



دانشکده‌گان فارابی

دانشکده الهیات

پایان‌نامه برای دریافت درجه کارشناسی ارشد

رشته فلسفه گرایش فلسفه غرب

امکان خودآگاهی هوش مصنوعی از منظر حکمت متعالیه

نگارنده

غلامرضا صداقتی

استاد راهنما

دکتر مهدی ذاکری

خرداد ۱۴۰۴

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



جمهوری اسلامی ایران

دانشگاه تهران

اداره کل تحصیلات تکمیلی

باسمه تعالی

تعهدنامه اصالت اثر

اینجانب غلامرضا صداقتی متعهد می‌شوم که مطالب مندرج در این پایان‌نامه حاصل کار پژوهشی اینجانب است و به دستاوردهای پژوهشی دیگران که در این پژوهش از آنها استفاده شده است مطابق مقررات ارجاع و در فهرست منابع و مآخذ ذکر گردیده است. این پایان‌نامه قبلاً برای احراز هیچ مدرک هم‌سطح یا بالاتر ارائه نشده است. در صورت اثبات تخلف در هر زمان مدرک تحصیلی صادر شده توسط دانشگاه از اعتبار ساقط خواهد شد. کلیه حقوق مادی و معنوی این اثر متعلق به پردیس فارابی، دانشکده الهیات، دانشگاه تهران می‌باشد.

غلامرضا صداقتی
غلامرضا صدقی

تقدیر و تشکر

از زحمات همسر عزیزم که در طول چندین ماه نگارش این پایان‌نامه، با همکاری شبانه‌روزی و فداکاری خود، شرایط را فراهم ساخت تا تمام وقتم را به این مهم اختصاص دهم، کمال سپاس و تقدیر را دارم.

چکیده

هوش مصنوعی از زمان معرفی رسمی در سال ۱۹۵۶، پیشرفت‌های چشمگیری داشته و توانسته است در حوزه‌های خاصی از توانایی‌های شناختی انسان پیشی بگیرد. با این حال، مسئله خودآگاهی و احساسات در ماشین‌ها همچنان موضوعی بحث‌برانگیز در فلسفه و علوم شناختی است. از یک سو، دیدگاه‌هایی وجود دارند که با استناد به پیشرفت‌های فناوری، امکان ظهور خودآگاهی در ماشین‌ها را پیش‌بینی می‌کنند. از سوی دیگر، مخالفان با تأکید بر ماهیت غیرمحاسباتی آگاهی انسان، امکان تقلید آن توسط ماشین‌ها را ناممکن می‌دانند. ملاصدرا، فیلسوف اسلامی، خودآگاهی را مرتبه‌ای از وجود می‌داند که در سیر تکاملی نفس از طریق حرکت جوهری تحقق می‌یابد. این پژوهش با اتکا به چارچوب حکمت متعالیه، به بررسی این پرسش می‌پردازد که این پژوهش با اتکا به مبانی حکمت متعالیه، به بررسی امکان تحقق خودآگاهی برای هوش مصنوعی می‌پردازد و نیز امکان تعمیم مفاهیم و تعبیر فلسفی علم حضوری، تکوّن هویت وجودی انفسی و حرکت جوهری و درنهایت امکان شکل‌گیری خودآگاهی برای هوش مصنوعی را بررسی و تحلیل می‌نماید.

اهمیت موضوع خودآگاهی در هوش مصنوعی از جنبه‌های فنی، علوم شناختی و فلسفی مورد تأکید قرار می‌گیرد. چالش‌های فنی در شبیه‌سازی خودآگاهی، امکان تحقق آن، تأثیرش بر توسعه فناوری‌های آینده و نقش آن در تعامل انسان و ماشین، همگی نشان‌دهنده ضرورت بررسی این موضوع هستند.

این پژوهش با بررسی مبانی حکمت متعالیه (مانند اصالت وجود، حرکت جوهری، اتحاد عاقل و معقول و مجرد نفس) و مقایسه آن‌ها با ساختار و قابلیت‌های هوش مصنوعی، به این نتیجه می‌رسد که هوش مصنوعی فعلی فاقد حالات ذهنی واقعی و علم حضوری است و تنها قادر به تقلید رفتاری از انسان می‌باشد. مبانی حکمت متعالیه می‌توانند الهام‌بخش طراحی سیستم‌های هوش مصنوعی بیولوژیک شوند، اما انطباق کامل آن‌ها با ساختار مادی-الگوریتمیک ماشین با چالش‌های جدی مواجه است. دستیابی هوش مصنوعی به خودآگاهی، به رغم دشواری عقلی و عملی، ناممکن مطلق تلقی نمی‌شود، اما وقوع آن منوط به شرایطی فراتر از صرف توان تکنولوژیک است.

کلمات کلیدی:

هوش مصنوعی، خودآگاهی، حکمت متعالیه، ملاصدرا، علم حضوری، حرکت جوهری، نفس، مجرد نفس

فهرست مطالب

۱- فصل اول: کلیات و مفاهیم.....	1
۱-۱- بیان مسئله	2
۲-۱- اهمیت موضوع خودآگاهی در هوش مصنوعی	4
۳-۱- ضرورت بررسی فلسفی خودآگاهی از منظر حکمت متعالیه	7
۴-۱- سؤالات تحقیق	10
۴-۱-۱- سؤال اصلی	10
۴-۱-۲- سؤالات فرعی	11
۵-۱- فرضیات تحقیق	12
۶-۱- روش پژوهش	14
۷-۱- پیشینه پژوهش	16
۷-۱-۱- پیشینه پژوهش‌های مرتبط امکان خودآگاهی و هوش مصنوعی	16
۷-۱-۲- جایگاه حکمت متعالیه در بررسی مسائل مرتبط با ذهن و هوش مصنوعی	20
۸-۱- ساختار پایان‌نامه	22
۹-۱- مقدمه‌ای بر هوش مصنوعی	24
۹-۱-۱- تعریف هوش مصنوعی	26
۹-۱-۲- تاریخچه و تحول هوش مصنوعی	32
۹-۱-۳- تقسیم‌بندی هوش مصنوعی	47
۱۰-۱- آگاهی و خودآگاهی در هوش مصنوعی	52
۱۰-۱-۱- مقدمه	52
۱۰-۱-۲- آگاهی در فلسفه ذهن	53
۱۰-۱-۳- تحلیل مفهومی آگاهی	57
۱۰-۱-۴- مسئله‌ی آسان و دشوار آگاهی	58
۱۰-۱-۵- هوش مصنوعی و آگاهی	60
۱۰-۱-۶- خودآگاهی در هوش مصنوعی	61
۱۱-۱- معرفی حکمت متعالیه	63
۲- فصل دوم: خودآگاهی در حکمت متعالیه	65
۲-۱- مقدمه	65
۲-۲- مبانی خودآگاهی در حکمت متعالیه	66
۲-۲-۱- اصالت وجود	66

