّه‌های محاسباتی هستند یا خیر؟ فرض هدایت‌کنندهٔ [پروژهٔ پژوهشیِ] هوش مصنوعی این است که آن‌ها این گونه‌اند (Paul Churchland,2013:167).

چرچلند که مغز آدمی را هم جز نوعی ماشین به‌شمار نمی‌آورد، در راستای پاسخ به سوال مهم فوق و درواقع درمورد امکان شبیه‌سازی کامل تمامی فعالیت‌های مغز انسان توسط ماشین، بیان می‌دارد که «هیچ دلیل الزام‌آوری نیست که فکر کنیم ماشین‌ها [ی به‌نحو مناسب برنامه‌ریزی‌شده] قادر به شبیه‌سازیِ [تمام‌عیار] چنین فعالیت[های]ی نیستند» (ibid.:178). از نظر او اگر هم فعلاً مشکلی دراین‌خصوص هست، این مشکل درواقع «نه به محدودیّت ذاتی در توانمندی ماشین‌ها، بلکه به محدودیّت فعلی ما در درک آن چیزی مربوط می‌شود که می‌خواهیم ماشین‌ها شبیه‌سازی کنند» (ibid.:178). مشکلی که از نظر او امید می‌رود علوم اعصاب در ادامه آن را نیز از میان بردارد (یعنی کل فعالیّت مغز با تمام ابعادش را که علی‌الادعا به‌کلی محاسباتی است و لذا آنچه ماشین باید شبیه‌سازی کند مشخص نماید).

چرچلندها در مخالفت خود با استدلال سِرل (علیه هوش مصنوعیِ آگاه و قادر به تفکّر)، بعد از این جمع‌بندی که «سِرل در حال اتّخاذ یک تست غیررفتاری برای آگاهی است: این‌که عناصر هوشِ آگاه باید دارای محتوایِ سمانتیکیِ واقعی باشند» (Churchland & Churchland,1990:34)، اقدام به رد این اَکسیوم از استدلال سِرل می‌کنند که «نحو به‌خودی‌خود برای سِمانتیک نه لازم است و نه کافی» (ibid.:34). از نظر چرچلندها این اکسیوم صدق بدیهی ندارد و لذا لازم بود سِرل برای آن دلیل اقامه کند که نکرده است. درواقع از نظر ایشان استدلال اتاق چینیِ سِرل قادر به نشان دادن صدق اَکسیوم یا اصل مزبور نیست. چرچلندها بعد از نقد استدلال سِرل، له این عقیدهٔ خویش استدلال می‌کنند که ماشین‌ها می‌توانند فکر کنند. آن‌ها می‌گویند ماشین‌های کلاسیک به این دلیل تناسب لازم را برای این منظور نداشتند که از نوع سِری بودند و نتیجتاً چندان «مغزمانند» نبودند. اما وقتی به سیستم‌های عصبی بیولوژیکی مراجعه می‌کنیم می‌بینیم

که آن‌ها درواقع نوعی «ماشین‌های موازی هستند[۱]. به این معنا که در آن‌ها سیگنال‌ها به‌طور همزمان در میلیون‌ها مسیر مختلف پردازش می‌شوند» (ibid.:35). بنابراین، می‌توان گفت که ماشین‌های جدید که همانند این نوع «ماشین‌های موازی» بوده باشند، قادر به تفکر خواهند بود. به‌عبارت دیگر، طبق استدلال چرچلندها، سامانه‌های مصنوعیِ غیربیولوژیکیِ موازی که بر پایهٔ سیستم‌های عصبی (بیولوژیکی) مدل شده باشند، تعاملشان دیگر با محوریّت نمادها و بر پایهٔ قواعدِ حساس به ساختار (قواعد صرفاً صوری) نخواهد بود: «تعاملِ نمادی صرفاً یکی از مهارت‌های شناختی بسیاری خواهد بود که شبکه ممکن است یاد بگیرد یا نگیرد که به نمایش بگذارد. تعاملِ نمادیِ قاعده‌محور، تشکیل‌دهندهٔ عملیات مبنایی و اصلی آن نخواهد بود» (ibid.:36). بدین‌ترتیب، آن‌ها نتیجه می‌گیرند که استدلال سِرل صرفاً علیه ماشین‌های دارای تعامل نمادی قاعده‌محور طراحی شده است. به‌نحوی که سامانه‌های موازی موردنظر ایشان (یعنی از نوع مغزمانند)، «از جانب استدلال اتاق چینی سِرل، حتّی اگر صحیح هم می‌بود، تهدید نمی‌شوند» (ibid.:36). چرچلندها منشاء اختلاف خود با سِرل را این‌گونه جمع‌بندی می‌کنند که «سِرل موضع خویش را بر شهودِ فهم عرفی دربارهٔ حضور یا غیبت محتوای سمانتیکی مبتنی می‌سازد. اما ما موضعمان را بر ناکامی رفتاریِ ویژهٔ ماشین‌های کلاسیکِ تعامل‌کننده با نمادها و مزایای ویژهٔ ماشین‌های واجد معماریِ مغزمانندتر بنا می‌کنیم» (ibid.:37). نهایتاً آنها ضمن نسبت دادنِ مزایای تجربیِ تعیین‌کننده و فراوان (از جهت کارکردها و وظایف شناختی)، به استراتژی محاسباتی خویش در مقایسه با سایر استراتژی‌ها اظهار می‌دارند: «واضح است که مغز به‌طور سیستماتیک از این مزیّت‌های محاسباتی استفاده می‌نماید. اما ضروری نیست که این تنها سامانهٔ فیزیکی باشد که قادر به انجام چنین کاری است. هوش مصنوعی در یک ماشین غیربیولوژیکی اما در حد زیاد موازی، افق و امکان روشن و متقاعدکننده‌ای را به نمایش می‌گذارد» (ibid.:37).

