Islamic Maaref University

Scientific Journal

ANDISHE-E-NOVIN-E-DINI

Vol. 20, Summer 2024, No. 77

The Feasibility of the Theory of "Achieving Strong Artificial Intelligence" in Hawkins' Thought with an Emphasis on the Physicalism of the Mind and the Category of Consciousness

Mojtaba Dehghani[1] \ Mohammad Javad Asghari[2]

1. Phd student, Islamic Maaref University, Qom, Iran (corresponding author).
mojtaba.dehghani.1995@gmail.com
2. Assistant professor, Baqer al-Olum University, Qom, Iran.
asgharii51@gmail.com

Abstract Info

Article Type:
Research Article

Received:
2024.04.23
Accepted:
2024.12.02

Abstract

Strong artificial intelligence is the title of a program with the aim of producing systems with human cognitive capabilities such as awareness and perception. The problem of this research is that most researchers in this field consider mindfulness and consciousness to be the product of physical processes of the brain. Affected by this point of view, Jeff Hawkins has claimed to discover the algorithm of brain consciousness and intelligence by presenting a physical analysis of the nature of the mind in a hypothesis called the thousand brain. In Hawkins' view, the category of consciousness has a deep connection with the ability to recognize a place. According to him, the systems designed with this algorithm will have awareness like a human. In this article, with an analytical-critical method, the possibility of achieving such systems was investigated, and the findings of the research show that due to problems such as absence of perception, problem of qualia, subjectivity, free will, altruism and Chinese room argument, Hawkins' hypothesis has not been able to solve the mystery of consciousness, and the theory of achieving strong artificial intelligence with the help of a thousand-brain algorithm is not possible.

Keywords Hawkins, Mind, Strong Artificial Intelligence, Physicalism, Qualia, Thousand Brains.

Cite this article: Dehghani, Mojtaba & Mohammad Javad Asghari (2024). The Feasibility of the Theory of "Achieving Strong Artificial Intelligence" in Hawkins' Thought with an Emphasis on the Physicalism of the Mind and the Category of Consciousness. *Andishe-E-Novin-E-Dini*. 20 (2). 91-104. DOI: https://doi.org/10.22034/20.77.89

DOI: https://doi.org/10.22034/20.77.89

Publisher: Islamic Maaref University, Qom, Iran.

مجلة علمية
الفكر الديني الجديد
جامعة المعارف الإسلامية
السنة ٢٠ / الصيف عام ١٤٤٦ / العدد ٧٧

مدي جدوي نظرية «تحقيق الذكاء الاصطناعى القوى» فى فكر هوكينز مع التركيز علي مادية العقل وفئة الوعى

مجتبي دهقاني [١] / محمدجواد اصغري [٢]

١. طالب دكتوراه في قسم تدريس المعارف الإسلامية، فرع الأسس النظرية للإسلام، جامعة المعارف الإسلامية، قم، ايران (الكاتب المسؤول). *mojtaba.dehghani.1995@gmail.com*

٢. أستاذ مساعد في قسم الفلسفة والكلام، جامعة باقر العلوم عليه السلام، مدرس المعارف الإسلامية، قم، ايران. *asgharii51@gmail.com*

معلومات المادة

نوع المقال: بحث

تاريخ الاستلام: ١٤٤٥/١٠/١٤

تاريخ القبول: ١٤٤٦/٠٥/٣٠

ملخّص البحث

الذكاء الاصطناعي القوي هو عنوان برنامج يهدف إلى إنتاج أنظمة ذات قدرات معرفية شبيهة بالقدرات المعرفية البشرية، مثل الوعي والإدراك. وتكمن مسألة البحث في أن معظم الباحثين في هذا المجال يعتبرون اليقظة والوعي نتاج العمليات الفيزيائية للدماغ. متأثرًا بوجهة النظر هذه، فادعى جيف هوكينز أنه اكتشف خوارزمية الوعي والذكاء في الدماغ من خلال تقديم تحليل فيزيائي لطبيعة العقل في فرضية تسمى ألف عقل. من وجهة نظر هوكينز، فإن فئة الوعي لها علاقة عميقة بالقدرة على التعرف على مكان ما. ووفقا له، فإن الأنظمة المصممة بهذه الخوارزمية سيكون لديها وعي مثل الإنسان. في هذه البحث، وبأسلوب تحليلي نقدي، تم دراسة إمكانية تحقيق مثل هذه الأنظمة، وأظهرت نتائج البحث أنه بسبب مشاكل مثل؛ غياب الإدراك، مسألة الكواليا،الذاتية، الإرادة الحرة، الإيثار وحجة الغرفة الصينية، فرضية هوكينز لم تتمكن من حل لغز الوعي، ونظرية تحقيق ذكاء اصطناعي قوي بمساعدة خوارزمية الألف دماغ غير ممكن.

الألفاظ المفتاحية هوكينز،العقل، الذكاء الاصطناعي القوي، الفيزيائية، الكواليا، الف عقل.

الاقتباس: دهقاني، مجتبي و محمدجواد اصغري (١٤٤٦). مدى جدوى نظرية «تحقيق الذكاء الاصطناعي القوي» في فكر هوكينز مع التركيز على مادية العقل وفئة الوعي. *مجلة الفكر الديني الجديد*. ٢٠ (٢). ١٠٤ ـ ٩١. DOI: https://doi.org/10.22034/20.77.89

رمز DOI: https://doi.org/10.22034/20.77.89

الناشر: جامعة المعارف الإسلامية، قم، ايران.

نشریه علمی
اندیشه نوین دینی
سال ۲۰، تابستان ۱۴۰۳، شماره ۷۷

امکان سنجی تئوری «دستیابی به هوش مصنوعی قوی» در اندیشه «هاوکینز» با تأکید بر فیزیکالیسم ذهن و مقوله آگاهی

مجتبی دهقانی[۱] / محمدجواد اصغری[۲]

۱. دانشجوی دکتری گروه مبانی نظری اسلام، دانشگاه معارف اسلامی، قم، ایران (نویسنده مسئول).
mojtaba.dehghani.1995@gmail.com

۲. استادیار گروه فلسفه و کلام دانشگاه باقرالعلوم علیه السلام، مدرس معارف اسلامی، قم، ایران.
asgharii51@gmail.com

اطلاعات مقاله

نوع مقاله: پژوهشی (۱۰۴ ـ ۹۱)