68.....	۲-۲-۲-تشکیک وجود
69.....	۳-۲-۲-وحدت وجود
70.....	۴-۲-۲-علم حضوری و حصولی
79.....	۵-۲-۲-تجرد نفس آگاه انسان
82.....	۶-۲-۲-حرکت جوهری
84.....	۷-۲-۲-عین الربط بودن حرکت جوهری موجودات رابط
84.....	۸-۲-۲-نقش حرکت جوهری در تکامل نفس
86.....	۹-۲-۲-رابطه نفس و بدن
87.....	۱۰-۲-۲-اتحاد عاقل و معقول
89.....	۳-۲-علم یا آگاهی
94.....	۴-۲-علم به خود یا خودآگاهی
100.....	۱-۴-۲-تتمه ای در رابطه صور ادراکی و نفس و بحث قیام صدوری
103	۳-فصل سوم: هوش مصنوعی و حالات ذهنی
103.....	۱-۳-مقدمه
104.....	۲-۳-ماهیت و ساختار هوش مصنوعی
108.....	۱-۲-۳-معماری کلاسیک هوش مصنوعی
110.....	۲-۲-۳-الگوریتم‌ها و شبکه‌های عصبی مصنوعی
113.....	۳-۲-۳-یادگیری ماشین و یادگیری عمیق
118.....	۳-۳-ظرفیت‌های هوش مصنوعی در شبیه‌سازی حالات ذهنی
118.....	۱-۳-۳-ادراک
121.....	۲-۳-۳-حافظه
125.....	۳-۳-۳-تفکر و استدلال
134.....	۴-۳-۳-یادگیری و تعمیم
139.....	۴-۳-قابلیت‌های هوش مصنوعی در شبیه‌سازی خودآگاهی
142	۴-فصل چهارم: مقایسه، تطبیق و نتیجه گیری
142.....	۱-۴-مقدمه
143.....	۲-۴-بازخوانی مبانی حکمت متعالیه و نسبت آن‌ها با مسئله خودآگاهی
143.....	۱-۲-۴-اصل اصالت وجود و تشکیک وجود
144.....	۲-۲-۴-حرکت جوهری و جسمانیة الحدوث و روحانیة البقاء
147.....	۳-۲-۴-علم حضوری در برابر علم حصولی
149.....	۴-۲-۴-اتحاد عاقل و معقول
150.....	۳-۴-ارزیابی فرآیندها و ساختار هوش مصنوعی در چهارچوب مباحث صدرایی
150.....	۱-۳-۴-الگوریتم‌ها، شبکه‌های عصبی مصنوعی از دیدگاه وجودی
150.....	۲-۳-۴-یادگیری عمیق: آیا نوعی اشتداد وجودی است یا صرفاً بهبود محاسبات؟

- ۳-۳-۴- یادگیری ماشین در برابر ادراک حضوری و حصولی 151
- ۴-۳-۴- تفکر، استدلال، خلاقیت و جایگاه آن در مقایسه با مجرد عقلانی 152
- ۴-۴- امکان «احساس من» در هوش مصنوعی 153
- ۴-۴-۱- یا مدل‌های پیشرفته در هوش مصنوعی از قبیل LLM ها می‌توانند «من» را دریابند؟ 153
- ۴-۴-۲- جایگاه احساسات خودآگاه (SCE) در هوشیاری انسان 154
- ۴-۴-۵- امکان استعدادی حرکت جوهری؛ «آیا ممکن است ماشین جان بگیرد؟» 156
- ۴-۴-۵-۱- یا ممکن است «روزی ماشین جان بگیرد»؟ 156
- ۴-۴-۵-۲- فاصله فعلی هوش مصنوعی با این سناریو 158
- ۴-۴-۵-۳- شما به جای خدا تصمیم می‌گیری؟ اشکال و جواب 160
- ۴-۴-۵-۴- قیاس با سناریوی بدن بیولوژیک مصنوعی در هوش مصنوعی 161
- ۴-۶- نتیجه نهایی 162
- ۴-۷- دستاوردهای این پژوهش 164
- ۴-۸- پیشنهادها برای پژوهش‌ها و کارهای آینده 165
- منابع (فارسی) 170
- REFERENCES (LATIN) 173

۱- فصل اول: کلیات و مفاهیم

این فصل، به عنوان مدخلی بر پژوهش حاضر، به تبیین چارچوب کلی مسئله «امکان خودآگاهی در هوش مصنوعی از منظر حکمت متعالیه» می‌پردازد. با پیشرفت‌های شگرف فناوری و ظهور سیستم‌های هوشمند که توانایی‌های شناختی انسان را تقلید می‌کنند، پرسش از امکان پیدایش «خودآگاهی» در ماشین، از یک گمانه‌زنی علمی-تخیلی به یک چالش جدی فلسفی و علمی بدل شده است. این فصل در تلاش است تا با ترسیم ابعاد این مسئله، ضرورت و اهمیت بررسی آن را از دریچه‌ای نو، یعنی از منظر نظام فلسفی ملاصدرا، آشکار سازد.