چنان‌که اشاره شد، چرچلندها به پیشرفت علوم اعصاب و کشف اَسرار و ابعاد گوناگون ذهن (مغز، از منظر آن‌ها) و شناخت و آگاهی توسط این علوم قویاً امیدوارند. به اعتقاد آن‌ها اساساً باید صبر نمود و منتظر پیشرفت‌های علمی و تکنولوژیکی ماند تا شناخت ما از مغز کامل‌تر گردد. آن وقت می‌توان جملگی این مسائل (فلسفی) را به‌نحو علمی توضیح داد. پاتریشا چرچلند که از پایه‌گذران و توسعه‌دهندگان نوروفلسفه نیز به‌شمار می‌آید، با همین رویکردِ جدّی گرفتنِ علوم اعصاب در آثار خود، تلاش صریحی را برای برقراری ارتباط میان دستاوردهای تجربی علوم اعصاب و مسائل فلسفی ترتیب می‌دهد. او تا جایی پیش می‌رود که مسئلهٔ دشوار آگاهی موردنظر چالمرز را صرفاً یک «مغالطه» و «فریفتن مخاطب» در نظر می‌گیرد. چراکه به‌اعتقاد وی، استدلال مربوط جز این نیست که از مجهول بودنِ کنونیِ چیزی بر ما نتیجه می‌گیرد که آن چیز همواره کشف‌ناپذیر خواهد بود. در حالی‌که مسئلهٔ دشوار آگاهی صرفاً یک مسئلهٔ تجربی است که روزی کشف خواهد شد (Pat Churchland,1996). پاول چرچلند هم درمورد این‌که آیا برنامه‌نویسان طی تلاش‌های مختلفِ خویش جهت هوشمندسازی ماشین (شبیه‌سازی ذهن انسان)، موفق به فراهم نمودن «خودآگاهی» برای ماشین خواهند شد یا نه، با آن‌که برداشتن خیز کامل برای این منظور را فعلاً قدری زود می‌شمارد ولی در کل چنین چیزی را امری کاملاً ممکن و شدنی در نظر می‌گیرد (Paul Churchland,2013:185-189).

باید دانست که تکیه‌گاه اصلی امیدواری و کلاً نوع نگاه و نکات فوق‌الذکر چرچلندها، درواقع جز موضع فلسفی آن‌ها تحت عنوان «ماتریالیسم حذف‌گرا» نیست: «اگر ماتریالیسم نهایتاً صحیح باشد، پس این [صرفاً] چارچوب مفهومی یک علمِ اعصابِ کمال‌یافته است که نمایانگر دانش اساسی مربوط به طبیعت درونی ما خواهد بود» (ibid.:281). ماتریالیسم در این بستر به‌معنی انکار و نفی این ادعای امثال دکارت است که ذهن به‌عنوان بخشی از آدمی که عهده‌دار تفکّر و اندیشیدن است، یک جوهر مستقل و ماهیتاً متفاوت از بدن (ماده) می‌باشد. ماتریالیسمِ حذف‌گرا نیز انسان و اساساً هر نوع حیوانی را «یک منظومهٔ فیزیکیِ صرف» (ibid.:43) با یک طبیعت و ماهیّت «کاملاً فیزیکی» (ibid.:44) درنظر گرفته و ذهن و

---

[۲۸]. چرچلندها صراحتاً می‌گویند «مغز نوعی کامپیوتر است، هرچند اکثر ویژگی‌های آن هنوز کشف نشده است» (Churchland & Churchland 1990:37).

حالات یا امور ذهنی از قبیل باورها، آرزوها، خوف و امیدها، لذّت یا درد، نیّت و... را که وجود داشتنشان بر پایهٔ فهم متعارف یا به‌تعبیر چرچلند «روان‌شناسی عامه» روشن و بدیهی تلقّی می‌شود، عاری از وجود واقعی اعلام می‌کند. به‌عبارت دیگر، طبق دیدگاه حذف‌گرا، هرچند اکنون ما خیال می‌کنیم این امور واقعی هستند، ولی همچنان‌که پیشرفت علم نشان داده هویاتی نظیر فلوژیستون، کالریک، فلک تدویر، ساحره‌ها و ... موجود نیستند، پیشرفت علوم اعصاب آشکار خواهد ساخت که این امور، یعنی مفاهیم روان‌شناسی عامه نیز از وجود واقعی برخوردار نیستند (ibid.:74-77).

