

**An Examination of Functionalism as the Foundation for
Computational Models of Free Will in Artificial Intelligence
from the Perspective of Transcendent Philosophy
(Hikmat-i Muta'aliyah)**



Mostafa Mokhtari¹ 

Kazim Fouladi Qaleh² 

1. Ph.D. in Islamic Philosophy, Faculty of Philosophy and Ethics, Bāqir al-`Ulūm University, Qom, Iran (**Corresponding Author**).

mokhtare.m14@gmail.com

2. Assistant Professor, Department of Computer Engineering, Faculty of Engineering, Fārābī Campus, University of Tehran, Qom, Iran.

kfouladi@ut.ac.ir

Abstract

The computational approach to the mind and mental properties, with various formulations including Machine Functionalism, has provided a suitable ground for proposing computational models of free will and decision-making in the field of Artificial Intelligence (AI). This approach has achieved relative successes along this path; however, the question remains: what are the deficiencies and challenges faced by the

This article is extracted from a PhD dissertation entitled "Analytical study of the relationship between human will and intelligent agents with emphasis on Mulla Sadra's principles" (Supervisor: Hassan Abdi, Advisor: Kazim Fouladi Qale) Department of Philosophy and Theology, Faculty of Philosophy and Ethics, Baqir-ul-Ulum University, Qom, Iran.

Mokhtari, M. & Fouladi Qaleh, K. (2025). An Examination of Functionalism as the Foundation for Computational Models of Free Will in Artificial Intelligence from the Perspective of Transcendent Philosophy (Hikmat-i Muta'aliyah). *Philosophia et Theologia: Dialogues in Criticism and Reflection*, 30(4), pp. 34-59.

<https://doi.org/10.22081/jpt.2025.71987.2239>

▣ **Article Type:** Research; **Publisher:** Islamic Sciences and Culture Academy

▣ **Received:** 2025/05/25 • **Revised:** 2025/06/15 • **Accepted:** 2025/07/08 • **Online Publication:** 2025/12/29

© 2025

authors retain the copyright and full publishing rights



computational model regarding human will and its philosophical foundations when confronted with deeper philosophical perspectives such as Transcendent Philosophy (Hikmat-i Muta'aliyah)? This research, using a descriptive-analytical method, examines the computational models proposed for free will in AI (focusing on the Symbolic Approach). Initially, Machine Functionalism is analyzed as the philosophical basis for this **approach. Then, relying on the principles and foundations of Sadrā's** Transcendent Philosophy, a comprehensive and critical explanation of these models is presented, answering the research's central question. The findings indicate that computational models lack the necessary ontological capacity to explain the truth of the will. This understanding neglects fundamental components such as consciousness (being conscious), phenomenal experience, and intentionality (التفاتی/*iltifātī status*), **which play a central role in Sadrā's explanation of the will.** The philosophical examination of this type of issue in the field of AI, apart from its philosophical effects and consequences, directly influences policy-making and legislation in this area, particularly concerning moral and legal responsibility. Therefore, evaluating the proposed ideas from the perspective of the foundations of Islamic philosophy is a philosophical and civilizational necessity.

Keywords

Human Will, Artificial Intelligence (AI), Symbolic Approach, Machine Functionalism, Soul (Nafs), Transcendent Philosophy (Hikmat-i Muta'aliyah), Mollā Sadrā.



تجزیه و تحلیل

بررسی کارکردگرایی به مثابه مبنای مدل‌های محاسباتی اراده آزاد در هوش مصنوعی ...

پرتال جامع علوم انسانی

بررسی کارکردگرایی به مثابه مبنای مدل‌های محاسباتی اراده آزاد در هوش مصنوعی از منظر حکمت متعالیه

ID^۱ مصطفی مختاری
 ID^۲ کاظم فولادی قلعه

۱. دکتری فلسفه اسلامی، دانشکده فلسفه و اخلاق، دانشگاه باقرالعلوم علیه السلام، قم، ایران (نویسنده مسئول).
 mokhtare.m14@gmail.com

۲. استادیار، گروه مهندسی کامپیوتر دانشکده مهندسی دانشکدهگان فارابی، دانشگاه تهران، قم، ایران.
 kfouladi@ut.ac.ir



چکیده

رویکرد محاسباتی به ذهن و ویژگی‌های ذهنی با تقریرهای مختلفی، از جمله کارکردگرایی ماشینی زمینه مناسبی را برای طرح مدل‌های محاسباتی از اراده و تصمیم‌گیری آزاد در حوزه هوش مصنوعی فراهم آورده است. این رویکرد در این مسیر، موفقیت‌های نسبی نیز به دست آورده است؛ اما جای این پرسش است که مدل محاسباتی درباره اراده انسانی و مبانی فلسفی آن، در مواجهه با نگرش‌های فلسفی عمیق‌تری همچون حکمت متعالیه دچار چه کاستی‌ها و چالش‌هایی است؟ این پژوهش با روش توصیفی - تحلیلی، به بررسی مدل‌های محاسباتی ارائه شده برای اراده آزاد در هوش مصنوعی (با تمرکز بر رهیافت نمادین) می‌پردازد. در ابتدا، کارکردگرایی ماشینی به مثابه مبنایی فلسفی برای این رویکرد تحلیل می‌شود. سپس با اتکا به اصول و



سال سی‌ام، شماره ۴، ۱۴۰۴

* این مقاله مستخرج از رساله دکتری با عنوان: «بررسی تحلیلی نسبت اراده انسانی با عامل‌های هوشمند با تاکید بر مبانی ملاصدرا (استاد راهنما: حسن عبدی، و استاد مشاور: کاظم فولادی قلعه)، گروه فلسفه و کلام، دانشکده فلسفه و اخلاق، دانشگاه باقرالعلوم علیه السلام، قم، ایران می‌باشد.

مختاری، مصطفی؛ فولادی قلعه، کاظم. (۱۴۰۴). بررسی کارکردگرایی به مثابه مبنای مدل‌های محاسباتی اراده آزاد در هوش مصنوعی از منظر حکمت متعالیه، نقدونظر، ۳۰(۴)، صص ۳۴-۵۹.

<https://doi.org/10.22081/jpt.2025.71987.2239>

▣ نوع مقاله: پژوهشی؛ ناشر: پژوهشگاه علوم و فرهنگ اسلامی.

▣ تاریخ دریافت: ۱۴۰۴/۰۲/۰۵ • تاریخ اصلاح: ۱۴۰۴/۰۳/۲۵ • تاریخ پذیرش: ۱۴۰۴/۰۴/۱۸ • تاریخ انتشار آنلاین: ۱۴۰۴/۱۰/۰۸

© ۲۰۲۵ «نویسندگان دارنده حق مؤلف مقاله خود بدون محدودیت هستند»



مبانی حکمت متعالیه، تبیینی جامع و انتقادی از این مدل‌ها ارائه و به پرسش محوری پژوهش پاسخ داده خواهد شد. یافته‌ها نشان می‌دهد که مدل‌های محاسباتی، فاقد ظرفیت وجودشناختی لازم برای تبیین حقیقت اراده‌اند. این تلقی، مؤلفه‌های بنیادینی همچون آگاهانه‌بودن، تجربه پدیداری و حیثیت التفاتی را که در تبیین صدرایی از اراده نقش محوری دارند، نادیده می‌گیرد. بررسی فلسفی این سنخ مسائل در حوزه هوش مصنوعی، جدا از آثار و پیامدهای فلسفی، در حوزه مسئولیت اخلاقی و حقوقی، به‌طور مستقیم در سیاست‌گذاری و قانون‌گذاری در این حوزه نیز تأثیرگذار است؛ از این‌رو ارزیابی ایده‌های مطرح‌شده از منظر مبانی فلسفه اسلامی، ضرورتی فلسفی و تمدنی است.

کلیدواژه‌ها

اراده انسانی، هوش مصنوعی، رهیافت نمادین، کارکردگرایی ماشینی، نفس، حکمت متعالیه، ملاصدرا.



پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



۳۷

محققان هوش مصنوعی، به‌ویژه کسانی که دغدغه طراحی هوش مصنوعی عمومی را دارند، با هدف درک بهتر از نحوه عملکرد و پیاده‌سازی ویژگی‌های ذهنی انسان در سیستم‌های مصنوعی، در پژوهش‌های خود کوشیده‌اند تا حد امکان ضمن تبیین مفهومی مقولات ذهنی، در مرحله فنی و عملیاتی نیز بسترهای آزمایشی را برای کاوش، اعتبارسنجی و اصلاح مدل‌های هوشمند برآمده از تعاریف خود مهیا کنند. پژوهش‌های موجود در این زمینه به‌طور عمده، از دو بخش نظری و فنی تشکیل شده است. در بخش نظری محققان تلاش کرده‌اند به‌طور عمده، با استفاده از روش فیلسوفان تحلیلی، توصیفی از چیستی ویژگی‌های ذهنی با تأکید بر مؤلفه‌های اصلی آن ارائه دهند و در بخش فنی، مدل‌های پیشنهادی خود را با استفاده از رهیافت‌های موجود برای پیاده‌سازی مطرح کنند. بر این اساس، نوع مواجهه متخصصان هوش مصنوعی با چیستی ذهن و ویژگی‌های ذهنی، به‌طور مستقیم در مدل‌سازی و شبیه‌سازی آن در سامانه‌های هوشمند تأثیر می‌گذارد.

یکی از جدی‌ترین این پژوهش‌ها که در سالیان اخیر، مورد توجه محققان هوش مصنوعی و فیلسوفان ذهن قرار گرفته است، «اراده آزاد» و امکان تحقق آن در سیستم‌های هوش مصنوعی است. از آنجا که این بحث در چارچوب یک فهم فیزیکالیستی از انسان مطرح شده است، به‌طور طبیعی این پرسش مطرح می‌شود که چگونه ممکن است یک ذهن کاملاً فیزیکی که هر تحول آن به‌شدت با قوانین فیزیکی اداره می‌شود، همچنان هرگونه آزادی انتخاب داشته باشد. فیلسوفان برای رفع مانع اراده آزاد در چارچوب یک ذهن کاملاً فیزیکی، راهکارهای مختلفی پیشنهاد داده‌اند. این راهکارها عموماً در تلاش‌اند تا بازتعریفی از این مفهوم بر اساس جدیدترین یافته‌های علمی، به‌ویژه مبتنی بر یافته‌های عصب‌شناسی و علوم رایانه‌ای ارائه دهند (Russell & Norvig, 2003, p. 955). در همین راستا، برخی پژوهشگران هوش مصنوعی در تلاش برای طراحی ربات‌هایی با توان اراده آزاد، مفهومی ابتدایی از آن را صورت‌بندی



کرده و مدل‌هایی محاسباتی برای آن ارائه داده‌اند. این مفهوم‌سازی که به‌طور عمده، تحت تأثیر رویکرد فنی و مهندسی شکل گرفته است، با استفاده از تحلیل مفهومی و ساده‌سازی‌های هدفمند مفاهیم پیچیده انسانی، بیشتر بر جنبه‌های عملیاتی و رفتاری اراده متمرکز بوده‌اند. در این رویکرد تلاش شده است اراده به‌صورت کارکردی و مستقل، از مؤلفه‌های پیچیده شناختی و انگیزشی آن در ذهن انسان بازنمایی شود. به همین دلیل، این مدل‌ها درگیر شدن با لایه‌های متافیزیکی و ساحت‌های وجودی اراده را رها کرده و بیشتر بر قابلیت‌های بازتولیدپذیر در سامانه‌های مصنوعی متمرکز نموده‌اند. در یک چارچوب فلسفی می‌توان برخی از آن مدل‌های محاسباتی را با دیدگاه کارکردگرایی ماشینی^۱ همسو دانست؛ دیدگاهی که به‌نوبه خود، هم‌نقطه عطفی در نظریه‌های مرتبط با ذهن به شمار می‌رود و هم برای محققان و طراحان هوش مصنوعی راهگشا بوده است تا آنجا که در متون رسمی و تخصصی هوش مصنوعی و ذیل بخش مبانی فلسفی آن، از میان انبوه نظریه‌های ارائه‌شده درباره ذهن و حالات ذهنی بر آن تأکید ویژه‌ای شده است (Russell & Norvig, 2003, p. 954). کارکردگرایی ماشینی با به‌چالش کشیدن دیدگاه‌های پیش از خود، به‌ویژه رفتارگرایی، زیربنای مفهومی بسیاری از تحقیقات در علوم شناختی - به‌ویژه هوش مصنوعی - را فراهم کرده است (Block, 1994, p. 171). بر این اساس، ذهن به‌مثابه یک برنامه رایانه‌ای پیچیده در نظر گرفته می‌شود که بر سخت‌افزار خاصی قابل اجراست. نوع سخت‌افزار (مغز یا سیلیکون) چندان مهم نیست؛ آنچه مهم است، ساختار و سازمان‌دهی این برنامه است (Putnam, 1988, p. 73).

شناخت دقیق از چیستی اراده انسانی و نسبتی که ممکن است با هوش مصنوعی داشته باشد، اهمیت این سنخ مسائل را مضاعف کرده است؛ زیرا «این فناوری در حال تغییر تفکر انسان، دانش، ادراک و واقعیت و از این رهگذر تغییر مسیر تاریخ انسان است. در عصری که ماشین‌ها به‌طور فزاینده‌ای وظایفی را بر عهده می‌گیرند که پیش‌تر،

1. Functionalism





انسان‌ها قادر به انجام آنها بودند، چه چیزی هویت ما را به‌عنوان انسان تشکیل خواهد داد؟» (کسینجر و همکاران، ۱۴۰۱، ص ۱۲). به همین دلیل، ضروری است به هوش مصنوعی نه تنها نگاه ابزاری، بلکه نگاهی به بنیادها و چارچوب‌های فلسفی آن نیز داشت. به تعبیر کسینجر و همکاران او، «زمانی که دنیای دیجیتال آغاز به گسترش کرد، هیچ انتظاری وجود نداشت که سازندگان می‌بایست چارچوبی فلسفی ایجاد کنند؛ اما امروزه این مسئله تبدیل به ضرورتی اجتناب‌ناپذیر شده است؛ به گونه‌ای که جان مک‌کارتی به‌مثابه یک مهندس هوش مصنوعی اذعان می‌کند که «برای محقق در زمینه هوش مصنوعی مهم است که به آنچه فیلسوفان گفته‌اند توجه کند» (McCarthy & Hayes, 1969, p. 5). نوآوری این پژوهش در پیوند زدن مبانی نظری کارکردگرایی، به‌ویژه کارکردگرایی ماشینی با دیدگاه حکمت متعالیه در باب اراده و نفس است. این بررسی با تکیه بر مبانی فلسفه صدرایی می‌کوشد تا سنجشی فلسفی از امکان و حدود تحقق اراده آزاد در سامانه‌های هوشمند ارائه دهد. این تحلیل افزون بر توجه بر نقایص رویکردهای صرفاً مهندسی محور می‌تواند در تبیین نسبت انسان و ماشین و نیز ترسیم افق‌های اخلاقی و حقوقی هوش مصنوعی در تمدن اسلامی نقشی مؤثر ایفا کند.

۱. مدل‌های محاسباتی از اراده آزاد در هوش مصنوعی

به اذعان محققان، اراده در زمره مفاهیم خاص، پیچیده و حتی مبهمی است که فقدان اجماع واحد در تعریف آن، حتی در میان فیلسوفان نیز مشهود است. به سبب همین ابهام است که تفسیرهای گوناگونی از آن صورت گرفته است؛ از این‌رو در تحقیقات هوش مصنوعی با انبوهی از مدل‌سازی‌ها و رویکردهای تفسیری متفاوت مواجه هستیم. برخی مهندسان هوش مصنوعی با هدف عملیاتی‌سازی مفهوم اراده، آن را در چارچوب مفاهیمی چون انتخاب و تصمیم‌گیری تعریف کرده‌اند؛ بدین معنا که اراده آزاد به‌مثابه الگوریتمی برای انتخاب میان گزینه‌های ممکن بازنمایی می‌شود، بی‌آنکه ضرورتی به بازسازی لایه‌های پیچیده‌تر روان‌شناختی یا متافیزیکی آن وجود داشته باشد. این فرایند شامل انتخاب گزینه‌ها و پردازش اطلاعات داخلی و خارجی است. این نوع نگاه به اراده