تاریخ دریافت:
۱۴۰۳/۰۲/۰۴

تاریخ پذیرش:
۱۴۰۳/۰۴/۱۷

چکیده

هوش مصنوعی قوی عنوان برنامه‌ای است با هدف تولید سیستم‌هایی با قابلیت‌های شناختی انسان نظیر؛ آگاهی و ادراک. مسئله تحقیق این است که اکثر محققانِ این عرصه، ذهنمندی و آگاهی را محصول فرایندهای فیزیکی مغز می‌دانند. جف هاوکینز متأثر از این دیدگاه، در فرضیه‌ای به نام هزارمغز، با ارائه تحلیلی فیزیکال از ماهیت ذهن، مدعی کشف الگوریتم آگاهی و هوشمندی مغز شده است. در نگاه هاوکینز مقوله آگاهی با توانایی تشخیص مکان، ارتباط عمیقی دارد. به باور او سیستم‌هایی که با این الگوریتم طراحی شوند، همانند یک انسان، واجد آگاهی خواهند بود. در این مقاله با روشی تحلیلی ـ انتقادی، امکان وقوعی دستیابی به چنین سیستم‌هایی مورد بررسی قرار گرفت و یافته‌های تحقیق نشان می‌دهد که با توجه به معضلاتی نظیر؛ تجرد ادراک، مسئله کوالیا، سابجکتیویتی، اراده آزاد، حیث التفاتی و استدلال اتاق چینی، فرضیه هاوکینز از عهده حل معمای آگاهی بر نیامده، و تئوری دستیابی به هوش مصنوعی قوی به کمک الگوریتم هزارمغز امکان وقوعی ندارد.

واژگان کلیدی هاوکینز، ذهن، هوش مصنوعی قوی، فیزیکالیسم، کوالیا، هزارمغز.

استناد: دهقانی، مجتبی و محمدجواد اصغری (۱۴۰۳). امکان سنجی تئوری «دستیابی به هوش مصنوعی قوی» در اندیشه «هاوکینز» با تأکید بر فیزیکالیسم ذهن و مقوله آگاهی. *اندیشه نوین دینی*. ۲۰ (۲). ۱۰۴ ـ ۹۱.
DOI: https://doi.org/10.22034/20.77.89

کد DOI: https://doi.org/10.22034/20.77.89

ناشر: دانشگاه معارف اسلامی، قم، ایران.

طرح مسئله

بدون تردید برنامه جامع اِی‌آی یا همان «هوش مصنوعی»[1] مهم‌ترین و البته پرسروصداترین مسئله علمی در دهه‌های اخیر بوده است. هوش مصنوعی به‌عنوان یکی از رشته‌های شش‌گانه «علوم شناختی»،[2] تاریخچه‌ای کوتاه اما پرفرازوفرود دارد. اولین بار در سال ۱۹۵۵ جان مک‌کارتی، استاد ریاضی کالج دارتموث، عبارت هوش مصنوعی را اختراع کرد (داون‌پورت، ۱۴۰۰: ۱۳). این رشته از زمان پیدایش تاکنون چندین بار دچار چرخه خوش‌بینی و به‌دنبال آن بدبینی شده است. محققان به این چرخه‌ها «تابستان هوش مصنوعی» و «زمستان هوش مصنوعی» می‌گویند (Hawkins, 2021: 136).

محققان هوش مصنوعی دو مسیر متفاوت را دنبال می‌کنند: برخی پژوهش‌ها با این هدف صورت می‌گیرد که روبات‌ها و رایانه‌ها فقط در برخی امور مثل تشخیص سلول‌های سرطانی، عملکردی بهتر از انسان داشته باشند؛ اما مسیر دوم به‌دنبال ساخت ماشین‌هایی با توانایی ادراک، آگاهی، فهم، استدلال و تعقل است. این مسیر برنامه «هوش مصنوعی قوی»[3] نامیده می‌شود (Ibid: 138).

در مسئله هوش مصنوعی قوی، اصلی‌ترین مبنا «فیزیکالیسم ذهن»[4] یا به تعبیر دیگر، تفسیر مادی و فیزیکی از ذهن و حالات ذهنی نظیر هوشمندی، آگاهی، ادراکات حسی و غیرحسی، عواطف و احساسات است. برخی محققان این عرصه با پیش‌فرض گرفتن تک‌ساحتی بودن انسان در تلاش‌اند هوش مصنوعی را به هوش انسانی نزدیک کنند.

در مقابلِ مادی‌گرایی و فیزیکالیسم، دیدگاهی موسوم به «دوئالیسم»[5] یا «دوگانه‌انگاری» قرار دارد. در جهان غرب «رنه دکارت» مشهورترین دوگانه‌انگار است. دکارت معتقد بود که انسان دارای دو جوهر کاملاً متمایزازهم است: جوهر نفس و جوهر بدن. او می‌گفت:

> من مفهوم واضح و متمایزی از خودم دارم. من فقط چیزی هستم که می‌اندیشد و امتداد ندارد، و از سوی دیگر مفهوم متمایزی هم از بدن دارم، از آن حیث که دارای امتداد و عاری از فکر است. پس بدون تردید، این من به‌کلی از بدنم متمایز است و می‌تواند بدون بدن وجود داشته باشد (دکارت، ۱۳۹۱: ۹۹).

دکارت درباره تمایز نفس از بدن این‌گونه استدلال می‌کرد:

---

1. Artificial intelligence.
2. Cognitive science.
3. Artificial General Intelligence.
4. Physicalism of the mind.
5. Dualism.

> میان نفس و بدن تفاوتی عظیم وجود دارد، از آن جهت که جسم بالطبع همواره قسمت‌پذیر است و نفس به‌هیچ‌روی قسمت‌پذیر نیست (همان: ۱۰۸ ـ ۱۰۷).

طولی نکشید که دیدگاه وی با معضلی به نام «معضل تعامل ذهن ـ بدن» مواجه شد. چگونه نفس مجردِ فاقد امتداد می‌تواند بر بدن فیزیکیِ دارای امتداد تأثیر بگذارد و خود نیز از آن تأثیر بپذیرد؟ چگونه بین دو جوهر کاملاً متمایز که در منتهای تباین و جدایی قرار دارند، رابطه علّی برقرار است؟ (مسلین، ۱۳۹۹: ۲۳)

این معضل با برخی یافته‌های علمی عصب‌شناختی در بین سال‌های ۱۸۴۸ تا ۱۹۳۰، که در خصوص تأثیر آسیب‌های مغزی بر تحولات روحی و شخصیتی افراد صورت پذیرفت قوت گرفت. یکی از عصب‌پژوهان معاصر دراین‌باره می‌نویسد:

> دکارت هیچ‌وقت به بخش بیماران اعصاب پا نگذاشته بود. اگر این کار را می‌کرد، متوجه می‌شد که وقتی مغز آدم‌ها تغییر پیدا کند، شخصیتشان نیز عوض می‌شود. برخی از آسیب‌های مغزی سبب پیدایش افسردگی یا شیدایی می‌شوند. انواع دیگر آسیب مغزی فرد را به‌شدت مذهبی یا شوخ‌طبع می‌کنند یا سبب ایجاد گرایش به قماربازی می‌شوند. برخی نیز فرد را به بی‌تصمیمی، هذیان‌گویی یا پرخاشگری گرفتار می‌کنند. بنابراین، در نظر گرفتن چارچوبی که در آن مسائل ذهنی از جنبه‌های فیزیکی و جسمی جدا باشند دشوار است (ایگلمن، ۱۳۹۷: ۳۶).

