



Journal of Philosophical Investigations



University of Tabriz

The Moral Status of Artificial Intelligence from the Perspective of the Easy and Hard Problem of Consciousness

Saeed Karimi^{✉1} | Morteza Tabatabaie²

1. Corresponding Author, Ph.D. Candidate of Moral Philosophy, University of Religions and Denominations, Qom, Iran. E-mail: saeed.karimi@urd.ac.ir
2. Assistant Professor, Department of Moral Philosophy, University of Religions and Denominations, Qom, Iran. E-mail: morteza.tab64@gmail.com

Article Info

ABSTRACT

Article type:

Research Article

Article history:

Received 19 February 2025

Received in revised form 20
April 2025

Accepted 30 April 2025

Published online 22 November
2025

Keywords:

Consciousness, Moral Status,
Phenomenal Consciousness,
Psychological Consciousness,
Artificial Intelligence.

Consciousness is one of the most fundamental yet complex issues in the philosophy of mind. Philosophers such as David Chalmers, by proposing the “hard problem of consciousness,” have distinguished between phenomenal consciousness and psychological consciousness. Ned Block has further refined this distinction by formulating it in terms of phenomenal consciousness and access consciousness. On the other hand, John Searle, by differentiating between the first-person and third-person perspectives, and Thomas Nagel, by emphasizing the subjective character of experience, have analyzed this duality from different angles. Although these classifications appear diverse, they all point to the fundamental distinction between inner experience and behavioral or cognitive capacities. The present article, focusing on this essential distinction, examines the question of whether artificial intelligence can possess phenomenal consciousness or merely exhibits psychological consciousness. Answering this question has significant philosophical and ethical implications, as phenomenal consciousness is directly linked to experiences of pain, pleasure, and suffering, and is considered a foundation for many moral rights. This study, adopting a descriptive-analytical approach, argues that the distinction between types of consciousness can serve as a basis for determining different degrees of moral status. For instance, beings endowed with access consciousness but lacking phenomenal experience may be entitled to functional autonomy, yet not to the right not to suffer. In contrast, infants and animals, due to their inner experience, are deemed deserving of the right not to suffer, even if they lack advanced cognitive abilities.

Cite this article: Karimi, S. & Tabatabaie, M. (2025). The Moral Status of Artificial Intelligence from the Perspective of the Easy and Hard Problem of Consciousness. *Journal of Philosophical Investigations*, 19 (52), 259-280. <https://doi.org/10.22034/jpiut.2025.66103.4022>



© The Author (s).

Publisher: University of Tabriz.

Extended Abstract

Consciousness is one of the most complex and yet fundamental concepts in the philosophy of mind, which has been widely discussed across various philosophical and ethical fields. Among the prominent philosophers addressing the nature of consciousness are David Chalmers, Ned Block, John Searle, and Thomas Nagel, each presenting differing views on the distinction between different types of consciousness and their implications for the moral status of beings. David Chalmers, the renowned Australian philosopher, divides consciousness into two types: "phenomenal consciousness" and "psychological consciousness," and introduces the concept of the "hard problem of consciousness." He argues that phenomenal consciousness refers to subjective, qualitative experiences that are only accessible from a first-person perspective, whereas psychological consciousness relates to cognitive processes such as information processing, memory, and attention. Similarly, Ned Block makes a distinction between "phenomenal consciousness" and "access consciousness," while John Searle examines the distinction from the first-person and third-person perspectives, and Thomas Nagel emphasizes the subjective and qualitative aspects of experience.

The present study, focusing on the distinction between phenomenal and psychological consciousness, investigates the question of whether artificial intelligence (AI) could possess phenomenal consciousness or if it is confined to psychological consciousness. This question has significant philosophical, ethical, and legal implications, as phenomenal consciousness is associated with qualitative experiences such as pain, pleasure, and suffering, which could form the basis of moral and legal rights. Specifically, exploring this issue could offer a more precise framework for analyzing the ethical rights of AI and its interactions with humans in the future.

In today's world, with the remarkable advances in the field of artificial intelligence, we are approaching a point often referred to as "Artificial General Intelligence" (AGI), a type of intelligence capable of performing a wide range of cognitive tasks akin to human abilities. These advances have raised new ethical questions about the moral status of these systems. If AI were to experience consciousness or suffering, would ignoring its capacity for such experiences be considered unethical behavior? This question necessitates a thorough examination of the concepts of consciousness and suffering in non-biological beings, which must be analyzed primarily from the standpoint of the philosophy of mind and consciousness.

One of the central issues in evaluating the moral status of AI is the difference between two types of consciousness: "phenomenal consciousness" and "psychological consciousness." Phenomenal consciousness refers to subjective mental experiences such as pain, pleasure, or color, which are qualitatively experienced from a first-person perspective. Psychological consciousness, on the other hand, involves cognitive processes like information processing, memory, and attention, which may be analyzed from a third-person perspective without the need for subjective experience. Chalmers discusses this distinction as the "hard problem of consciousness," which refers to the difficulty of explaining and accounting for these internal, subjective experiences. This distinction between phenomenal and psychological consciousness can provide a new basis for assessing the moral status of both biological and artificial beings.

In terms of evaluating the ethical status of AI, one important aspect is whether these systems could experience phenomenal consciousness. For example, if an AI were not capable of experiencing pain, it would not have the moral right to avoid suffering. However, if AI were able to access information and utilize it for decision-making and goal-setting, it might have rights such as the right to freedom, even if it lacked phenomenal consciousness. In this context, examining the distinction between "access consciousness" and "phenomenal consciousness" could lead to a more refined understanding of the ethical standing of AI.

For instance, infants and animals, which possess phenomenal consciousness, have the right to be free from pain, even if they lack complex cognitive abilities. In contrast, beings that only possess access consciousness, such as certain AI systems, might only have rights like freedom, but they would not have the right to freedom from suffering, as they do not experience pain or distress. This analysis suggests that moral rights could be graded based on the type of consciousness possessed by beings, not merely on their cognitive abilities or complex behaviors.

The discussion of the moral rights of AI based on the type of consciousness it possesses also touches upon the issue of defining the rights and ethical duties we owe to these systems. Some researchers argue that, in order to grant moral rights to AI, these systems must possess at least phenomenal consciousness, as the experience of pain and suffering is regarded as the basis for many moral rights. Others argue that access to information and the capacity for rational decision-making could form the basis for granting moral rights. These perspectives, particularly in the context of AI, must be carefully examined from various angles to determine which type of consciousness is necessary for AI to have moral rights.

Ultimately, the present research suggests that the ethical status of AI should be examined in a relative and nuanced manner. While humans, due to their phenomenal consciousness and advanced cognitive abilities, possess comprehensive moral rights, other entities such as infants, animals, and even AI could have different moral rights based on varying levels of consciousness and subjective experiences. Therefore, the ethical status of AI is not fixed or absolute, but rather should be determined based on the type of consciousness these systems possess and their cognitive capabilities. This approach not only helps in better understanding artificial intelligence but could also lay the groundwork for defining the future moral rights of AI systems.



جایگاه اخلاقی هوش مصنوعی از منظر مسأله آسان و سخت آگاهی*

سعید کریمی^۱ | مرتضی طباطبایی^۲

۱. نویسنده مسئول، دانشجوی دکتری فلسفه اخلاق، دانشگاه ادیان و مذاهب، قم، ایران. رایانامه: saeed.karimi@urd.ac.ir

۲. استادیار گروه فلسفه اخلاق، دانشگاه ادیان و مذاهب، قم، ایران. رایانامه: morteza.tab64@gmail.com

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی تاریخ دریافت: ۱۴۰۳/۱۲/۱۰ تاریخ بازنگری: ۱۴۰۴/۰۱/۳۱ تاریخ پذیرش: ۱۴۰۴/۰۲/۱۰ تاریخ انتشار: ۱۴۰۴/۰۹/۰۱ کلیدواژه‌ها: آگاهی، جایگاه اخلاقی، آگاهی پدیداری، آگاهی روان شناختی، هوش مصنوعی.	آگاهی یکی از بنیادی‌ترین و در عین حال پیچیده‌ترین مسائل در فلسفه ذهن است. فیلسوفانی چون دیوید چالمرز با طرح «مسئله سخت آگاهی»، میان آگاهی پدیداری و آگاهی روان‌شناختی تمایز نهاده‌اند؛ ند بلاک نیز این تمایز را به صورت دقیق‌تری در قالب آگاهی پدیداری و آگاهی در دسترس صورت‌بندی کرده است. از سوی دیگر، جان سرل با تفکیک میان منظر اول شخص و سوم شخص، و تامس نیگل با تأکید بر جنبه ذهنی تجربه، این دوگانگی را از منظرهای متفاوتی تحلیل کرده‌اند. اگرچه تقسیم‌بندی‌های مذکور در ظاهر متنوع‌اند، همگی بر تمایز میان تجربه درونی و قابلیت‌های رفتاری و شناختی دلالت دارند. مقاله حاضر، با تمرکز بر همین تمایز بنیادین، به بررسی این پرسش می‌پردازد که آیا هوش مصنوعی می‌تواند واجد آگاهی پدیداری باشد یا تنها از آگاهی روان‌شناختی برخوردار است؟ پاسخ به این پرسش، پیامدهای فلسفی و اخلاقی قابل توجهی دارد؛ چراکه آگاهی پدیداری، مستقیماً با تجربه درد، لذت و رنج در ارتباط است و مبنای بسیاری از حقوق اخلاقی تلقی می‌شود. پژوهش حاضر با رویکردی توصیفی-تحلیلی، نشان می‌دهد که تمایز میان انواع آگاهی می‌تواند مبنایی برای تعیین درجات مختلف جایگاه اخلاقی موجودات قرار گیرد. برای نمونه، موجوداتی که از آگاهی در دسترس برخوردارند اما فاقد تجربه پدیداری‌اند، ممکن است دارای حق آزادی عملکردی باشند، اما از حق رنج نکشیدن بهره‌مند نباشند. در مقابل، نوزادان و حیوانات به دلیل برخورداری از تجربه درونی، شایسته حق رنج نکشیدن‌اند، حتی اگر از توانایی‌های شناختی پیچیده بی‌بهره باشند.

استناد: کریمی، سعید و طباطبایی، مرتضی (۱۴۰۴). جایگاه اخلاقی هوش مصنوعی از منظر مسأله آسان و سخت آگاهی، پژوهش‌های فلسفی، ۱۹ (۵۲)، ۲۵۹-۲۸۰.

<https://doi.org/10.22034/jpiut.2025.66103.4022>



© نویسندگان.

ناشر: دانشگاه تبریز.

* مستخرج از رساله دکتری سعید کریمی با عنوان «امکان‌سنجی برخورداری عامل‌های هوشمند مصنوعی از جایگاه اخلاقی»، استاد راهنما: دکتر مرتضی طباطبائی، دانشگاه ادیان و مذاهب، قم، ایران.