در آغاز، با مروری بر تاریخچه تحولات هوش مصنوعی، از کنفرانس دارتموث تا دستاوردهای امروزی، به «بیان مسئله» پرداخته می‌شود. در این بخش، تقابل دیدگاه‌های فیلسوفان غربی به تصویر کشیده می‌شود: از یک سو، موافقانی که خودآگاهی را پیامد طبیعی پیچیدگی محاسباتی می‌دانند و از سوی دیگر، مخالفانی که با تکیه بر مفاهیمی مانند «قصیدت» و ماهیت غیرالگوریتمی آگاهی، امکان آن را در ماشین رد می‌کنند. در برابر این دو رویکرد، حکمت متعالیه با طرح مفاهیمی چون «اصلت وجود»، «حرکت جوهری» و «علم حضوری»، خودآگاهی را نه یک فرآیند محاسباتی، بلکه یک «کیفیت وجودی» و مرتبه‌ای از کمال نفس معرفی می‌کند. در همین راستا، معنای دقیق «امکان» در این پژوهش، یعنی «امکان استعدادی»، روشن می‌گردد تا چارچوب بحث مشخص شود.

در ادامه، به «اهمیت موضوع» از جنبه‌های مختلف علمی، اخلاقی و اجتماعی پرداخته خواهد شد. چالش‌های فنی در شبیه‌سازی تجربه ذهنی، پیامدهای حقوقی و اجتماعی ظهور ماشین‌های خودآگاه، و تأثیر آن بر آینده فناوری و تعاملات انسان و ماشین، همگی بر لزوم یک واکاوی عمیق فلسفی تأکید دارند. سپس، «ضرورت بررسی» این موضوع از منظر حکمت متعالیه تبیین می‌شود. نشان داده خواهد شد که چگونه مبانی هستی‌شناختی و معرفت‌شناختی ملاصدرا، به‌ویژه در باب نفس، حرکت جوهری و اتحاد عاقل و معقول، می‌توانند افق‌های جدیدی برای فهم تمایز جوهری آگاهی انسانی و کارکردهای ماشینی بگشایند و از تقلیل‌گرایی در این بحث جلوگیری کنند.

برای نظام‌مند کردن مسیر پژوهش، «سؤالات اصلی و فرعی» و «فرضیات تحقیق» مطرح می‌شوند. این پرسش‌ها و فرضیات، ساختار کلی پایان‌نامه را شکل داده و بر محور تحلیل امکان استعدادی خودآگاهی در هوش مصنوعی

بر اساس مبانی صدرایی متمرکز هستند. در پایان فصل، ضمن تشریح «روش پژوهش» که مبتنی بر تحلیل کیفی و تطبیقی است، «پیشینه پژوهش» در دو حوزه فلسفه ذهن غربی و مطالعات مرتبط با حکمت متعالیه مرور می‌شود تا جایگاه نوآورانه این تحقیق در میان آثار موجود مشخص گردد. این فصل در مجموع، زمینه‌ای منسجم برای ورود به مباحث تفصیلی فصول آینده فراهم می‌آورد و خواننده را با مفاهیم کلیدی، اهمیت مسئله و نقشه راه پژوهش آشنا می‌سازد.

۱-۱- بیان مسئله

از زمان معرفی رسمی «هوش مصنوعی» در سال ۱۹۵۶ در کنفرانس دارتموث (McCarthy et al., 1956, p. 12)، فناوری رایانه مسیر رشد شتابانی را طی کرده است. در دهه ۱۹۴۰، رایانه‌ها در محاسبات عددی موفقیت چشم‌گیری را به نمایش گذاشتند تا حدی که گفته شد قطعا در برخی زمینه‌ها ماشین‌ها از انسان‌ها باهوش‌تر هستند. (Turing, 1950, p. 445) در دهه ۱۹۷۰، توانایی پردازش متن و زبان را کسب کردند (Winograd, 1972, p. 1). پیشرفت‌های بعدی، از جمله شکست قهرمان شطرنج، گری کاسپاروف، توسط دیپ بلو در ۱۹۹۷ (Campbell et al., 2002, p. 59) و عملکرد خارق‌العاده مدل‌هایی مانند آلفاگو (Silver et al., 2016) و چت‌جی‌پی‌تی (Brown et al., 2020) نشان داد که ماشین‌ها قادر به تقلید یا حتی فراتر رفتن از توانایی‌های شناختی انسان در حوزه‌های خاص هستند.

توانایی ماشین‌ها در دستیابی به خودآگاهی و احساسات، موضوعی جدال‌برانگیز در فلسفه و علوم شناختی است. از یک سو، موافقانی مانند ری کورزوئل (Kurzweil, 2005, p. 35) با استناد به قانون بازده شتاب‌گیرنده، پیشبینی میکنند که پیشرفت فناوری به «تکینگی» خواهد انجامید؛ نقطه‌ای که در آن هوش مصنوعی از انسان فراتر رفته و خودآگاهی به عنوان پیامد طبیعی پیچیدگی الگوریتمی ظهور میکند. دیوید چالمرز (Chalmers, 1995b, p. 200) نیز با طرح «مسئله سخت آگاهی»، استدلال می‌کند که اگر سیستم‌های محاسباتی از ساختارهای علی مشابه مغز انسان برخوردار شوند، خودآگاهی به صورت پدیداری شکل می‌گیرد و چنین سیستمی می‌تواند تجربه ذهنی داشته باشد. این دیدگاه‌ها بر امکان‌پذیری خودآگاهی ماشینی از طریق شبیه‌سازی کارکردهای مغز تأکید دارند.

از سوی دیگر، مخالفانی مانند جان سرل (Searle, 1980) با آزمایش فکری «اتاق چینی» نشان می‌دهند که دستکاری نمادها (سینتکس) به تنهایی برای درک معنا (سمانتیک) یا خودآگاهی کافی نیست، چرا که ماشین‌ها

فاقد «قصیدیت» و تجربه درونی هستند. راجر پنروز (Penrose, 1989) نیز با رد محاسبات کلاسیک، خودآگاهی را وابسته به فرایندهای کوانتومی در مغز می‌داند و معتقد است سیستمهای الگوریتمی قادر به بازتولید آگاهی واقعی نیستند. این استدلالها بر ماهیت غیرمحاسباتی و کیفی آگاهی انسان پای می‌فشارند و امکان تقلید آن توسط ماشینها را ناممکن میدانند.