در چنین شرایطی، آنچه در بدو امر مد نظر اندیشمندان قرار گرفت، نظریه «این‌همانی نوعی»[1] بود. بر اساس این نظریه، نوعی از حالات ذهنی با نوعی از فرایندهای مغزی یکی هستند. به‌عنوان مثال، حالت ذهنی درد همان فرایند شلیک عصب C در مغز و سیستم اعصاب مرکزی است؛ اما این دیدگاه نیز به‌رغم شواهد و مؤیدات، مورد تردید واقع شد؛ زیرا از آنجا که رابطه این‌همانی رابطه وجودی است، نه رابطه زبانی و معنایی، لازم است این رابطه تحت هر شرایطی برقرار باشد، درحالی‌که چنین نیست (همتی مقدم، ۱۳۹۱: ۲۸). برخی از حیوانات حالت ذهنی درد را ادراک می‌کنند، درحالی‌که فاقد عصب C هستند. این معضل سبب شد نظریه «این‌همانی نوعی» به «این‌همانی مصداقی»[2] تعدیل شود؛ به این معنا که مصداقی از یک حالت ذهنی با مصداقی از یک حالت فیزیکی این‌همان است. به‌عنوان مثال، مصداقی از درد، مانند درد در ناحیه‌ای از بدن انسان، با مصداقی از حالت فیزیکی، مانند شلیک رشته عصبی C در مغز انسان، این‌همان است، نه اینکه نوع درد همان نوعِ شلیک رشته عصبی C باشد (مسلین، ۱۳۸۷: ۱۲۸).

---

1. Type identity.
2. Token identity.

با مقبولیت این‌همانی مصداقی و نیز هم‌زمان با گسترش علوم رایانه، دیدگاه جدیدی از سوی برخی فیزیکالیست‌ها ارائه شد که زمینه‌ساز طرح کلی برنامه هوش مصنوعی است. بر اساس این دیدگاه، ذهن برنامه‌ای است که روی سخت‌افزار مغز اجرا می‌شود و ورودی‌های حسی را به خروجی‌های رفتاری تبدیل می‌کند. بنابراین، تنها مغز انسان کارکرد هوش را پدید نمی‌آورد، بلکه می‌توان این کارکرد را در مغزهایی با ساختار سلولی متفاوت یا سیستم‌های رایانه‌ای پدید آورد. این نظریه «کارکردگرایی»[1] نام دارد (کرباسی‌زاده، ۱۳۹۹: ۷۴ ـ ۷۱).

«جف هاوکینز» از محققانی است که با نگرش فیزیکالیستی به ذهن و اتخاذ رویکردی کارکردگرایانه، بر این باورند که سلول‌های مغزی بر مبنای الگوریتم خاص، ذهنمندی و آگاهی را به وجود می‌آورند. لذا ذهنمندی محدود به مغز و ساختار زیستی انسان نیست، بلکه می‌توان از طریق شبیه‌سازی الگوریتم مغز در دستگاه‌های فیزیکی متفاوتی مانند رایانه‌ها ذهنمندی و آگاهی را به وجود آورد. او دراین‌باره می‌نویسد:

> من از همان زمانی که به مطالعه مغز علاقه‌مند شدم، احساس کردم که اول باید طرز کار آن را بفهمیم تا بعد بتوانیم ماشین‌های باهوش بسازیم. این مطلب به نظر من بدیهی بود؛ چون مغز تنها چیز هوشمندی است که می‌شناسیم (Hawkins, 2021: 141).

تبیین دیدگاه هاوکینز

هاوکینز متأثر از «داروینیسم»[2] و همسو با جریان «عصب‌پژوهی»[3] معاصر بر این باور است که در فرایند «فرگشت»،[4] مغز انسان نیز همانند گونه انسان، به تکامل رسیده و از مغز سایر جانداران متمایز شده است؛ منتها با این تفاوت که در فرگشت گونه‌ها غالباً با ظاهر شدن گونه جدید، گونه‌های قدیمی از بین می‌روند، اما در فرگشت مغز بخش‌های جدید روی بخش‌های قدیمی افزوده می‌شوند (باربور، ۱۳۹۷: ۵۶۶). مغز قدیم یا همان ساقه مغز مسئول کنترل رفتار ابتدایی و غیرهوشمندانه انسان مانند گوارش، تنفس و رابطه جنسی است و تمام قابلیت‌هایی که از نظر ما به‌عنوان هوش تعبیر می‌شود، از قبیل بینایی، زبان، ریاضیات، علم و مهندسی، به وسیله مغز جدید یا «نوقشر»[5] ایجاد می‌شود. بدیهی است که ابتدا باید نحوه عملکرد نوقشر را بفهمیم تا بعد بتوانیم الگوریتم هوش را به دست آوریم (Ibid: 26).

هاوکینز می‌گوید: معلومات ما صرفاً به صورت تلّی از دانسته‌ها ذخیره نمی‌شود، بلکه به‌گونه‌ای

1. Functionalism.
2. Darwinism.
3. Neuroscience.
4. Evolution.
5. Neocortex.

سازمان‌دهی می‌شود که منعکس‌کننده ساختار جهان و اشیای درون آن باشد. مثلاً برای دانستن اینکه دوچرخه چیست، لیستی از دانستنی‌ها درباره دوچرخه به ذهنمان نمی‌آید، بلکه مغز ما مدلی از دوچرخه ایجاد می‌کند که شامل قطعات مختلف آن، آرایش آنها در برابر یکدیگر و چگونگی حرکت و کار کردن آنها با هم است. این مدل مبنای ادراکات ماست. از نظر هاوکینز هوش ارتباط نزدیکی با مدل مغز از جهان دارد. ازاین‌رو، برای فهمیدن اینکه مغز چگونه هوش را ایجاد می‌کند، باید دریابیم که مغز چگونه مدلی از جهان و اشیای درون آن را یاد می‌گیرد. درواقع پرسش اصلی این است: بخش جدید مغز انسان یا همان نوقشر که متشکل از هزاران ستون قشری تقریباً یکسان است، چگونه یک مدل از جهان را یاد می‌گیرد؟ (Ibid: 46 - 49)

او برای پاسخ به این سؤال ابتدا دو انگاره علوم اعصاب را بیان می‌کند: انگاره اول این است که افکار، ایده‌ها و ادراکات فعالیت‌های نورون‌ها هستند. او می‌نویسد هر فکری که دارید، فعالیت نورون‌هاست. هر چیزی هم که می‌بینید یا می‌شنوید یا حس می‌کنید، فعالیت نورون‌هاست. فعالیت‌های ذهنی ما و فعالیت نورون‌ها هر دو یک چیز هستند؛ انگاره دوم می‌گوید هر چیزی که می‌دانیم، در اتصالات بین نورون‌ها ذخیره می‌شود. اتصالات درون مغز مدل ما از جهان را ذخیره می‌کنند (Ibid: 51 - 54). سپس هاوکینز انگاره سومی را مطرح می‌کند: نورون‌ها از طریق ایجاد یک «چارچوب مرجع»،[1] قدرت پیش‌بینی ورودی‌های بعدی را پیدا می‌کنند و از این طریق، مغز انسان هوشمند و آگاه می‌شود. چارچوب مرجع ابزاری فرضی است که فیزیک‌دانان به کمک آن، موقعیت مکانی و زمانی یک ذره متغیر را معیّن می‌کنند. هاوکینز معتقد است سلول‌های مغز از طریق ایجاد چارچوب‌های مرجع، موقعیت مکانیِ خودِ جانور و موقعیت مکانی اشیای دیگری را که پیرامون آن جانور وجود دارند معیّن می‌کنند و درنتیجه جانور به وسیله مغز هوشمند می‌شود (Ibid: 64 - 67).