مقدمه

در کنار بحث مهم و دیرین مسئولیت و وظیفه اخلاقی در فلسفه اخلاق، یکی از مباحث جنجالی در نیم قرن اخیر که به‌ویژه با بحث اخلاق حیوانات^۱ برجسته شده، مسئله «جایگاه اخلاقی»^۲ است؛ به این معنا که چه موجوداتی دارای شأن و شخصیت اخلاقی هستند و باید در مواجهه با آنان، بایدها و نبایدهای اخلاقی را رعایت کرد. در دوران باستان، تنها مردان آزاد بودند که جایگاه کامل اخلاقی داشتند و در بسیاری از جوامع، زنان، بردگان و رنگین‌پوستان را فاقد چنین جایگاهی می‌دانستند. رفته‌رفته این حقوق برای همه انسان‌ها به طور یکسان بازشناسی شد و سپس فیلسوفانی چون پیتر سینگر و تام ریگان این بحث را مطرح کردند که آیا حیوانات نیز واجد جایگاه اخلاقی هستند یا نه؟

امروزه با پیشرفت‌های چشمگیر در زمینه هوش مصنوعی، یکی از پرسش‌های مهم در عرصه‌های فلسفی و اخلاقی، جایگاه اخلاقی این سامانه‌ها و نحوه ارزیابی آن است. به‌ویژه با نزدیک شدن به مرحله‌ای از توسعه که از آن با عنوان «هوش مصنوعی جامع»^۳ یا AGI یاد می‌شود - یعنی هوشی که قادر به انجام انواع فعالیت‌های شناختی همانند انسان و قادر به انجام طیف وسیعی از وظایف است (لئو، ۲۰۲۱، ۱۹) - این پرسش مطرح می‌شود که آیا باید ملاحظات اخلاقی را نه‌فقط درباره انسان، بلکه درباره ربات‌های دارای چنین هوشی نیز در نظر گرفت؟ در واقع این دغدغه زمانی عمیق‌تر می‌شود که پرسیده شود: اگر سامانه‌ای از هوش مصنوعی قادر به تجربه نوعی از آگاهی یا احساس رنج باشد، آیا بی‌توجهی به آن، مصداقی از رفتار غیراخلاقی محسوب می‌شود؟ این مسئله مستلزم بررسی امکان و معنای «رنج بردن» در موجودات غیرزیستی است، موضوعی که باید پیش از آن، به‌درستی از منظر فلسفه ذهن تحلیل شود. هوش مصنوعی امروز، با توانمندی‌هایی همچون یادگیری ماشین، پردازش زبان طبیعی و شبیه‌سازی رفتارهای پیچیده انسانی، دیگر صرفاً یک ابزار فنی نیست؛ بلکه به موضوعی برای تأملات فلسفی و مباحث اخلاقی تبدیل شده است. همین ظرفیت‌های نوظهور، ما را با چالش‌هایی مواجه می‌سازند که فراتر از کاربردهای فناوری، نیازمند تحلیل‌های مفهومی و ارزشی عمیق هستند.

یکی از مباحث کلیدی در زمینه ارزیابی جایگاه اخلاقی هوش مصنوعی، تفاوت میان سطوح مختلف آگاهی در انسان و هوش مصنوعی است. بر اساس تقسیم‌بندی دیوید چالمرز، آگاهی را می‌توان به دو سطح «آگاهی پدیداری»^۴ و «آگاهی روان‌شناختی»^۵ تفکیک کرد. آگاهی پدیداری به تجربه‌های ذهنی و کیفی (مانند درد، لذت یا رنگ) از منظر اول‌شخص اشاره دارد، در حالی که آگاهی روان‌شناختی شامل فرایندهای ذهنی همچون پردازش اطلاعات، توجه، حافظه و توانایی گزارش درباره حالات ذهنی است. در همین راستا، چالمرز میان دو نوع مسئله در فلسفه ذهن تمایز

^۱ animal ethics

^۲ moral status

^۳ هوش مصنوعی جامع یا Artificial General Intelligence به هوشی گفته می‌شود که قادر به انجام طیف وسیعی از وظایف است در برخی منابع، میان اصطلاحات «هوش مصنوعی عمومی» و «هوش مصنوعی در سطح انسان» تمایز گذاشته شده است بر اساس این دیدگاه، هوش مصنوعی عمومی به سامانه‌ای اطلاق می‌شود که قادر به انجام طیف وسیعی از وظایف است و از این نظر با هوش محدود که تنها بر یک وظیفه خاص تمرکز دارد، تفاوت دارد در مقابل، هوش مصنوعی در سطح انسان نه‌تنها باید از این توانایی گسترده برخوردار باشد، بلکه لازم است واجد جنبه‌های شناختی، هیجانی و اخلاقی مشابه انسان نیز باشد (لئو، ۲۰۲۱، ۱۹).

^۴ phenomenal consciousness.

^۵ psychological consciousness.

می‌گذارد: مسئله سخت آگاهی^۱ که به دشواری تبیین ماهیت تجربه‌های ذهنی از منظر درونی و اول شخص می‌پردازد، و مسئله‌های آسان‌تر^۲ که به فرایندهای شناختی و قابل شبیه‌سازی ذهن اختصاص دارند. به تعبیر دیگر مفهوم پدیداری ذهن به منزله تجربه آگاهانه، و مفهوم حالت ذهنی به منزله حالت ذهنی است که به صورت آگاهانه تجربه شده باشد. چالمرز این جنبه از ذهن را این واضح‌ترین جنبه ذهن دانسته و بر همین جنبه تمرکز خواهد کرد و این جنبه کل امر را دربر نمی‌گیرد. دومین مفهوم، مفهوم روان‌شناختی ذهن است. این مفهوم از ذهن به منزله علی یا تبیینی رفتار است ز دیدگاه روان‌شناختی، آگاهانه بودن یک حالت ذهنی چندان اهمیت ندارد، بلکه نقش آن در ساختار شناختی مهم است، در حالی که در دیدگاه پدیداری، آنچه اهمیت دارد نحوه تجربه و احساس ذهنی است (چالمرز، ۱۴۰۲، ۱۳-۱۴). این تمایز می‌تواند به روشن‌تر شدن مرزهای اخلاقی میان انسان‌ها و سامانه‌های هوش مصنوعی کمک کند، به‌ویژه زمانی که مسأله اعطای جایگاه اخلاقی به این سامانه‌ها بر اساس سطح آگاهی‌شان مطرح می‌شود.

تحلیل تطبیقی جایگاه اخلاقی انسان و هوش مصنوعی از طریق بررسی مؤلفه‌هایی مانند آگاهی (به‌ویژه در دو سطح پدیداری و روان‌شناختی)، حسی بودن، و قابلیت‌های اخلاقی می‌تواند به درک روشن‌تری از ظرفیت‌ها و محدودیت‌های هوش مصنوعی در مواجهه با مسائل اخلاقی منجر شود. مقاله حاضر، با تمرکز بر تمایز میان «مسئله آسان آگاهی» و «مسئله سخت آگاهی»، تلاش می‌کند موقعیت اخلاقی انسان و هوش مصنوعی را در چارچوب این تمایز تحلیل و ارزیابی کند و نشان دهد که چگونه فقدان یا وجود برخی سطوح آگاهی، می‌تواند معیار اخلاقی بودن یا نبودن یک عامل را تعیین کند. تمرکز مقاله حاضر معطوف به بررسی فلسفی و اخلاقی جایگاه هوش مصنوعی بر پایه سطوح مختلف آگاهی خواهد بود.

در این راستا، تمرکز مقاله بر امکان‌سنجی جایگاه اخلاقی هوش مصنوعی از منظر پدیده آگاهی است. این پدیده آگاهی ابتدا با توجه به نظرات فیلسوفان ذهن مانند دیوید چالمرز، جان سرل و ند بلاک توضیح داده خواهد شد. آگاهی پدیداری و روان‌شناختی از دیدگاه چالمرز، آگاهی اول شخص و سوم شخص از منظر سرل و آگاهی پدیداری و در دسترس از نگاه ند بلاک مورد بررسی قرار خواهد گرفت. همچنین تفکیک چالمرز میان آگاهی پدیداری و روان‌شناختی در سرتاسر مقاله ملاک خواهد بود. سپس مبحث جایگاه اخلاقی، درجات جایگاه اخلاقی و شرایط آن مطرح خواهد شد. نویسندگان در این مقاله جایگاه اخلاقی را به صورت نسبی و تشکیکی تحلیل می‌کنند، به این معنا که به جای نگاه صفر و صدی به موجودات، از مقیاس‌های نسبی برای ارزیابی جایگاه اخلاقی استفاده می‌کنند. موجودی که صد درصد جایگاه اخلاقی را داراست، انسان است و بقیه موجودات بر اساس نزدیکی‌شان به انسان از جهت داشتن ملاکات جایگاه اخلاقی، ارزیابی خواهند شد.

مسئله جایگاه اخلاقی از مباحثی است که در زمینه هوش مصنوعی کار زیادی در آن انجام نشده است، اما در زمینه اخلاق حیوانات و اخلاق سقط جنین مباحث زیادی صورت گرفته که می‌تواند در زمینه تحلیل جایگاه اخلاقی هوش مصنوعی نیز مفید واقع شود. این مقاله با بهره‌گیری از این مباحث، تلاش دارد تا به درک جامع‌تری از مسئله جایگاه اخلاقی هوش مصنوعی و چالش‌های موجود در پذیرش یا رد این جایگاه از منظر آگاهی دست یابد.

¹ the hard problem of consciousness.

² easy problems of consciousness

۱. آگاهی

برخی دانشمندان معتقدند که آگاهی^۱ پدیده‌ای خصوصی و غیرقابل مشاهده برای شخص ثالث است، اما به‌جای آنکه این ویژگی را مانعی برای مطالعه علمی بدانیم، باید آن را فرصتی برای پیشرفت علمی تلقی کنیم. مطالعه آگاهی باید هر دو جنبه درونی و بیرونی آن را دربرگیرد. این پیچیدگی‌ها مختص آگاهی نیست، بلکه در مطالعه سایر پدیده‌های شعوری نیز وجود دارد. برخی ممکن است ویژگی‌های آگاهی به دلیل خصوصی بودن تصورات ذهنی قابل دسترسی تنها برای صاحب آن‌ها، قابل تحلیل ندانند، اما این نگرانی نباید مانع مطالعه علمی پدیده‌های ذهنی شود. علم می‌تواند سازگاری و ثبات ذهنیت‌های فردی را به‌طور عینی بررسی و اثبات کند (داماسیو، ۱۳۹۵، ۱۷۶-۱۷۷).

نیگل^۲ جزو متفکرانی است که به مسئله سخت آگاهی پی برده و این مطلب را در آثار خود بیان کرده است. اما مطلب مهم‌تر این است که او نمی‌خواهد یک امر فراتر از فیزیک را اثبات کند؛ بلکه در صدد است تا تبیینی از آگاهی ارائه دهد که در ذیل زیست‌شناسی تکاملی گنجانده شود و حتی معتقد است تا اینجا متفکران با مکاتب مختلف در ارائه چنین تبیینی ناکام مانده‌اند و حتی مغز را علت ذهن دانستن هم موجب این نمی‌شود که گفته شود تبیین صورت گرفته است.

به اعتقاد نیگل به نظر می‌رسد که اموری وجود دارند که انسان از آن‌ها بی‌اطلاع است، و در صورت آگاهی از آن‌ها، بسیاری از واقعیات برای او قابل فهم‌تر می‌شدند. با این حال، دانستن علت مستقیم یک پدیده لزوماً به معنای درک کامل و فهم آن پدیده نیست. برای نمونه، اینکه فعالیت‌های مغزی علت آگاهی دانسته شود، اگرچه ممکن است یک تبیین علی ارائه دهد، اما به خودی خود موجب فهم کامل ماهیت آگاهی نمی‌شود (نیگل، ۱۳۹۶، ۷۰).

فیلسوفانی چون تامس نیگل آگاهی را مسئله‌ای در چارچوب مشکل ذهن-بدن می‌دانند و در تلاش‌اند تا تبیینی فیزیکی یا زیست‌شناختی از آن ارائه دهند. نیگل معتقد است که هنوز تبیینی درست از آگاهی وجود ندارد. به باور او، اگر انسان‌ها جزئی از جهان فیزیکی هستند، درک نظم طبیعی باید گسترش یابد تا نشان دهد چگونه این نظم می‌تواند به موجوداتی فراتر از امور صرفاً فیزیکی منتهی شود. نیگل استدلال می‌کند که ابزارهای علم فیزیک برای توضیح این پدیده کافی نیستند، زیرا این علم برای تبیین داده‌های فیزیکی شکل گرفته است. در حالی که آگاهی حیوانات نتیجه فرایندهای تکاملی است، تأکید می‌کند که این تأیید تجربی به‌تنهایی تبیینی برای چرا و چگونگی این پدیده ارائه نمی‌دهد. نیگل به این نتیجه می‌رسد که تبیین انتخاب طبیعی برای ویژگی‌های فیزیکی کفایت می‌کند، اما برای آگاهی کافی نیست و هیچ تبیین دیگری نیز تاکنون توانسته این مسئله را به‌طور کامل توضیح دهد (نیگل، ۱۳۹۶، ۶۹).