در مقابل، هیوبِرت درِیفوس این رویکرد که ذهن را ماشینی از نوع کامپیوترهای دیجیتالی در نظر گرفته و مدعی شویم که «رفتار هوشمند انسان حاصل پردازش اطلاعات توسط [ذهن به‌مثابهٔ] یک کامپیوتر دیجیتال است و... این‌که انسان و پردازش اطلاعات مکانیکی درنهایت فرآیندهای مبنایی یکسانی را شامل می‌شوند» (Dreyfus,1992:155) رد می‌کند. دریفوس می‌گوید این فرض که انسان همچون یک دستگاه همه‌منظورهٔ تعامل‌کننده با نمادها عمل می‌کند درواقع بر پیش‌فرض‌های گوناگونِ زیست‌شناختی، روان‌شناختی، معرفت‌شناختی و هستی‌شناختیِ توجیه‌نشده مبتنی است. گرچه دریفوس تک‌تکِ این پیش‌فرض‌ها را طی مباحث مفصّل (ibid.:159-227) مورد شناسایی، معرفی و سپس فرض قرار می‌دهد، این‌جا اشاره به یک یا دو مورد از آن‌ها و ابعاد اصلی ادلهٔ وی علیه هوش مصنوعی مکفی خواهد بود.

دریفوس می‌گوید حامیان پروژهٔ هوش مصنوعی این پیش‌فرض روان‌شناختی را دارند که «ذهن را می‌توان به‌عنوان دستگاهی در نظر گرفت که برمبنای بیت‌های اطلاعات و مطابق قواعد صوری کار می‌کند» (ibid.:156) و همچنین امید آن‌ها به موفقیّت این پروژه مبتنی است بر این پیش‌فرض معرفت‌شناختی که «کل دانش را می‌توان صوری‌سازی نمود» (ibid.). اما از نظر دریفوس این‌گونه پیش‌فرض‌های هوش مصنوعی همگی ناموجّه هستند: «هوشمندی نیازمند فهمیدن است و فهمیدن هم نیازمند این است که به کامپیوتر پیش‌زمینهٔ فهم متعارف را که انسان‌های بالغ به‌سبب داشتن بدن، تعامل مهارت‌آمیز با جهانِ مادی و تربیت شدنشان در یک فرهنگ واجد آن می‌شوند بدهیم» (ibid.:3). چنان‌که از این مقوله پیداست، دریفوس «فهمیدن» را جزء شروط هوشمندیِ حقیقی دانسته و هوش مصنوعی را به این سبب که قادر به تصاحب یکی از لوازم مهم فهمیدن در موقعیت‌های واقعی یعنی دانش پیش‌زمینه‌ایِ فهم متعارف نیست، عاجز از ارضاء این شرط می‌داند. به‌عبارت دیگر، سخن و استدلال دریفوس این است که «هوش مصنوعی به این دلیل محکوم به‌شکست است که هوش حقیقی به پیش‌زمینهٔ فهم متعارف نیاز دارد، ولی کامپیوترها هرگز قادر به‌تصاحب آن نیستند» (Spitzer,2016:430). درمورد منظور دریفوس از «پیش‌زمینهٔ فهم متعارف»، مثال خود وی از امر ساده‌ای چون هدیه دادن کمک‌کننده است: «برای این‌که بدانید چگونه به‌کسی هدیه‌ای مناسب، در زمانی مناسب و به‌شیوه‌ای مناسب تقدیم کنید، نیازمند تبحّر فرهنگی(cultural *savoir faire*) هستید» (Dreyfus,1992:xxiii). مثلاً وقتی ما در مغازه‌ای قدم می‌زنیم و به‌دنبال هدیه‌ای برای بردن به مهمانی شام دوستمان می‌گردیم، برای این منظور ممکن است بردن یک دسته‌گل را در نظر بگیریم، ولی بردن۲۰ قوطی ماهی تن یا یک قوطی سوسک‌کُش را لحاظ نمی‌کنیم (Spitzer 2016:430). علت این‌که دریفوس پیش‌زمینهٔ فهم متعارف را قابل‌کسب توسط کامپیوتر نمی‌داند این است که او (و بردارش استوارت)، «معتقدند که چنین پیش‌زمینه‌ای شامل مجموعه‌ای از مهارت‌ها است، نه مجموعه‌ای از دانش و آگاهی‌ها. یعنی پیش‌زمینهٔ فهم متعارف نیازمند دانشِ مهارتی است و نه دانشِ گزاره‌ای» (ibid.:430). دریفوس‌ها بر این باورند که این نوع اطلاعات توسط کامپیوتر قابل‌بازنمایی نیست: «اگر فهم پیش‌زمینه در واقع یک مهارت است و اگر مهارت‌ها نه بر قواعد بلکه بر الگوهای کامل مبتنی‌اند، انتظار این است که بازنمایی‌هایِ نمادی قادر به کسب فهم پیش‌زمینه‌ای ما نباشند» (Dreyfus & Dreyfus,1988:33). بنابراین گزاره‌ای و لذا قابل‌صوری‌سازی نبودن کاردانی و معرفت مهارتی موجود در هوشمندی طبیعی بشری مانعی است برای هوش مصنوعی از این‌که بتواند به‌حد هوش طبیعی برسد.