او می‌نویسد هوشمندی یک موجود با دانستن مکانِ خود و مکان موجودات و اشیای دیگر ارتباط مستقیم دارد. در تنازع بقا، جانوری که هوش بیشتری نسبت به سایرین داشته باشد، بخت بیشتری برای بقا خواهد داشت. هوشمندی جانور مبتنی بر این است که بتواند اطلاعات هر مکان را در مغز خود یادداشت کند. مثلاً اگر در مکانی توانست غذا پیدا کند، بتواند دوباره به آن مکان بازگردد یا اگر در مکانی با خطری مواجه شد، از رفتن به آن مکان خودداری کند. جانوری می‌تواند چنین قابلیتی داشته باشد که چارچوب مرجعی برای جهان خود داشته باشد. الگوریتم چارچوب مرجع برای بقای جانوران مفید بوده است. جانوری که می‌داند کجاست، جانوری که همیشه مکان خود را در محیطش می‌داند، می‌تواند به یاد آورد که دفعه قبل کجا غذا یافته و از چه جاهایی به‌عنوان سرپناه استفاده کرده است. این جانور می‌تواند از مکان فعلی خود به این مکان‌ها و دیگر جاهایی که قبلاً رفته است برود. چنین جانوری هوشمند و آگاه است (Ibid: 74).

1. Frame of reference.

هاوکینز چارچوب مرجع را به‌عنوان روشی برای سازماندهی هر نوع دانش در نظر می‌گیرد. از نظر او از سازوکار چارچوب‌های مرجع برای سازمان‌دهی اشیایی که قابل حس نیستند هم می‌توان استفاده کرد؛ زیرا در ابتدا چارچوب مرجع در مغز برای یاد گرفتن ساختار محیط تکامل یافت تا انسان بتواند در جهان اطراف حرکت کند. سپس مغز تکامل پیدا کرد و از همین سازوکار برای یاد گرفتن ساختار اشیای فیزیکی بهره برد تا ما بتوانیم آنها را تشخیص دهیم و با آنها کار کنیم. در آخر، مغز یک بار دیگر هم تکامل یافت تا از آن سازوکار برای یادگیری و بازنمایی اشیای مفهومی مانند مردم‌سالاری و ریاضیات استفاده کند. لذا چارچوب‌های مرجع برای مدل‌سازی همه دانسته‌های ما کاربرد دارند، نه‌فقط اشیای فیزیکی و نه‌فقط تجربیات حسی (Ibid: 87 - 89).

هاوکینز با مقدمه قرار دادن نظریه این‌همانیِ فعالیت‌های ذهنی با فرایندهای مغزی، فرضیه‌ای را پیشنهاد می‌کند که بر اساس آن بتواند مهم‌ترین کارکرد ذهن، یعنی آگاهی و ادراک را تبیین کند. در نگاه او، از آنجا که فعالیت‌های ذهنی همان فعالیت‌های مغزی هستند، آگاهی ما به‌عنوان فعالیت ذهنی، فرایند مغزی است. آگاهی درحقیقت کارکردی ذهنی است که توسط مغز و به وسیله چارچوب‌های مرجع ایجاد می‌شود. لذا می‌توان با استفاده از این الگوریتم، و مهندسی معکوس آن روی سیستم‌های رایانه‌ای، رؤیای هوش مصنوعی قوی را به واقعیت تبدیل کرد.

سیستم‌هایی که بتوانند همانند جانوران، موقعیت مکانی خود و اشیای دیگر را تشخیص دهند و اطلاعات هر موقعیت را به‌درستی ذخیره کنند و این اطلاعات را مبنای تصمیم‌گیری بعدی خود قرار دهند، واجد سطحی از آگاهی و ادراک خواهند بود. با تکامل این الگوریتم، روبات‌ها به درک مفاهیم و امور غیرحسی نیز نائل می‌شوند و به عالی‌ترین سطح آگاهی دست می‌یابند.

هاوکینز دراین‌باره می‌نویسد:

> من معتقدم که آینده هوش مصنوعی مبتنی بر اصول مغز خواهد بود. ماشین‌های حقیقتاً هوشمند درست مانند نوقشر مدل‌هایی از جهان را با استفاده از چارچوب‌های مرجعِ نقشه‌مانند یاد خواهند گرفت. به نظر من، این گریزناپذیر است. فکر نمی‌کنم راه دیگری برای ایجاد ماشین‌های حقیقتاً باهوش وجود داشته باشد (Ibid: 143).

بررسی دیدگاه هاوکینز

دیدگاه هاوکینز مبتنی بر فیزیکالیسم ذهن و آگاهی است. ازاین‌رو، ابتدا به بررسی و نقد این رویکرد می‌پردازیم و سپس امکان وقوعی دستیابی به هوش مصنوعی قوی را کنکاش می‌کنیم.

۱. نقد فیزیکالیسم ذهن و مقوله آگاهی

یک. معضل «کوالیا»[1]

فرضیه هاوکینز بر این اصل استوار است که ذهن چیزی جز مغز نیست. پدیده‌ها و حالات ذهنی نظیر هوشمندی و آگاهی نیز چیزی جز فرایندهای نورونی مغز و سیستم اعصاب نیستند. اگر مدعای فیزیکالیستی هاوکینز درست باشد، گزارشی کامل از نحوه کارکرد مغز و دستگاه عصبی گزارش نهایی، و نیز تبیین فیزیکیِ کامل از واقعیت تبیینی کامل و نهایی خواهد بود؛ زیرا بنا بر فرضِ فیزیکالیسم، اولاً هیچ امر غیرفیزیکی وجود ندارد و ثانیاً با یک تبیین فیزیکی کامل از پدیده‌ها، به هرچه فیزیکی است شناخت کامل پیدا کرده‌ایم. اما مسئله کوالیا این مدعا را رد می‌کند. معضل کوالیا تمام اَشکال فیزیکالیسم را تضعیف می‌کند (مسلین، ۱۳۸۷: ۲۲۳ ـ ۲۲۲).