را می‌توان در کامپیوترهای مدرن نیز یافت؛ به این معنا که کامپیوترها نیز مانند انسان‌ها می‌توانند تصمیم بگیرند و این انتخاب‌ها ممکن است به آنها اعتبار اراده آزاد دهد (Simons, 1985, p. 108). این شیوه نگرش به مقولات ذهنی، به نگرش مکانیکی - محاسباتی شناخته می‌شود. این ایده در تلاش است تا با پیاده‌سازی الگویی از اراده آزاد در چارچوب هوش مصنوعی دترمینیستی، هماهنگی بنیادین اراده آزاد با دیدگاه سازگارگرایی^۱ فلسفی را به اثبات برساند (McCarthy, 2000, p. 3). این دیدگاه تحت تأثیر روش تحلیل منطقی - زبانی با تأکید بر بررسی‌های مفهومی و منطقی، کارآمدی مفاهیم را به جای تحلیل سنتی ماهیت آنها دنبال می‌کند؛ از این رو به جای تمرکز بر تحلیل ذات و ماهیت اراده و بحث‌های متافیزیکی درباره آن، بر فرایندها و مکانیزم‌های عملیاتی تمرکز دارد. این ایده، تحدید فیزیکیالیستی ذهن و امور ذهنی را در چشم‌انداز جدیدی قرار می‌دهد (خاتمی، ۱۴۰۰، صص ۱۳۲-۱۳۳) و با پذیرش اصل تبیین مکانیکی و فیزیکی ذهن، چارچوبی برای بررسی ویژگی‌های ذهنی و از جمله اراده آزاد فراهم می‌کند که در آن، ذهن و حالات ذهنی، به‌طور عمده به مکانیزم‌های فیزیکی و کارکردی فروکاسته می‌شوند.

این تلقی از اراده آزاد به معنای آن است که اگر انسان یا یک کامپیوتر بتوانند از میان گزینه‌های مختلف انتخابی داشته باشند، آنها صاحب اراده آزاد خواهند بود. بر اساس این رویکرد، مکانیسمی که عمل را تعیین می‌کند به دو بخش تقسیم می‌شود: بخش اول، اقدامات احتمالی و پیامدهای آنها را محاسبه می‌کند. سپس، بخش دوم تصمیم می‌گیرد که کدام عمل ارجحیت بیشتری دارد و آن را انتخاب می‌کند و انجام می‌دهد و می‌کوشد با یک مدل زبانی منطبق ریاضی که اغلب در هوش مصنوعی کلاسیک استفاده می‌شود، این مفهوم را رسمی کند (McCarthy, 2005, p. 1). در این رویکرد، تصمیم‌گیری به‌مثابه یک فرایند محاسباتی در نظر گرفته می‌شود که در آن ماشین‌ها با استفاده از مدل‌های داخلی خود از جهان و داده‌های موجود، گزینه‌های

۱. به آن دسته از دیدگاه‌هایی اطلاق می‌شود که قائل به سازگاری آزادی با جبر هستند (خاتمی، ۱۴۰۰، ص ۸۶).





مختلف و نتایج احتمالی هریک را ارزیابی می‌کنند. این فرایند محاسباتی که بر اساس ورودی‌ها و الگوریتم‌های مشخصی عمل می‌کند، به‌مثابه علت نهایی تصمیمات سیستم به‌شمار می‌رود.

مدل محاسباتی اراده، به‌مثابه یک مدل شناخته‌شده و مورد پذیرش بسیاری از مهندسان هوش مصنوعی مطرح شده است. در ادامه، جهت آشنایی، به معرفی برخی از مصادیق برجسته آن می‌پردازیم.

۱-۱. مدل «اراده آزاد قطعی ساده»^۱ مک‌کارتی

در سال ۱۹۸۸م، جان مک‌کارتی اظهار داشت که ربات‌هایی که ما ایجاد می‌کنیم، کاملاً سیستم‌های قطعی هستند و از این رو فرض اراده آزاد برای ربات بی‌معنا خواهد بود. بعدها، از این مدعا فاصله گرفت و مدلی از «اراده آزاد قطعی ساده» را معرفی کرد (McCarthy, 2000, p. 1). او اراده آزاد را به‌مثابه مؤلفه‌ای در نظر گرفت که هر عاملی حتی ربات‌ها می‌توانند سهمی از آن داشته باشند. او به‌جای تمرکز بر جنبه‌های فلسفی یا الهیاتی اراده آزاد، بیشتر به کاربردهای عملی و تکنولوژیک آن در زمینه طراحی ربات‌ها پرداخت (McCarthy, 2006, p. 1) و دو مؤلفه اصلی یعنی اقدامات (گزینه‌های) ممکن و انتخاب گزینه ارجح را برای آن در نظر گرفت (McCarthy, 2005, p. 1). عامل از میان اقدامات جایگزین انتخابی دارد و عملی را انتخاب می‌کند که دارای رجحان بیشتری است.

این تلقی از اراده مبتنی بر اصل ساده‌سازی بود که در نیمه قرن بیستم متأثر از موضع نئوپوزیتیویستی در علوم پوزیتیو شایع بود. بر اساس این اصل، همه امور باید به ساده‌ترین حالت ممکن تحویل روند و فروکاسته شوند (اصل تحویل). اگر قابل تحویل نباشند، باید تبیین علمی موجهی داشته باشند (اصل تبیین)، و اگر قابل تبیین نباشند، باید حذف شوند (اصل حذف) (خاتمی، ۱۴۰۱، ص ۴۸۰). مک‌کارتی بر این باور بود که برای

1. Simple deterministic free will

بازنمایی^۱ مفاهیم پیچیده به خصوص آن سنخ مفاهیمی که برخاسته از هوش انسانی است، لازم است مفاهیم به گونه‌ای ساده‌سازی شوند که قابلیت تبدیل شدن به زبان رایانه را پیدا کنند. بر همین اساس، در این مدل، مفاهیمی مانند انتخاب و تصمیم‌گیری را که در سنت فلسفی کلاسیک همراه نوعی آگاهی به خود و شهود درونی است، به گونه‌ای تفسیر می‌شود که مستلزم شهود درونی عامل نباشد. برای نزدیک شدن به چنین درکی، اراده در سیستم‌های هوش مصنوعی به مثابه فرایندی محاسباتی تعریف می‌شود که در آن ماشین‌ها با استفاده از مدل‌های خود از جهان و اطلاعات موجود، اعمال ممکن و پیامدهای آنها را ارزیابی می‌کنند. بر این اساس، هوش مصنوعی می‌تواند بدون اتکا به شهود درونی، اقدام به تصمیم‌گیری کند و نیازمند آگاهی از فرایند تصمیم‌گیری خود نیست (McCarthy, 2002, p. 2).

۲-۱. مدل «تولید و آزمون»^۲ سایمون

مدلی را که توسط هربرت سایمون معرفی شد نیز می‌توان نوعی دیگر از مدل‌های محاسباتی به شمار آورد. این مدل که از دسته مدل‌های مبتنی بر قوانین^۳ است، به مثابه روشی کارآمد برای حل مسئله و شبیه‌سازی فرایندهای تصمیم‌گیری انسانی معرفی شده است. در این مدل، فرایند حل مسئله به دو مرحله اصلی تقسیم می‌شود: (۱) تولید قوانین:^۴ در این مرحله، مجموعه‌ای از قوانین و راه‌حل‌های ممکن برای مسئله تولید می‌شود. این راه‌حل‌ها می‌توانند از راه‌های مختلفی، از جمله استفاده از قوانین تجربی، تصادفی یا بر اساس دانش قبلی تولید شوند. (۲) آزمون:^۵ در این مرحله، راه‌حل‌های تولیدشده ارزیابی

۱. بازنمایی (Representation) یکی از مفاهیم کلیدی در تحقق ایده مک‌کارتی و دو مدل بعدی است و از ارکان اصلی در رهیافت نمادین به هوش مصنوعی به شمار می‌رود. هدف از آن، تبدیل مفاهیم پیچیده به ساختارهایی است که رایانه بتواند آنها را پردازش کند. در ادامه، با رهیافت نمادین، بیشتر آشنا خواهیم شد.