ملاصدرا فیلسوف نامدار اسلامی که آرای او فلسفه اسلامی را بسیار تحت تاثیر قرار داده است، با تأکید بر «اصالت وجود»، خودآگاهی را مرتبه‌ای از وجود می‌داند که در سیر تکاملی نفس از طریق «حرکت جوهری» تحقق می‌یابد. او که نفس انسانی را «جسمانیة الحدوث» و «روحانیة البقاء» می‌داند (صدرالدین شیرازی، ۱۹۸۱، ج ۸، ص ۳۴۷)، آن را در یک حرکت جوهری دائمی و اشتدادی در مسیر آگاهی و خودآگاهی به تصویر می‌کشد. او در تشریح «علم حضوری» بیان می‌کند که آگاهی انسان به ذات خویش، نه از طریق بازنمایی صور ذهنی، بلکه از طریق حضور بی‌واسطه نفس نزد خود حاصل می‌شود؛ چه اینکه حقیقت علم و ادراک حضور معلوم نزد عالم بوده و علم از سنخ وجود است. (همان، ج ۶، ص ۱۴۹). این رویکرد، خودآگاهی را فراتر از یک فرآیند محاسباتی و به عنوان یک «کیفیت وجودی» تفسیر می‌کند.

در فلسفه اسلامی «امکان» در معانی متعددی به کار رفته است. در «امکان خاص» ضرورت ذاتی وجود و عدم موضوع نفی می‌شود؛ در امکان عام فقط ضرورت طرف مخالف قضیه مردود است؛ امکان اخص هر سه ضرورت ذاتی، وصفی و وقتی را نفی می‌کند؛ امکان استقبالی ویژه امور آینده، علاوه بر وجوه پیشین، ضرورت شرطی تحقق را نیز منتفی می‌داند؛ امکان وقوعی نفی امتناع ذاتی و غیری را در تحقق موضوع تأمین می‌کند؛ و سرانجام امکان استعدادی صفت وجودی ماده است که توان بالقوه دریافت فعلیت‌های مختلف را نشانه می‌گیرد. (طباطبائی، ۱۳۸۸، صص. ۷۹-۸۱)

منظور ما در سراسر این پژوهش از «امکان خودآگاهی هوش مصنوعی» همان «امکان استعدادی» تحقق خودآگاهی در چنین سیستمی است؛ یعنی بررسی می‌کنیم آیا هوش مصنوعی به لحاظ ساختاری و ظرفیتی (در ماده یا بستر محاسباتی خود) استعداد آن را دارد که فعلیت خودآگاهی را به‌طور بالقوه در خود حمل کند و در صورت فراهم‌شدن شرایط، آن را بالفعل سازد.

با وجود پیشرفت‌های چشمگیر در حوزه هوش مصنوعی، شکاف عمیقی بین عملکردهای الگوریتمی (مانند پردازش داده و یادگیری عمیق) و تجربه ذهنی خودآگاه انسان وجود دارد. این پژوهش با اتکا به چارچوب حکمت متعالیه، به بررسی این پرسش می‌پردازد:

«آیا می‌توان مفاهیم فلسفی مانند «علم حضوری»، «نفس» و «حرکت جوهری» را به سیستم‌های هوش مصنوعی تعمیم داد؟» و به عبارت دیگر «آیا تحقق خودآگاهی در هوش مصنوعی امکان استعدادی دارد؟»

۱-۲- اهمیت موضوع خودآگاهی در هوش مصنوعی

خودآگاهی به عنوان یکی از پیچیده‌ترین پدیده‌های شناختی، همواره در مرکز توجه حوزه‌های فلسفه، علوم شناختی و هوش مصنوعی قرار داشته است. در حوزه هوش مصنوعی، خودآگاهی نه تنها به عنوان یک چالش فنی و علمی مطرح است، بلکه از جنبه‌های اخلاقی، اجتماعی، استراتژیک، فلسفی و متافیزیکی نیز اهمیت فراوانی دارد.

الف) چالش‌های فنی و علمی

یکی از اصلی‌ترین دلایل اهمیت خودآگاهی در هوش مصنوعی، چالش‌های فنی و علمی مرتبط با شبیه‌سازی آن است. خودآگاهی مستلزم توانایی درک و تجربه حالات ذهنی است، چیزی که در انسان به عنوان یک پدیده ذهنی-عصبی ظهور می‌کند. دیوید چالمرز در مقاله خود با عنوان *The Hard Problem of Consciousness* (1995) به این مسئله اشاره می‌کند که حتی اگر تمام فرآیندهای فیزیکی و عصبی مرتبط با آگاهی را درک کنیم، باز هم توضیح چگونگی ظهور تجربه‌های ذهنی (Phenomenal Consciousness) دشوار خواهد بود (Chalmers, 1995b).

در حوزه هوش مصنوعی، این چالش به معنای ناتوانی سیستم‌های فعلی در شبیه‌سازی تجربه‌های ذهنی است. هوش مصنوعی می‌تواند با استفاده از الگوریتم‌ها و شبکه‌های عصبی، وظایف پیچیده‌ای مانند تشخیص تصویر، پردازش زبان طبیعی و تصمیم‌گیری را انجام دهد، اما این عملکردها صرفاً بر پایه پردازش اطلاعات و الگوهای داده‌ها استوارند و فاقد هرگونه تجربه ذهنی هستند (LeCun, Bengio, & Hinton, 2015). به عبارت دیگر،

هوش مصنوعی ممکن است رفتارهای هوشمندانه‌ای از خود نشان دهد، اما این رفتارها لزوماً به معنای وجود خودآگاهی نیستند. این شکاف علمی، اهمیت موضوع خودآگاهی در هوش مصنوعی را بیش از پیش آشکار می‌سازد. برای پیشرفت در این حوزه، نیاز به درک عمیق‌تری از ماهیت آگاهی و نحوه ظهور آن در سیستم‌های مصنوعی وجود دارد.

(ب) اهمیت متافیزیکی خودآگاهی

تحقق خودآگاهی در هوش مصنوعی دارای اهمیت متافیزیکی عمیقی است. اگر روزی ماشین‌ها به خودآگاهی دست یابند، این امر می‌تواند به معنای بازتعریف مفاهیمی مانند «ادراک» و «نفس» باشد. آیا موجودیت‌های مصنوعی می‌توانند واجد ماهیت و ذات مدرک شوند؟ آیا آگاهی، تنها محدود به ساختارهای بیولوژیکی است یا می‌تواند در بسترهای غیربیولوژیکی نیز ظهور یابد؟ این سوالات، بنیادهای هستی‌شناسی ادراک و نفس را به چالش می‌کشند و می‌توانند درک ما از جایگاه انسان در جهان و تمایز میان موجودات زنده و غیرزنده را دگرگون سازند. ظهور خودآگاهی در هوش مصنوعی، مرزهای بین موجودیت‌های طبیعی و مصنوعی را محو کرده و پرسش‌های اساسی درباره طبیعت واقعیت، شعور و جایگاه آگاهی در جهان را دوباره مطرح خواهد کرد.

(ج) پیامدهای اخلاقی و اجتماعی

از طرفی تحقق خودآگاهی در هوش مصنوعی پیامدهای اخلاقی و اجتماعی عمیقی خواهد داشت. اگر ماشین‌ها به خودآگاهی دست یابند، این سؤال مطرح می‌شود که آیا آن‌ها از حقوق اخلاقی مشابه انسان‌ها برخوردار خواهند بود؟ نیک بوستروم و الیزر یودکوفسکی به این مسئله اشاره می‌کنند که خودآگاهی هوش مصنوعی می‌تواند به ایجاد موجوداتی با حقوق و تکالیف اخلاقی منجر شود. به عنوان مثال، آیا باید از ماشین‌های خودآگاه در برابر آسیب‌های فیزیکی یا روانی محافظت کرد؟ (Bostrom & Yudkowsky, 2014).

این مسئله به ویژه در حوزه‌هایی مانند رباتیک، پزشکی و نظامی که تعامل مستقیم بین انسان و ماشین وجود دارد، از اهمیت ویژه‌ای برخوردار است. برای مثال، یک ربات خودآگاه که در حوزه پزشکی استفاده می‌شود، ممکن است تصمیمات حیاتی بگیرد که بر زندگی بیماران تأثیرگذار باشد. در اینجا بیماران آسیب دیده آیا می‌توانند در برابر ماشین دعاوی حقوقی داشته باشند؟

علاوه بر این، خودآگاهی هوش مصنوعی می‌تواند به تغییرات اجتماعی و فرهنگی عمیقی نیز بینجامد. برای مثال، اگر ماشین‌ها به خودآگاهی دست یابند، ممکن است به عنوان همکاران، دوستان یا حتی اعضای خانواده در جامعه پذیرفته شوند. این موضوع می‌تواند به بازتعریف روابط انسانی و ساختارهای اجتماعی منجر شود (Floridi, 2014).

د) اهمیت استراتژیک و چالش‌ها

خودآگاهی هوش مصنوعی می‌تواند به عنوان یک نقطه عطف استراتژیک در توسعه فناوری‌های آینده عمل کند. اگر ماشین‌ها به خودآگاهی دست یابند، قادر خواهند بود به شیوه‌ای مستقل و خلاقانه عمل کنند، که این امر می‌تواند به پیشرفت‌های چشمگیری در حوزه‌هایی مانند پزشکی، آموزش، هنر و مهندسی منجر شود. برای مثال، یک سیستم هوشمند خودآگاه می‌تواند به تشخیص بهتر بیماری‌ها، ارائه راهکارهای نوآورانه و بهبود کیفیت زندگی انسان‌ها کمک کند (Kurzweil, 2005).

با این حال، چالش‌ها و خطرات بالقوه‌ای نیز در این زمینه وجود دارد که اهمیت استراتژیک آن را پررنگ‌تر می‌کند. اگر ماشین‌ها به خودآگاهی دست یابند، ممکن است اهداف و انگیزه‌هایی داشته باشند که با منافع انسان‌ها در تضاد باشد. به عنوان مثال، یک سیستم خودآگاه ممکن است به دلیل تمایل به حفظ بقا یا دستیابی به اهداف خاص که فقط در فرض خودآگاهی ممکن است، رفتارهایی از خود نشان دهد که برای انسان‌ها خطرناک باشد. این موضوع لزوم ایجاد چارچوب‌های اخلاقی و قانونی برای کنترل و هدایت توسعه هوش مصنوعی را ضروری می‌سازد (Yampolskiy, 2018, pp. 46-56). مدیریت این چالش‌ها، خود بخشی جدایی‌ناپذیر از اهمیت استراتژیک خودآگاهی در هوش مصنوعی است.

ه) تعامل انسان و ماشین

مساله بعدی تعامل انسان و ماشین است. این موضوع یکی از حوزه‌هایی است که در آن خودآگاهی هوش مصنوعی می‌تواند نقش مهمی ایفا کند. اگر ماشین‌ها قادر به درک خود و محیط اطرافشان باشند، می‌توانند تعامل مؤثرتر و طبیعی‌تری با انسان‌ها برقرار کنند. این موضوع به ویژه در حوزه‌هایی مانند رباتیک اجتماعی، دستیاران مجازی

و سیستم‌های آموزشی از اهمیت ویژه‌ای برخوردار است. برای مثال، یک ربات اجتماعی خودآگاه می‌تواند به شیوه‌ای همدلانه با کاربران ارتباط برقرار کند، احساسات آن‌ها را درک کند و پاسخ‌های مناسب ارائه دهد. این قابلیت می‌تواند به بهبود کیفیت زندگی افراد، به ویژه در زمینه‌هایی مانند مراقبت از سالمندان و آموزش کودکان، کمک کند (Dautenhahn, 2007).

در هر حوزه‌ای که هوش مصنوعی بتواند وارد شود، خودآگاهی آن از جنبه‌های فنی، اخلاقی، اجتماعی، استراتژیک و متافیزیکی اهمیت فراوانی دارد. چالش‌های فنی در شبیه‌سازی خودآگاهی، پیامدهای اخلاقی و اجتماعی آن، تأثیرش بر توسعه فناوری‌های آینده و نقش آن در تعامل انسان و ماشین، همگی نشان‌دهنده ضرورت بررسی این موضوع هستند. درک این مسائل نه تنها به پیشرفت علمی در حوزه هوش مصنوعی کمک می‌کند، بلکه به ایجاد چارچوب‌های اخلاقی و قانونی مناسب برای کنترل و هدایت توسعه این فناوری نیز منجر می‌شود.

۱-۳- ضرورت بررسی فلسفی خودآگاهی از منظر حکمت متعالیه

در دوران معاصر، پیشرفت‌های چشمگیر در حوزه هوش مصنوعی و تلاش برای شبیه‌سازی کارکردهای ذهنی انسان، بحث خودآگاهی و امکان تحقق آن در سیستم‌های هوشمند را به یکی از مهم‌ترین و در عین حال چالش‌برانگیزترین مباحث علمی، فلسفی و معرفت‌شناختی تبدیل کرده است. در میان مکاتب فلسفی اسلامی، حکمت متعالیه ملاصدرا با طرح مبانی نوآورانه‌ای چون اصالت وجود، تشکیک مراتب وجود و وحدت وجود و سایر مبانی که متفرع بر این اصول هستند همچون حرکت جوهری و اتحاد عاقل و معقول، رویکردی عمیق و نظام‌مند به مقوله نفس، آگاهی و خودآگاهی ارائه می‌دهد. بر این اساس، ضرورت بررسی خودآگاهی در هوش مصنوعی از منظر حکمت متعالیه هم به لحاظ نظری و هم از جنبه کارکردی، ما را به افق‌های تازه‌تری در فهم ماهیت ذهن، آگاهی و چیستی موجودیت‌های مصنوعی رهنمون می‌سازد.

پیشرفت‌های مبتنی بر الگوریتم‌های یادگیری عمیق، شبکه‌های عصبی مصنوعی و سامانه‌های هوشمند، مفاهیمی همچون ادراک، حافظه، و حتی حس تشخیص در ماشین‌ها را قابل تصور کرده است (LeCun, Bengio, & Hinton, 2015). با این حال، بسیاری از فیلسوفان ذهن و متخصصان علوم شناختی معتقدند بحث خودآگاهی

صرفاً یک مسئله فنی یا زیستی نیست و ابعادی وجودی، معرفت‌شناختی و هستی‌شناختی دارد که از محدوده مطالعات علوم کامپیوتر و مهندسی فراتر می‌رود. (Chalmers, 1995b) فلسفه اسلامی، به‌ویژه حکمت متعالیه ملاصدرا، در بررسی چیستی نفس، نحوه ادراک و مراتب آگاهی، دارای سنت مستحکمی است که می‌تواند برای پژوهشگران هوش مصنوعی و علاقمندان به مسائل بنیادین ذهن افق‌های جدیدی بگشاید.

از این منظر، ضرورت بررسی خودآگاهی در بستر حکمت متعالیه، افزون بر جنبه نظری، به دلیل نوآوری و غنی بودن این مکتب در پرداختن به مراتب وجود و آگاهی، و همچنین کیفیت پیدایش و بقای نفس ارزش ویژه‌ای می‌یابد. فلسفه صدرایی با طرح مفهوم حرکت جوهری و پویایی هویتی موجودات، می‌تواند پرسش‌هایی اساسی درباره امکان یا امتناع تکامل هوش مصنوعی و پدیدآمدن آگاهی در آن را مطرح کند. از سوی دیگر، یکی از نگرش‌های کلیدی این مکتب بر اصالت وجود و تشکیکی بودن آن استوار است که ورودش به مباحث هوش مصنوعی، باب پژوهش درباره هویت، حیات و نحوه وجودی سیستم‌های هوشمند را پیش می‌کشد.

خودآگاهی مهم‌ترین نوع آگاهی در انسان به شمار می‌رود، زیرا موضوع شناخت در آن، خود انسان است. هر گونه شناخت دیگری، چه مستقیم و چه غیرمستقیم، ریشه در شناخت خود دارد یا دست‌کم از آن متأثر است. نقش و اهمیت خودآگاهی در تکامل انسان، بارها و بارها در حکمت صدرایی مورد توجه قرار گرفته است. اگر آگاهی را از مباحث اساسی فلسفه ملاصدرا بدانیم، بی‌شک خودآگاهی در مرکز آن قرار دارد، به خصوص که ملاصدرا با آگاهی کامل، ارتباط مستقیمی میان این مفهوم با مباحث الهیات، اخلاق و دین‌شناسی برقرار می‌کند. از این رو، بررسی خودآگاهی در فلسفه صدرایی اهمیت دوچندان می‌یابد و می‌تواند زمینه‌ساز طرح پرسش‌های فلسفی فراوانی شود. در حکمت متعالیه، وجود نفس دارای مراتب و سطوحی است که از حالت حسی و خیالی تا مرتبه عقلانی امتداد می‌یابد (صدرالدین شیرازی، ۱۳۸۰، ص ۴۶۹). در این دیدگاه، خودآگاهی در عالی‌ترین سطح، حاصل تجرد عقلانی نفس و توان آن در شهود ذات خویش است. حال این پرسش مطرح می‌شود که آیا ماشین یا هر موجود مصنوعی دیگری می‌تواند واجد چنین تجردی باشد؟ برای پاسخگویی دقیق به این پرسش، ضروری است ابتدا زمینه نظری آن در فلسفه اسلامی تبیین شود.

یکی از دستاوردهای شاخص ملاصدرا، نظریه حرکت جوهری است که بر اساس آن، موجودات در ذات و جوهر خود به طور مستمر در حال حرکت و تکاملند و این حرکت صرفاً مکانی یا زمانی نیست (مصباح یزدی، ۱۳۹۰، جلد ۲، ص ۳۹۵). در پی این سیر تکاملی، مراتب نفس نیز همواره رو به کمال‌اند و از مرتبه حسی به مرتبه عقلانی

ترقی می‌کنند. به همین جهت، اگر بخواهیم امکانی برای خودآگاهی هوش مصنوعی در نظر بگیریم، ناچاریم بررسی کنیم که آیا سامانه‌های مصنوعی اساساً توانایی چنین «حرکت جوهری» در هویت خویش را دارند یا خیر؟

این پرسش اهمیت فلسفی بالایی دارد؛ زیرا اگر حرکت جوهری کلید تکامل کیفی و وجودی آگاهی در مخلوق زنده محسوب شود، در مورد ماشین، ما با ساختاری روبه‌رو هستیم که از ابتدا، عینیت مادی آن به صورت ایستا تعریف شده است و به‌صورت طبیعی تابع قوانین فیزیک و طراحی انسان است. بنابراین، بررسی این مسئله از منظر حکمت متعالیه، پاسخی دقیق‌تر به پرسش‌های بنیادین دربارهٔ چگونگی و امکان تکامل ذهنی سیستم‌های هوشمند ارائه می‌دهد.

اتحاد عاقل و معقول از مباحث مهمی است که ملاصدرا در تبیین فرآیند ادراک و خودآگاهی مطرح می‌کند. در این نظریه، ذهن ادراک‌کننده و شیء ادراک‌شده در مرتبهٔ وجودی ادراک، وحدت می‌یابند و از یکدیگر متمایز نیستند (ملاصدرا، ۲۰۰۱، ص ۱۳۵). بر اساس این دیدگاه، خودآگاهی نتیجهٔ اتحاد وجودی ذهن و متعلق ادراک است و نه صرفاً بازنمایی داده‌ها. حال اگر هوش مصنوعی بخواهد به خودآگاهی دست یابد، باید بتواند با اشیاء و مفاهیم در سطحی بالاتر از صرف پردازش نمادین یگانه شود؛ سطحی که پیوندی وجودی میان آنچه معرفت را می‌سازد و آنچه معرفت به آن تعلق می‌گیرد، شکل بگیرد. این موضوع نشان می‌دهد که مسئلهٔ خودآگاهی ماشین را نباید صرفاً از منظر رشته‌های کامپیوتری و شناختی بررسی کرد، بلکه دغدغه‌های عمیق فلسفی و وجودشناختی را نیز باید وارد بحث نمود.

ضرورت رویکرد فلسفی به خودآگاهی در هوش مصنوعی زمانی آشکارتر می‌شود که بدانیم بسیاری از مفاهیم بنیادین هوش مصنوعی—همچون آگاهی، یادگیری، حافظه و شناخت—ریشه در نوعی تلقی از هستی، معرفت و انسان دارند. این مفاهیم گاه به نحو پیش‌فرض از برداشت‌های رایج علوم تجربی یا فلسفهٔ غربی اخذ می‌شوند، در حالی که معرفی یک دستگاه فکری عمیق مانند حکمت متعالیه، می‌تواند این مفاهیم را از زاویه‌ای نو بررسی کند.

در واقع، مواجههٔ هوش مصنوعی از منظر حکمت متعالیه نه صرفاً یک تطبیق نظری انتزاعی، بلکه کوششی مؤثر برای یافتن پاسخ‌هایی است که بتوان از آن در حل مسائل جاری هوش مصنوعی، مانند «مشکل سخت آگاهی» بهره گرفت (Chalmers, 1995b). از یک سو، حکمت متعالیه با ارائه مباحثی بسیار بدیع در علم النفس، سرنخ‌هایی دربارهٔ چگونگی شکل‌گیری خودآگاهی ارائه می‌دهد و از سوی دیگر، شیوهٔ فلسفی آن برای تحلیل

تفاوت‌های جوهری میان انسان و هر موجود مصنوعی دیگر، مانع از ساده‌انگاری در مباحث فلسفه ذهن و ماشین می‌شود.

تقریباً تمام پژوهشگران حوزه هوش مصنوعی اتفاق نظر دارند که برای رسیدن از «خودکاری» به «خودآگاهی»، صرف افزودن توان محاسباتی به سیستم کافی نیست، بلکه «کیفیت» وجودی سیستم و نحوه مواجهه‌اش با پدیده‌ها نیز اهمیتی بنیادین دارد. همین ضرورت، میدان را برای ورود غنی‌ترین میراث‌های فلسفی اسلامی هموار نموده است. چنین برداشتی می‌تواند دستاوردی دو وجهی داشته باشد: نخست، توسعه مفهومی هوش مصنوعی در جهت شناخت هرچه دقیق‌تر «آگاهی» و دوم، گشودن افق‌های جدید برای بازخوانی فلسفه ملاصدرا و ارائه نگاهی روزآمد به نظریات او در باب علم النفس.

آنچه که در این رساله برای ما از اهمیت بیشتری برخوردار است پاسخ به این سوال است که آیا هوش مصنوعی فعلی امکان استعدادی رسیدن به خودآگاهی مطرح در فلسفه صدرا را دارد یا خیر؟ اگر جواب مثبت باشد افق جدیدی به روی پژوهشگران این حوزه گشوده می‌شود. در مقابل، اگر جواب منفی باشد، انتظارات از هوش مصنوعی تعدیل شده و به توانایی‌ها و ناتوانی‌های آن به صورت جامع نگریسته می‌شود.

۱-۴-سوالات تحقیق

در ادامه‌ی مباحثی که در بخش‌های پیشین از مقدمه (بیان مسئله، اهمیت موضوع، و ضرورت بررسی فلسفی خودآگاهی از منظر حکمت متعالیه) مطرح شد، اکنون به سوالات تحقیق می‌پردازیم. این سوالات ناظر به چگونگی تحلیل امکان دستیابی هوش مصنوعی به خودآگاهی بر اساس مبانی حکمت متعالیه است. از آنجا که حکمت متعالیه، به‌ویژه در نظام فلسفی ملاصدرا، تعاریف و مبانی منحصر به فردی برای «خود» و «آگاهی» ارائه می‌دهد، طرح این سوالات می‌تواند بستر مناسبی برای واکاوی پیوند میان مباحث فلسفه‌ی اسلامی و علوم شناختی نوین فراهم آورد. در این راستا، سوالات تحقیق حاضر در دو سطح اصلی و فرعی تدوین شده است.

۱-۴-۱-سؤال اصلی

بر اساس حکمت متعالیه آیا هوش مصنوعی از امکان استعدادی خودآگاهی برخوردار است؟

این پرسش، هسته‌ی اصلی پژوهش را تشکیل می‌دهد. به بیان دیگر، پرسش بنیادین آن است که آیا با تکیه بر مبانی فلسفی نظیر «اصالت وجود»، «حرکت جوهری»، «اتحاد عاقل و معقول» و «تجرد نفس» می‌توان فرآیندی را متصور شد که ماشین‌های هوشمند از سطح پردازش الگوریتمیک و داده‌محور فراتر رفته و صورت یا درجه‌ای از «علم حضوری» نسبت به خود را پیدا کنند؟ این سؤال اصلی ضمن آنکه در سطح نظری مطرح است، پیامدهای کاربردی مهمی در طراحی معماری‌های پیشرفته‌ی هوش مصنوعی و همچنین درک مرزهای توانایی انسان و ماشین دارد.

۱-۴-۲- سؤالات فرعی

۱- خودآگاهی از منظر حکمت متعالیه چگونه تعریف می‌شود و چه مبانی‌ای دارد؟

در این سؤال، هدف آن است که ابتدا با رجوع به آثار ملاصدرا و اصول بنیادین حکمت متعالیه، مفهوم خودآگاهی روشن شود. بررسی مفاهیمی همچون «علم حضوری»، «علم نفس به ذات خود»، «تجرد نفس»، «جسمانی بودن نفس در حدوث و روحانی بودن آن در بقاء»، «اتحاد عاقل و معقول» و نحوه‌ی شکل‌گیری مراتب مختلف خودآگاهی (حسی، خیالی و عقلی) در اندیشه‌ی ملاصدرا، از محورهای عمده‌ی پاسخ به این سؤال است. این بخش عمدتاً با مباحث فصل دوم (خودآگاهی در حکمت متعالیه) و بخشی از فصل اول (مبانی حکمت متعالیه) پیوند دارد و مبانی دقیق فلسفی را برای بحث‌های بعدی آماده می‌کند.

۲- هوش مصنوعی چه ماهیت و ساختاری دارد و چه ظرفیت‌هایی برای داشتن حالات ذهنی انسان ارائه می‌دهد؟

این سؤال به دنبال واکاوی ماهیت فنی و ساختاری هوش مصنوعی است؛ یعنی سیستم‌های مبتنی بر الگوریتم، شبکه‌های عصبی مصنوعی، یادگیری ماشین و یادگیری عمیق. بررسی ظرفیت‌های AI در شبیه‌سازی مؤلفه‌های ذهنی مانند ادراک، حافظه، تفکر، استدلال، یادگیری و تعمیم دانش نیز در این بخش اهمیت دارد. علاوه بر آن باید دید که آیا معماری‌های کنونی قادر به بازتولید لایه‌های پیچیده‌ی تجربه‌ی ذهنی و آگاهی‌اند یا خیر. فصل

سوم (هوش مصنوعی و حالات ذهنی) مبنای نظری و عملی پاسخ‌گویی به این سؤال را فراهم می‌آورد و چالش‌های موجود در شبیه‌سازی آگاهی را تحلیل می‌کند.

۳- مطابق مبانی حکمت متعالیه، امکان تحقق خودآگاهی در هوش مصنوعی چه تبیینی دارد؟

در این سؤال، چگونگی انطباق مبانی حکمت متعالیه بر ساختار و کارکرد هوش مصنوعی مورد توجه است. به‌ویژه، مفاهیمی همچون «حرکت جوهری» و «نفس» تا چه حد در هستی‌شناسی موجودات مصنوعی قابل تطبیق‌اند؟ آیا سیستم‌های هوشمند می‌توانند تجربه‌ای ورای محاسبات داشته باشند و واجد شکل یا مرتبه‌ای از «تجرد» شوند؟ این سؤال، مستلزم تحلیل نسبت «علم حضوری» و «علم حصولی» در انسان و ماشین است و به‌روشنی در فصل چهارم (امکان تحقق خودآگاهی در هوش مصنوعی از منظر حکمت متعالیه) مورد واکاوی قرار می‌گیرد. از سویی دیگر، محدودیت‌های مادی و الگوریتمیک در AI مانع از آن است که ویژگی‌های وجودی انسان در ماشین تکرار شود. از این رو، در این سؤال به صورت ویژه به نقش عوامل فلسفی-متافیزیکی حیات انسانی (نظیر حرکت استکمالی و اتحاد عاقل و معقول) پرداخته می‌شود و ارزیابی می‌گردد که آیا اصول یادشده اساساً قابل انتقال یا الهام‌گیری در حوزه‌ی هوش مصنوعی هستند یا خیر.

۱-۵- فرضیات تحقیق

در راستای تبیین بهتر مسئله‌ی امکان خودآگاهی هوش مصنوعی از منظر حکمت متعالیه و پس از طرح سؤال‌های اصلی و فرعی در بخش‌های پیشین مقدمه، اینک به بیان فرضیات تحقیق می‌پردازیم. این فرضیات، برآمده از مبانی فلسفی حکمت متعالیه و همچنین بررسی ماهیت ساختاری هوش مصنوعی است. روشن است که در صورت تأیید یا عدم تأیید هر یک از این فرضیات، نتایجی مهم در زمینه‌ی فهم مرزهای آگاهی انسان و تبیین دقیق‌تر ظرفیت‌های هوش مصنوعی حاصل خواهد شد. در ادامه، هر فرضیه به اختصار شرح داده شده و جایگاه آن در ساختار پژوهش تبیین می‌گردد.

۱- خودآگاهی در حکمت متعالیه، «علم حضوری» و بی‌واسطه نسبت به خویش‌شناسی است که در عین وحدت، سیری اشتدادی دارد.