کوالیا یا همان کیفیات ذهنی در اصطلاح فلاسفه به معنای کیفیتِ پدیداریِ تجربه است؛ یعنی جلوه و احساسی که فرد از ادراکات حسی‌اش دارد. کیفیت پدیداری مزه مانند شیرینی شکر و کیفیت پدیداری رنگ مانند قرمزیِ رنگ قرمز مثال‌هایی از کوالیا هستند (تامپسون، ۱۳۹۴: ۲۰۰). به تعبیر برخی از فلاسفه، کوالیا سازنده آگاهی است. وقتی ما به چیزی آگاهی داریم، مثلاً آگاهی ما به اینکه این سیب قرمز است، این تجربه آگاهانه ما از دیدن قرمزی، یک کوالیا یا یک کیفیتِ پدیداری خاص دارد که در واقعیت فیزیکی این رنگ موجود نیست. نحوه‌ای که رنگ قرمز به نظر می‌رسد، چیزی نیست که در طول موج این رنگ وجود داشته باشد. ما صرفاً با دانستن اطلاعات مربوط به طول موج رنگ قرمز، نحوه تأثیرگذاری آن بر شبکیه و نحوه عملکرد نورون‌ها نمی‌توانیم یک تبیین نهایی از دیدن رنگ قرمز ارائه کنیم. درواقع چیزی هست که ما آن را تبیین نکرده‌ایم و آن نحوه به نظر رسیدن رنگ قرمز است (کمبل، ۱۳۹۹: ۱۸۹).

فرض کنید شخصی به نام مری دانشمندی است که کل زندگی خود را در اتاقی سیاه‌وسفید گذرانده است. او جهان را از طریق نمایشگر سیاه‌وسفیدی می‌بیند که اطلاعات را از دوربین‌های بیرونی دریافت می‌کند. مری دانشمندی خبره در علوم اعصاب و متخصص مکانیسم‌های مغزی مربوط به بینایی است. او تمامی اطلاعاتی را که به لحاظ فیزیکی و کارکردی در انسان‌ها هنگام دیدن رنگ‌های گوناگون رخ می‌دهد می‌داند. او می‌داند که نور چگونه از سطوح اشیا منعکس می‌شود، چگونه به شبکیه تأثیر می‌گذارد، تغییرات گوناگونی که در عصب بینایی روی می‌دهد کدام‌اند و... . او تمام طول موج رنگ‌ها و اتفاقات درون شبکیه چشم و فعالیت‌های مغز و نورون‌ها را هنگام ادراک بصری رنگ می‌شناسد. به‌عنوان مثال هنگامی که شخصی گوجه‌فرنگی را می‌بیند، مری می‌تواند از آنچه در دستگاه بیناییِ آن شخص رخ می‌دهد،

1. Qualia.

توصیف نوروفیزیولوژیکیِ کاملی ارائه کند؛ اما مری تاکنون هیچ رنگی را ندیده است. روزی مری از دنیای تک‌رنگ خود خارج می‌شود و قرمزی گوجه‌فرنگی را تجربه می‌کند. مری با دیدن رنگ قرمز چیز تازه‌ای فراتر از جزئیات صرفاً فیزیکیِ ادراک رنگ فرامی‌گیرد؛ یعنی تجربه دیدن قرمزی. این آزمایش فکری به ما می‌گوید خانم مری چیز جدیدی درباره ادراک بصری رنگ قرمز می‌داند که قبلاً از تمام آن واقعیات فیزیکی نمی‌توانست بداند (مسلین، ۱۳۸۷: ۲۲۶). به عبارت دیگر، ادراک حسی رنگ قرمز ویژگی‌ای دارد که مری به آن دسترسی ندارد، و از آنجا که مری به تمام ویژگی‌های فیزیکی این ادراک حسی دسترسی دارد، آنچه از دسترس او دور می‌ماند باید ویژگی‌ای غیرفیزیکی باشد (کمبل، ۱۳۹۹: ۱۹۹).

همچنین کودکی را تصور کنید که بدون توانایی احساس درد متولد شده و در محیطی کاملاً تحت مراقبت زنده مانده است. این کودک هرگز نمی‌تواند بداند تجربه درد چه کیفیتی دارد. هر اندازه هم که درباره الگوی شلیک عصب‌ها در مغز و فرایندهای زیست شیمیایی اطلاعات داشته باشد، باز نمی‌تواند کاملاً درک کند که این نوع از احساس چیست (تای، ۱۳۹۳: ۳۵).

بنابراین، صرف احساس کردن رویدادها در ماده یا انرژی وجود ندارد. نحوه‌ای که رنگ قرمز به نظر می‌رسد چیزی نیست که در طول موج این رنگ وجود داشته باشد؛ همچنین نحوه‌ای که درد احساس می‌شود، چیزی نیست که در شلیک عصب‌های مغزی وجود داشته باشد. پس از آنجا که ماده و انرژی از کوالیا بیگانه هستند و ذهن ما مملو از کوالیاست، ذهن ما مادی نیست. حال که چنین است، آیا هوش مصنوعی که بر پایه الگوریتم مغز طراحی شده است، قادر به ادراک کیفیت رنگ قرمز خواهد بود؟

دو. معضل «سابجکتیویتی»[1]

معضل دشوار دیگری که فیزیکالیست‌هایی نظیر هاوکینز با آن مواجه‌اند، سابجکتیو بودن حالات آگاهانه ذهن است. از آنجا که حالات آگاهانه ذهن ماهیتی آبجکتیو ندارند، اولاً فیزیکی نیستند و ثانیاً از حوزه مطالعه علم خارج‌اند.

می‌توان فیزیکی نبودن حالات ذهنی را در قالب یک قیاس استثنایی به این صورت تبیین کرد که اگر حالات آگاهانه ذهن فیزیکی باشند، باید همچون تمام واقعیات فیزیکی ماهیتی آبجکتیو داشته باشند؛ زیرا آن چیزی که عینی نیست و ماهیتی آبجکتیو ندارد، نمی‌تواند فیزیکی باشد. این در حالی است که حالات آگاهانه ذهن ماهیتی آبجکتیو ندارند، بلکه ماهیتی سابجکتیو دارند. به‌عنوان مثال، آگاهی من به دندان‌دردی که دارم، یک حالت ذهنی است که کاملاً درونی و خصوصی است و تا زمانی که این درد را در چهره یا رفتار خود نشان ندهم، کسی از آن مطلع نخواهد شد (همتی مقدم، ۱۳۹۱: ۳۸ ـ ۳۷).

1. Subjectivity.

پس حالات آگاهانه ذهن فیزیکی نیستند. هر اندازه که درباره شلیک نورون‌های عصبی اطلاعات کسب کنیم، باز به نظر می‌رسد چیزی هست که تبیین نکرده‌ایم؛ یعنی اینکه چرا و چگونه این تغییرات عصبی آن احساس سابجکتیو را پدید می‌آورد. به تعبیر بهتر، فرایندهای مغزی هیچ شباهتی به تجارب آگاهانه ذهنی ندارند و تجارب ذهنی نیز هیچ شباهتی به فرایندهای مغزی ندارند. این تمایز نشان از غیرفیزیکی بودن آگاهی دارد.