2. Generate and Test
3. Rule-Based Models
4. Generate Rules
5. Test





می‌شوند تا تعیین شود که آیا آنها مسئله را حل می‌کنند یا خیر. راه‌حل‌های نامناسب حذف می‌شوند و راه‌حل‌های مناسب به‌مثابه گزینه‌های قابل پذیرش نگه داشته می‌شوند (Simon, 1973, p. 16). سایمون معتقد بود که ذهن انسان در مواجهه با مسائل پیچیده، از روشی مشابه همین مدل استفاده می‌کند؛ به این معنا که ابتدا بر اساس ورودی‌های داده‌شده به سیستم، تعدادی راه‌حل (گزینه) ممکن در نظر گرفته می‌شود و سپس این راه‌حل‌ها بر اساس معیارهای مشخصی، ارزیابی و آزمون می‌شوند تا بهترین گزینه انتخاب شود.^۱

۲. مدل محاسباتی و رهیافت نمادین^۲

منظور از «محاسباتی» بودن در اینجا آن است که مؤلفه‌های اصلی مفهوم اراده، همچون انتخاب، آگاهی از گزینه‌ها و تصمیم‌گیری، همگی در قالب ساختارهای صوری و الگوریتمی قابل بازنمایی، پیاده‌سازی و تحلیل هستند. این بدان معناست که رفتار یک عامل (خواه انسانی و خواه مصنوعی) می‌تواند بر اساس منطق، الگوریتم‌ها تعریف شود و قابلیت اجرا در سیستم‌های واقعی را داشته باشد. بر همین اساس، مدل محاسباتی اراده آزاد، در نسبت مستقیم و بنیادی با رهیافت نمادین در هوش مصنوعی قرار دارد. تفسیرهایی که اراده را به‌مثابه یک فرایند بسیط و مکانیکی تحلیل می‌کنند، در مرحله مدل‌سازی، از مدل‌های نمادین و تئوری‌های منطقی بهره می‌برند. از منظر علوم شناختی که به دنبال شناخت ذهن و کارکردهای شناختی آن است، رهیافت نمادین را می‌توان مدلی در مورد پردازش اطلاعات در ذهن بررسی کرد. این مدل که توسط دو دانشمند علم کامپیوتر و برنده جایزه تورینگ، هربرت سیمون و آلن نیوول، پیشنهاد شد، با عنوان «فرضیه سیستم نماد فیزیکی» در سال ۱۹۷۵م مطرح گردید (برمودز، ۱۳۹۶، ص ۲۱۱). این

۱. باب دوئل در فصل دوازدهم از کتاب اراده آزاد، رسوایی در فلسفه، به انواعی از مدل‌های اراده آزاد با عنوان «مدل‌های دومرحله‌ای» اشاره کرده است که برخی از آنها ذیل مدل‌های محاسباتی قابل تحلیل است (Doyle, 2011, pp. 160-184).

2. Symbolic

فرضیه هر چند ابتدا از دامن آموزه تورینگ برخواست، در محدوده این آموزه باقی نماند (خاتمی، ۱۴۰۱، ص ۴۵۹). رایانه کلاسیک را می توان نمونه‌ای از سیستمی دانست که مطابق با این فرضیه عمل می کند. یک رایانه کلاسیک ماشینی است که نمادهای زبان گونه ساختارمند را به مثابه ورودی می گیرد و نمادهای زبان گونه ساختارمند را به مثابه خروجی تولید می کند. این سیستم ورودی - خروجی تحت تأثیر و وساطت کاربرد قوانینی است که در مجموع برنامه مورد اجرای ماشین را تشکیل می دهند. کاربرد قوانین دست کاری نمادها - اجرای برنامه - فرایندی مکانیکی است؛ زیرا به هیچ گونه هوش یا خلاقیت از سوی ماشین نیازمند نیست. این امر مستلزم آن نیست که ماشین معنای نمادهای مورد دست کاری را درک کند. ماشین تنها به ویژگی‌های نحوی و صوری نمادها حساس است، و قوانین مورد کاربرد آن برای دست کاری نمادها به چنین ویژگی‌هایی مربوط می شود (کین، ۱۳۹۶، ص ۵۰). این نگاه به رایانه‌های کلاسیک که بر نحوه پردازش نمادها بدون درک معنای آن‌ه تأکید دارد، زمینه‌ای را فراهم می کند تا هوشمندی را به مثابه نتیجه بازنمایی‌های ذهنی پیچیده در انسان نیز در نظر بگیریم. بنابراین نظریه نمادین به طور غیرمستقیم، به مسئله مهم چستی حالات ذهنی، از جمله اراده انسان نیز دلالت دارد؛ به این معنا که اگر هوشمندی را توانایی حل مسئله و تصمیم گیری بدانیم، می توان آن را به مثابه مبنایی برای فهم نحوه تحقق اراده در انسان و امکان شبیه سازی آن در ماشین‌ها در نظر گرفت. تصمیم گیری در چارچوب سیستم‌های نماد فیزیکی، حاصل پردازش نمادها و انتخاب گزینه بهینه از میان چندین امکان است. بنابراین درک نحوه پردازش نمادها در ماشین‌ها و انسان‌ها می تواند گامی در جهت فهم چگونگی تحقق اراده و شرایط امکان آن در مصنوعات هوشمند باشد. بر اساس این دیدگاه، همه رفتارهایی که از انسان به مثابه موجودی هوشمند صادر می شود، برون داد بازنمایی‌های پیچیده‌ای است که در ذهن انجام گرفته است (کین، ۱۳۹۶، ص ۵۰).

در این رویکرد، تأکید اصلی روی برنامه‌ها یا الگوریتم‌هایی است که عملیات نمادین را انجام می دهند. این برنامه‌ها حاوی مجموعه کاملی از قوانین و دستورالعمل‌هایی هستند که تعیین می کنند چگونه نمادها می توانند ترکیب شوند، تغییر





کنند و برای انجام وظایف مختلف مورد استفاده قرار گیرند. سخت‌افزار یا ماشین، تنها بستری برای اجرای این برنامه‌ها است. این برنامه‌ها که حاوی قوانین دست‌کاری نمادها هستند می‌توانند روی انواع مختلفی از سخت‌افزار اجرا شوند؛ از این‌رو رویکرد نمادین در اصل ماهیتی محاسباتی دارد و بر قوانین رسمی و فرایندهای دست‌کاری نماد تمرکز می‌کند، بدون اینکه لزوماً مکانیسم‌های بیولوژیکی زیربنایی این فرایندها در مغز را در نظر بگیرد؛ اما با این حال، توضیحات بیولوژیکی را رد نمی‌کند (خاتمی، ۱۴۰۱، صص ۴۶۶-۴۷۰).

به‌طور خلاصه می‌توان گفت این رهیافت بر اساس این اصول بنا شده است: (۱) ذهن در اصل پردازشگر اطلاعات است. (۲) اطلاعات را می‌توان در قالب نمادها نشان داد. (۳) نمادها از طریق مجموعه‌ای از قوانین با یکدیگر ترکیب می‌شوند. (۴) عملکرد ذهنی (یا مغزی) مشابه عملکرد ماشین تورینگ است (خاتمی، ۱۴۰۱، صص ۳۴۸)؛ از این‌رو رهیافت نمادین ذیل قوی‌ترین ایده مطرح‌شده در علوم شناختی یعنی کارکردگرایی ماشینی صورت‌بندی می‌شود و فیزیکیسم با تقریر مکانیکی - ماشینی ذهن، از اصول اولیه این رهیافت است (خاتمی، ۱۴۰۱، صص ۴۶۲-۴۶۳). بسیاری از فلاسفه و دانشمندان علوم شناختی بر این باورند که عملیاتی که مغز روی بازنمایی‌ها انجام می‌دهد، به‌طور کامل فیزیکی و مکانیکی است (برمودز، ۱۳۹۶، صص ۲۲۸).

۳. کارکردگرایی ماشینی به‌مثابه مبناي فلسفی مدل‌های محاسباتی از اراده

در قرن هفدهم، دکارت معتقد بود که ذهن و بدن دو ماهیت جداگانه دارند و بدن انسان یک ماشین پیچیده است. این دیدگاه دوگانه‌گرا تا قرن نوزدهم بر بسیاری از اندیشمندان تأثیر‌گذار بود. این ایده که انسان می‌تواند به‌مثابه یک ماشین پیچیده در نظر گرفته شود، در قرن هجدهم مطرح شد و با پیشرفت‌های علمی و فناوری، این ایده پررنگ شد که «مغزها ماشین‌های بسیار پیچیده‌ای هستند و رفتار آنها در نهایت، به همان صورت مکانیکی درک خواهد شد که هر ماشین دیگری درک می‌شود» (Newsome, 2009, p. 54) و همین نگاه، زمینه ظهور هوش مصنوعی در قرن بیستم را سرعت

بخشید. این سیر تاریخی نشان می‌دهد که ایده کارکردگرایی ماشینی در فلسفه ذهن معاصر، ریشه در اندیشه‌های فلسفی قرن هفدهم و هجدهم دارد (خانی، ۱۴۰۰، صص ۱۹۳-۱۹۷).