بدین‌ترتیب سخن دریفوس این است که اولاً هوشمندی واقعی مستلزم فهمیدن، از جمله فهمیدنِ دانش پیش‌زمینه‌ای مندرج در فهم متعارف انسانی است. ثانیاً چنین دانشی که از قضا بخش بزرگی از معرفت بشری (شامل اغلب امورات روزمره مثل دوچرخه‌سواری و شنا و...) را شکل می‌دهد، نه یک دانش گزاره‌ای، نمادی و قابل‌صوری‌سازی، بلکه دانشی مهارتی و ضمنی و غیرقابل‌صورت‌بندی صریح است. پس این پیش‌فرض کلیدی پروژهٔ هوش مصنوعی که «هر دانشی را می‌توان صوری ساخت» درواقع کاذب است: دانش بشری (که تسلّط بر آن لازمهٔ هوشمندی است) تماماً قابل پیاده‌سازی، بیان، فهم و استفاده توسط کامپیوتر و برنامه‌های کامپیوتری (که ماهیتی صرفاً صوری‌نمادی دارند) نیست. این امر گذشته از پیش‌فرض معرفت‌شناختی، هم‌زمان بی‌اعتباریِ پیش‌فرض روان‌شناختی پروژهٔ هوش مصنوعی را نیز عیان می‌سازد: وقتی ذهن بشری (هوشمندی طبیعی) عملاً دارندهٔ دانش غیرقابل‌صوری‌سازی است، پس ذهن در حقیقت ماشینی نبوده که صرفاً بر مبنای بیت‌های اطلاعات و قواعد صوری کار می‌کند.

وقوع هوش مصنوعی: ممکن یا ناممکن؟

با مباحث فوق و تحلیل‌ها از زوایای گوناگون، اکنون زمینه برایمان مهیّاست که ببینیم نهایتاً درخصوص موضوع این مقاله، یعنی امکان تحقّق عملی هوش مصنوعی و حدود آن، چه قضاوتی می‌توان داشت. برای این منظور قبل از هرچیز باید (به‌پیروی از سِرل) هوش مصنوعی ضعیف و قوی را از هم تمییز داده و تکلیف هرکدام را مستقلاً در نظر بگیریم. درمورد «هوش مصنوعی ضعیف» باید گفت که نسبت به وقوع آن هیچ اختلاف و تردیدی در میان نیست: کما این‌که این نوع هوشمندی همین الان و بلکه مدت‌هاست که در ماشین محقّق شده و همچنان نیز در مسیر بهبود و ارتقاء قرار دارد. دراین‌خصوص همچنین باید توجه داشت که هوش مصنوعی ضعیف به‌معنی این نیست که ماشینِ مجهز به این نوع هوشمندی هرگز قادر نیست عملکردی بهتر و قوی‌تر از انسان (به‌عنوان صاحب هوش طبیعی) داشته باشد. اتفاقاً بنا به مباحث فوق، چون این نوع هوشمندی بر عملیّات صرفاً صوری و محاسباتی مبتنی است و همچنین به‌جهت ماشینی و غیربیولوژیکی بودن می‌تواند عاری از برخی محدودیّت‌های مختص گونه‌هایِ زیستی باشد، می‌تواند فعالیت‌های خود در زمینه‌های تعریف‌شده را با خطای کمتر و دقّت و سرعتِ بیشتر و در سطح وسیع‌تر انجام داده و در محدودهٔ خود از هوش طبیعی انسانی پیشی نیز بگیرد.

اما آنچه که محل بحث و اختلاف است «هوش مصنوعی قوی» است: هوشی که از تمام ابعاد در حد هوش انسانی یا فراتر باشد. تحلیل‌های لارسون، سِرل، چالمرز و دریفوس همگی از موانع جدّی و احتمالاً غلبه‌ناپذیر بر سر راه وقوع چنین هوشی در ماشین (غیربیولوژیکی) حکایت می‌کنند. این تحلیل‌ها به‌نظر می‌رسد در عین تفاوت و بیان اختصاصی در این نکتهٔ مبنایی با هم مشترکند که برای هوش مصنوعی‌ماشینی فراروَی از عملکردِ صوری‌نحویِ صرف و دست‌یابی به سمانتیکس و محتوا به‌منظور نزدیکی به هوش طبیعی به‌سادگی مقدور نیست. درواقع طبق استدلال‌های این نظریه‌پردازن، از طرفی هوشمندی حقیقی (در حد طبیعی و انسانی) نیازمند اموری چون فهمیدن، IBE، تفکّر و آگاهی است و از طرف دیگر، برای اِسناد قابلیّت چنین اموری به ماشین، مَحمِل یا مُستمسَکِ مشخّص و موجّهی دیده نمی‌شود. براساس استراتژی بدیع و جالب لارسون، هوش مصنوعیِ مجهزنشده به استنتاج ابداکتیو، هوش مصنوعی قوی نخواهد بود. اما طبق توسعه و تقویتی که ما برای این استراتژی (از طریق صراحت‌بخشی به نقش توانمندی‌هایِ کیفی همچون تخیّل، تفکّر، خلاقیّت، فهم، آگاهی، شمّ خوب، و امثالهم دراین‌خصوص) تدارک دیدیم، بهره‌مندی از استنتاج ابداکتیو خود در گرو برخورداری از چنین قابلیّت‌هایِ شناختیِ کیفی است: بدون چنین قابلیت‌های کیفی استنتاج ابداکتیو، حداقل در یک معنای اصیل و حقیقی قابل اجراء نیست. امکان برخورداری ماشین از «آگاهی» را به‌عنوان پایهٔ حداقلی قابلیّت‌های کیفی مزبور یا به‌هرحال به‌عنوان لازمهٔ یک هوش مصنوعیِ درصددِ نیل به‌حد هوش انسانی مورد برّرسی قرار دادیم. در تردیدناپذیری توقف هوشمندی بر آگاهی، گذشته از موارد تاکنون ارائه‌شده، همین بس که هیچ‌یک از استراتژی‌های بررسی‌شده، از جمله استراتژی چرچلندها، انکار این توقف را در دستور کار خود قرار نمی‌دهند. به‌نحوی که چرچلندها به‌جای انکار به‌دنبال امکان‌پذیری ظهور آگاهی