اما به چه دلیل ادعا کردیم که حالات ذهنی از حوزه مطالعه علم خارج‌اند؟ به چه دلیل هرچه آبجکتیو نباشد نمی‌تواند موضوع مطالعات علمی قرار گیرد؟

در پاسخ به این سؤال ابتدا مقدمه‌ای کوتاه را عرض می‌کنیم. حالات ذهنی ما از این جهت که درونی و خصوصی هستند، از «منظر اول‌شخص» یعنی بدون واسطه و به نحو حضوری درک می‌شوند، اما حالات فیزیکی به جهت عینی و عمومی بودن، «از منظر سوم‌شخص» یعنی باواسطه و به نحو حصولی درک می‌شوند. بر این اساس، تمام مطالعات علمی از منظر سوم‌شخص صورت می‌گیرند؛ زیرا علم ما به حقایق عینی و فیزیکی، یک علم حصولی و باواسطه است. به‌عنوان مثال، ما صاعقه را نوری ناگهانی و درخشان در آسمان می‌بینیم. ممکن است موجودات دیگر آن را کاملاً متفاوت مشاهده کنند. بالفرض ممکن است موجودات فضایی صاعقه را به صورت نور فرابنفش ببینند که کاملاً با ادراک بصری ما متفاوت است. اما با وجود این، هم ما و هم آدم‌فضایی‌ها می‌توانیم توصیف آبجکتیو واحدی از صاعقه ارائه کنیم، یعنی هر دو توافق داریم که صاعقه «تخلیه بار الکتریکی» است، بی‌آنکه در خصوص ماهیت سابجکتیو و نحوه ادراک درونی آن توافقی داشته باشیم (کمبل، ۱۳۹۹: ۱۹۷ ـ ۱۹۰).

حال اگر بخواهیم همانند آقای هاوکینز و سایر فیزیکالیست‌ها، توصیف فیزیکی و آبجکتیو از آگاهی ارائه کنیم، به‌ناچار باید دیدگاه اول‌شخص درباره آگاهی را رها کنیم؛ زیرا تمام توصیفات فیزیکی از منظر سوم‌شخص رخ می‌دهند. با این حساب، مطالعه فرایندهای فیزیکی مغز هنگام ادراک آگاهانه، مستلزم رها کردن دیدگاه اول‌شخص است که معرِّف آگاهی است. هرچه دیدگاهمان به آگاهی آبجکتیوتر و تعبیری علمی‌تر باشد، از سابجکتیویته‌ای که معرِّف آگاهی است فاصله بیشتری می‌گیریم. درنتیجه امکان مطالعه علمی آگاهی منتفی است. مطالعه علمی منظر اول‌شخص را به نفع دیدگاهی آبجکتیو رها می‌کند (همان: ۱۹۸).

سه. معضل «اراده آزاد»[1]

مفهوم علیت در علم فیزیک به این معناست که حوادث طبیعی به‌طور یگانه متعیّن هستند. به تعبیری بهتر، قوانینی تخلف‌ناپذیر در طبیعت وجود دارند که حالت آینده هر سیستم را از روی وضعیت فعلی آن

1. Free will.

به‌طور یگانه تعیین می‌کنند. در عرف فیزیک‌دانان وقتی علیت به این مفهوم خاص به کار می‌رود، آن را «موجبیت»[1] می‌نامند (گلشنی، ۱۳۸۵: ۲۲۹ ـ ۲۲۷).

در نگاه فیزیکالیست‌ها، جهانی که ما در آن زندگی می‌کنیم واجد یک نظام موجبیتی است. پدیده‌های طبیعی جهان تبیین علّی دارند و این تبیین‌ها شروط علّی کافی به دست می‌دهند. به‌عنوان مثال، در تبیین پدیده‌ای مانند زلزله می‌گویند با توجه به نیروهایی که بر صفحات پوسته زمین عمل می‌کنند، امکان دیگری وجود ندارد؛ یعنی در آن شرایط چنان رویدادی می‌بایست اتفاق می‌افتاد. به عبارت دیگر، علل در آن شرایط، برای تعین بخشیدن به این رویداد کافی بوده‌اند. بنابراین، پیش‌فرض ما در مواجهه با طبیعت این است که هر اتفاقی که روی می‌دهد، پیامد شروطِ علّیِ کافیِ پیشین است (سرل، ۱۳۹۲ الف: ۵۱ ـ ۴۶).

حال اگر کنش‌های انسان صرفاً وقایعی فیزیولوژیک باشند که در مغز و سلسله اعصاب مرکزی رخ می‌دهند، باید همانند دیگر پدیده‌های فیزیکی معلول شروط علّی پیشین، و متعین و غیراختیاری باشند. این در حالی است که ما شهوداً و وجداناً خود را موجودی مختار می‌دانیم (پترسون، ۱۴۰۰: ۳۲۹). اختیار و اراده آزاد مقوله‌ای انکارناپذیر است. حتی نمی‌توان از به‌کارگیری آن اجتناب کرد؛ زیرا خودِ این اجتناب کردن نمونه‌ای از به‌کارگیری آن است.

در دانشگاه هاروارد[2] آزمایشی صورت گرفت که مؤید این استدلال است. گروهی داوطلب برای انجام آزمایش دعوت شدند. داوطلبان جلو صفحه‌های نمایش رایانه نشستند و دست‌های خود را به حالت کشیده روی میز قرار دادند. وقتی صفحه به رنگ قرمز درمی‌آمد، آنها می‌بایست تصمیم می‌گرفتند که کدام دست خود را حرکت دهند. بعد صفحه به رنگ زرد درمی‌آمد و سرانجام وقتی سبز می‌شد، داوطلبان اجازه داشتند تصمیم خود را برای به حرکت درآوردن دست راست یا چپ اجرا کنند. سپس محققان با استفاده از تحریک مغناطیسی از راه جمجمه، پالسی مغناطیسی برای تحریک ناحیه خاصی از مغزِ داوطلبان می‌فرستادند و به این ترتیب کورتکس حرکتی چپ یا راست را تحریک می‌کردند تا فرد حرکت را در دست چپ یا راست آغاز کند. وقتی نوار زرد ظاهر می‌شد و داوطلبان آماده بودند تا تصمیمی را که گرفته‌اند اجرا کنند، محققان پالس را وارد می‌کردند. ارسال پالس سبب می‌شد که داوطلبان صرفاً تمایل بیشتری برای حرکت دادن یکی از دو دست خود داشته باشند، نه اینکه به صورت جبری و غیراختیاری فقط و فقط یکی از دو دست را حرکت دهند. مثلاً تحریک کورتکس چپ باعث می‌شد که شرکت‌کنندگان به احتمال بیشتری دست راست خود را حرکت دهند، اما هرگز این احتمال تعیّن پیدا نکرد. به اعتراف محققان، بیشتر داوطلبان تصمیم خود را عوض کردند، نه همه آنها و آنهایی که تصمیم

1. Determinism.
2. Harvard University.

خود را عوض کردند در اینکه با اختیار و اراده آزاد این کار را کرده‌اند هیچ تردیدی نداشتند (ایگلمن، ۱۳۹۷: ۹۳).