هیلاری پاتنم در مجموعه مقالاتی با عنوان «ذهن و ماشین‌ها» در سال ۱۹۶۰م و در مقاله «ماهیت حالات ذهنی» که درباره شناسایی حالات ذهنی با رویکرد کارکردگرایی در سال ۱۹۶۷م نوشته بود، ایده ذهن ماشینی را به اوج خود رساند (Shagrir, 2010, pp. 220-250). او در این شناسایی پیشنهاد داد «دانستن اینکه یک انسان دارای یک باور خاص، یا ترجیح یا هر چیز دیگری است، مستلزم دانستن چیزی در مورد سازمان عملکردی انسان است» (Putnam, 1979, p. 424). پاتنم معتقد بود که مطالعه ماشین‌ها می‌تواند به روشن شدن مسائل مربوط به ذهن کمک کند. او پیشنهاد کرد که مقایسه ذهن با ماشین‌های محاسباتی، فهم بسیاری از پرسش‌های پیچیده درباره ذهن و هوشیاری را تسهیل می‌کند؛ از همین رو، او با استناد به این قیاس، به سراغ حل مسائل ذهن و بالتبع رابطه ذهن و بدن رفت (Putnam, 1979, p. 362). در این تلقی، حالات ذهنی به مثابه حالت‌های محاسباتی مغز تفسیر می‌شد. این مدل، مغز را به مثابه یک رایانه دیجیتال توصیف می‌کرد و روان‌شناسی را به مثابه نرم‌افزار حاکم بر سخت‌افزار زیستی در نظر می‌گرفت (Putnam, 1988, p. 73). پاتنم از ادبیات علوم کامپیوتر و دو کلیدواژه نرم‌افزار و سخت‌افزار کمک گرفت و تمایز میان این دو را برای توضیح حالات ذهنی مورد توجه قرار داد (Westphal, 2016, p. 68). او می‌خواست با این کار نشان دهد که می‌توان با استفاده از مفاهیم و ابزارهای علوم کامپیوتر، به درک بهتری از ماهیت ذهن و هوشیاری دست یافت. او معتقد بود به دلیل شباهت‌های کارکردی میان ذهن و ماشین تورینگ، همان‌طور که دو توصیف ممکن برای یک ماشین تورینگ وجود دارد، دو توصیف ممکن برای ذهن انسان نیز وجود دارد. یکی از این توصیف‌ها، به ساختار فیزیکی و شیمیایی آن اشاره می‌کند که با توصیف سخت‌افزار ماشین مطابقت دارد. دیگری، توصیف انتزاعی‌تر از فرایندهای ذهنی انسان است که بر قواعد و ساختارهای منطقی حاکم بر کنترل ذهن تکیه دارد؛ از این رو آنچه مهم است، ماهیت فیزیکی مغز نیست، بلکه سازمان‌دهی عملکردی آن





است. به عبارت دیگر، مهم نیست که مغز ما از چه ماده‌ای ساخته شده است، بلکه نحوه پردازش اطلاعات در آن است که تعیین می‌کند آیا ما دارای حالات ذهنی هستیم یا خیر (Shagrir, 2010, pp. 7-8). به تعبیر پاتنم، «ما می‌توانیم از پنیر سوئیسی ساخته شده باشیم» (Putnam, 1980, p. 134). این جمله بیانگر آن است که حالات ذهنی ما به‌جای اینکه به ماهیت فیزیکی مغز (سخت‌افزار) وابسته باشند، به نحوه پردازش اطلاعات در مغز بستگی دارند. به عبارت دیگر، مغز ما مانند یک کامپیوتر عمل می‌کند و حالات ذهنی ما برنامه‌های نرم‌افزاری هستند که روی این سخت‌افزار اجرا می‌شوند. حتی اگر مغز ما از پنیر سوئیسی ساخته شده باشد، به شرطی که سازمان‌دهی عملکردی آن به همان شکلی باشد که اکنون هست، می‌توانیم همچنان همان حالات ذهنی را داشته باشیم. این یکی از آموزه‌های مهم کارکردگرایی ماشینی است که نشان می‌دهد که ذهن از ماده مستقل است و به نوع ماده‌ای که از آن ساخته شده است، وابسته نیست.^۱

در کارکردگرایی ماشینی، از ماشین تورینگ به‌مثابه یک مدل انتزاعی استفاده می‌شود تا نشان دهد که «حالات ذهنی» چگونه به‌صورت تابع‌هایی از داده‌های ورودی، حالات داخلی و پاسخ‌های خروجی قابل توصیف هستند. این رویکرد به‌طور مشخص، به‌دنبال ارائه یک توصیف محاسباتی از فرایندهای ذهنی است (Putnam, 1988, p. 73). ماشین تورینگ به‌منزله یک دستگاه دست‌کاری نماد، ورودی را دریافت می‌کند و آن را به‌مثابه نماد روی نوار حافظه می‌نویسد (یا پاک می‌کند). همه این عملیات مطابق با مجموعه‌ای از دستورالعمل‌ها که «جدول ماشین» نامیده می‌شود، صورت می‌گیرد. جدول ماشین یک مسئله پیچیده را به وظایف کوچک تقسیم می‌کند و به این وظایف یا حالت‌های کوچک، حالت‌های ماشین می‌گویند. ذهن هر موجودی را می‌توان با یک ماشین تورینگ مدل‌سازی کرد.

کارکردگرایی ماشینی با معرفی مفهوم ماشین تورینگ، این ایده را مطرح کرد که

۱. آزمایش «مغز در خمیره» و «پروتز مغزی» در تأیید این دیدگاه به ذهن انجام گرفت (See: Russell & Norvig, 2003, pp. 956-957).

یک حالت ذهنی واحد می‌تواند در سیستم‌های فیزیکی مختلف با ساختارهای متفاوت تحقق یابد. این مفهوم که به قابلیت تحقق چندگانه معروف است، با تحقق چندگانه ذهن سازگار است. از آنجاکه «هر ماشین تورینگ که بتواند به صورت فیزیکی محقق شود، می‌تواند به طرق کاملاً متفاوتی محقق شود» (Putnam, 1980, p. 418)، از نظر فیزیکی باید ممکن باشد که حالات ذهنی نیز چندگانه محقق شوند. کارکردگرایی با انتزاع کردن حالات ذهنی از جزئیات فیزیولوژیکی مغز، امکان مطالعه مکانیسم‌های شناختی و رفتاری را در سطحی انتزاعی فراهم می‌کند. این رویکرد، مطالعه سیستم‌های هوشمند مصنوعی را نیز تسهیل می‌کند و به محققان اجازه می‌دهد تا با شبیه‌سازی سیستم‌های کارکردی انتزاعی، مکانیسم‌های تصمیم‌گیری و اراده را در این سیستم‌ها بررسی کنند (چرچلند، ۱۳۸۶، ص ۷۰).

بر اساس این تعریف، اراده به مثابه یک حالات ذهنی بر اساس ورودی‌های محیطی، خروجی‌های رفتاری و روابطش با دیگر حالات ذهنی توصیف می‌شود؛ یعنی اراده به مثابه یک ویژگی مستقل عارض بر جوهر ذهن - مثل نظریات دوگانه‌انگاران - در نظر گرفته نمی‌شود، بلکه تنها به مثابه یک عملکرد معین در ساختار ذهنی و رفتاری عامل عمل می‌کند. نکته مهم دیگر آن است که این نوع از جدول ماشین، نوعی تعیین‌گرایی را در عملکرد ماشین توصیف می‌کند که در این صورت، دراصل چنین مدلی در مقابل با آزادی تصمیم در ماشین و مفهوم اراده آزاد در یک چارچوب محدود برای ماشین است؛ از همین رو، بیشتر قائلان به کارکردگرایی ماشینی مدل مناسب ذهن را مدل «ماشین احتمال‌گرایانه» می‌دانند؛ مدلی که در آن، برنامه به ازای هر حالت و هر مجموعه‌ای از ورودی‌ها، احتمالی را مشخص می‌کند که مطابق با آن، وارد حالت بعدی می‌شود و خروجی خاصی را تولید می‌کند (گراهام و لوین، ۱۳۹۴، ص ۸۴-۸۵). این دیدگاه می‌تواند مبنای مدل‌های ساده‌تری برای تبیین اراده آزاد باشد. بر اساس این مدل، یک عامل می‌تواند از میان احتمالات مختلف، گزینه‌ای را انتخاب کند و عملی را که دارای اولویت یا رجحان بیشتری است، انجام دهد؛ از این رو کارکردگرایی پیش‌بینی می‌کند که می‌توان اراده را در سیستم‌های مصنوعی نیز شبیه‌سازی کرد؛ زیرا اراده یک حالت





ذهنی است و از ایده‌آلا مستثنی نیست. این به معنای آن است که ممکن است بتوان ماشین‌هایی ساخت که دارای اراده و تصمیم‌گیری مستقل باشند.

۴. برخی چالش‌های فلسفی رویکرد محاسباتی به اراده انسانی

اگرچه رویکرد محاسباتی و کارکردگرایانه به اراده در هوش مصنوعی توانسته است نمود بیرونی این ویژگی ذهنی انسان را با تکیه بر بازنمایی حالات ذهنی توجیه و بازتولید کند، اما به دلیل اتکای بنیادین بر فیزیکیالیسم تحویل‌گرا از یک سو، و اتخاذ روش شناختی سوم‌شخص از سوی دیگر، با چالش‌های فلسفی عمیقی روبه‌رو شده است. در رویکرد ماشینی پاتم، کارکردگرایان استدلال می‌کنند که - برخلاف فیزیکیالیسم نوعی - حالات ذهنی را نمی‌توان تنها به ویژگی‌های فیزیکی خاصی تقلیل داد؛ اما آنها را تنها باید بر اساس روابط کارکردی‌شان درک کرد (Block, 2006, pp. 35-37). بنابراین اراده چیزی است که دارای یک نقش علی خاص است، بدون اینکه به‌طور مستقیم، به یک وضعیت فیزیکی خاص وابسته باشد؛ از این رو می‌تواند مستقل از بستر زیستی‌اش و از طریق بازنمایی‌های نمادین یا شبکه‌های محاسباتی در سامانه‌های غیرزیستی مانند هوش مصنوعی نیز تحقق یابد. بنابراین در مجموع کارکردگرایی ماشینی را می‌توان نظریه‌ای فیزیکیالیستی از نوع تحویل‌گرای کارکردی نامید. در این دیدگاه، اراده فاقد هرگونه حیثیت درونی، آگاهانه و پدیداری است. افزون بر آن، رویکرد محاسباتی به سبب استفاده از روش مطالعه سوم‌شخص، توان درک و تبیین «تجربه اول‌شخص از اراده» را ندارد؛ امری که یکی از ارکان اساسی در فهم اراده انسانی به شمار می‌رود. این دو عامل (یعنی فروکاست‌گرایی کارکردی و حذف سابیجکتیویته) سبب می‌شوند که این رویکرد، با همه توان مدل‌سازی و شبیه‌سازی خود، در برابر ابعاد وجودشناختی و معرفتی اراده، ناتوان جلوه کند. در ادامه، این کاستی‌ها را بر اساس چهار اصل بنیادین فلسفه ملاصدرا یعنی اصالت وجود، تشکیک در وجود، حدود جسمانی نفس و حرکت جوهری و نیز آموزه‌های او درباره اراده انسانی بررسی می‌کنیم.

۴-۱. جایگاه جوهری نفس در تحلیل اراده انسانی

دیدگاه کارکردگرایی در تحلیل ذهن، برخلاف حکمت متعالیه بر نقش و عملکرد حالات ذهنی تمرکز دارد و ذات ذهن را مورد بررسی قرار نمی‌دهد. کارکردگرایی، ذهن را نه یک ذات یا جوهر مستقل، بلکه یک ویژگی کارکردی می‌داند که می‌تواند در سیستم‌های مادی و حتی غیرمادی تحقق یابد. این دیدگاه در اصل فرض ذاتی بودن ذهن را نادرست می‌شمارد (Westphal, 2016, pp. 69-70)؛ به همین دلیل، کارکردگرایان معتقدند که حتی اگر جوهری غیرمادی نیز وجود داشته باشد، آنچه اهمیت دارد، نقش‌های کارکردی آن است، نه حقیقت وجودی آن. در مقابل، در نگرش صدرایی در اصل تحقق ویژگی‌هایی همچون اراده بدون در نظر گرفتن جوهری به نام نفس، ناممکن است. دلیل این مطلب آن است که اراده انسانی که به معنای عزم، اجماع و تصمیم است و به مثابه کیف نفسانی در نظر گرفته شده است (صدرالمتألهین، ۱۳۶۸، ج ۶، ص ۳۳۸؛ صدرالمتألهین، ۱۳۶۶، ج ۳، ص ۲۲۶)، ماهیتی عرضی و قائم به غیر دارد که وجود لافسه آن عین وجود لغیره‌اش است و به همین دلیل، تحقق اراده بدون تحقق جوهر نفس ممکن نیست (صدرالمتألهین، ۱۳۶۶، ج ۴، ص ۱۰۱). نفس که در ابتدای پیدایش جوهری جسمانی است، در خلال حرکت جوهری، پیوسته سیر تکاملی خود را از طریق وصول به مراتب برتر و شدیدتر وجود ادامه می‌دهد و با حفظ وحدت شخصی، در نهایت، به مرتبه تجرد راه می‌یابد و موجودی مجرد می‌شود (مصباح یزدی، ۱۳۷۵، ج ۱، صص ۶۰-۶۱). در فرایند حرکت اشتدادی، اراده و دیگر کیفیات نفسانی به سبب آنکه وجودشان عین وجود نفس و نه مستقل از آن است، تابع مراتب وجودی نفس هستند و به تدریج در سیر تکاملی آن پدیدار می‌شوند؛ از این رو این کیفیات تنها به مثابه کارکردهایی از یک سیستم فیزیکی یا غیرمادی قابل تبیین نیستند، بلکه نحوه‌ای از تجلی وجودی نفس در مراتب مختلف کمال آن به شمار می‌آیند و سخن گفتن از این کیفیات بدون لحاظ جایگاه جوهری نفس، صورت‌بندی نادرستی از آنها خواهد بود. بر این اساس، اراده تنها یک حالت ذهنی یا نرم‌افزار نیست که در هر بستر مادی یا غیرمادی قابل تحقق باشد، بلکه ویژگی‌ای نفسانی است که در انسان، تابع وجود او است و در متن حرکت





اشتدادی نفس تحقق می‌یابد. حالات نفس نه تنها عرضی و کارکردی محض نیستند، بلکه شأنی از شئون نفس و حکایت از نحوه وجودی نفس در مراتب مختلف آن دارند. بنابراین برای فهم حالات ذهنی نمی‌توان از تحلیل جوهر نفس و کیفیت‌های وجودی آن غفلت کرد، بلکه باید این حالات را بر پایه شناخت نفس و نحوه وجود آن در هر مرتبه بررسی نمود.