از مادهٔ محض می‌روند. به‌هرحال، نکات و ادلهٔ اتخاذشده در اغلب رویکردهای بحث‌شده حاکی از این موضع‌اند که آگاهی به‌موجب جنس و ماهیت ویژه‌اش و حداقل به‌حسب ظاهر قابل‌اعطاء به‌ماشین به‌عنوان یک هویّت فیزیکی و مادی محض نمی‌باشد: اعطاءِ چنین قابلیّتی به‌ماشین، (گذشته از نرم‌افزار) نیازمند سخت‌افزار ویژه در معنای سِرلی با قوای علّی ذهن یا چیزی همچون جوهر ذهنی دکارتی (جوهری نفسانی و متفاوت از جوهر مادّیِ ممتد در جهات سه‌گانه) است.

بااین‌حال، در برابر موضع فوق، با رویکرد امثال چرچلندها مواجهیم که قائل به امکان ماشینِ آگاه و متعاقبِ آن هوش مصنوعی قوی است. خط سیر استدلالی این اندیشمندان برای این منظور درواقع به قرار زیر بود. ابتدا این اصل سِرل را که «نحو و ابعاد صوری به‌خودی‌خود برای سِمانتیکس یا ابعاد معنایی نه لازم است و نه کافی» با این گفته رد کردند که چنین مدعا یا اصلی صدق بدیهی ندارد (نیازمند دلیل بوده ولی ارائه نشده است). سپس استدلال چالمرز حول «مسئلهٔ دشوار آگاهی» را به اتهام این‌که از مجهول بودنِ کنونیِ چیزی بر ما کشف‌ناپذیری همیشگی آن را نتیجه می‌گیرد، مغالطه شمردند و متعاقب آن مسئلهٔ مزبور را یک مسئلهٔ صرفاً تجربی در نظر گرفتند که روزی کشف خواهد شد. نهایتاً با توسّل به فرضیهٔ «ماتریالیسم حذفی»، هوش ماشینی غیربیولوژیکیِ متفکّر و آگاه (یعنی هوش مصنوعی قوی، به‌معنای سِرلی) را ممکن اعلام نمودند. ولی ما درخصوص خط سیر استدلالی چرچلندها بر این اعتقادیم که ردّیه‌هایشان در دو گام نخست از دقّت و استحکام لازم برخوردار نبوده و بیشتر بازتابی از مفروضات دیدگاهِ فلسفی مورد حمایتشان (ماتریالیسم حذفی) هستند تا ردّیه‌ای مستدل؛ زیرا درخصوص مورد اول به‌نظر می‌رسد صِرفِ تصوّر صحیح تمایز میان «امور صوری‌نحوی» و «امور معناشناختی‌محتوایی» و امکان تحقّق اولی بدون دومی که آزمایش فکری اتاق چینی نیز دقیقاً و کاملاً مویّد همین است، برای تصدیق ادعای سِرل کفایت می‌کند. وقتی این دو، مفهوماً و در وقوع منفک و مستقل از یکدیگرند و میان آن‌ها پیوندی جاری نیست که تحقّقِ یکی دیگری را الزام نماید، پس چرا نباید اولی برای دومی غیرلازم و غیرکافی شمرده شود؟ مثلاً و به‌عنوان یک وضعیّت مشابه، در فرهنگ اسلامی درخصوص نماز دو امر مستقل قابل‌تصوّر است: «اَذکار و حرکات ظاهری نماز» و چیزی به‌نام «حضور قلب» (که شرط مهمی هم در قبولی این عبادت شمرده می‌شود). حال آیا این دو را نمی‌توان مُنفک از هم در نظر گرفت، به‌نحوی که اَذکار و افعال صوری نماز را کسی یا حتّی رُباتی بی‌کم‌وکاست به اجرا بگذارد، بی‌آن‌که نسبت به معبود و این اَذکار و افعال حضور قلب یا آگاهی داشته باشد؟