۲. امکان‌سنجی تئوری هاوکینز

با توجه به کاستی‌های دیدگاه فیزیکالیستی هاوکینز، ایده دستیابی به هوش مصنوعی قوی امکان وقوعی نخواهد داشت؛ زیرا این ایده مبتنی بر تک‌ساحتی بودن انسان و نیز فیزیکی بودن مقوله آگاهی و ادراک است.

فلاسفه مسلمان درباره تجرد ادراک بحث‌های مفصلی را مطرح کرده‌اند. از نظر ایشان آگاهی‌ها و ادراکات ما به لحاظ وجودی مجرد از ماده هستند؛ زیرا خواص و ویژگی امور مادی نظیر تغیر و دگرگونی، انقسام‌پذیری، جزء داشتن، بُعد داشتن و ممتد بودن را ندارند (مصباح، ۱۳۹۰: ۱۶۲ ـ ۱۶۰).

موجودی که مجرد از ماده است محل حضور و تحققش نیز باید مجرد باشد. لذا سلول‌های مادی مغز، نمی‌توانند محل حضور ادراکات مجرد باشند (طباطبایی، ۱۴۴۳ ق: ۳۶۷). با این حساب، امکان پدید آوردن آگاهی روی سیستم‌های رایانه‌ای غیرممکن خواهد بود. علاوه بر این، برنامه هوش مصنوعی قوی با دو معضل دشوار مواجه است که به‌اختصار به آنها اشاره می‌کنیم.

یک. مسئله «حیث التفاتی»[1]

حالات آگاهانه ذهن حیث التفاتی دارند؛ یعنی به چیزی فراتر از خود اشاره می‌کنند، اما فرایندهای فیزیکیِ درون مغز قدرت بازنمایی کردن امور را ندارند (تامپسون، ۱۳۹۴: ۱۲۵ ـ ۱۲۴). به‌عنوان مثال، واژه‌ها و کلمات بدون معنادهی یک ذهن آگاه، که آنها را برای موجودات خارجیِ بدون معنا معیّن و وضع کند، به هیچ موجودی ارجاع نخواهد داد و این امر در امور فیزیکی نیز جریان دارد. یک تصویر به این دلیل به یک مرجع ارجاع می‌دهد که ما آن معنا را برای آن تصویر مشخص کرده‌ایم. مثلاً تصویر گُل بدون معنادهی ما، به هیچ گلی در جهان ارجاع نخواهد داد یا مهره‌های ستون اول چُرتکه که ارقام یکان را بازنمایی می‌کنند، بدون معنادهی از طرف ما، به‌خودی‌خود هیچ معنایی نمی‌دهد.

حال رایانه‌ای با هوش مصنوعی قوی را در نظر می‌آوریم. نمایشگر رایانه حروف یا اعدادی را به ما نشان می‌دهد که از طرف خودِ ما دارای معنا شده‌اند. این حروف و اعداد بدون به‌کارگیری معنای مشخص‌شده توسط ما، فقط الکترون‌های بی‌معنا خواهند بود. در درون یک رایانه جریان برق پالس‌های روشن ـ خاموش را طبق الگوریتم خاصی می‌فرستد که در زبان رایانه معنای ۰ و ۱ دارند؛ اما درست مانند مثال چرتکه، ما این معنا را برای پالس‌های برق معیّن کرده‌ایم. در غیر این صورت، خود پالسِ روشن ـ

1. Intentionality.

خاموش به معنای ۰ و ۱ نخواهد بود. جان سرل می‌نویسد:

> دلیل اینکه یک برنامه کامپیوتری هرگز نمی‌تواند یک ذهن باشد، صرفاً این است که یک برنامه کامپیوتری تنها برنامه‌ای دستور زبانی است و ذهن چیزی بیش از امری دستور زبانی است. ذهن‌ها اموری معناشناختی هستند؛ به این معنا که چیزی بیش از ساختار صوری دارند. آنها محتوا و مضمون دارند (سرل، ۱۳۸۲: ۶۰).

اگر مادی‌گرایی و فیزیکالیسم ذهن درست باشد، به‌عنوان مثال فکر ما درباره یک گُل ترکیب خاصی از الکترون‌ها، نورون‌ها و شاخه‌های عصبی مغز خواهد بود، منتهی برخلاف رایانه یا چرتکه، کسی معنادهی را برای این اجزا معیّن نکرده است. پس چگونه اجزایی بی‌معنا به چیزهایی دیگر ارجاع می‌دهند، بدون اینکه ذهنی دیگر لازم باشد تا این معنا را به کار گیرد؟ چگونه مغز به کمک چند نورون و ایجاد یک یا چند چارچوب مرجع، به چیز یا چیزهایی ارجاع می‌دهد که وجود خارجی ندارند یا در فاصله‌ای بسیار دورتر از ما قرار گرفته‌اند؟

پس از آنجا که رایانه‌ها ذاتاً قدرت بازنمایی کردن ندارند، مغز انسان نیز به‌تنهایی توانایی چنین کاری را نخواهد داشت. نورون‌ها و چارچوب‌های مرجعِ مغز بدون در نظر گرفتن ذهنی آگاه، هوشمند، مدرِک و مستقل از بدن، همانند الکترون‌ها و پالس‌های روشن ـ خاموشِ رایانه، نه خود واجد معنا هستند نه می‌توانند به چیزی معنا ببخشند. به‌عنوان مثال، فرض کنید جمعیت کشور چین تقریباً برابر با تعداد نورون‌های مغز باشد. حال با مدل‌سازی شبکه ارتباطی نورون‌ها روی مردم چین، به این صورت که هر فرد چینی را در حکم نورونی در نظر بگیریم که همان ارتباطات را با نورون‌های اطرافش از طریق تماس با تلفن‌های همراه برقرار می‌کند، آیا می‌توان نتیجه گرفت که کشور چین ذهن یا نفس دارد؟ آیا می‌توان نتیجه گرفت که کشور چین آگاهی یا خودآگاهی دارد؟ آیا کشور چین می‌تواند یک تصویر را بازنمایی کند؟ (دی‌هارت، ۱۳۸۱: ۹۰ ـ ۸۹) بر این اساس، اولاً انسان صرفاً موجودی مادی و فیزیکی نیست؛ ثانیاً در هوش مصنوعی خبری از آگاهی و هوش نیست.