۴-۲. منشأ حدوث اراده

در هر دو دیدگاه کارکردگرایی و صدرایی، اراده امری حادث است؛ اما در منشأ حدوث آن اختلاف است. در فلسفه صدرایی، حدوث اراده در انسان یک فرایند تدریجی، دارای ابعاد شناختی و انگیزشی و چندمرحله‌ای است که درون نفس رخ می‌دهد. تصور، تصدیق به فایده، شوق و شوق اکید مراحل پیشین تحقق اراده هستند که پس از پشت سر نهادن آنها، اراده‌ای که به آن «اجماع» گفته می‌شود، به دست می‌آید. این اراده در انسان، برخاسته از قوه‌ای برتر از قوه شوقیه حیوانی است - که خود به دو شاخه شهوت و غضب تقسیم می‌شود - و آن قوه برتر عقل عملی است (صدرالمتألهین، ۱۳۶۸، ج ۶، صص ۳۵۴-۳۵۶). هریک از مراحل این فرایند به وسیله یکی از قوای نفس انجام می‌شود؛ اما بر اساس قاعده «النفس فی وحدتها کل القوی»، نفس به مثابه یک حقیقت واحد، همه این مراتب را در یک پیوستار وجودی در بر دارد (صدرالمتألهین، ۱۳۶۸، ج ۸، ص ۲۲۱). در مقابل، کارکردگرایی بر آن است که فرایند اراده و تصمیم از تجمع تدریجی داده‌ها و پردازش آنها تا رسیدن به آستانه تصمیم ناشی می‌شود؛ از این رو تفاوت اصلی میان این دو نگرش در این است که یکی، اراده را حقیقتی برخاسته از جوهر نفس و در بستر حرکت جوهری نفس و تعامل قوای آن با یکدیگر تبیین می‌کند، و دیگری، آن را حاصل عملیات پردازشی سیستم عصبی می‌داند. بنابراین منشأ حدوث اراده تنها تغییراتی در سطح اطلاعات و پردازش نمادها نیست، بلکه از لوازم نفس است و نمی‌توان منشائی خارج از نفس برای آن متصور بود.

۳-۴. تبیین ناپذیری خاصیت پدیداری اراده

ملاصدرا در تبیین اراده، به خاصیت پدیداری آن توجه دارد و آن را مانند لذت و درد حقیقتی وجدانی می‌داند که حضور آن در نفس، همچون دیگر وجدانیات، از راه تجربه بی‌واسطه درک می‌شود (صدرالمآلهین، ۱۳۶۸، ج ۸، ص ۳۳۶). این امر بدان معناست که اراده در لحظه تحقق خود، به‌طور مستقیم بر آگاهی فرد ظهور می‌یابد و فاعل نسبت به آن، تجربه آگاهانه دارد، و تنها یک کارکرد نیست، بلکه امری وجودی است که ریشه در نحوه تحقق و تشخیص نفس به‌مثابه فاعل شناسا دارد؛ درحالی که در رویکرد کارکردگرایی، به‌دلیل رویکرد سوم‌شخصی به حالات ذهنی، این خاصیت اراده لحاظ نشده و تنها به‌مثابه یک سازوکار محاسباتی برای تصمیم‌گیری در نظر گرفته شده است (مسلمین، ۱۴۰۰، ص ۹۳)؛ یعنی یک سیستم مصنوعی ممکن است بر اساس ورودی‌ها و قواعد نوشته‌شده، یک تصمیم یا انتخاب داشته باشد؛ اما این تصمیم‌گیری فاقد حس و تجربه درونی برای سیستم است؛ زیرا دراصل این رویکرد فاقد جنبه‌های شناختی و درونی است (خاتمی، ۱۴۰۰، ص ۴۲۹). برای درک بهتر این مطلب، فودور این فرض را مطرح می‌کند که می‌توان به‌راحتی دو سامانه کارکردی، محاسباتی یا ماشین تورینگ را تصور کرد که در وضعیت یکسانی قرار دارند؛ اما ورودی‌های آنها دارای طیف‌های رنگی معکوس هستند و بر اساس این ورودی‌ها، محاسبه انجام می‌دهند؛ از این رو تجربه کیفی طیف رنگی نمی‌تواند تنها به وضعیت کارکردی یا محاسباتی فروکاسته شود؛ زیرا یک ماشین تورینگ ممکن است درحال محاسبه باشد بی‌آنکه کوچک‌ترین درکی از آنچه درحال محاسبه آن است، داشته باشد (Westphal, 2016, pp. 70-71). آزمایش «ذهن چینی» ند بلاک نیز بیانگر این مطلب است (Block, n.d., p. 7).^۱ بر همین اساس، دریفوس نیز استدلال

۱. فرض کنید که کل جمعیت چین با استفاده از رادیوهای دوسویه و یک سیستم ارتباطی پیچیده، به‌گونه‌ای سازمان‌دهی شوند که بتوانند رفتار یک مغز انسانی را شبیه‌سازی کنند و هر فرد در این سیستم، نقش یک نورون را ایفا کند و مطابق یک جدول ماشین تورینگ، ورودی‌ها را دریافت و خروجی‌های مناسب را ارسال نماید. این سیستم به یک بدن مصنوعی متصل است که از طریق سیگنال‌های رادیویی با آن تعامل دارد. در نتیجه از دیدگاه کارکردگرایانه، این مجموعه باید یک ذهن واقعی داشته باشد؛ زیرا عملکرد آن از نظر ورودی‌ها،



می‌کند که کامپیوترها در حقیقت نمی‌توانند تجربه ذهنی و زیستی انسان را بازآفرینی کنند (خاتمی، ۱۳۸۰، ص ۱۵۰).

۴-۴. تبیین ناپذیری حیثیت التفاتی اراده

در حکمت متعالیه، التفاتی بودن اراده انسانی در چارچوب نظام غایت‌گرایانه تبیین شده است. بر اساس اصل حرکت جوهری، موجودات اعم از طبیعی و غیرطبیعی (ارادی) به‌طور وجودی، به‌سوی غایات خاص خود متوجه‌اند (صدرالمتألهین، ۱۳۶۸، ج ۲، ص ۲۵۶). این جهت‌گیری ذاتی به‌سوی غایت، ناظر به‌نوعی شعور و آگاهی در فاعل است که در موجودات طبیعی به‌صورت ناآگاهانه و در موجودات دارای نفس، به‌ویژه انسان، به‌صورت آگاهانه تحقق می‌یابد. از آنجا که انسان در قوس صعودی حرکت جوهری خود به‌سوی فعلیت‌های بالاتر سوق داده می‌شود، اراده او نیز همواره با شعور و التفات همراه است. این اراده نمی‌تواند تنها نتیجه یک فرایند مکانیکی و محاسباتی فاقد محتوا باشد. به تعبیر دیگر، اراده امری نیست که بدون متعلق در نفس تحقق یابد، بلکه همواره دارای متعلق است (صدرالمتألهین، ۱۴۲۲ق، ص ۲۴۴) و نفس هنگام اراده، هم حالت درونی اراده را می‌یابد و هم به متعلق آن علم دارد. این آگاهی امری حضوری است، نه تنها یک بازنمایی ذهنی از علل و نتایج. در مقابل، دیدگاهی که نه تنها اراده، بلکه دراصل حالات ذهنی را به سازوکارهای محاسباتی تقلیل می‌دهد، نمی‌تواند این جنبه شناختی را تبیین کند؛ زیرا در کارکردگرایی، این حالات به‌صورت مکانیکی تفسیر می‌شود و در نتیجه، اراده به یک خروجی قطعی از پردازش‌های پیشین فروکاسته می‌شود که فاقد محتوا و معناست (مسلمین، ۱۴۰۰، ص ۸۷)؛ از این‌رو یکی دیگر از چالش‌های اساسی کارکردگرایی این است که نمی‌تواند حیث التفاتی اراده را توضیح دهد؛ زیرا در این



۵۴

نظر
سال سی‌ام، شماره ۴، ۱۴۰۴

خروجی‌ها و روابط میان اجزاء، معادل یک مغز انسانی است. اشکالی که از این استدلال متوجه کارکردگرایی است آن است که این سیستم با وجود تحقق ساختاری عملکردی مغز، هیچ تجربه آگاهی یا حالات ذهنی کیفی ندارد؛ یعنی چیزی نیست که «بودن آن» مانند «بودن یک انسان» احساس شود. این مسئله را بلاک (استدلال کیفیات غایب) (Absent Qualia Argument) می‌نامد (Block, n.d., p. 7).