درمورد استدلال چالمرز نیز گفتنی است که اصل سخن این استدلال این است که تأمّل در ماهیّت و ابعاد شهودشدهٔ آگاهی ما را برحذر می‌دارد از این‌که چنین هویّتی را به‌سادگی جزء امور فیزیکی محض (حداقل بدان‌گونه که ما این امور را می‌شناسیم یا تعریف می‌کنیم) در نظر بگیریم. حال مغالطه نامیدن آن (با توسّل به این نکتهٔ بدیهی که عدم‌شناخت کنونی ما از چیزی به‌معنی شناخت‌ناپذیری مطلق و همیشگی آن نیست)، ضمن این‌که صحیح و وارد نیست، حتّی به‌فرض صحّت نیز نتیجه نمی‌دهد که پس مسئلهٔ آگاهی در حقیقت یک مسئلهٔ تجربی است و روزی با روش‌های کاملاً تجربی مکشوف خواهد شد. اگر استدلال چالمرز مغالطه است، این طرز نتیجه‌گیری چرچلندها نیز در مغالطه بودن دست کمی از آن نخواهد داشت؛ زیرا اگر عدم‌شناخت کنونی چیزی به‌معنی شناخت‌ناپذیری مطلق آن نیست، به هیچ وجه به‌معنی شناخت‌پذیری آن در آینده نیز نیست: در چنین حالتی، موضع صحیح صرفاً می‌تواند لاادری‌گرایی باشد نه انتخاب طرف دوم. اما اساساً نکتهٔ اصلی در این است که در صورت فهم و گزارش درست، استدلال چالمرز معلوم می‌گردد که آن را نمی‌توان به این سادگی به مغالطه منسوب ساخت. سخن چالمرز درواقع این است که ما به‌عنوان یکی از دارندگان آگاهی و کسانی که به‌نحو درونی و ذهنی و شهودی به آن راه داریم، آگاهی را به‌گونه‌ای در می‌یابیم که با توجه به شناخت و تعریفی که از فرایندهای فیزیکی محض داریم نمی‌توانیم آن را اساساً قابل تبیین با این نوع فرایندها در نظر بگیریم. بنابراین این‌جا استدلال در واقع لهِ تبیین‌ناپذیری همیشگی (و نه موقت) آگاهی بر پایهٔ امور فیزیکی محض و فروکاست‌ناپذیری آن به این امور است. به‌عبارت دیگر، چنین نیست که چالمرز از صرف نیافتن تبیین در زمان کنونی تبیین‌ناپذیری مطلق آن براساس امور فیزیکی را نتیجه می‌گیرد، بلکه او از بررسی و در نظر گرفتن اوصاف ماده و اوصاف آگاهی نتیجه می‌گیرد که آگاهی اساساً برمبنای

ماده قابل‌تبیین نیست. اگر چرچلندها بخواهند چنین استنتاجی را به‌معنی واقعی رد کنند، لازم است نحوه یا لااقل امکان‌پذیری تبیین آگاهی براساس امور فیزیکی را نشان دهند، نه این‌که چنین چیزی را صرفاً فرض گرفته و تأکید کنند که با صبرکردن مطمئناً راه چنین تبیینی در آینده یافت خواهد شد!

اما گام سوم مسیر استدلالی چرچلندها، ماتریالیسم حذفی، مهمترین گامی است که درواقع هم مبنای بقیهٔ مدعیّات آن‌ها (از جمله در دو گام نخست) است و هم به‌تنهایی می‌تواند شرایط را از هر نظر به‌نفع تبیین آگاهی براساس امور فیزیکی (یا فروکاهی آن به‌چنین اموری) تغییر دهد، فقط به این شرط که فرض مبنایی آن صحیح بوده باشد: این‌که هیچ عنصر یا جوهری جز جوهر مادی و فیزیکی در این جهان (یا لااقل در خصوص ذهن) در کار نیست. در صورت صحّت این فرض، به‌راحتی می‌توان نتیجه گرفت که (برخلاف همهٔ استدلال‌ها از طرف امثال چالمرز یا سِرل) آگاهی نهایتاً قابل اعطاء به هوش مصنوعی است؛ زیرا در این صورت وقوع آگاهی در انسان که طبق ماتریالیسم حذفی سراسر وجودش (یا حداقل ذهن او) به‌کلی مادّی و فیزیکی است، خود بهترین شاهد برای امکان وقوع آگاهی و بقیهٔ حالات ذهنی و در یک کلام، هوش مصنوعی قوی در هویات برخوردار از ساختار صرفاً فیزیکی خواهد بود. البته چون ما در اینجا مرادمان از هوش مصنوعی هوشی است که در یک ماشین فیزیکی و همچنین بدون ابعادِ زیستی (و تکامل زیستی) شکل گرفته باشد، لذا علاوه‌بر فرض ماتریالیسم حذفی باید این فرض دوم را نیز اتّخاذ نمائیم که در پدیدآمدن هوش طبیعی در انسان، بدن فیزیولوژیکی و زیستی او دارای هیچ دخالت و نقش تعیین‌کننده و غیرقابل‌دستیابی از کانال غیرزیستی نیست (یا به‌هرحال تفکّر و هوش طبیعی چیزی همچون یک خاصیت نسبتاً فرامادی نیست که طی تکامل زیستی به انحائی چون نوخاسته‌گرایی از پیچیدگی ماده ظهور یافته باشد). آری در این صورت همچون چرچلندها به‌راحتی می‌توان نتیجه گرفت که ذهن ما چیزی جز مغزِ سراسر فیزیکی نیست و لذا اگر روزی به مدل کامل عملکرد مغز دست‌یابیم و آن را در ماشین پیاده بسازیم، به‌معنای دقیق کلمه یک ماشین با هوش مصنوعی قوی خواهیم داشت: در این صورت هرآنچه اکنون هوش انسانی واجد آن است، قابل فراخوانی و بازسازی در ماشین نیز خواهد بود.

اما آیا دو فرض بزرگ و سنگین فوق بالاخص اولی، از چنان مبنای محکمی برخوردارند که با خیال راحت و قاطعانه به امکان وقوع هوش مصنوعی قوی در ماشین حکم کنیم؟ آیا برای صحّت موضع فلسفی ماتریالیسم حذفی، دلایل کافی در اختیار داریم؟ دلیل اصلی که معمولاً امثال چرچلند برای این منظور ارائه می‌کنند این است که در تاریخ علم هویّات فراوانی نظیر افلاک بلورین، فلوژیستون، کالریک، اِتِر و... بوده که بعدها معلوم شده‌اند عاری از وجود واقعی‌اند؛ پس ذهن و حالات ذهنی از قبیل باورها، خوف و امیدها، لذّت یا درد، نیّت و... نیز (که ما وجودشان را بر پایهٔ فهم متعارف قطعی و نشانهٔ برخودرای از یک جوهر یا هویّت غیرمادی در نظر می‌گیریم) روزی معلوم خواهد شد که عاری از وجود واقعی بوده‌اند. ولی از نظر ما چنین استدلالِ صرفاً تمثیلی ضعیف‌تر از آن است که قادر به تمهید مبنای لازم برای نتیجه‌گیری فوق باشد. زیرا ما احساس می‌کنیم میان امور یا مفاهیم علمی چون فلوژیستون و اِتِر از یک طرف و امور ذهنی و درونی چون لذّت، درد، نیّت و امثالهم از طرف دیگر فرقی بنیادی نهفته است و قیاس آن‌ها جز قیاس مع‌الفارق نیست: دستیابی معرفتی ما به دوّمی‌ها، برخلاف اوّلی‌ها که غیرمستقیم و حصولی و حدسی است، مستقیم و شهودی و به‌تعبیر فلاسفهٔ اسلامی «حضوری» است. اگر قرار باشد این نوع امور حاضر در قوای معرفتی ما نیز به‌همان اندازهٔ اولی‌ها که طبق فرض هویّاتی بیرون از ما بوده و به‌همین سبب علم به آن‌ها مصون از خطا نیست، بتوانند غیرواقعی و خطا از آب درآیند، دیگر تقریباً هیچ گزاره و حکم معرفتی، از جمله و به‌نحو اولی‌تر خود همین فرض ماتریالیسم حذفی مبنی‌بر این‌که چیزی جز ماده وجود ندارد، از سوی آدمی قابل بیان و صدور نخواهد بود. جالب است که چرچلندها که درواقع حاضر نمی‌شوند یافته نشدنِ امکان‌ناپذیریِ تبیین آگاهی بر پایهٔ ماده را به‌معنی نبود چنین تبیینی در نظر بگیرند، به‌سهولت این‌که آن‌ها چیزی جز مادهٔ محض نمی‌یابند را به این معنی می‌گیرند که پس مطلقاً و ابداً هیچ‌گونه جوهر غیرمادی در کار نیست.

بدین‌ترتیب، به‌اعتقاد ما مدعایی که درخصوص موضوع تحت بررّسی این مقاله می‌توان صادر کرد، صرفاً به‌نحو مشروط خواهد بود: می‌توان حکم به امکان وقوع هوش مصنوعی قوی در ماشین (غیربیولوژیکی) داد. مشروط‌بر این‌که چیزی

همچون فرض مبنایی ماتریالیسم حذفی یا در صورت صحّت دوگانه‌گرایی (اعتقاد به دو نوع جوهر مادی و غیرمادی) لااقل این فرض صحیح بوده باشد که امور و حالات ذهنی بشری (هوش طبیعی) تماماً متکی به بخش مادی بوده و این امور جز یک‌سری فرایندها و حالات فیزیکی محض نیستند. روشن است که خروج این حکم از حالت مشروط منوط به موضع فلسفی خواهد بود که فرد مایل است نسبت به چیزی چون ماتریالیسم حذفی و ساختار و سخت‌افزار تشکیل‌دهندهٔ ذهن اتّخاذ نماید (هرچند با توجه به جمیع مباحث و تحلیل‌های فوق و همچنین شهودات قوی که ما نسبت به حالات ذهنی خود داریم، ماتریالیسم حذفی به‌نظر ما موضعی محکم و مُدلَّل نمی‌آید). ضمن این‌که بی‌تردید مسیر بالفعل پیشرفت آتی داستان هوش ماشینی نیز خود ممکن است به قضاوت قطعی‌تر ما در این زمینه یا حداقل در خصوص ساختار (مادی یا غیرمادی) ذهن کمک نماید.

نتیجه‌گیری

طبق آنچه گذشت هوش مصنوعی عمومی یا قوی رقم نخواهد خورد، مگر این‌که برخی ملزومات (معرفتی) آن محقّق شده باشند. یکی از این ملزومات مجهز شدنِ ماشین به استنتاج بر پایهٔ بهترین تبیین است. خود این محقّق نخواهد شد مگر آن‌که توانایی‌های کیفی همچون تخیّل، خلاقیّت، فهمیدن، آگاهی، تفکّر، شهود و شمّ خوب به ماشین (فیزیکیِ محض و البته غیربیولوژیکی) اعطاء گردد. اما بررسی مهمترین نمایندهٔ این توانمندی‌ها یعنی «آگاهی» نشان داد که نگاه خوش‌بینانه به هوش مصنوعی (قوی) با مسائل صعب‌العبوری مواجه است: حکم پیشینی به امکان وقوع ماشینِ آگاه و هوش مصنوعی قوی منوط به صحّت مواضع فلسفی چون «ماتریالیسم (حذفی)» یا حداقل این فرض است که «ذهن انسان به‌کلی از عناصر و فرایندها و امور فیزیکی محض شکل یافته و در عملکرد آن هیچ امر غیرفیزیکی و غیرمادی دخالت ندارد.» اما چنین پیش‌فرض‌هایی چنان سنگین هستند که اساساً اثبات یا حتّی پشتیبانی آن‌ها با ادلهٔ نسبتاً محکم هم مقدور به‌نظر نمی‌آید. لذا گویی حکم پیشینی به امکان وقوع هوش مصنوعی قوی نمی‌تواند چندان با قطعیّت صادر شود و بیشتر به مواضع فلسفی فرد نسبت به ذهن انسانی و ساختار آن وابسته است. البته ناگفته پیداست که اگر توسعهٔ هوش مصنوعی نهایتاً در مسیر بالفعل خود خبر از تجهیز ماشین به توانمندی‌های کیفی مورد بحث دهد، و ما نیز آزمونی برای احراز چنین موفقیّتی داشته باشیم، ابهام و تصمیم‌ناپذیری قاطعانه این قصه رخت برخواهد بست.

به‌هرحال، چنان‌که معلوم گشت، مسائل اعطاءِ توانمندهای کیفی به ماشین تأثیری بر وقوع هوش مصنوعی ضعیف نداشته و درواقع این نوع هوش اینک نیز به وقوع پیوسته و هرروز هم در حال ارتقاء و اوج‌گیری است. این نوع هوش حتی می‌تواند در ابعاد صوری‌محاسباتی از هوش انسانی پیشی نیز بگیرد، که این را تا همین الان عملاً هم نشان داده است.

منابع

- Bonjour, L. (1998), In Defence of Pure Reason, Cambridge: Cambridge University Press.
- Chakravartty, A. (2017), Scientific Ontology: integrating naturalized metaphysics and voluntarist epistemology, Oxford University Press.

- Chalmers, D. J. (1997), The Conscious Mind: In Search of a Fundamental Theory, Oxford Paperbacks.
- Chalmers, D. (2017), "The Hard Problem of Consciousness", in S. Schneider and M. Velmans (eds.), The Blackwell Companion to Consciousness, Wiley-Blackwell, 32-42.
- Chalmers, D. (2022), The Mystery of Consciousness: A Dialogue Between a Neuroscientist and a Philosopher (D. Chalmers and A. Damasio), in M. Gleiser (2022), 1-25.
- Churchland, Paul (2013), Matter and Consciousness, 3rd ed., Cambridge (Massachusetts), London: Mit Press.
- Churchland, Patricia (1986), Toward a Unified Science of the Mind-Brain, Cambridge, MA.
- Churchland, Patricia (1996), "The Hornswoggle Problem", Journal of Consciousness Studies, 3(5-6), 402-408.
- Churchland, Paul, and Churchland, Patricia (1990), "Could a Machine Think?", Scientific American 262, 32-37.
- Damasio, A. (2022), The Mystery of Consciousness: A Dialogue Between a Neuroscientist and a Philosopher (D. Chalmers and A. Damasio), in M. Gleiser (2022), 1-25.
- Dreyfus, Hubert (1992), What Computers Still Can't Do: A Critique of Artificial Reason, The MIT Press.
- Dreyfus, Hubert & Dreyfus, Stuart (1986), Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer, Simon and Schuster.
- Dreyfus, Hubert, & Dreyfus, Stuart (1988), "Making a Mind Versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint", Daedalus, vol. 117 (1), 15–43.
- Gleiser, M. (ed.) (2022), Great Minds Don't Think Alike: Debates on Consciousness, Reality, Intelligence, Faith, Time, AI, Immortality, and the Human, Columbia University Press.
- Harman, G. H. (1965), "The Inference to the Best Explanation", The Philosophical Review, 74(1), 88-95.
- Larson, E. J. (2021), The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do, Harvard University Press.
- Launchbury, J. (2017, February 15), A DARPA perspective on artificial intelligence, Retrieved September 8, 2023, from https://www.youtube.com/watch?v=-O01G3tSYpU&t=313s
- Nagel, T. (1974), "What Is It Like to Be a Bat?", The Philosophical Review, 83(4), 435-450.
- Peirce, C. S. (1992). The Essential Peirce: Selected Philosophical Writings, Vol. 1 (1867–1893), edited by Nathan Houser and Christian Kloesel, Indiana University Press.
- Peirce, C. S. (1998). The Essential Peirce: Selected Philosophical Writings, Vol. 2 (1893–1913), edited by the Peirce Edition Project: Indiana University Press.
- Searle, J. R. (1980), "Minds, Brains, and Programs", Behavioral and Brain Sciences 3(3), 417-424 and 450-457.
- Spitzer, E. (2016), "Tacit Representations and Artificial Intelligence: Hidden Lessons from an Embodied Perspective on Cognition", in V. Müller (ed.), Fundamental Issues of Artificial Intelligence, Springer, 425-441.
- Turing, A. M. (1950), "Computing Machinery and Intelligence", Mind 59, 433-460.