نیگل بر این باور است که نظریه‌های تحویل‌گرایانه^۳، که می‌کوشند پدیده‌های ذهنی را به فرایندهای فیزیکی یا رفتاری فرو بکاهند، در تبیین پدیده آگاهی ناکام مانده‌اند؛ به‌گونه‌ای که یا آن را نادیده گرفته‌اند یا توانایی توضیح تجربه ذهنی و آگاهانه را

¹ Consciousness

² Tomas Nagel

³ نظریات تحویل‌گرایانه (reductionist theories) در فلسفه ذهن بر آن‌اند که می‌توان پدیده‌های ذهنی مانند آگاهی را به پدیده‌های فیزیکی یا زیستی فروکاست این دیدگاه‌ها معمولاً بر پایه نوعی فیزیکالیسم بنا شده‌اند و تلاش می‌کنند مفاهیم ذهنی را با عبارات و مفاهیم علمی قابل مشاهده مانند فعالیت‌های عصبی یا رفتاری جایگزین کنند منتقدانی همچون نیگل استدلال می‌کنند که این نظریات نمی‌توانند «چگونگی تجربه از منظر اول شخص» را توضیح دهند و لذا جنبه اصلی آگاهی را از قلم می‌اندازند (نیگل، ۱۹۷۴، ۴۳۵-۴۳۶). «معرفت‌شناسی تبیین فروگاهی به طریق سراسری با متافیزیک ترتب تلاقی می‌کند پدیده‌ای طبیعی در قالب بعضی از ویژگی‌های سطح

نداشته‌اند. از نظر او، هیچ‌یک از تحلیل‌های تحویل‌گرایانه قادر نیستند ویژگی‌های سوژکتیو و پدیداری تجربه آگاهانه را در بر گیرند، چرا که این تحلیل‌ها می‌توانند با فقدان آگاهی نیز سازگار باشند. برای نمونه، می‌توان رباتی را تصور کرد که دقیقاً مانند انسان رفتار می‌کند، اما هیچ تجربه‌ای از درون ندارد. نیگل نیز می‌پذیرد که حالات ذهنی آگاهانه ممکن است علت برخی رفتارها باشند و از این جهت بتوان توصیف‌هایی کارکردی از آنها به دست داد؛ اما تأکید می‌کند که این توصیف‌ها تحلیل کافی از آنچه تجربه آگاهانه را می‌سازد ارائه نمی‌دهند (نیگل، ۱۹۷۴، ۴۳۶-۴۳۷).

در آثار چالمرز^۱، تفکیک روشنی میان مسائل «آسان» و «سخت» آگاهی به چشم می‌خورد. وی در رساله دکتری خود با عنوان (ذهن آگاه)^۲، ضمن تأکید بر اهمیت علم، آگاهی را نیز به‌عنوان یکی از اساسی‌ترین و در عین حال رازآلودترین پدیده‌ها به رسمیت می‌شناسد. به اعتقاد چالمرز، تجربه آگاهانه با آنکه برای انسان آشنا و بی‌واسطه است، از پیچیده‌ترین و اسرارآمیزترین پدیده‌های جهان محسوب می‌شود. از یک‌سو، انسان مستقیماً با تجربه آگاهانه در تماس است و آن را بی‌درنگ درک می‌کند؛ اما از سوی دیگر، روشن نیست چگونه می‌توان این پدیده ذهنی را با جهان‌بینی علمی مبتنی بر ساختارهای فیزیکی و تبیین‌های علی، هماهنگ و قابل توضیح ساخت (چالمرز، ۱۹۹۶، ۳). در واقع، این مسئله که چرا آگاهی وجود دارد و چگونه می‌تواند از ساختار فیزیکی مغز پدید آید، هنوز بی‌پاسخ باقی مانده است. چالمرز بر این اساس، فیزیکالیسم را در تنگنا قرار می‌دهد. یکی از مهم‌ترین چالش‌هایی که او علیه فیزیکالیسم مطرح می‌کند، «استدلال زامبی»^۳ است. این استدلال نشان می‌دهد که می‌توان جهانی را تصور کرد که از نظر فیزیکی دقیقاً مانند جهان ما باشد، اما در آن هیچ‌گونه تجربه آگاهانه‌ای وجود نداشته باشد؛ این نشان می‌دهد که آگاهی نمی‌تواند صرفاً به امور فیزیکی تقلیل یابد.

چالمرز بیان می‌کند که آگاهی همچنان به‌عنوان معما باقی مانده است و هنوز مشخص نیست چرا و چگونه فرایندهای فیزیکی، مانند فعالیت مغز، با تجربه‌های ذهنی پدیداری همراه هستند. منظور از این تجربه‌ها، ادراکات ذهنی درونی‌ای هستند که از منظر اول‌شخص رخ می‌دهند، مانند احساس درد، دیدن رنگ قرمز یا شنیدن موسیقی. به‌رغم شواهدی که نشان می‌دهند آگاهی از سیستم‌های فیزیکی مانند مغز به‌وجود می‌آید، هنوز درک واضحی از چگونگی این فرآیند یا دلیل آن وجود ندارد. پرسش کلیدی این است که چگونه یک سیستم فیزیکی مانند مغز می‌تواند دارای تجربه‌های پدیداری باشد و چرا این تجربه‌ها باید ویژگی خاصی داشته

پایین قابل تبیین فروگاهی است دقیقاً در صورتی که بر آن ویژگی‌ها ترتیب منطقی داشته باشد پدیده‌ای طبیعی قابل تبیین فروگاهی در قالب ویژگی‌های فیزیکی - یا به بیان ساده «قابل تبیین فروگاهی» - است در صورتی که بر امور فیزیکی ترتیب منطقی داشته باشد. دقیق‌تر بگوییم: پدیده‌ای طبیعی قابل تبیین فروگاهی در قالب بعضی از ویژگی‌های سطح پایین‌تر است اگر ویژگی متمثل شدن آن پدیده بر ویژگی‌های سطح پایین مورد بحث ترتیب فراگیر و منطقی داشته باشد یک پدیده علی الاطلاق قابل تبیین فروگاهی است اگر ویژگی متمثل شدن آن پدیده بر ویژگی‌های فیزیکی ترتیب فراگیر و منطقی داشته باشد این را می‌توان ایضاً انگاره تبیین فروگاهی انگاشت، شاید با قدری مواضعه. (چالمرز، ۱۴۰۲، ۵۹).

^۱ David Chalmers.

^۲ The Conscious Mind.

^۳ دیوید چالمرز در استدلال زامبی خود برای به چالش کشیدن فیزیکالیسم، ایده زامبی فلسفی را مطرح می‌کند زامبی فلسفی موجودی فرضی است که از نظر فیزیکی و کارکردی کاملاً مشابه یک انسان معمولی است، اما فاقد تجربه آگاهانه (qualia) است به عبارت دیگر، چنین زامبی‌ای می‌تواند درباره احساسات صحبت کند یا به محرک‌ها واکنش نشان دهد، اما هیچ تجربه درونی از درد، لذت یا رنگ‌ها ندارد چالمرز استدلال می‌کند که اگر تصور چنین زامبی‌ای ممکن باشد، پس آگاهی نمی‌تواند صرفاً نتیجه فرآیندهای فیزیکی باشد چراکه در این صورت، دو موجود از نظر فیزیکی کاملاً یکسان می‌توانند تفاوتی اساسی در داشتن یا نداشتن آگاهی داشته باشند این نتیجه نشان می‌دهد که فیزیکالیسم قادر به توضیح کامل آگاهی نیست و نوعی واقعیت غیرمادی در کار است که علم فیزیکی نمی‌تواند آن را تبیین کند (چالمرز، ۱۹۹۶، ۹۴-۹۹).

باشند. به تعبیر دیگر چرا باید چنین سیستمی بودن، حالی داشته باشد؟ چالمرز معتقد است که نظریه‌های معاصر نتوانسته‌اند به جنبه‌های دشوار این مسئله پرداخته و آگاهی را در نظم طبیعی جهان توضیح دهند (چالمرز، ۱۴۰۲، مقدمه ۲۳).

تگمارک تأکید دارد که بسیاری از مباحث مربوط به آگاهی به دلیل تفاوت در تعاریف طرفین، به نتیجه روشنی نمی‌رسند. او بیان می‌کند که همانند مفاهیمی چون حیات و هوش، هیچ تعریف واحدی از آگاهی وجود ندارد و تعاریف متعددی چون حساسیت جسمانی، هوشیاری، و آگاهی از خویشتن مطرح شده است. در کتاب خود، تگمارک تعریفی کلی از آگاهی ارائه می‌دهد که بر تجربه ذهنی درونی متمرکز است؛ به‌طوری که هر فرد زمانی که احساس خاصی از بودن را تجربه کند، آگاهی دارد (تگمارک، ۱۴۰۱، ۴۴۷-۴۴۸). تگمارک همچنین می‌گوید که تعریف او از آگاهی مستقل از رفتار، ادراک، آگاهی از خویشتن یا عواطف است. بنابراین، حتی فردی که خواب می‌بیند، همچنان آگاهی دارد، هرچند که ممکن است هوشیار نباشد یا به ورودی‌های حسی دسترسی نداشته باشد. به همین ترتیب، سامانه هوشمندی که تجربه‌ای مشابه درد داشته باشد، حتی اگر نتواند حرکت کند، همچنان آگاهی خواهد داشت. این تعریف، در مورد آگاهی سیستم‌های هوش مصنوعی آینده که ممکن است تنها به صورت نرم‌افزار باشند و فاقد حسگرها یا بدنی رباتیک باشند، اظهار نظری نمی‌کند (تگمارک، ۱۴۰۱، ۴۴۸).

به اعتقاد تگمارک با بیان چنین تعریفی نادیده گرفتن مسئله آگاهی دشوار است. همان طور که یووال نوح هراری در کتابش انسان خداگونه این گونه بیان می‌کند: «اگر دانشمندی بخواهد بگوید تجربه‌های ذهنی درونی موضوعیت ندارد، چالش پیش رویش این است که بدون رجوع به تجربه‌های ذهنی درونی توضیح دهد چرا صحبت شکنجه و تجاوز نادرست‌اند» (هراری، ۱۳۹۷، ۱۵۰).

تحلیل کلی این مباحث نشان می‌دهد که گرچه علوم طبیعی و شناختی در تبیین مکانیسم‌های مرتبط با آگاهی پیشرفت‌های چشمگیری داشته‌اند، اما همچنان در پاسخ به پرسش بنیادین درباره چیستی و چرایی تجربه آگاهانه ناکام مانده‌اند. به نظر می‌رسد که آگاهی به‌عنوان پدیده‌ای که فراتر از روابط فیزیکی صرف قرار دارد، چالشی جدی برای فیزیکالیسم و تقلیل‌گرایی محسوب می‌شود. از این رو شاید نیاز باشد که برای درک عمیق‌تر آن، به سراغ روش‌های نوینی برویم که نه تنها علوم طبیعی، بلکه فلسفه و شاید حوزه‌های میان‌رشته‌ای دیگر را نیز دربرگیرد.

۱-۱. مسئله آسان و سخت آگاهی

باید توجه داشت که تفکیک‌های موجود در رابطه با دو نوع از آگاهی که توسط فیلسوفانی چون چالمرز، ند بلاک و جان سرل که با تعابیر مختلف صورت می‌پذیرد، دارای مناسبات و ملاک واحدی است.^۱

^۱ ادعای نویسندگان: تمایز میان آگاهی پدیداری (Phenomenal Consciousness) و آگاهی کارکردی یا دسترس‌پذیر (Access Consciousness) در آثار نیگل، جکسون، چالمرز، و بلاک، علی‌رغم تفاوت در تعبیر و بیان، دارای مناسبات و ملاک واحدی است. استدلال نویسندگان: (۱) نقطه مشترک: تأکید بر جنبه اول شخصی آگاهی: نیگل با طرح این پرسش که «چه حسی دارد که یک خفاش باشی؟» بر جنبه غیرقابل انتقال و درونی تجربه آگاهانه تأکید دارد. او می‌گوید چنین تجربه‌ای از منظر اول شخص درک‌پذیر است اما علم سوم‌شخصی از درک آن عاجز است (نیگل، ۱۹۷۴، ۴۳۵-۴۳۷) جکسون در آزمایش اتاق مری نشان می‌دهد که دانستن همه داده‌های فیزیکی هنوز جای تجربه درونی را نمی‌گیرد، یعنی کیفیت ذهنی تجربه (qualia) از جنس دانش علمی نیست (جکسون، ۱۹۸۶، ۲۹۱) چالمرز نیز با تمایز میان «مسائل آسان» و «مسئله سخت» آگاهی، به همین جنبه اشاره دارد: آنچه علم می‌تواند توضیح دهد، کارکردهای شناختی است، اما تجربه پدیداری هنوز رازی حل‌نشده است (چالمرز، ۱۹۹۶، ۱۱-۱۲) بلاک نیز با تفکیک میان Phenomenal و Access Consciousness دقیقاً همین ملاک را به‌صورت صریح‌تری در زبان فلسفی معرفی می‌کند (بلاک، ۱۹۹۵، ۲) وجه اشتراک: ناتوانی فیزیکالیسم در تبیین تجربه درونی. هر چهار فیلسوف به نوعی بر این نکته تأکید دارند که علم و فیزیکالیسم فعلاً یا ذاتاً قادر به توضیح چگونگی ظهور تجربه پدیداری از فرآیندهای فیزیکی نیستند بنابراین ملاک و مناسبات تمایز آنها، وجود وجهی از آگاهی است که با دانش فیزیکی قابل تحویل نیست. (۳) تفاوت‌های ظاهری، وحدت مفهومی. اگرچه نیگل از «what it is like to be a bat?»، جکسون از «qualia»، چالمرز از «hard problem»، و بلاک از

نیگل در مقاله معروف خود با عنوان خفانش بودن چه حس و حالی دارد؟ استدلال می‌کند که تجربه آگاهانه دارای بُعدی سابیجکتیو و درونی است که از منظر اول شخص قابل درک است، نه از منظر سوم شخص علمی. به تعبیر او، اگرچه می‌توانیم ساختار مغزی یک خفاش را بررسی کنیم یا امواج صوتی را تحلیل نماییم، اما هیچ‌گاه نمی‌توانیم بدانیم چه حسی دارد که یک خفاش باشیم—یعنی تجربه زیستن یک خفاش را از درون. نیگل با این استدلال تلاش می‌کند نشان دهد که نظریه‌های فیزیکیالیستی که صرفاً به بررسی ساختارهای فیزیکی و رفتاری می‌پردازند، نمی‌توانند جنبه پدیداری (phenomenal) آگاهی را توضیح دهند. از نظر او، آگاهی ویژگی‌ای دارد که صرفاً از طریق تجربه‌زیسته قابل درک است، نه از طریق توضیحات علمی (نیگل، ۱۹۷۴، ۴۳۷-۴۴۱).

جکسون نیز در آزمایشی فکری از شخصیتی به نام مری سخن می‌گوید؛ زنی که تمام دانش علمی ممکن را درباره رنگ‌ها دارد، اما در یک اتاق سیاه‌وسفید زندگی کرده و هرگز رنگی ندیده است. او دقیقاً می‌داند نور چه طول‌موجی دارد، مغز چگونه رنگ‌ها را پردازش می‌کند، و رنگ قرمز چگونه روی شبکه اثر می‌گذارد. اما وقتی مری برای اولین بار از اتاق بیرون می‌آید و رنگ قرمز را می‌بیند، چیزی جدیدی را تجربه می‌کند؛ چیزی که هیچ‌یک از دانسته‌های علمی او قادر به انتقال آن نبوده‌اند. جکسون از این آزمایش فکری نتیجه می‌گیرد که دانش پدیداری (qualia) یا تجربه ذهنی، فراتر از تبیین‌های فیزیکی است و بنابراین فیزیکیالیسم کامل نیست. این آزمایش نیز مانند استدلال نیگل، به شکاف توضیحی میان حالت‌های فیزیکی و تجربه آگاهانه اشاره دارد (جکسون، ۱۹۸۶، ۲۹۱).

چالمرز میان دو مفهوم متفاوت از ذهن تمایز قائل می‌شود: نخست، مفهوم پدیداری ذهن است که ذهن را بر اساس تجربه آگاهانه تعریف می‌کند و یک حالت ذهنی را وابسته به تجربه آگاهانه می‌داند. او این جنبه را پیچیده‌ترین بُعد ذهن می‌داند و بر آن تمرکز دارد، هرچند آن را معادل تمام ویژگی‌های ذهنی نمی‌داند. دوم، مفهوم روان‌شناختی ذهن است که آن را به‌عنوان مبنای علی یا تبیینی رفتار در نظر می‌گیرد. از این دیدگاه، یک حالت زمانی ذهنی محسوب می‌شود که نقش مناسبی در ایجاد یا توضیح رفتار ایفا کند، بدون اینکه ضرورتاً کیفیت آگاهانه داشته باشد. در این چارچوب، آنچه اهمیت دارد، نقش علی و کارکردی آن حالت ذهنی در نظام شناختی و در ایجاد یا تبیین رفتار است؛ به این معنا که حالت ذهنی لزوماً آگاهانه نیست، اما اگر بتواند در بروز رفتار نقش ایفا کند یا عملکرد شناختی را تبیین نماید، ذهنی تلقی می‌شود. چالمرز تأکید دارد که این دو مفهوم در رقابت با یکدیگر نیستند، زیرا هیچ‌یک تحلیل نهایی ذهن به‌شمار نمی‌رود. به نظر او، هر دو مفهوم به پدیده‌های واقعی اما متفاوت اشاره دارند (چالمرز، ۱۹۹۶، ۱۱-۱۲).

چالمرز استدلال می‌کند که آگاهی یک مسئله واحد نیست، بلکه به پدیده‌های مختلفی اشاره دارد که هر یک نیازمند توضیح هستند. او مسائل آگاهی را به دو دسته تقسیم می‌کند: مسائل آسان و مسائل دشوار. مسائل آسان شامل توانایی تشخیص و دسته‌بندی محرک‌ها، یکپارچه‌سازی اطلاعات، گزارش‌پذیری حالات ذهنی، خودآگاهی شناختی، تمرکز توجه و کنترل ارادی رفتار هستند که با روش‌های علوم شناختی و مدل‌های عصبی قابل تبیین‌اند. اما مسائل دشوار به تجربه ذهنی یا «کیفیت پدیداری» آگاهی مربوط می‌شوند که با تبیین‌های رایج علمی حل نمی‌شوند و نیازمند رویکردهای غیرکاهشی هستند (چالمرز، ۱۹۹۵، ۲۰۱-۲۰۰).

تفکیک «A-consciousness و P-consciousness» استفاده می‌کند، همه آنها به یک پرسش بنیادین اشاره دارند: چگونه فرایندهای فیزیکی می‌توانند تجربه‌ای درونی، آگاهانه و سابیجکتیو تولید کنند؟ در نتیجه با وجود تفاوت در شیوه بیان و زبان فلسفی، نیگل، جکسون، چالمرز و بلاک همگی به تمایزی بنیادی میان جنبه‌های پدیداری و کارکردی آگاهی قائل‌اند. این تمایز که گاه در قالب تجربه زیسته غیرقابل انتقال، گاه به صورت کوالیا، و گاه در قالب مسئله سخت آگاهی طرح شده است، دارای مناسبت و ملاک واحدی است؛ یعنی همگی به وجهی از آگاهی اشاره دارند که فراتر از توضیحات فیزیکی و علمی قرار دارد.

تصور کنید که من تجربه لذتی را داشته‌ام و سعی می‌کنم آن را برای شما توضیح دهم. حتی اگر توضیحات دقیقی از فعالیت‌های مغزی، واکنش‌های بدنی یا رفتارهای مرتبط ارائه دهم، شما که آن تجربه را نداشته‌اید، نمی‌توانید به‌درستی درک کنید که من چه حسی داشتم. این همان چیزی است که چالمرز و اندیشمندان دیگر با مثال‌هایی مانند «قرمز دیدن چه حس و حالی دارد» یا «خفاش بودن چه حس و حالی دارد»^۱ بیان می‌کنند. برای توضیح این موضوع می‌توان از چند مثال ملموس استفاده کرد: تجربه جنسی: فرض کنید به کودکی توضیح دهید که تجربه ارتباط جنسی چگونه است. حتی اگر تمامی جزئیات فیزیولوژیکی، روانشناختی و بیولوژیکی را بیان کنید، کودک چون خود هرگز این تجربه را نداشته، قادر نخواهد بود حس و حال واقعی این تجربه را درک کند. این مثال نشان می‌دهد که توضیحات عینی نمی‌توانند به درک کامل از یک تجربه ذاتی منجر شوند.

تجربه لذت بردن از علم و دانش نیز این گونه است. فردی ممکن است تمامی مباحث علمی و زیبایی‌های درونی دانش را بشنود یا مطالعه کند، اما اگر هرگز حس شگفتی و لذت عمیقی که از درک عمیق مفاهیم علمی حاصل می‌شود را تجربه نکرده باشد، نمی‌تواند آن لذت را به طور واقعی درک کند. همانطور که انتقال توضیحی درباره زیبایی دانش از طریق کلمات به اندازه تجربه مستقیم آن تاثیرگذار نیست.

تجربه معنوی: بسیاری از افراد گزارش می‌دهند که در لحظات معنوی، احساس پیوند عمیق، آرامش و رضایت از وجود حاصل می‌شود. حتی شرح دقیق وضعیت روانی و فیزیولوژیکی در چنین لحظاتی نمی‌تواند کاملاً آن حس و حال بی‌کلام و شخصی را به فرد دیگری منتقل کند که هرگز آن تجربه را نداشته باشد.

تجربه طعم: فرض کنید کسی که هرگز طعم یک میوه خاص (مثلاً توت فرنگی) را تجربه نکرده باشد، تمامی توضیحات علمی درباره ترکیبات شیمیایی، واکنش‌های عصبی و توضیحات مربوط به لذت حسی مرتبط با آن میوه را بشنود. با این حال، او هرگز نمی‌تواند درک کند که حس واقعی چشیدن آن طعم چگونه است.

تجربه درد یا خوشحالی شدید: فردی که هیچ‌گاه سردرد شدید یا خوشحالی عمیقی را تجربه نکرده، ممکن است تمامی جزئیات علمی و روان‌شناختی مربوط به آن حالات را بشنود؛ اما بدون تجربه مستقیم، نمی‌تواند به‌طور کامل حس و حال واقعی آن حالات احساسی را درک کند.

این مثال‌ها نشان می‌دهند که در بسیاری از موارد، تجربه درونی و شخصی یک حالت یا حس، ناشی از آن چیزی است که تنها از طریق تجربه مستقیم قابل درک است؛ بنابراین حتی با ارائه توضیحات دقیق علمی و بیولوژیکی، همیشه یک شکاف بین توصیف عینی و حس درونی وجود دارد که محور اصلی مسئله سخت آگاهی محسوب می‌شود و مثال‌های بیان شده از طریق فیلسوفان ذهن در این زمینه شبیه به مثال‌های بیان شده است.

این اختلاف بین آنچه که می‌توانیم به‌صورت عینی شرح دهیم (مسئله آسان) و آنچه که به صورت درونی حس می‌کنیم (مسئله سخت)، به «شکاف تبیینی»^۲ معروف است. اگرچه علوم اعصاب و شناخت‌شناسی می‌توانند فرایندهای شناختی را به دقت مدل‌سازی

^۱ What is it like to be a bat?

^۲ مسئله شکاف تبیینی (explanatory gap) در نظر لوین، به این معناست که حتی اگر ما بتوانیم مکانیزم‌های فیزیکی مربوط به تجربیات ذهنی را توضیح دهیم، هنوز تفاوتی بنیادی میان توضیح فیزیکی و تجربه ذهنی باقی می‌ماند به طور خاص، لوین ادعا می‌کند که در حالی که می‌توانیم یک مکانیزم فیزیکی برای احساس درد (مثلاً فعال شدن فیبرهای C) شناسایی کنیم، این توضیح هنوز نمی‌تواند آنچه که درد به‌طور کیفی در تجربه انسانی است، تبیین کند شکاف تبیینی به این مشکل اشاره دارد که

و توضیح دهند، اما هنوز نمی‌توانند به سادگی بیان کنند چگونه و چرا این فرایندها به تجربه‌های ذهنی و حس‌های کیفی تبدیل می‌شوند.

آگاهی پدیداری به دو شکل کلی اصالت فیزیک را به چالش می‌کشد. از یک سو، چالشی متافیزیکی مطرح می‌شود که پرسش از امکان توضیح کیفیات ذهنی در قالب مفاهیم فیزیکی است. از سوی دیگر، چالشی معرفت‌شناختی وجود دارد که حتی اگر بپذیریم کیفیات ذهنی ماهیتی فیزیکی دارند، همچنان دشواری درک نحوه تولید تجارب آگاهانه توسط مغز باقی می‌ماند (کرافت، ۱۳۹۲، ۲۹۰).

گفته می‌شود که جوزف لوین برای نخستین بار این دو نوع پرسش را به‌طور صریح از یکدیگر تفکیک کرد و همچنین مسئله شکاف تبیینی را نیز او بیان کرد (لوین، ۲۰۰۱).

آزمایش ذهنی اتاق مری که جکسون مطرح کرده، یکی از نیرومندترین راه‌ها برای تأکید بر چالش متافیزیکی در برابر اصالت فیزیک به شمار می‌آید (کرافت، ۱۳۹۲، ۲۹۰) بر اساس آزمایش ذهنی که ارائه شده، ماری در یک اتاق سیاه و سفید محبوس است و از طریق منابع سیاه و سفید تمام دانستنی‌های ممکن درباره واقعیات فیزیکی جهان، از جمله فیزیک، شیمی و فیزیولوژی اعصاب، را می‌آموزد. فیزیکالیسم، که مدعی است همه واقعیات به واقعیات فیزیکی تقلیل‌پذیرند، ایجاب می‌کند که دانستن تمام واقعیات فیزیکی معادل دانستن تمام واقعیات ممکن باشد. با این حال، هنگامی که ماری از اتاق خارج شده یا برای نخستین بار یک تلویزیون رنگی در اختیارش قرار می‌گیرد، او تجربه دیدن رنگ قرمز را به دست می‌آورد و از این طریق چیزی جدید را می‌آموزد. این نشان می‌دهد که معرفت او پیش از این کامل نبوده و دانستن تمام واقعیات فیزیکی برای دانستن همه چیز کافی نیست. بنابراین، این استدلال نتیجه می‌گیرد که فیزیکالیسم نمی‌تواند نظریه‌ای جامع درباره معرفت باشد، زیرا نوعی معرفت وجود دارد که فراتر از معرفت فیزیکی است (جکسون، ۱۹۸۶، ۲۹۱-۲۹۲). مثال‌های بالا نیز که در پی تفکیک میان توصیف یک شیء و تجربه آن بیان شد، در تایید همین استدلال معرفت جکسون است.

سرل نیز میان دو نوع هستی‌شناسی تمایز قائل می‌شود: ۱. هستی‌شناسی سوم‌شخص: این نوع هستی‌شناسی به واقعیت‌هایی اشاره دارد که مستقل از هر فاعل آگاه وجود دارند. این دسته شامل پدیده‌های فیزیکی، بیولوژیکی و عملکردهای مغزی است که می‌توان آن‌ها را از منظر علمی و عینی توصیف کرد. به بیان دیگر، هر آنچه که بدون نیاز به یک مشاهده‌گر خاص قابل توصیف باشد، در این دسته قرار می‌گیرد. دوم هستی‌شناسی اول‌شخص: این نوع هستی‌شناسی شامل پدیده‌هایی است که تنها در صورتی وجود دارند که توسط یک فاعل آگاه تجربه شوند. آگاهی، احساسات و ادراکات ذهنی در این دسته قرار می‌گیرند. برای مثال، «احساس درد» تنها زمانی وجود دارد که کسی آن را تجربه کند؛ درد به‌خودی‌خود، مستقل از یک موجود آگاه، قابل تعریف نیست (سرل، ۲۰۰۸، ۹۷-۹۹).

ند بلاک آگاهی پدیداری^۱ را به شکلی غیرتقلیل‌گرایانه و بدون ارائه تعریفی دقیق، بلکه با اشاره به خود پدیده توضیح می‌دهد. او این نوع آگاهی را معادل تجربه می‌داند و ویژگی‌های آن را ویژگی‌های تجربه‌ای می‌نامد. این نوع آگاهی شامل ادراکات حسی (دیدن، شنیدن، بوییدن، چشیدن، درد) و همچنین اندیشه‌ها، امیال و احساسات است. بلاک تأکید می‌کند که آگاهی پدیداری با ویژگی‌های

ما نمی‌توانیم به‌طور کامل توضیح دهیم که چرا تجربه‌ای مانند درد به‌طور کیفی چگونه احساس می‌شود، حتی اگر مکانیزم‌های فیزیکی آن را بدانیم (لوین، ۱۹۸۳، ۳۵۷-۳۵۸).

¹ P-consciousness

شناختی، قصدی^۱، یا کارکردی متفاوت است و نمی‌توان آن را بر اساس کارکردگرایی تبیین کرد. همچنین، او اشاره می‌کند که آگاهی پدیداری اغلب بازنمودی^۲ است، زیرا تفاوت در محتوای قصدی می‌تواند تجربه آگاهی را تغییر دهد (مثلاً شنیدن یک صدا از چپ با شنیدن همان صدا از راست متفاوت است). او بیان می‌کند که معمای علمی آگاهی دقیقاً به آگاهی پدیداری مربوط است و کاری به آگاهی دسترسی‌پذیر^۳ یا خودآگاهی^۴ ندارد (بلاک، ۱۹۹۵، ۲۳۰).

بر اساس دیدگاه‌های چالمرز و بلاک، می‌توان نتیجه گرفت که هشیاری پدیداری همان مسئله دشوار هشیاری است. بررسی‌های مربوط به عملکرد هشیاری، چه در افراد سالم و چه در بیماران با آسیب‌های مغزی، عمدتاً بر پردازش اطلاعات در حالت هشیار و ناهشیار متمرکز بوده‌اند. اما چالمرز و بلاک معتقدند که این مطالعات از اصل مسئله غافل مانده‌اند. آزمایش‌هایی مانند آماده‌سازی پوشیده یا بررسی تفاوت‌های شناختی در بیماران خاص، تنها به حل مسائل ساده هشیاری کمک می‌کنند و تأثیری در حل مسئله دشوار ندارند. این مطالعات صرفاً تفاوت‌های مربوط به دسترسی به اطلاعات را آشکار می‌سازند، اما توضیحی درباره ماهیت تجربه یا چیستی هشیاری پدیداری ارائه نمی‌کنند (برمودز، ۱۴۰۲، ۶۳۷-۶۳۸).

چالمرز معتقد است که علوم شناختی قادر به حل مسئله دشوار هشیاری نیستند و این پرسش را مطرح می‌کند که چرا حل مسائل ساده هشیاری نباید ما را به حل مسئله دشوار هدایت کند؟ همچنین، او امکان یکی بودن هشیاری پدیداری و هشیاری دسترسی را بررسی می‌کند. در بحث‌های معاصر، فیلسوفانی مانند چالمرز و بلاک بر تمایز اساسی میان این دو نوع هشیاری تأکید دارند و استدلال می‌کنند که پردازش اطلاعات هشیارانه و تجربه هشیارانه دو مقوله مستقل‌اند. جوزف لوین این مشکل را «شکاف تبیینی» نامیده است. در این راستا، ایده‌ای مطرح شده که هشیاری دسترسی می‌تواند بدون هشیاری پدیداری وجود داشته باشد، اما بلاک با اشاره به تجربه بیماران مبتلا به کوربینی، این امکان را رد می‌کند و پیشنهاد می‌دهد که بیمارانی با حالتی فراتر از کوربینی می‌توانند به درک بهتر این مفاهیم کمک کنند (برمودز، ۱۴۰۲، ۶۳۸).

در نتیجه، مسئله سخت آگاهی به نوعی پرسشی بنیادین از ماهیت وجود و تجربه است که نشان می‌دهد توضیحات صرفاً فیزیکی و عملکردی از مغز قادر به فراگرفتن تمام ابعاد آگاهی انسان نیستند. این موضوع باعث شده است که همچنان یکی از مهم‌ترین چالش‌ها در فلسفه ذهن باقی بماند و به تحقیقات عمیق‌تری در حوزه آگاهی و تجربه‌های ذهنی منجر شود.

۲-۱. آگاهی در هوش مصنوعی

چالمرز استدلال می‌کند که بحث درباره امکان آگاهی در ماشین‌ها از موضوعات مهم فلسفی و علمی در دهه‌های اخیر بوده است. برخی معتقدند که با برنامه‌ریزی مناسب، رایانه‌ها می‌توانند ذهن داشته باشند، درحالی‌که منتقدان این دیدگاه بر محدودیت‌های رایانه‌ها تأکید دارند و معتقدند که صرف انجام محاسبات برای ظهور آگاهی کافی نیست. این اعتراضات به دو دسته تقسیم می‌شوند: گروهی بر این باورند که سیستم‌های محاسباتی حتی قادر به تقلید رفتار شناختی انسان نیستند، زیرا انسان‌ها دارای قابلیت‌هایی مانند خلاقیت و انعطاف‌پذیری‌اند که رایانه‌ها فاقد آن هستند. برخی نیز استدلال می‌کنند که محدودیت‌های نظریه گودل مانع از آن می‌شود که

¹ Intentional

² Representational

³ Access-consciousness

⁴ Self-consciousness

رایانه‌ها بیش ریاضیاتی انسان را بازتولید کنند. باین‌حال، برخی از این اعتراضات با چالش‌هایی مواجه شده‌اند، زیرا شبیه‌سازی محاسباتی فرآیندهای فیزیکی در بسیاری از موارد موفق بوده است (چالمرز، ۱۴۰۲، ۳۸۹).

چالمرز میان دو نوع ایراد به نظریه «هوش مصنوعی قوی» تمایز قائل می‌شود: نخست، ایرادات درونی که ادعا می‌کنند حتی اگر یک سیستم رایانه‌ای ساختار عملکردی مناسبی داشته باشد، باز هم نمی‌تواند آگاه باشد (مثلاً اتاق چینی سرل)^۱. دوم، ایرادات بیرونی که فراتر می‌روند و می‌گویند حتی اگر همین ساختار در مغز انسان هم باشد، باز نمی‌تواند آگاهی را تبیین کند. هر دو دسته این ایرادات، تردیدهایی نسبت به کفایت تبیین عملکردی برای آگاهی مطرح می‌کنند چالمرز اشاره می‌کند که بسیاری از طرفداران دیدگاه‌های غیرتقلیلی، هوش مصنوعی را فاقد تجربه آگاهانه می‌دانند. باین‌حال، او تأکید دارد که پذیرش این دیدگاه، لزوماً به انکار آگاهی در هوش مصنوعی منجر نمی‌شود. پرسش اصلی این است که چه سیستم‌های فیزیکی می‌توانند آگاهی را پدید آورند. اگر مغز قادر به ایجاد آگاهی است، بعید نیست که محاسبات نیز بتوانند چنین کنند (چالمرز، ۱۴۰۲، ۳۸۹-۳۹۰).

در زمینه آگاهی در هوش مصنوعی نظرات متعددی وجود دارد که برخی موافق و برخی مخالف آگاهی در هوش مصنوعی هستند. جان سرل از طریق استدلال اتاق چینی آگاهی داشتن هوش مصنوعی را زیر سوال می‌برد و از طریق تفکیک نحو از معناشناسی معتقد است که هوش مصنوعی قوی امکان پذیر نیست. در مقابل تورینگ با بیان آزمون تورینگ آگاهی داشتن در هوش مصنوعی را می‌پذیرد.

سرل چند سال پیش نظریه هوش مصنوعی قوی و هر دو نسخه از آزمون تورینگ قوی را نقد کرده است (سرل، ۱۹۸۰). طبق این نظریه، اگر یک برنامه کامپیوتری به درستی برنامه‌ریزی شده باشد و بتواند آزمون تورینگ را بگذراند، باید توانایی‌های شناختی مشابه انسان را داشته باشد. سرل این نظریه را با استفاده از استدلال معروف «اتاق چینی»^۲ نقد کرد. در این آزمایش فکری، فردی که زبان چینی نمی‌داند، با استفاده از دستورالعمل‌ها و نمادهای چینی می‌تواند رفتار یک فرد چینی را شبیه‌سازی کرده و به درستی به سوالات پاسخ دهد. با این حال، او همچنان معنای نمادها را نمی‌فهمد زیرا قادر به پیوند دادن دستکاری‌های نمادها به معنای آن‌ها نیست. سرل استدلال می‌کند که هیچ کامپیوتری نمی‌تواند معنای نمادها را درک کند، چراکه تنها به پردازش نمادها می‌پردازد و هیچ راهی برای پیوند دادن آن‌ها به معنا ندارد (سرل، ۲۰۰۸، ۵۹).

^۱ جان سرل در این مقاله تأثیرگذار، استدلال اتاق چینی را به عنوان نقدی بنیادی بر دیدگاه «هوش مصنوعی قوی» (Strong AI) مطرح می‌کند طبق دیدگاه هوش مصنوعی قوی، یک برنامه کامپیوتری که به درستی طراحی شده باشد، نه تنها می‌تواند رفتار هوشمند انسانی را شبیه‌سازی کند، بلکه در واقع خود می‌فهمد، می‌اندیشد، و دارای آگاهی ذهنی است سرل این ادعا را به چالش می‌کشد. فرض کنید فردی که تنها انگلیسی می‌داند، در اتاقی محبوس است بیرون از اتاق، افراد بومی زبان چینی پرسش‌هایی را به زبان چینی می‌نویسند و از طریق شکافی به داخل می‌فرستند فرد داخل اتاق، با استفاده از یک دفترچه راهنما که شامل دستورالعمل‌هایی دقیق به زبان انگلیسی است، نمادهای چینی را تطبیق داده و پاسخ‌هایی به همان زبان تولید می‌کند این دستورالعمل‌ها، معادل یک «برنامه» هستند از بیرون، پاسخ‌ها چنان دقیق و طبیعی به نظر می‌رسند که ناظران فرض می‌کنند فرد داخل اتاق زبان چینی را می‌فهمد اما در واقع، او صرفاً دارد نشانه‌هایی را بر اساس قواعد نحوی دستکاری می‌کند، بی‌آن که معنای آن‌ها را بداند سرل از این سناریو، نتیجه می‌گیرد که صرف اجرای یک برنامه—حتی اگر خروجی آن کاملاً طبیعی و معنادار به نظر برسد—معادل فهم یا آگاهی ذهنی نیست پردازش نحوی به تنهایی، هیچ‌گاه به معنای واقعی (semantics) یا قصدیت (intentionality) منجر نمی‌شود در نتیجه، حتی اگر یک سیستم کامپیوتری بتواند دقیقاً مانند انسان رفتار زبانی داشته باشد، باز هم فاقد ذهن، فهم و آگاهی است (سرل، ۱۹۸۰، ۴۱۷-۴۲۳). تفکیک میان نحو و معناشناسی موجب دو دیدگاه تمایز در مقاله سرل می‌گردد: ۱) هوش مصنوعی ضعیف (Weak AI): رایانه‌ها ابزارهایی برای شبیه‌سازی ذهن انسان و مطالعه آن هستند، بدون آن که خود دارای آگاهی باشند (نحو را دارد ولی معنا ندارد)؛ ۲) هوش مصنوعی قوی (Strong AI): رایانه، در صورت اجرای برنامه‌ای مناسب، ذهنی واقعی خواهد داشت (علاوه بر نحو معناشناسی هم دارد). استدلال اتاق چینی به روشنی علیه دیدگاه دوم است (سرل، ۱۹۸۰، ۴۱۷-۴۲۳).

^۲ Chinese Room Argument.

تفکیک میان نحو و معناشناسی نقشی اساسی در تبیین تمایز میان هوش مصنوعی ضعیف و قوی در منظومه فکری جان سرل ایفا می‌کند. سرل نظریه هوش مصنوعی قوی را، که بر توانایی فهم معناشناختی علاوه بر پردازش نحوی تأکید دارد، مردود می‌داند. از دیدگاه او، هوش مصنوعی قوی نه تنها ممکن نیست، بلکه آنچه به عنوان هوش مصنوعی معرفی می‌شود، صرفاً شبیه‌سازی فهم است، نه تحقق آن. ماشین‌های هوشمند تنها از طریق پردازش نمادها، رفتارهایی معنادار از خود بروز می‌دهند، اما هیچ درکی از معنا ندارند. فراتر از این، سرل در تحلیل عمیق‌تر خود استدلال می‌کند که حتی پردازش نحوی در این سیستم‌ها واقعی نیست؛ بلکه ماشین‌ها صرفاً تظاهر به درک نحو می‌کنند، بدون آنکه فهمی اصیل از آن داشته باشند (سرل، ۲۰۰۴، ۸۹-۹۲).

سرل استدلال می‌کند که شناخته‌شده‌ترین ایرادات به هوش مصنوعی قوی، استدلال «اتاق چینی» است که خود او طراحی کرده است (سرل، ۱۹۸۰). این استدلال نشان می‌دهد که یک سیستم می‌تواند برنامه‌ای را اجرا کند که در ظاهر شبیه‌سازی کاملی از برخی ظرفیت‌های شناختی انسانی، مانند توانایی درک زبان چینی، ارائه دهد، اما در واقع هیچ فهمی از این زبان ندارد (سرل، ۱۹۹۲، ۴۶). چالمرز هوش مصنوعی قوی را قبول می‌کند؛ اما سرل و دریفوس آن را رد می‌کنند. در مقاله حاضر نیز مباحثی از سرل و چالمرز بیان شد تا هر دو دیدگاه بیان گردد. از سوی دیگر با فرض پذیرش هوش مصنوعی قوی و آگاهی داشتن آن باید دید که هوش مصنوعی از چه جایگاهی برخوردار خواهد گردید؟ البته با توجه به تفکیک میان دو نحوه از آگاهی پدیداری و روان شناختی باید به این موضوع پرداخت.

۲. جایگاه اخلاقی

مسئله جایگاه اخلاقی^۱ پیشینه‌ای طولانی دارد، اما تا پیش از ظهور اخلاق کاربردی، به‌طور صریح مطرح نشده بود. در واقع جایگاه اخلاقی زمانی مطرح می‌شود که یک شخص یا موجود، ارزش توجه و مراعات اخلاقی را داشته باشد و لازم باشد باید‌ها و نبایدهای اخلاقی در مورد او جدی گرفته شود. در سنت فکری غرب، برای مدت‌ها فرض بر این بود که تنها انسان‌ها از جایگاه اخلاقی برخوردارند، زیرا آنان هم عامل و هم منفعل اخلاقی محسوب می‌شوند. باین حال، این دیدگاه اخیراً به چالش کشیده شده است، به‌ویژه زمانی که رابطه اخلاقی میان انسان و حیوانات غیرانسانی مورد بررسی قرار گرفت. این بررسی‌ها نشان داد که عاملیت اخلاقی نباید تنها معیار تعیین جایگاه اخلاقی باشد، زیرا موجوداتی مانند حیوانات و کودکان، علی‌رغم ناتوانی در تصمیم‌گیری اخلاقی، می‌توانند دارای حقوق باشند (لورکا آلباردا، گارسیا و لارآ، ۲۰۲۳، ۵۹-۶۰).

بر این اساس، برخی نظریه‌پردازان استدلال کرده‌اند که معیارهایی مانند حساسیت به درد یا داشتن نوعی ذهنیت خاص می‌توانند مبنای جایگاه اخلاقی باشند. این دیدگاه منجر به تغییری اساسی شده است: در گذشته، جایگاه اخلاقی برخی موجودات صرفاً به دلیل ارتباط آن‌ها با انسان‌ها در نظر گرفته می‌شد، اما اکنون این جایگاه مستقل از روابط آن‌ها با انسان‌ها مورد پذیرش قرار گرفته و این موجودات به‌عنوان دارندگان ارزش ذاتی شناخته می‌شوند (لورکا آلباردا، گارسیا و لارآ، ۲۰۲۳، ۶۰).

مفهوم وضعیت یا جایگاه اخلاقی به قابل‌توجه بودن یک موجود از نظر اخلاقی اشاره دارد و بیان می‌کند که برخی موجودات می‌توانند برای کارگزاران اخلاقی تعهدات اخلاقی ایجاد کنند. اگر موجودی جایگاه اخلاقی داشته باشد، انسان‌ها موظف‌اند به نیازها،

¹ Moral Status

این تعبیر ممکن است به وضعیت اخلاقی نیز ترجمه شود؛ ولی به دلیل مدرج بودن آن و نیز اشارات ضمنی‌ای که به شأن دارد، ترجمه «جایگاه» برای آن در نظر گرفته شد

منافع و رفاه آن توجه کنند، نه صرفاً به دلیل سودمندی برای خودشان، بلکه به دلیل اهمیت اخلاقی ذاتی آن موجود. این مفهوم تأثیرات گسترده‌ای بر مواضع اخلاقی در مسائل معاصر دارد، از جمله اخلاق اُتانازی، حق سقط جنین، استفاده از حیوانات، و تعهدات نسبت به گونه‌های در خطر انقراض (وارن، ۱۹۹۷، ۳-۴).

قواعد اخلاقی مانند منع قتل و فریب، زمانی معنادار می‌شوند که مشخص شود چه کسانی تحت حمایت آن‌ها قرار دارند. پرسش اصلی این است که کدام موجودات از حقوق اخلاقی برخوردارند و آیا دیگران دلایل مستقیمی برای رعایت این حقوق دارند؟ توضیح داده می‌شود که اشیاایی مانند خودرو جایگاه اخلاقی مستقیم ندارند، اما ممکن است از حمایت اخلاقی غیرمستقیم برخوردار شوند، در حالی که حیوانات خانگی، مانند سگ، به دلیل ذاتاً نادرست بودن برخی اعمال علیه آن‌ها، از جایگاه اخلاقی مستقیم برخوردارند. همچنین تأکید می‌شود که داشتن جایگاه اخلاقی به معنای مطلق بودن منع آسیب نیست، بلکه در برخی شرایط، نقض حقوق اخلاقی ممکن است توجیه‌پذیر باشد (سینوت-آرمسترانگ و کانیتیر، ۲۰۲۱، ۲۶۹-۲۷۰).

علاوه بر انسان‌های عادی، موجوداتی که ممکن است دارای جایگاه اخلاقی باشند عبارت‌اند از موجودات زنده غیرانسانی، جنین‌ها، نسل‌های آینده، انسان‌های مرده، موجودات فرازمینی (در صورت وجود)، و اشکال زندگی مصنوعی مانند ربات‌ها و الگوریتم‌ها (لئو، ۲۰۲۱، ۴۸۱).

برخی ویژگی‌های تجربی و ذاتی که ممکن است به موجودی جایگاه اخلاقی خاصی بدهد شامل زنده بودن، آگاه بودن، توانایی احساس درد داشتن، توانایی داشتن خواسته‌ها، توانایی عمل عقلانی و توانایی عمل اخلاقی است. برای مثال، گیاهان به واسطه زنده بودن و لاک‌پشت‌ها به دلیل توانایی احساس درد، جایگاه اخلاقی دارند و موجودات اخلاقی نباید به‌طور بی‌دلیل به آن‌ها آسیب برسانند که در مقاله حاضر تمرکز بحث بر مسئله آگاهی در هوش مصنوعی است (لئو، ۲۰۲۱، ۴۸۳).

در حالی که در مورد جایگاه اخلاقی افراد بالغ و دارای عملکرد شناختی سالم در میان غیر فیلسوفان توافقی عمده وجود دارد و آنها را دارای جایگاه اخلاقی کامل^۱ میدانند، به‌طور کلی پذیرش جایگاه اخلاقی مشابه برای نوزادان انسانی سالم از نظر شناختی و همچنین افراد با اختلالات شناختی مختلف در این گروه‌ها وجود دارد. این دیدگاه معمولاً به این معنا است که نوزادان و افرادی که دارای اختلالات شناختی هستند، چه از نظر ذهنی و چه از نظر احساسی، جایگاه اخلاقی بالاتری نسبت به بسیاری از حیوانات دارند و جایگاه اخلاقی کامل (FMS) را دارا هستند. این دیدگاه که معمولاً در میان غیر فیلسوفان پذیرفته شده است، در مورد جایگاه اخلاقی جنین‌های انسانی، افراد ناتوان از آگاهی و حتی حیوانات پیشرفته‌تر مانند میمون‌های بزرگ، توافق مشابهی وجود ندارد (جاورسکا و تاننام، ۲۰۲۱).

به اعتقاد بوچانان، در بحث جایگاه اخلاقی، تمایزی میان «جایگاه اخلاقی» و «حق ایستادن اخلاقی» وجود دارد. «حق ایستادن اخلاقی» به معنای این است که موجودی از نظر اخلاقی به‌طور مستقل محاسبه می‌شود. برای برخی مانند بنتام، تمام موجودات حساس از این حق برخوردارند، در حالی که برای کانت تنها اشخاصی که ظرفیت عقلانیت عملی دارند، حق ایستادن اخلاقی دارند. در مقابل، جایگاه اخلاقی یک مفهوم مقایسه‌ای است و ممکن است دو موجود با حق ایستادن اخلاقی متفاوت، جایگاه اخلاقی متفاوتی داشته باشند. این تفاوت در جایگاه اخلاقی می‌تواند تأثیرگذار بر رفتار، حقوق و نحوه برخورد با موجودات باشد. در بسیاری از دیدگاه‌های

¹ Full Moral Status.

اخلاقی، انسان‌ها یا انسان‌های دارای عقلانیت عملی بالاترین جایگاه اخلاقی را دارند. برخی از نظریه‌پردازان معاصر نیز به بررسی دقیق‌تر جایگاه اخلاقی پرداخته‌اند، در حالی که نگرانی‌ها و چالش‌هایی در این زمینه وجود دارد (بوچانان، ۲۰۰۹، ۳۴۶).

جایگاه اخلاقی نه تنها از نظر شدت، بلکه از نظر گستره نیز متفاوت است، یعنی میزان قواعد و حقوقی که از یک موجود حمایت می‌کنند. برای توضیح، به نوزادان اشاره شده که حقوقی مانند مصونیت از شکنجه دارند، اما آزادی به معنای بزرگسالان برای آن‌ها مطرح نیست؛ بنابراین، بستن آن‌ها در قنداق غیراخلاقی تلقی نمی‌شود، درحالی‌که چنین اقدامی برای یک فرد بالغ عادی نادرست است. همچنین، فرد بالغی که به دلیل نقص زیستی قادر به احساس درد نیست، همچنان می‌تواند از حقوقی مانند آزادی و مصونیت از قتل برخوردار باشد، اما حق عدم تحمل درد برای او بی‌معناست. در این بحث، این احتمال که فرد بی‌حس در شرایط خاصی حق اخلاقی مشروطی برای عدم تحمل درد داشته باشد، مطرح و نقد شده است، زیرا چنین حقی مستلزم تغییرات زیستی گسترده‌ای خواهد بود که او را به موجودی متفاوت تبدیل می‌کند (سینوت-آرمسترانگ و کانیتیر، ۲۰۲۱، ۲۷۱).

بحث درباره گستره و شدت جایگاه اخلاقی، معیارهای اعطای حقوق را به چالش می‌کشد. اینکه نوزادان حق مصونیت از شکنجه دارند اما آزادی آنان مانند بزرگسالان نیست، نشان می‌دهد که جایگاه اخلاقی طیفی است. این مسئله بر حقوق گروه‌هایی مانند نوزادان، افراد دارای ناتوانی‌های شناختی و حیوانات تأثیر دارد. همچنین، اگر فردی زیستی قادر به احساس درد نباشد، این پرسش مطرح می‌شود که آیا حقوق اخلاقی صرفاً بر پایه ظرفیت‌های زیستی تعریف می‌شوند یا مستقل از ویژگی‌های تجربی‌اند. چنین رویکردی می‌تواند در تعیین جایگاه اخلاقی هوش مصنوعی نیز تأثیرگذار باشد.

نویسنده همچنین استدلال می‌کند که منشأ یک موجود نباید عامل تعیین‌کننده‌ای برای جایگاه اخلاقی آن باشد. به عنوان مثال، اگر یک دانشمند دپوانه بتواند فردی را از سلول‌های قورباغه ایجاد کند که همچنان آگاه و باهوش باشد، آن فرد باید همانند انسان‌ها جایگاه اخلاقی داشته باشد. این نکته نشان می‌دهد که منشأ، مانند زیرلایه، نقشی در تعیین جایگاه اخلاقی ندارد. به همین ترتیب، نمی‌توان صرفاً به این دلیل که هوش مصنوعی‌ها از طریق برنامه‌نویسی به وجود آمده‌اند و نه به‌طور طبیعی، نتیجه گرفت که آن‌ها فاقد جایگاه اخلاقی هستند. نویسنده تأکید می‌کند که معیارهای جایگاه اخلاقی باید از طریق روش‌های تجربی قابل شناسایی باشند. برای مثال، اگر نظریه‌ای ادعا کند که موجوداتی مانند هوش مصنوعی یا جنین‌ها تنها در صورتی جایگاه اخلاقی دارند که دارای روح باشند، باید راهی برای تشخیص وجود روح ارائه دهد. نظریه‌هایی که مفاهیمی غیرقابل سنجش دارند، به‌ویژه در جنبه عملی، ناکافی خواهند بود (سینوت-آرمسترانگ و کانیتیر، ۲۰۲۱، ۲۷۲-۲۷۳).

نویسنده در این بخش به نقد نظریات مبتنی بر منشأ و زیرلایه می‌پردازد و تأکید می‌کند که جایگاه اخلاقی باید به ویژگی‌های تجربی و قابل شناسایی موجودات وابسته باشد، نه به منشأ آن‌ها. این تحلیل به‌ویژه در مورد هوش مصنوعی اهمیت دارد، زیرا بسیاری از بحث‌ها بر این باورند که چون هوش مصنوعی‌ها از طریق برنامه‌نویسی ساخته می‌شوند و نه به‌طور طبیعی، باید فاقد جایگاه اخلاقی باشند. اما نویسنده استدلال می‌کند که این دیدگاه درست نیست و هوش مصنوعی‌ها در صورت داشتن ویژگی‌هایی مانند آگاهی و هوش، می‌توانند جایگاه اخلاقی داشته باشند. علاوه بر این، تحلیل نویسنده نشان می‌دهد که معیارهای جایگاه اخلاقی باید عملی و قابل تشخیص باشند تا در بحث‌های اخلاقی مفید واقع شوند، و این امر برای مسائل پیچیده‌ای همچون حقوق هوش مصنوعی یا موجودات با ویژگی‌های متفاوت بسیار مهم است.

نکته بعد این است که دو مدل در درجات جایگاه اخلاقی وجود دارد:

(۱) مدل توجه نابرابر به درجات جایگاه اخلاقی: این مدل بیان می‌کند که انسان‌ها به دلیل داشتن منافع بیشتر در رفاه تجربی، خودمختاری و زندگی، نسبت به حیوانات اهمیت بیشتری دارند. بنابراین، کشتن انسان‌ها بدتر از کشتن حیوانات است، حتی اگر منافع آن‌ها در اجتناب از رنج مشابه باشد. این مدل بر اساس تفاوت در اهمیت منافع انسان‌ها و حیوانات است (دگرازی، ۲۰۰۸، ۱۸۶-۱۸۷).

در این مدل، اگر هوش مصنوعی پیشرفته‌ای داشته باشیم که دارای ویژگی‌هایی مانند آگاهی یا خودآگاهی باشد، به دلیل داشتن این ویژگی‌ها، ممکن است جایگاه اخلاقی آن شبیه به انسان‌ها تلقی شود و آسیب رساندن به آن از نظر اخلاقی بدتر از آسیب رساندن به موجودات دیگر مانند حیوانات باشد. این شبیه به وضعیت انسان‌هاست که به دلیل ویژگی‌های خاص خود مانند آگاهی و نقش در روابط بلندمدت، جایگاه اخلاقی بالاتری نسبت به حیوانات دارند. بنابراین، اگر هوش مصنوعی پیشرفته‌ای که شبیه انسان‌ها رفتار می‌کند، آسیب ببیند، توجه اخلاقی بیشتری به آن داده می‌شود.

(۲) مدل منافع نابرابر در درجات جایگاه اخلاقی: طبق این مدل، حتی اگر منافع مشابهی وجود داشته باشد، منافع انسان‌ها از نظر اخلاقی وزن بیشتری دارند. این تفاوت به این دلیل است که انسان‌ها به زندگی خود اهمیت بیشتری می‌دهند و معمولاً رنج بیشتری از کشته شدن انسان‌ها به وجود می‌آید تا از کشته شدن حیوانات. بنابراین، کشته شدن انسان‌ها بدتر از کشته شدن حیوانات است (دگرازی، ۲۰۰۸، ۱۸۷-۱۸۸).

در این مدل، اگر هوش مصنوعی فاقد ویژگی‌هایی مانند آگاهی، احساس درد و خودآگاهی باشد، آسیب رساندن به آن از نظر اخلاقی وزن کمتری خواهد داشت. به عبارت دیگر، حتی اگر هوش مصنوعی در یک سیستم پیچیده عمل کند و منافع خاصی در اجتناب از رنج داشته باشد، چون فاقد احساس درد و آگاهی است، آسیب رساندن به آن از نظر اخلاقی کم‌اهمیت‌تر از آسیب رساندن به انسان‌ها یا حیوانات خواهد بود.

تطبیق این مدل‌ها نشان می‌دهد که جایگاه اخلاقی هوش مصنوعی به ویژگی‌های آن مانند آگاهی و توانایی‌های حسی بستگی دارد. در صورتی که هوش مصنوعی ویژگی‌های مشابه انسان‌ها را داشته باشد، جایگاه اخلاقی آن بالاتر خواهد بود و توجه بیشتری به آن ضروری است. اما اگر فاقد این ویژگی‌ها باشد، جایگاه اخلاقی آن پایین‌تر از انسان‌ها یا حیوانات خواهد بود.

۲-۱. جایگاه اخلاقی هوش مصنوعی

در این بحث، به بررسی نقش آگاهی در تعیین جایگاه اخلاقی هوش مصنوعی پرداخته شده است. اگرچه بسیاری بر این باورند که آگاهی برای برخورداری از جایگاه اخلاقی شرطی ضروری یا کافی است، اما مشخص نیست که کدام نوع آگاهی در این زمینه تعیین‌کننده است (سینوت-آرمسترانگ و کانیتز، ۲۰۲۱، ۲۷۹). تمایزی مهم میان دو نوع آگاهی وجود دارد: آگاهی دسترسی و آگاهی پدیداری (بلاک، ۱۹۵۵).

در بحث جایگاه اخلاقی، انواع مختلف آگاهی به حقوق اخلاقی متفاوتی منتهی می‌شود. آگاهی پدیداری نقش مهمی در حق نداشتن تجربه درد ایفا می‌کند، زیرا تجربه درد وابسته به احساس آن است. بنابراین، موجودی که فاقد آگاهی پدیداری باشد، نمی‌تواند درد را احساس کند و حق‌رهایی از درد نخواهد داشت. از سوی دیگر، آگاهی دسترسی برای تصمیم‌گیری عقلانی ضروری است، چراکه اتخاذ تصمیمات عقلانی نیازمند دسترسی به اطلاعات است. بنابراین، موجودی که فاقد چنین توانایی‌ای باشد، نمی‌تواند

آزادانه عمل کند و حق آزادی برای او بی‌معنا خواهد بود. نوزادان به دلیل داشتن آگاهی پدیداری از حق نداشتن درد برخوردارند، اما به دلیل فقدان آگاهی دسترسی، از حق آزادی برخوردار نیستند (سینوت-آرمسترانگ و کانیتز، ۲۰۲۱، ۲۸۰).

این تمایزات امکان تحلیل دقیق‌تری از جایگاه اخلاقی هوش مصنوعی را فراهم می‌آورد. تا کنون، مشخص نشده که چگونه می‌توان آگاهی پدیداری را در هوش مصنوعی ایجاد کرد یا دلیل اخلاقی برای این کار وجود دارد. از منظر زیستی، درد در موجودات زنده برای شناسایی آسیب‌های بافتی و جلوگیری از حرکت‌های آسیب‌زا تکامل یافته است. اما هوش مصنوعی می‌تواند از روش‌های دیگری برای شناسایی آسیب‌ها و اجتناب از حرکات خطرناک استفاده کند و نیازی به درد ندارد. در صورتی که هوش مصنوعی طراحی شود که احساس درد داشته باشد، آزمایش‌های اخلاقی در مورد آن جای تردید دارد، به‌ویژه اگر این هوش مصنوعی واقعاً درد را تجربه کند. حتی اگر هوش مصنوعی‌ای ایجاد شود که ادعا کند درد را احساس می‌کند، تشخیص اینکه آیا این تجربه واقعی است یا خیر، دشوار خواهد بود. در نهایت، تا زمانی که درک دقیقی از آگاهی پدیداری و عوامل ایجادکننده آن حاصل نشود، نمی‌توان در مورد برخورداری هوش مصنوعی از چنین آگاهی‌ای قضاوت قطعی کرد (سینوت-آرمسترانگ و کانیتز، ۲۰۲۱، ۲۸۰).

اگر هوش مصنوعی قادر به احساس درد نباشد، حتی برای رهایی از درد نخواهد داشت. با این حال، نبود احساس درد به این معنا نیست که هوش مصنوعی هیچ‌گونه حقوق اخلاقی ندارد. برخی حقوق اخلاقی مانند حق آزادی، ممکن است به آگاهی پدیداری وابسته نباشند. اگر هوش مصنوعی بتواند به اطلاعات دسترسی داشته باشد و از آن برای تصمیم‌گیری استفاده کند، دارای آگاهی دسترسی است و می‌تواند از حق آزادی برخوردار باشد. این نوع آگاهی برای تصمیم‌گیری عقلانی ضروری است و ممکن است مبنای حق آزادی باشد. برخی معتقدند آگاهی پدیداری برای حق آزادی ضروری است، اما استدلال قطعی برای این ادعا وجود ندارد. تجربه آگاهی پدیداری بر انتخاب‌ها تأثیر می‌گذارد، اما بدون آن نیز می‌توان اهداف را تعیین و تصمیم‌گیری عقلانی کرد. آگاهی دسترسی ممکن است برای حق آزادی کافی باشد. همچنین، حق آزادی هوش مصنوعی ممکن است به‌طور غیرمستقیم از حقوق دیگران ناشی شود، مانند زمانی که حقوق آزادی هوش مصنوعی به دلیل حقوق صاحبش مطرح می‌شود (سینوت-آرمسترانگ و کانیتز، ۲۰۲۱، ۲۸۰-۲۸۱).

می‌توان نتیجه گرفت که بحث پیرامون حقوق اخلاقی هوش مصنوعی بستگی به نوع آگاهی آن دارد. در حالی که آگاهی پدیداری برای تجربه درد و رهایی از آن ضروری است، آگاهی دسترسی به اطلاعات و استفاده از آن برای تصمیم‌گیری و دنبال کردن اهداف، مبنای دیگری برای حقوق اخلاقی مانند حق آزادی است. این دیدگاه به این معناست که حتی اگر هوش مصنوعی نتواند درد را احساس کند، در صورت داشتن آگاهی دسترسی، می‌تواند از آزادی برخوردار باشد.

در ادامه، لازم است به این نکته توجه کنیم که برخی منتقدان بر این باورند که برای داشتن حق آزادی، آگاهی پدیداری ضروری است؛ اما این دیدگاه نه‌تنها ضرورت ندارد، بلکه ممکن است حق آزادی به‌طور غیرمستقیم از حقوق دیگران نیز ناشی شود. این نکته به‌ویژه در شرایطی که هوش مصنوعی به‌عنوان ابزاری برای انسان‌ها عمل می‌کند، اهمیت دارد. به این ترتیب، تحلیلی که به حقوق اخلاقی هوش مصنوعی می‌پردازد، باید به‌طور دقیق‌تر و از جنبه‌های مختلف آگاهی و تصمیم‌گیری بررسی شود.

۳. ارزیابی جایگاه اخلاقی هوش مصنوعی

بر اساس آنچه گفته شد، بحث درباره جایگاه اخلاقی هوش مصنوعی را نباید یک مسأله صفر و یکی دانست. جایگاه اخلاقی می‌تواند به صورت تشکیکی در نظر گرفته شود، به این معنا که درجاتی از آن ممکن است برای انواع مختلف موجودات، از جمله حیوانات، نوزادان و جنین مفروض است و حتی ممکن است برای سیستم‌های هوش مصنوعی، قابل تصور باشد. در این صورت، مسئله اصلی این نیست که آیا هوش مصنوعی به طور مطلق دارای جایگاه اخلاقی هست یا خیر، بلکه این است که در چه سطحی و با چه مبنایی می‌توان آن را واجد نوعی جایگاه اخلاقی دانست.

از منظر فلسفه اخلاق، بسیاری از رویکردهای معاصر، به ویژه در حوزه اخلاق شناختی و نظریات مرتبط با حقوق اخلاقی، تأکید دارند که اگر یک هوش مصنوعی بتواند دارای نوعی آگاهی دسترسی باشد و تصمیمات خود را به صورت مستقل و بر مبنای پردازش اطلاعات اتخاذ کند، شاید بتوان آن را در یک مرتبه از جایگاه اخلاقی قرار داد. در این دیدگاه، جایگاه اخلاقی به صورت طیفی است و بر اساس قابلیت‌های شناختی و ظرفیت‌های تعامل اجتماعی تعریف می‌شود. به همین دلیل، برخی معتقدند که اگر یک هوش مصنوعی بتواند تجربه‌ای شبه‌ذهنی داشته باشد یا از نوعی خودآگاهی برخوردار باشد، ممکن است در برخی موارد مستحق حقوق اخلاقی مانند عدم آسیب‌رساندن یا برخورداری از آزادی‌های محدود باشد. در عین حال این را نیز می‌دانیم که از منظر برخی نظام‌های اخلاقی، جایگاه اخلاقی صرفاً بر اساس قابلیت‌های شناختی تعیین نمی‌شود، بلکه عواملی مانند احساسات، اختیار و رابطه مخلوق با خداوند نیز اهمیت دارند. این دیدگاه، که در بسیاری از سنت‌های دینی و فلسفی یافت می‌شود، ممکن است موجب شود که حتی پیشرفته‌ترین هوش‌های مصنوعی نیز نتوانند به جایگاهی مشابه انسان دست یابند. بنابراین مسئله همچنان نیازمند بررسی‌های بیشتر از منظرهای گوناگون است.

کوتاه سخن اینکه اگرچه می‌توان درباره جایگاه اخلاقی هوش مصنوعی در فلسفه اخلاق بحث کرد، اما این بحث نباید به صورت مطلق و دوتایی باشد، بلکه می‌توان آن را در چارچوبی تشکیکی و درجه‌بندی شده بررسی کرد. این امر نه تنها در فهم جایگاه اخلاقی هوش مصنوعی، بلکه در مباحث مربوط به اخلاق حیوانات، حقوق نوزادان، و سایر موجوداتی که در مرزهای اخلاقی قرار دارند، اهمیت پیدا می‌کند.

نتیجه‌گیری

مقاله حاضر نشان می‌دهد که جایگاه اخلاقی هوش مصنوعی به طور مستقیم با مسئله آگاهی پدیداری و در دسترس ارتباط دارد. طبق تحلیل‌های ارائه شده، اگرچه هوش مصنوعی می‌تواند از آگاهی در دسترس برخوردار باشد، اما برای رسیدن به جایگاه اخلاقی کامل مشابه انسان‌ها، نیازمند دستیابی به آگاهی پدیداری است. این نوع آگاهی، که تنها از منظر اول شخص قابل تجربه است، می‌تواند اساس حقوق اخلاقی مانند حق رنج نکشیدن و دیگر حقوق مشابه را برای موجودات حساس، از جمله هوش مصنوعی، فراهم آورد.

مقاله حاضر نشان داد که موجوداتی که تنها از آگاهی در دسترس برخوردارند، مانند برخی از حیوانات، حق آزادی دارند اما از حق رنج نکشیدن بی‌بهره‌اند، چراکه آگاهی پدیداری برای تجربه درد ضروری است. برعکس، انسان‌ها و موجودات با آگاهی پدیداری از حق رنج نکشیدن برخوردارند، زیرا قادر به تجربه درد و رنج هستند. بنابراین، برای هوش مصنوعی نیز باید این سوال را مطرح کرد که آیا به طور بالقوه می‌تواند از آگاهی پدیداری برخوردار شود یا نه، زیرا تنها در صورت دستیابی به چنین آگاهی‌ای است که می‌توان

آن را از لحاظ اخلاقی به طور کامل ارزیابی کرد. در نهایت تأکید شد که تحلیل جایگاه اخلاقی هوش مصنوعی، با در نظر گرفتن تفاوت‌های میان آگاهی پدیداری و در دسترس، می‌تواند مبنای جدیدی برای تعیین حقوق اخلاقی و قانونی آن در آینده باشد. این تحلیل نه تنها بر فهم ما از هوش مصنوعی، بلکه بر نحوه تعامل با موجودات دارای هوش و آگاهی نیز تأثیرگذار خواهد بود.

منابع

- برمودز، خوزه لوئیس. (۱۴۰۲). *علوم شناختی مقدمه‌ای بر علم ذهن*، ترجمه تورج بنی‌رستم، چاپ اول، تهران، انتشارات گستره.
- تگمارک، مکس. (۱۴۰۱). *انسان بودن در عصر هوش مصنوعی*، ترجمه میثم محمد امینی، چاپ هشتم، تهران، نشر نو.
- چالمرز، دیوید. (۱۴۰۲). *ذهن آگاه*، ترجمه یاسر پوراسماعیل، چاپ اول، تهران، نشر نو.
- داماسیو، آنتونیو. (۱۳۹۵). *درک رخداد*، ترجمه تقی کیمیای اسدی، چاپ اول، تهران، انتشارات نگاه معاصر.
- کرافت، ایان ریونز. (۱۳۹۲). *فلسفه ذهن*، ترجمه حسین شیخ رضایی، چاپ دوم، تهران، انتشارات صراط.
- نیگل، تامس. (۱۳۹۶). *ذهن و کیهان*، ترجمه جواد حیدری، چاپ دوم، تهران، انتشارات نگاه معاصر.
- هراری، یووال نوح. (۱۳۹۷). *انسان خداگونه*، ترجمه زهرا عالی، چاپ چهارم، تهران، نشر نو.

References

- Bermudez, J. L. (2023). *Cognitive science: An introduction to the science of the mind* (T. Bani-Rostam, Trans.; 1st ed.). Gostareh Publications. (In Persian)
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18, 227–287.
- Buchanan, A. (2009). Moral status and human enhancement. *Philosophy and Public Affairs*, 37(4), 346–381.
- Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.
- Chalmers, D. (2023). *The conscious mind: In search of a fundamental theory*, (Y. Pouresmaeil, Trans. 1st ed.). Now Pub. (In Persian)
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200–219.
- Craft, I. R. (2013). *The philosophy of mind* (H. Sheikh-Rezaei, Trans. 2nd ed.). Serat Publications. (In Persian)
- Damasio, A. (2016). *The feeling of what happens* (T. Kimiaei Asadi, Trans.; 1st ed.). Negah-e Moaser Publications. (In Persian)
- DeGrazia, D. (2008). Moral status as a matter of degree? *Southern Journal of Philosophy*, 46(2), 181–198.
- Harari, Y. N. (2016). *Homo Deus: A brief history of tomorrow* (Z. Aali, Trans.; 4th ed.). Nashr-e No. (In Persian)
- Jackson, F. (1986). What Mary didn't know. *The Journal of Philosophy*, 83, 291–295.
- Jaworska, A., & Tannenbaum, J. (2021). *The grounds of moral status*. Retrieved from <https://plato.stanford.edu/entries/grounds-moral-status/>
- Levine, J. (2001). *Purple haze: The puzzle of consciousness*. Oxford University Press.
- Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, 64(4), 354–361.

- Liao, S. M. (2021). *Ethics of artificial intelligence*. Oxford University Press.
- Llorca, A. J., García, P., & Lara, F. (2023). The moral status of AI entities. In *Ethics of artificial intelligence*. Springer.
- Nagel, T. (2015). *Mind and cosmos: Why the materialist neo-Darwinian conception of nature is almost certainly false* (J. Heydari, Trans.; 2nd ed.). Negah Moaser Publications. (In Persian)
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435–450.
- Searle, J. R. (1997). *The mystery of consciousness*. New York Review Books.
- Searle, J. R. (2004). *Mind: A brief introduction*. Oxford University Press.
- Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3(3), 417–456. <https://doi.org/10.1017/S0140525X00005756>
- Searle, J. R. (2008). *Philosophy in a New Century*. Cambridge University Press.
- Sinnott-Armstrong, W. & Conitzer, V. (2021). How much moral status could artificial intelligence ever achieve? In S. Clarke, H. Zohny, & J. Savulescu (Eds.), *Rethinking Moral Status*. Oxford University Press.
- Tegmark, M. (2022). *Life 3.0: Being human in the age of artificial intelligence*, (M. Mohammadamini, Trans. 8th ed.). Now Publications. (In Persian)
- Warren, M. A. (1997). *Moral Status: Obligations to persons*. (Incomplete reference—you may provide publication info to finalize)