چارچوب، حالات ذهنی از حیث کارکردی تعریف می‌شوند و هیچ سطحی از خودآگاهی یا التفات ذاتی در این سیستم‌ها فرض نمی‌شود. بنابراین اگر حیث التفاتی را به‌مثابه شرط ضروری اراده در نظر بگیریم، تبیین آن در چارچوب کارکردگرایی نه‌تنها ناتمام، بلکه دراصل ناممکن خواهد بود. کیت مسلین شکست کارکردگرایی در تبیین ویژگی پدیداری - که در مورد قبل بیان شد - و حیث التفاتی را ناشی از تبیین صرفاً فیزیکی‌الستی و مطالعهٔ سوم‌شخصی واقعیات ذهنی می‌داند. او در نقد روش کارکردگرایان می‌گوید: «مادام که معرفت آنها محدود به واقعیت‌های عینی، سوم‌شخص و فیزیکی باشد، پدیده‌هایی واقعی وجود دارند که خارج از قلمرو معرفت آنها قرار می‌گیرند؛ پدیده‌های واقعی که پدیده‌هایی ساجکتیو، اول‌شخص و آگاهانه‌اند. آن تجربه را نمی‌توان با متناظرهای کارکردی عصبی سوم‌شخص و عینی آن یکی گرفت» (مسلین، ۱۴۰۰، ص ۹۳).

نتیجه‌گیری

کارکردگرایی که به‌طور عمده، بر پایهٔ تحلیل رفتار و نقش‌های علی در سیستم‌های شناختی استوار است، اراده را به‌مثابهٔ پدیده‌ای رفتاری و برآمده از تعامل ورودی‌ها و سازوکارهای محاسباتی در نظر می‌گیرد و زمینهٔ نظری لازم را برای بازنمایی صورتی اراده در سامانه‌های مصنوعی فراهم می‌سازد؛ ازاین‌رو پیوندی منطقی و فلسفی میان ساختارهای مهندسی، الگوهای شناختی نمادین و مبانی فلسفی اراده برقرار می‌شود؛ اما چنین نگاهی، در قیاس با تحلیل‌های وجودی و جوهری در حکمت صدرایی، فهمی تقلیل‌گرایانه از اراده ارائه می‌دهد؛ زیرا حقیقت اراده نه‌تنها یک واکنش علی یا پدیده‌ای رفتاری، بلکه جلوه‌ای از سیر وجودی نفس است که در بستر حرکت جوهری و با اتکا بر آگاهی، شعور و فاعلیت وجودی پدیدار می‌شود. با در نظر گرفتن تفکیک سه‌گانهٔ ذهن به سطوح محاسبه‌گری، ادراک و احساس، و اراده و عاطفه می‌توان نتیجه گرفت که رویکرد کارکردگرایی تنها توانایی تبیین سطح اول، یعنی محاسبه‌گری را دارد. بر این اساس، نسبت‌دادن حقیقی اراده به عامل‌های هوشمندی که تنها رفتارهایی



شبيه به رفتارهای ارادی انسان بروز می دهند، غیردقیق و معرفتی ناتمام خواهد بود؛ زیرا اراده، افزون بر وجه کارکردی، مستلزم مؤلفه های بنیادینی چون آگاهانه بودن، تجربه پدیداری، حیثیت التفاتی و نوعی خودآگاهی درونی فاعل نسبت به آن است که در ساختارهای فعلی هوش مصنوعی و الگوهای محاسباتی آن به چشم نمی خورد.



پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



۵۶

نظر
سال سیام، شماره ۴، ۱۴۰۴

فهرست منابع

- برمودز، خوزه لویس. (۱۳۹۶). علوم شناختی: مقدمه‌ای بر علم ذهن (مترجم: تورج بی‌رستم). تهران: کیان افراز.
- چرچلند، پاول. (۱۳۸۶). ماده و آگاهی؛ درآمدی به فلسفه ذهن امروز (مترجم: امیر غلامی). تهران: نشر مرکز.
- خاتمی، محمود. (۱۳۸۰). مدل کامپیوتری ذهن و هوش مصنوعی. فلسفه (دانشگاه تهران)، ۳(۲)، صص ۱۴۵-۱۵۵.
- خاتمی، محمود. (۱۴۰۰). فلسفه ذهن. تهران: انتشارات تمدن علمی.
- خاتمی، محمود. (۱۴۰۱). فلسفه هوش مصنوعی. تهران: نشر علم.
- صدرالمآلهین شیرازی، محمد. (۱۳۶۶). شرح اصول کافی (ج ۳، ۴). تهران: وزارت فرهنگ و آموزش عالی، مؤسسه مطالعات و تحقیقات فرهنگی.
- صدرالمآلهین شیرازی، محمد. (۱۳۶۸). الحکمه المتعالیه فی الأسفار العقلیه الأربعة (ج ۲، ۶ و ۸). قم: مکتبه المصطفی.
- صدرالمآلهین شیرازی، محمد. (۱۴۲۲ق). شرح الهدایه الاثیریة. بیروت: مؤسسه التاریخ العربی.
- کیسینجر، هنری؛ اشمیت، اریک؛ و هوتنلوچر، دانیل. (۱۴۰۱). عصر هوش مصنوعی و آینده ما انسان‌ها (مترجم: هادی صفرزائی). تهران: کتاب پارسه.
- کین، ام. ج. (۱۳۹۶). فلسفه علم شناختی (مترجم: مصطفی تقوی). تهران: ترجمان علوم انسانی.
- گراهام، جورج؛ لوین، جنت. (۱۳۹۴). رفتارگرایی و کارکردگرایی (مترجم: یاسر پوراسماعیل). تهران: ققنوس.
- مسلین، کیت. (۱۴۰۰). فلسفه ذهن (مترجم: مهدی ذاکری). تهران: انتشارات علمی و فرهنگی.
- مصباح یزدی، محمدتقی. (۱۳۷۵). شرح جلد هشتم اسفار اربعه (ج ۱). قم: انتشارات مؤسسه آموزشی و پژوهشی امام خمینی علیه السلام.



- Block, N. (1978). Troubles with functionalism. *Minnesota Studies in the Philosophy of Science*, 9(1), pp. 261–325. <https://philpapers.org/rec/BLOTWF>.
- Block, N. (1994). Functionalism. (S. D. Guttenplan, Ed.). *A companion to the philosophy of mind* (pp. 171–182). Blackwell.
- Block, N. (n.d.). *Troubles with functionalism*. Retrieved October 25, 2024, from <http://w3.uniroma1.it/cordeschi/Articoli/block.htm>.
- Block, N. (2006). *What is functionalism? Readings in philosophy of psychology* (Vol. 1, pp. 27-44). MIT Press.
- Doyle, B. (2011). *Free Will: The Scandal in Philosophy*. Cambridge, MA: I-Phi Press.
- Westphal, J. (2016). *The mind-body problem*. The MIT Press.
- McCarthy, J., & Hayes, P. J. (1969). *Some philosophical problems from the standpoint of artificial intelligence*. Stanford University, Computer Science Department. <http://www-formal.stanford.edu/jmc/>
- McCarthy, J. (2000). *Free will even for robots* [Technical Report]. Stanford University, Computer Science Department. <http://www-formal.stanford.edu/jmc/freewill.pdf>
- McCarthy, J. (2005). *Simple deterministic free will*. Retrieved January 11, 2024, from <http://www-formal.stanford.edu/jmc/freewill>.
- McCarthy, J. (2006). *The philosophy of AI and the AI of philosophy*. Retrieved January 10, 2024, from <http://jmc.stanford.edu/articles/aiphil2.html>.
- Newsome, W. T. (2009). Human freedom “Emergence” (N. Murphy, G. F. R. Ellis, & T. O’Connor, Eds.). *Downward causation and the neurobiology of free will* (pp. 53–62). Springer-Verlag. https://doi.org/10.1007/978-3-642-03205-9_3
- Putnam, H. (1980). Philosophy and our mental life (N. Block, Ed.). *Readings in Philosophy of Psychology* (Vol. 1, pp. 134-143). Cambridge, MA: Harvard University Press.
- Putnam, H. (1988). *Representation and reality*. Cambridge, MA: MIT Press



Russell, S. J, & Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*, Upper Saddle River. New Jersey: Prentice Hall.

Simon, H. A., & Lea, G. (1973). *Problem solving and rule induction: A unified view*. Pittsburgh, PA: Carnegie-Mellon University.

Simons, G. (1985). *The biology of computer life*. Brighton: The Harvester Press.

Shagrir, O. (2010). *The rise and fall of computational functionalism* (Y. Ben-Menahem, Ed.). *Hilary Putnam* (pp. 220-250). Cambridge: Cambridge University Press. Doi: <https://doi.org/10.1017/CBO9780511614187.009>.



پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی



نظر

بررسی کارکردگرایی به مثابه مبنای مدل‌های محاسباتی اراده آزاد در هوش مصنوعی ...