دو. استدلال اتاق چینی

از آنجایی که آن شخص خودش اهل چین نیست، زبان چینی را براساس آن کتاب راهنما نمی‌فهمد، هرچند به کمک این کتاب، واژه‌های چینی را کنار هم می‌گذارد و جمله معنا داری را به وجود می‌آورد. اما از فهم معنای این جمله ناتوان است. در رایانه و هوش مصنوعی دقیقاً همین اتفاق رخ می‌دهد. هوش مصنوعی براساس یک الگوریتم (شبیه آن کتاب راهنما) اطلاعاتی را کنار یکدیگر گذاشته و در اختیار ما قرار می‌دهد اما هیچ گونه فهم و درکی از محتوایی که ارائه کرده ندارد (سرل، ۱۳۹۲ ب: ۱۲۴).

بر مبنای این استدلال، میان محاسبه‌گری و فهم تفاوت زیادی هست: هوش مصنوعی صرفاً سیستمی محاسبه‌گر است و طبق الگوریتم‌هایی که یک موجود آگاه برایش طراحی کرده است عمل می‌کند. هوش مصنوعی هرقدر هم که پیشرفت کند، موجود فهمنده، آگاه و مدرِک نخواهد شد.

نتیجه

ایده دستیابی به هوش مصنوعی قوی با معضلات دشواری دست به گریبان است و تحقق آن غیرممکن به نظر می‌رسد؛ زیرا هوشمندی، آگاهی، خودآگاهی، ادراک و سایر حالات ذهنی صرفاً برآمده از فرایندهای فیزیکیِ مغزی نیستند تا بتوان آنها را روی سیستم‌های رایانه‌ای شبیه‌سازی کرد، بلکه انسان واجد بُعدی فرامادی و غیرفیزیکی است که منشأ آگاهی و هوشمندی ماست.

تجرد ادراک، مسئله کوالیا، سابجکتیویتی و اراده آزاد مثبت این مدعا هستند. علاوه بر این، با تأمل در معضل حیث التفاتی و استدلال اتاق چینی و نیز تفاوتی که میان مقوله محاسبه‌گری و فهم وجود دارد، ماهیت حقیقی هوش مصنوعی عیان‌تر می‌شود. هوش مصنوعی چُرتکه‌ای پیشرفته است، نه موجودی فهمنده و آگاه.

منابع و مآخذ

۱. اکاشا، سمیر (۱۳۹۱). ***فلسفه علم***. ترجمه هومن پناهنده. تهران: فرهنگ معاصر.

۲. ایگلمن، دیوید (۱۳۹۷). ***مغز؛ داستان شما***. ترجمه محمداسماعیل فلزی. تهران: مازیار.

۳. باربور، ایان (۱۳۹۷). ***دین و علم***. ترجمه پیروز فطورچی. تهران: پژوهشگاه فرهنگ و اندیشه اسلامی.

٤. پترسون، مایکل (١٤٠٠). ***عقل و اعتقاد دینی***. ترجمه احمد نراقی و ابراهیم سلطانی. تهران: طرح نو.

۵. تامپسون، مل (١٣٩٤). ***آشنایی با فلسفه ذهن***. ترجمه مهران داور و نیلوفر خوش‌زبان. تهران: نقش و نگار.

٦. تای، مایکل (۱۳۹۳). ***فلسفه آگاهی***. ترجمه یاسر پوراسماعیل. تهران: حکمت.

۷. داون‌پورت، توماس (۱۳۹۹). ***هوش مصنوعی***. ترجمه علیرضا کاظمی‌نیا و شراره شبل الحکما. تهران: راه پرداخت.

۸. دکارت، رنه (۱۳۹۱). ***تأملات بر فلسفه اولی***. ترجمه احمد احمدی. تهران: سمت.

۹. دلمونت، شین (١٣٩٦). ***درآمدی کوتاه به مغز***. ترجمه محمد یوسفی. تهران: نشر نی.

۱۰. دی. هارت، ویلیام (۱۳۸۱). ***فلسفه نفس***. ترجمه امیر دیوانی. تهران و قم: سروش و طه.

۱۱. سرل، جان (۱۳۸۲). ***ذهن. مغز و علم***. ترجمه امیر دیوانی. قم: بوستان کتاب.

۱۲. سرل، جان (۱۳۹۲ الف). ***اختیار و عصب زیست‌شناسی***. ترجمه محمد یوسفی. تهران: ققنوس.

۱۳. سرل، جان (۱۳۹۲ ب). ***درآمدی کوتاه به ذهن***. ترجمه محمد یوسفی. تهران: نشر نی.

۱٤. سرل، جان (۱۳۹۹). ***فلسفه ذهن***. ترجمه مهدی ذاکری. تهران: علمی و فرهنگی.

۱۵. سولمز، مارک (۱۳۸۸). ***مغز و دنیای درون***. ترجمه علی فیروزآبادی و آرش مانی. تهران: نشر قطره.

۱٦. طباطبایی، سید محمدحسین (۱٤٤۳ ق). ***نهایة الحکمه***. قم: مرکز مدیریت حوزه علمیه.

۱۷. فرشاد، محسن (۱۳۸۹). ***تأملی بر اسرار ذهن***. تهران: علم.

۱۸. کرباسی‌زاده، امیراحسان و حسین شیخ‌رضایی (۱۳۹۹). ***آشنایی با فلسفه ذهن***. تهران: هرمس.

۱۹. کمبل، نیل (۱۳۹۹). ***درآمدی کوتاه به ذهن***. ترجمه محمد یوسفی. تهران: نشر نی.

۲۰. گلشنی، مهدی (۱۳۸۵). ***تحلیلی از دیدگاه‌های فلسفی فیزیک‌دانان معاصر***. تهران: پژوهشگاه علوم انسانی و مطالعات فرهنگی.

۲۱. لو، جاناتان (۱۳۸۹). ***فلسفه ذهن***. ترجمه امیر غلامی. تهران: نشر مرکز.

۲۲. مسلین، کیت (۱۳۸۷). ***درآمدی به فلسفه ذهن***. ترجمه مهدی ذاکری. قم: پژوهشگاه علوم و فرهنگ اسلامی.

۲۳. مسلین، کیت (۱۳۹۹). ***فلسفه ذهن***. ترجمه مهدی ذاکری. تهران: شرکت انتشارات علمی و فرهنگی.

۲٤. مصباح، محمدتقی (۱۳۹۱). ***آموزش فلسفه***. قم: مؤسسه آموزشی و پژوهشی امام خمینی.

۲۵. مندیک، پیت (۱۳۹۵). ***کلیدواژه‌های فلسفه ذهن***. ترجمه محمدحسن ترابی. تهران: نویسه پارسی.

۲٦. هاوکینز، جف (۱٤۰۰). ***هزار مغز؛ نظریه‌ای جدید برای هوش***. ترجمه قاسم کیانی‌مقدم. تهران: مازیار.

۲۷. همتی مقدم، احمدرضا و علی صبوحی (۱۳۹۱). ***نظریه‌های مادی‌انگارانه ذهن***. قم: پژوهشگاه علوم و فرهنگ اسلامی.

28. Hawkins, Jeff (2021), ***A Thousand Brains: A New Theory of Intelligence***. New York: Basic Books.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی