

**Journal of Legal Philosophy Studies**  
Volum 1, Consecutive Number 1, 2025

Journal Homepage: <https://philosophylawjournal.ir/>

This is an Open Access paper licensed under the Creative Commons License CC-BY 4.0 license.



## **The Philosophy of Artificial Intelligence; From the Basics And Origin to the Realm of Metamorphosis**

**Mahmoud Abbasi<sup>1</sup>; Mehrdad Teymouri<sup>2\*</sup>** 

1. Professor of Medical Law and Head of Medical Law and Ethics Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

2. Assistant Professor of Criminal Law and Criminology and Researcher of Medical Law and Ethics Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

mehrdadteymoori1@gmail.com

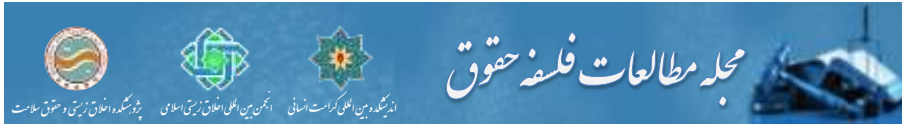
### **Abstract**

Artificial intelligence is increasingly infiltrating our lives; Therefore, relying only on traditional and technology-based approaches for development and deployment will not be enough because that technology is such that it increases the human experience. Artificial intelligence has close ties with philosophy as both share several concepts such as intelligence, action, self-awareness, epistemology and even agency. The ever-increasing development of technology is trying to create artificial life or artificial creatures, for this reason, philosophers pay great attention to artificial intelligence. These factors have caused the emergence of the philosophy of artificial intelligence. In addition, this technology is related to the creation of artificial beings, so this field can be of interest to philosophers. These factors contributed to the emergence of the philosophy of artificial intelligence. Philosophy of artificial intelligence is a branch of philosophy of mind and philosophy of computer science that examines artificial intelligence and its implications for the knowledge and understanding of intelligence, ethics, consciousness, epistemology, and free will; Therefore, the present article theoretically and in a descriptive and analytical way to recognize existential philosophy; The foundations, origins and scope of the philosophical developments of artificial intelligence have been discussed. The research findings strengthen the hypothesis that the philosophy of artificial intelligence is a set of issues that primarily deal with the issue of whether artificial intelligence is possible or not and whether it is possible to build an intelligent thinking machine or not. The most important issues of "possibility or not" lie in the intersection of the theories of the semantic content of thought and the nature of calculations. Another problem surrounds the nature of rationality and

- Abbasi, M., Teymouri M. (2025). The Philosophy of Artificial Intelligence; From the Basics And Origin to the Realm of Metamorphosis, *Journal of Legal Philosophy Studies*, 1(1), 1-17.

revolves around the apparently "superior" reasoning powers of the human mind. Another key point is related to issues related to the architecture of a smart machine. Once the understanding of a technology is provided, the appropriate response should be found according to the regulations and laws according to its social reactions. All the aforementioned features are also present in new AI technologies. Meanwhile, the philosophical validity of artificial intelligence is accompanied by the fear of ending the era of human control over the earth. Therefore, regarding the autonomy of artificial intelligence and decision-making capabilities, future research should be able to focus on the controlling moral frameworks.

**Keywords:** Artificial Intelligence, Philosophical Validity, Philosophy Of Mind, Philosophy Of Artificial Intelligence, Logic Of Artificial Intelligence



مجله مطالعات فلسفه حقوق

دوره اول - شماره اول - ۱۴۰۴

صفحات ۱-۱۷ (مقاله پژوهشی)

تاریخ: دریافت ۱۴۰۳/۰۷/۱۸ - پذیرش ۱۴۰۴/۰۴/۲۵ - انتشار ۱۴۰۴/۰۵/۲۰

## فلسفه هوش مصنوعی؛ از مبانی و خاستگاه تا قلمرو دگرذیسی

محمود عباسی<sup>۱</sup>، مهرداد تیموری<sup>۲\*</sup>

۱. استاد حقوق پزشکی و رییس مرکز تحقیقات اخلاق و حقوق پزشکی دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.

۲. استادیار حقوق جزا و جرم‌شناسی و پژوهشگر مرکز تحقیقات اخلاق و حقوق پزشکی دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران. Mehrdadteymoori1@gmail.com

### چکیده

هوش مصنوعی به طرق فزاینده‌ای در حال نفوذ در زندگی ما است؛ بنابراین با تکیه تنها در رویکردهای سنتی و مبتنی بر فناوری برای توسعه و استقرار کافی نخواهد بود زیرا آن فناوری به گونه‌ای است که تجربه انسان را افزایش می‌دهد. هوش مصنوعی پیوندهای نزدیکی با فلسفه دارد زیرا هر دو در مفاهیم متعددی مانند هوش، عمل، خودآگاهی، معرفت‌شناسی و حتی اختیار مشترک هستند. رشد روزافزون فناوری درصدد ایجاد حیات مصنوعی یا مخلوقات مصنوعی است، به همین دلیل فلاسفه توجه شایانی به هوش مصنوعی نشان می‌دهند. این عوامل باعث ظهور فلسفه هوش مصنوعی شده است. علاوه بر این، این فناوری با ایجاد موجودات مصنوعی مرتبط است، بنابراین این رشته می‌تواند مورد توجه فیلسوفان واقع شود. این عوامل به ظهور فلسفه هوش مصنوعی کمک کردند. فلسفه هوش مصنوعی شاخه‌ای از فلسفه ذهن و فلسفه علوم کامپیوتر است که هوش مصنوعی و پیامدهای آن را برای دانش و درک هوش، اخلاق، آگاهی، معرفت‌شناسی و اراده آزاد بررسی می‌کند؛ از همین رو مقاله حاضر به صورت نظری و به شیوه توصیفی و تحلیلی به بازشناسی فلسفه وجودی؛ مبانی، خاستگاه و قلمرو تحولات فلسفی هوش مصنوعی پرداخته است. ره‌آورد پژوهش این فرضیه را تقویت می‌نماید که فلسفه هوش مصنوعی مجموعه‌ای از مسائلی است که در درجه اول به این موضوع می‌پردازد که آیا هوش مصنوعی امکان‌پذیر است یا خیر و اینکه امکان ساخت یک ماشین تفکر هوشمند وجود دارد یا خیر. مهم‌ترین مسائل «ممکن بودن یا نه» در تلاقی نظریه‌های محتوای معنایی اندیشه و ماهیت محاسبات نهفته است. مسئله دیگر مربوط به ماهیت عقلانیت را احاطه

– عباسی، محمود؛ تیموری، مهرداد. (۱۴۰۴). فلسفه هوش مصنوعی؛ از مبانی و خاستگاه تا قلمرو دگرذیسی، مجله مطالعات فلسفه حقوق، ۱(۱)، صفحات ۱-۱۷.

کرده است و حول قوای استدلالی به ظاهر «متعالی» ذهن انسان می‌چرخد. نکته کلیدی دیگر مربوط به مسائل مربوط به معماری یک ماشین هوشمند می‌باشد. هنگامی که زمینه درک یک فناوری فراهم شد باید با توجه به واکنش‌های اجتماعی آن پاسخ مناسب آن را مطابق مقررات و قوانین پیدا کرد. همه ویژگی‌های یاد شده در مورد فناوری‌های جدید هوش مصنوعی نیز وجود دارند. در این میان اعتبار فلسفی هوش مصنوعی همراه با ترس از پایان دادن به دوره کنترل انسان بر روی زمین است. لذا نسبت به استقلال هوش مصنوعی و قابلیت‌های تصمیم‌گیری، تحقیقات آینده می‌باید می‌تواند بر چارچوب‌های اخلاقی کنترل‌کننده تمرکز کند.

کلیدواژه: هوش مصنوعی، اعتبار فلسفی، فلسفه ذهن، فلسفه هوش مصنوعی، منطق هوش مصنوعی

## مقدمه

توسعه هوش مصنوعی صرفاً یک کار نیست بلکه یک کار فنی اما عمیقاً فلسفی است. چراکه ماهیت هوش و آگاهی فراتر از محدوده است. علم و فلسفه اغلب با اشاره به این نکته متمایز می‌شوند که علم به دنبال توضیح است درحالی که فلسفه به دنبال توجیه است. برای مثال، پرسیدن اینکه علت شلیک عصبی سیناپس‌ها در مغز چیست، یک سؤال علمی است، درحالی که پرسیدن اینکه چه چیزی می‌تواند زمینه کافی برای باور اینکه یک شبکه عصبی ساخته شده مصنوعی آگاهانه است، یک سؤال فلسفی است؛ بنابراین، فلسفه به‌عنوان ارزیابی انتقادی باورها از طریق تحلیل مفاهیم در یک حوزه تحقیق مشخص شده است. البته علم به ارزیابی انتقادی باورها و تحلیل مفاهیم نیز می‌پردازد. با این حال، فلسفه یک رشته غیر تجربی یا پیشینی، در تضاد متمایز با علم است (Timothy, 2000: 24). به نظر می‌رسد که علوم کامپیوتر مانند هر علم دیگری از فلسفه متمایز است؛ اما علم کامپیوتر در بین علوم در انواع مدل‌هایی که ایجاد می‌کند منحصر به فرد است. علم در جستجوی توضیح، اغلب مدل‌هایی برای آزمایش فرضیه‌ها برای توضیح پدیده‌ها می‌سازد. این مدل‌ها، در قالب دستگاه‌های آزمایشی، البته اشیاء فیزیکی هستند. مدل‌های ساخته شده و دستکاری شده در علوم کامپیوتر، با این حال، به هیچ وجه فیزیکی نیستند. علوم کامپیوتر علمی است که به مطالعه فرآیندهای محاسباتی می‌پردازد. فرآیند محاسباتی از مثلاً یک فرآیند شیمیایی یا الکتریکی متمایز می‌شود، زیرا «به روشی که ماهیت فیزیکی آن را نادیده می‌گیرد» مورد مطالعه قرار می‌گیرد. (Hailperin, Kaiser, Knight, 1999: 3) جدای از ملاحظات ناشی از مدل‌سازی دانش و عقل، علم کامپیوتر برای تجزیه و تحلیل خوب قدیمی که فلسفه می‌تواند برای هر علمی ارائه کند، آماده است؛ بنابراین، نقش کاملاً معقول فلسفه تلاش برای قرار دادن علوم رایانه در طیف وسیعی از تحقیقات است که علم را تشکیل می‌دهد.

وسواس نسبت به فناوری پیشینه‌ای طولانی از تکامل موازی بین انسان و ماشین دارد. این وسواس زمانی غیر قابل برگشت شد که هوش مصنوعی بخشی از زندگی روزمره ما شد. با این حال، این ادغام هوش مصنوعی زمانی که ترس از پیشرفت هوش مصنوعی در کسب آگاهی در بین بشر رخنه کرد، موضوع بحث و جدل شد. هوش مصنوعی موضوعی است که مدت‌ها مورد بحث در زمینه هوش مصنوعی و علوم اعصاب است که چالش‌ها و

تهدیدات اخلاقی زیادی در جامعه از کارهای روزمره گرفته تا مأموریت‌های مریخ دارد. همچنان که با تجزیه و تحلیل مجموعه‌ای از فیلم‌های علمی تخیلی به منظور ارائه تفاوت واقعی بین هوش مصنوعی علمی تخیلی و هوش مصنوعی عملیاتی و بررسی مفهوم شگفت‌انگیز هوش مصنوعی در پرتو اصطلاحات پسانسانی، تکینگی فناوری و ابر هوشی نتیجه این تئوری‌ها برای آگاهی مصنوعی آینده‌نگر در هوش مصنوعی ساخته شده‌اند. این مطالعه مغالطه‌های پسانسانی را که پیرامون ترس از هوش مصنوعی به دست می‌آید و پیامد آن را افشا می‌کند (Prabhu, Premraj, 2024: 3). در سال ۱۹۸۰، عارف و متفکر ۸۵ ساله جیدو کریشنامورتی (۱۸۹۵-۱۹۸۶)، که توسط دانشمندان رایانه با موضوع هوش مصنوعی آشنا شد، در چالشی که با چشم‌انداز تصاحب ماشین برای ذهن انسان ایجاد می‌کرد، غرق شد. فرآیندها و توانایی‌های آن آنچه زاویه کریشنامورتی را شایسته بحث جداگانه می‌کند این است که نه پیامدهای اجتماعی و فرهنگی، بلکه بر یک بحران بالقوه فلسفی و روانی که عمدتاً نادیده گرفته شده است، تأکید می‌کند. کریشنامورتی به‌عنوان متفکر و عارفی که کار زندگی‌اش کاملاً بر دگرگونی ذهن انسان متمرکز بود، نگران بود که ذهنی که به‌اندازه کافی پرورش نیافته و صرفاً برای اهداف مادی و مکانیکی به کار گرفته شده بود، کاملاً قابل تقلید باشد و در نتیجه با رایانه‌ها و ماشین‌های دیگر قابل جایگزینی باش؛ بنابراین، نگرانی اصلی ما نباید ماشین‌هایی باشد که به ذهن‌های انسان مانند دست می‌یابند، بلکه باید افرادی با ذهن ماشین مانند باشند (Tubali, 2024: 1). در حال حاضر، به نظر می‌رسد توافق گسترده‌ای در میان فلاسفه وجود دارد که احتمال از ظهور یک هوش مصنوعی عمومی یا قوی یا ضعیف است یا وجود ندارد، و هوش مصنوعی یک آژانس مصنوعی تقلید یا شبیه‌سازی ظرفیت‌های خاصی از ذهن انسان با تکیه بر تعامل انسان و ماشین است (Krämer, 2022: 17-18). در نتیجه بعد خاص برخورد هوش مصنوعی و ذهن می‌تواند به‌طور قابل توجهی زمینه فلسفه هوش مصنوعی را گسترش دهد. اگر یک کامپیوتر متفکر از حالت‌های مجزا یا پیوسته محاسباتی و نمایشی استفاده می‌کند، داشتن بدنی لازم و آگاهانه ضروری است. از همین رو سؤالاتی که در ذهن به وجود می‌آورد این است که آیا رایانه می‌تواند هوشیار باشد؟ آیا هوش مصنوعی می‌تواند حس اخلاقی داشته باشد؟ آیا ما وظایفی در قبال رایانه‌های متفکر و ربات‌های هوشمند خواهیم داشت؟ به‌عنوان مثال، آیا برای انسان ساخت یک ماشین هوشمند اخلاقی است؟ اگر چنین ماشینی ساخته شود، آیا

خاموش کردن آن معادل قتل است؟ اگر ما یک نژاد از این گونه ماشین‌های هوشمند داشتیم، آیا این غیر اخلاقی نبود که آن‌ها را مجبور کنیم برای ما کار کنند؟

### ۱. ظهور هوش مصنوعی

هوش مصنوعی به نحوی نسبت به سایر فناوری‌ها به ما نزدیک‌تر می‌شود - بنابراین زمینه «فلسفه هوش مصنوعی» وجود دارد. شاید این به این دلیل است که پروژه هوش مصنوعی ایجاد ماشین‌هایی است که دارای یک ویژگی مرکزی برای نحوه تلقی ما انسان‌ها از خودمان هستند، یعنی به‌عنوان موجوداتی احساس، فکر و باهوش. اهداف اصلی یک عامل هوشمند مصنوعی احتمالاً شامل سنجش، مدل‌سازی، برنامه‌ریزی و اقدام است، اما کاربردهای فعلی هوش مصنوعی شامل ادراک، تجزیه و تحلیل متن، پردازش زبان طبیعی، استدلال منطقی، بازی، سیستم‌های پشتیبانی تصمیم، تجزیه و تحلیل داده‌ها، تجزیه و تحلیل پیش‌بینی و همچنین وسایل نقلیه خودمختار و سایر اشکال رباتیک (Stone et al, 2016: 25). هوش مصنوعی ممکن است شامل تعدادی تکنیک محاسباتی برای دستیابی به این اهداف باشد، چه هوش مصنوعی دست‌کاری نمادهای کلاسیک، با الهام از شناخت طبیعی، یا یادگیری ماشینی از طریق شبکه‌های عصبی (Silver et al, 2018: 22). هوش مصنوعی در حال حاضر طیف گسترده‌ای از زیرشاخه‌ها را دربر می‌گیرد که از عمومی (یادگیری و ادراک) به خاص، مانند بازی شطرنج، اثبات قضایای ریاضی، نوشتن شعر، رانندگی در یک خیابان شلوغ و تشخیص بیماری‌ها می‌توان اشاره نمود. همچنان که ربات‌ها در واقع اجسامی هستند که می‌توانند اشکال مختلف فیزیکی به خود گرفته و به‌طور مستقل با انسان‌ها ارتباط برقرار نمایند. آن‌ها قادرند از انسان‌ها نیازمند به کمک مثل سالمندان مراقبت کنند، هرچند که مطمئناً نمی‌توانند درک دلسوزانه‌ای از مراقبت داشته باشند. با این حال فناوری هوش مصنوعی هنوز در مراحل ابتدایی است و مطالعات بیشتری در حال انجام است، چراکه هر سال نسبت به سال قبل پیشرفته‌تر می‌شود (عباسی، تیموری، ۱۴۰۲: ۳)

در سال‌های اخیر، سیل اطلاعات بسیار زیاد شده است و ما دائماً بار اطلاعاتی را تجربه می‌کنیم. اضافه‌بار اطلاعات جدید و منحصر به عصر ما نیست. قبلاً در آثار کلاسیک سیمون (۱۹۷۱) با راه‌حل‌های فن‌آوری ارائه‌شده حتی قبل‌تر (بوش، ۱۹۴۵) به‌خوبی مستند شده است. برخی توانمندی‌ها در هوش مصنوعی، به‌ویژه الگوریتم‌های یادگیری ماشینی و

تجزیه و تحلیل داده‌های بزرگ، وعده افزایش انتخاب اطلاعات، پردازش و تصمیم‌گیری کاربر را برای ایجاد یک زندگی کم استرس و «بدون اصطکاک» می‌دهند به‌طور معمول، این راه‌حل‌ها الگوریتم‌های یادگیری ماشینی، مانند موتورهای توصیه (برای جستجوی اطلاعات، اخبار، موسیقی، خرید و غیره) را با هم ترکیب می‌کنند و آن‌ها را برای داده‌های دیجیتال کاربر اعمال می‌کنند. (Bar-Gil, 2024: 34) انجام این کار، شرکت‌های فناوری حوزه‌های سیاسی، جامعه‌شناسی و اقتصادی را متناسب با مدل سرمایه‌داری نظارتی می‌سازند. اگرچه بحث در مورد این پیامدها اهمیت بسیار زیادی دارد، آیا الگوریتم‌های پردازش اطلاعات مبتنی بر هوش مصنوعی، با هدف تقویت و حتی جایگزینی پردازش شناختی داخلی ما، می‌توانند هوش مصنوعی انسان‌محور در نظر گرفته شوند؟ برای پاسخ به این سؤال، اگرچه این فناوری‌ها ممکن است قابلیت‌های جدیدی را برای افزایش مشارکت کاربران انسانی در حوزه اطلاعات فراهم کنند، اما تأثیر قابل توجهی بر کاربران دارند عاملیت آن‌ها را کاهش می‌دهند، آن‌ها را مستقل‌تر می‌کنند، آن‌ها را تأیید می‌کنند و عقلانیت الگوریتمی را تأیید می‌کنند و در نتیجه تغییراتی در الگوهای حافظه آن‌ها ایجاد می‌شود.

مفهوم هوش مصنوعی به‌طور گسترده به‌عنوان هر نوع سیستم محاسباتی مصنوعی که رفتار هوشمندانه را نشان می‌دهد، یعنی رفتار پیچیده‌ای که منجر به دستیابی به اهداف می‌شود، درک می‌شود. به‌طور خاص، ما نمی‌خواهیم «هوش» را به آنچه اگر توسط انسان انجام می‌شد به هوش محدود کنیم، همان‌طور که مینسکی (۱۹۸۵) پیشنهاد کرده بود. این بدان معناست که ما طیف وسیعی از ماشین‌ها، از جمله ماشین‌هایی که در «هوش مصنوعی فنی» هستند که فقط توانایی‌های محدودی در یادگیری یا استدلال نشان می‌دهند، اما در خودکارسازی وظایف خاص برتری دارند و همچنین ماشین‌هایی در «هوش مصنوعی عمومی» که هدفشان ایجاد یک هوش مصنوعی عمومی است را ترکیب می‌کنیم. عامل هوشمند

## ۲. هوش مصنوعی و علوم شناختی در چارچوب‌های معرفتی

دیدگاه هوش مصنوعی این است که نظریه‌های فلسفی تنها در صورتی برای هوش مصنوعی مفید هستند که مانع از سیستم‌های مصنوعی در سطح انسانی نشوند و مبنایی برای طراحی سیستم‌هایی با اعتقادات، انجام استدلال و برنامه‌ریزی فراهم کنند. تحقیقات هوش

مصنوعی به‌ویژه بر رسمی کردن اقدامات موجود در یک موقعیت و پیامدهای انجام هر یک از چندین اقدام تأکید کرده است. برای انجام این کار، هوش مصنوعی عمدتاً با تقریب‌های ساده پدیده‌ها سروکار داشته است (McCarthy, 2006: 5). شاید هیچ مفهومی به‌اندازه مفهوم محاسبات در پایه‌های علوم شناختی مدرن مرکزی نباشد. جاه‌طلبی‌های هوش مصنوعی بر یک چارچوب محاسباتی استوار است و در سایر حوزه‌های علوم شناختی، مدل‌های فرآیندهای شناختی اغلب به‌صورت محاسباتی ارائه می‌شوند. نقش اساسی محاسبات را می‌توان در دو پایان‌نامه اساسی بیان کرد. اول، اساس باور به امکان هوش مصنوعی، نظریه کفایت محاسباتی وجود دارد که بیان می‌کند که نوع درست ساختار محاسباتی برای داشتن ذهن و برای داشتن طیف گسترده‌ای از ویژگی‌های ذهنی کافی است. دوم، تسهیل پیشرفت علوم شناختی به‌طور کلی‌تر، نظریه‌ای از تبیین محاسباتی وجود دارد که بیان می‌کند که محاسبات چارچوبی کلی برای توضیح فرآیندهای شناختی و رفتار فراهم می‌کند (Chalmers, 2011: 326). هر سیستمی متشکل از عناصری که سیگنال‌ها را با یکدیگر و گهگاه با سایر نقاط جهان مبادله می‌کنند، می‌توان گفت که درگیر محاسبات است. مغز متشکل از نورون‌ها نمونه‌ای از سیستمی است که محاسبه می‌کند و محاسباتی که نورون‌ها به‌طور جمعی انجام می‌دهند، ذهن مغز را تشکیل می‌دهد (Edelman, 2008: 11).

هوش مصنوعی و هوش انسانی هر دو بر اساس برون‌دادها و درون‌دادها کار می‌کنند. درون‌دادهای انسان علوم و دریافت‌های حضوری و بدیهیات است که تنها در حوزه تصدیق‌های بدیهی یعنی قضایا قیاسات‌ها معنا رخ می‌دهد. بدون توجه به برون‌دادها یا تصورات کسبی قضایا شکل نمی‌گیرد. همان‌طور که هوش مصنوعی نیاز به داده‌هایی دارد تا بر اساس نرم‌افزار موجود در آن عملیات تحلیل و طبقه‌بندی و سایر اموری که ذکر شد را انجام دهد، در انسان نیز عقل استعدادی به‌مثابه نرم‌افزاری تعبیه شده در یک ماشین است که انسان مطابق و در پرتو آن می‌تواند به نحوه عملکردهای عالی‌تری دست یابد و البته کارکرد آن بر اساس فلسفه مشروط به دریافت‌های حصولی و برون‌دادها هست و هر نوع نقص فطری در برنامه‌گریزی موجود یعنی عقل هیولانی منطبق با سیستم عصبی، باعث تغییر در ایصال به‌مراتب بالاتر از هوشمندی هست همان‌طور که انسان به‌خودی‌خود نمی‌تواند واجد آن چیزی شود که ندارد و بر اساس قاعده فاقد شیئی معطی شیئی نمی‌شود؛

لازم است که عقلی برتر و بالفعل باشد تا انسان به واسطه آن ارتقای وجودی را بر اساس هویت علم بیابد، به همین صورت هوش ماشین را نمی‌توان متعلق به خود ماشین دانست بلکه در حقیقت این هوشمندی منتسب به هوش برتر انسان است اما برنامه‌ریزی صورت گرفته در ماشین با امکانات موجود در ماشین می‌تواند نتایجی مانند تشخیص، بهینه‌سازی و ارزیابی را انجام دهد. در ماشین‌های هوشمند، نرم‌افزار کار اصلی محاسبه و تعمیر و ... را انجام می‌دهد (شریعتی، اکبرزاده توتونچی، ۱۴۰۱: ۱۰۲). چارچوب‌های معرفتی در هوش مصنوعی نگرانی‌های معرفت‌شناختی به‌ویژه عدم اطمینان در مورد ذهن‌های دیگر و محدودیت‌های درون‌نگری را برای تحقیقات میان‌رشته‌ای فیلسوفان، دانشمندان علوم شناختی به وجود می‌آورد و محققان هوش مصنوعی می‌توانند برای بررسی چگونگی هوش مصنوعی با یکدیگر همکاری کنند شاید از طریق پیشرفت در تئوری ذهن یا مدل‌های همدلی در هوش مصنوعی سیستم‌ها می‌تواند ذهن‌های دیگر را شبیه‌سازی یا درک کند (Tripathi, 2024: 1). در ادغام تجربیات حسی حرکتی با زبانی مدل‌ها، محیطی برای توسعه درک معنایی غنی‌تر شاید با استفاده از سیستم‌های هوش مصنوعی که با آن‌ها تعامل دارند. مطالعات برای کشف اینکه چگونه سیستم‌های هوش مصنوعی می‌توانند خودآگاهی انسان را تقلید یا الگوبرداری می‌کنند. این ممکن است منجر شود تحقیقات نوآورانه در فراشناخت و خود ارجاعی هوش مصنوعی گسترش پیدا کند.

### ۳. چارچوب‌های اخلاقی هوش مصنوعی

امروزه فناوری‌های نوین به سرعت در حال تکامل است. در این میان هوش مصنوعی با پیشرفت روزافزون خود مزایای بالقوه زیادی برای بشر ارائه می‌دهد. چنین فناوری‌ای نیاز به پردازش و تجزیه و تحلیل مقادیر زیادی از داده‌ها و نیز دسترسی به حجم بزرگی از کلان داده‌ها دارد. باین حال، باید مسائل اخلاقی هوش مصنوعی به‌منظور جلوگیری از سوءاستفاده شناسایی شود و مورد توجه قرار گیرد. اخلاق هوش مصنوعی شاخه‌ای از اخلاق است که به مجموعه قوانین و دستورالعمل‌های اخلاقی مربوط می‌شود که بر رفتار بین فردی میان افراد یا شرکت‌ها حاکم می‌شود که با میانجیگری فناوری رایانه، چه در داخل شرکت یا به‌طور گسترده‌تر در بازارها و جامعه انجام می‌شود. اخلاق رقومی (که اخلاق اطلاعات نیز نامیده می‌شود) را به‌عنوان شاخه‌ای از اخلاق تعریف می‌کنند که بر رابطه بین ایجاد، سازمان‌دهی،

انتشار و استفاده از اطلاعات و اخلاق تمرکز دارد. معیارها و قوانین اخلاقی حاکم بر رفتار انسان در جامعه است (عباسی، تیموری، ۱۴۰۳: ۳۴). درحالی که می توان از هوش مصنوعی برای هدایت ربات‌هایی استفاده کرد که اگر فرآیندها یا ظاهر آن‌ها شامل فریب، تهدید کرامت انسانی یا نقض اخلاق کانتی «احترام» باشد که برای بشریت مشکل ساز است. انسان‌ها خیلی راحت ویژگی‌های ذهنی را به اشیا نسبت می دهند و با آن‌ها همدلی می کنند، به خصوص زمانی که ظاهر بیرونی این اشیا شبیه به موجودات زنده باشد. این می تواند برای فریب دادن انسان‌ها (یا حیوانات) به منظور نسبت دادن اهمیت فکری یا حتی احساسی بیشتر به ربات‌ها یا سیستم‌های هوش مصنوعی، بیش از آنچه که شایسته است، مورد استفاده قرار گیرد (Müller, 2021: 17). اخلاق هوش مصنوعی و رباتیک اغلب بر «نگرانی‌های» از انواع مختلف متمرکز است که پاسخی معمولی به فناوری‌های جدید است. بسیاری از این نگرانی‌ها نسبتاً عجیب و غریب هستند (قطارها برای روح خیلی سریع هستند). برخی از آن‌ها به طور قابل پیش بینی اشتباه می کنند وقتی می گویند این فناوری اساساً انسان‌ها را تغییر خواهد داد (تلفن ارتباطات شخصی را از بین می برد، نوشتن حافظه را از بین می برد، کاست‌های ویدیویی بیرون رفتن را زائد می کند). برخی از آن‌ها به طور کلی صحیح هستند اما نسبتاً مرتبط هستند (فناوری دیجیتال صناعی را که فیلم‌های عکاسی، نوار کاست یا صفحات وینیل می سازند، نابود می کند). اما برخی از آن‌ها به طور کلی درست و عمیقاً مرتبط هستند (ماشین‌ها کودکان را می کشند و اساساً چشم‌انداز را تغییر می دهند). وظیفه مقاله‌ای از این دست، تحلیل مسائل و کم کردن مسائل غیر مسئول است. برخی از فناوری‌ها، مانند انرژی هسته‌ای، خودروها یا پلاستیک‌ها، باعث بحث‌های اخلاقی و سیاسی و تلاش‌های سیاسی قابل توجهی برای کنترل مسیر این فناوری‌ها شده‌اند، معمولاً تنها زمانی که آسیبی وارد شود. علاوه بر چنین «نگرانی‌های اخلاقی»، فناوری‌های جدید هنجارها و سیستم‌های مفهومی کنونی را به چالش می کشند که مورد توجه خاص فلسفه است.

#### ۴. منطق‌گرایی در هوش مصنوعی

به سختی می توان انکار کرد که ربات‌ها به طور فزاینده‌ای توانایی خواهند داشت و اینکه انسان‌ها به طور فزاینده‌ای از این قابلیت‌ها با استقرار آن‌ها در محیط‌های حساس اخلاقی مانند بیمارستان‌ها استفاده می کنند، جایی که رفتار نادرست ربات‌ها می تواند عواقب بدی

برای انسان داشته باشد. چگونه می‌توانیم اطمینان حاصل کنیم که چنین ربات‌هایی همیشه از نظر اخلاقی درست رفتار می‌کنند. چگونه می‌توانیم پیش از موعد، از طریق استدلال‌هایی که به وضوح به زبان طبیعی بیان می‌شوند، بدانیم که رفتار آن‌ها به‌ویژه توسط کدهای اخلاقی انتخاب شده توسط ناظران انسانی محدود می‌شود. به‌طور کلی، یک رویکرد این است که اصرار کنیم که ربات‌ها فقط اعمالی را انجام می‌دهند که می‌توان آن‌ها را از نظر اخلاقی در منطقی دیونیتی‌ک منتخب انسان اثبات کرد - یعنی منطقی که یک کد اخلاقی را رسمیت می‌دهد. خود علمای اخلاق با ارائه نظریه‌ها و معضلات اخلاقی به صورت اظهاری و استدلال بر این اطلاعات با استفاده از منطق غیر رسمی و رسمی کار می‌کنند (Bringsjord, Arkoudas, Bello, 2006: 39). نویسندگان یک روش منطقی‌گرا را به‌طور کلی، عاری از هرگونه تعهد به سیستم‌های خاص، توصیف می‌کنند و نشان می‌دهند که آن چالشی را در رابطه با رفتار ربات در بخش مراقبت‌های ویژه حل می‌کند. این مقاله بخشی از یک شماره ویژه در اخلاق ماشین است. مشارکت‌های فلسفی در پایه‌های رشته‌های علمی اغلب حول «سؤالات محوری» در رابطه با تلاش‌های تقلیل‌گرا متمرکز است. به‌عنوان مثال، در فلسفه زیست‌شناسی، سؤال این است که آیا علم ارگانیک را می‌توان به علم غیر آلی (تقلیل قوانین بیولوژیکی به فیزیکی) تقلیل داد. در ریاضیات، منطق‌گرایی ادعا می‌کند که همه ریاضیات را می‌توان به منطق تقلیل داد. برای علم به‌طور کلی، پوزیتیویسم منطقی از کاهش واژگان نظری به واژگان مشاهده‌ای حمایت می‌کرد. یک «سؤال محوری» اولیه در فلسفه علوم کامپیوتر این است که آیا علم کامپیوتر به شاخه‌ای از ریاضیات منحرف شده است یا خیر. اینکه چگونه یک دانشمند کامپیوتر به این سؤال پاسخ می‌دهد می‌تواند بر روش‌شناسی او تأثیر بگذارد.

از آنجا که این توجیهی است برای اینکه چگونه استدلال در مورد برنامه‌ها را می‌توان غیر ریاضی دانست، باین حال ادعای بسیار گسترده‌تری است که بگوییم علوم کامپیوتر شاخه‌ای از ریاضیات است یا باید آرزو داشته باشیم که باشد. زیرا هنوز این مسائل وجود دارد که آیا مشخص کردن، تولید یا نگهداری برنامه‌ها (به غیر از استدلال در مورد تمام شده‌ها) مانند یک فعالیت ریاضی است یا باید باشد. موضوعی که انگیزه و زیربنای بسیاری از تنش‌ها در بحث فلسفی علوم رایانه است، تأیید رسمی، استدلال ریاضی درباره نتیجه برنامه است (Timothy R, 2000: 3) موضوع علم کامپیوتر شامل تعداد زیادی از

فعالیت‌های مختلف است که از موضوعات ریاضی انتزاعی تا شیوه‌های مهندسی اصلی و تحقیقات علمی را شامل می‌شود. متعاقباً، فلسفه علوم کامپیوتر با فلسفه‌های ریاضیات، علوم فناوری همپوشانی دارد و دغدغه‌های فلسفی اصلی این رشته‌ها همگی مشابه‌های محاسباتی دارند. وضعیت هستی‌شناختی برنامه‌ها، ماهیت انتزاع محاسباتی و نوع دانش ارائه شده توسط اثبات صحت، نمونه‌های اصلی هستند. علاوه بر این، به دلیل تمرکز علم کامپیوتر بر زبان‌های رسمی و تفسیر معنایی آن‌ها، فلسفه علم کامپیوتر از مباحث و الهام‌گیری از فلسفه‌های زبان و ذهن الهام می‌گیرد.

### ۵. فلسفه هوش مصنوعی

به دنبال پدید آمدن و رشد چشمگیر رایانه‌ها از اواسط سده گذشته میلادی، این تصور به وجود آمد که پیشرفت‌های مذکور به نقطه‌ای خواهد انجامید که در آن با همین روش‌های مطرح در علوم رایانه، می‌توان هوش انسانی را به صورت هوش مصنوعی بازسازی کرد و رایانه‌ها قادر به انجام اموری خواهند بود که از موجود هوشمندی مانند انسان برمی‌آید. در میان دانشمندان علوم رایانه‌ای، این پروژه با عنوان «هوش مصنوعی» شهرت یافت و نظریات گوناگونی پیرامون امکان موفقیت آن مطرح شد؛ گستره وسیعی از نظریات که در میان آن‌ها اختلاف‌های شایان توجهی وجود دارد تا جایی که برخی امکان آن را انکار کردند و گروه دیگری به گونه‌ای افراطی معتقدند که این پروژه به موفقیت نائل شده است و حتی همین رایانه‌های امروزی دارای هوش هستند و به خوبی فکر می‌کنند (مطلبی کربکندی، مینایی، دیرباز، ۱۳۹۳: ۱۸۰). نظریه هوش مصنوعی قوی در عصر حاضر، یکی از نظریات پرچالش و بااهمیت است که طرفداران و منکران بسیاری دارد. پاسخ به این مسأله، مستلزم داشتن رویکردی درباره رابطه ذهن/نفس و بدن است. تفاوت رویکردها به تفاوت پاسخ به امکان و عدم امکان هوش مصنوعی قوی منجر می‌گردد (قاضوی، واعظی، ۱۴۰۲: ۱۵۸). در دهه گذشته، پیشرفت قابل توجهی در هوش مصنوعی باعث پذیرش الگوریتم‌های آن شده است. در کنار این گام‌های قابل توجه، افزایش هم‌زمان پیچیدگی مدل و اتکا به مدل‌های هوش مصنوعی غیر شفاف و فاقد شفافیت وجود داشته است. در سناریوهای متعدد، خود سیستم‌ها ممکن است نیاز به تصمیم‌گیری‌هایی داشته باشند که دارای ابعاد اخلاقی باشد. در نتیجه، ابداع راه‌حل‌هایی برای ادغام ملاحظات اخلاقی در شیوه‌های توسعه سیستم هوش مصنوعی و

تسهیل استفاده گسترده‌تر از سیستم‌های هوش مصنوعی در حوزه‌های مختلف ضروری شده است. تلاش‌های تحقیقاتی باید رویکردهای نوآورانه را برای تقویت اصول اخلاقی در سیستم‌های هوش مصنوعی، ترویج شفافیت، مسئولیت‌پذیری و اعتماد بیشتر بررسی کنند. حمایت از حقوق اساسی فردی، کرامت انسانی، خودمختاری، حریم خصوصی، برابری و انصاف، در عین کاهش آسیب احتمالی، از اهمیت بالایی برخوردار است. در نظر گرفتن ارزش‌های اخلاقی و اطمینان از انطباق با الزامات اخلاقی در طول چرخه عمر توسعه سیستم‌های مستقل و هوشمند، اعتماد و قابلیت اطمینان را در استفاده از آن‌ها پرورش می‌دهد (Jedličková, 2024: 3). از اواسط قرن بیستم و به دنبال رشد چشمگیر علوم رایانه‌ای این تصور پدید آمد که می‌توان با ابزارهای فیزیکی و روش‌های مطرح در فیزیک و علوم رایانه‌ای هوش انسانی را به‌طور مصنوعی بازسازی کرد. هنگامی که در دهه ۱۹۵۰ میلادی ماشین‌های محاسبه‌گر کارآیی زیادی از خود نشان دادند، تحقیقات در جهت ساخت ماشین‌های هوشمند آغاز گردید؛ از آن زمان تاکنون با انجام تحقیقات فیزیکی و رایانه‌ای تلاش‌های وسیعی در این جهت صورت گرفته است. پس از گذشت چند سال از آغاز این تحقیقات، اندک‌اندک امکان یا عدم امکان ساخت هوش مصنوعی به مسأله‌ای تبدیل شد که فیلسوفان را به‌سوی خود خواند افت‌وخیزهایی که در زمینه ساخت ماشین‌های هوشمند پدید آمد و نیز بحث‌های فلسفی در این باب، سبب شد تا عقاید مختلفی در خصوص هوش مصنوعی میان فیلسوفان بروز پیدا کند (طهماسبی، ۱۳۸۵: ۲۵). ملاحظات اخلاقی باید در رویه‌های سازمانی که فعالیت‌های تحقیقاتی هوش مصنوعی را هدایت می‌کنند، ایجاد چارچوب‌های قوی که به نگرانی‌های اخلاقی رسیدگی می‌کند و پیامدهای اخلاقی سیستم‌های مبتنی بر هوش مصنوعی را منعکس کند و با ادغام ملاحظات اخلاقی در شیوه‌های توسعه سیستم هوش مصنوعی چارچوب‌های اخلاقی در تقویت اجرای هوش مصنوعی اخلاقی و تضمین یکپارچگی اخلاقی فناوری‌های هوش مصنوعی را ارائه کند.

### نتیجه‌گیری

پیشرفت سریع سیستم‌های خودکار و هوشمند نویدبخش عصر جدیدی از نوآوری‌ها است. تأثیرات مثبت عمیق در جنبه‌های مختلف انسان‌زندگی در سراسر جهان شامل آموزش، مراقبت‌های بهداشتی، کسب‌وکار، صنعت، حمل‌ونقل و ساختارهای اجتماعی، فناوری‌های

هوش مصنوعی، از جمله یادگیری ماشینی یادگیری عمیق، پردازش زبان طبیعی و کامپیوتر چشم‌انداز، درهم‌تنیده شده تا فرصت‌های بی‌شماری را برای بهبود رفاه انسان ارائه دهد. سیستم‌های هوش مصنوعی. کار نه تنها درک ما را از هوش مصنوعی عمیق‌تر می‌کند، بلکه ما را به بازنگری در ماهیت ذهن، هوش و شناخت به چالش می‌کشد. هوش مدت‌هاست که با توجه به نقش دانش و استدلال در هوش مصنوعی شناخته شده است؛ و برعکس، نقش هوش مصنوعی در فلسفه، حتی توسط برخی، هرچند بحث‌برانگیز، برجسته شده است؛ اما جدای از ملاحظات ناشی از مدل‌سازی دانش و عقل، علم کامپیوتر برای تجزیه و تحلیل خوب قدیمی که فلسفه می‌تواند برای هر علمی ارائه کند، آماده است؛ بنابراین، نقش کاملاً معقول فلسفه تلاش برای قرار دادن علوم رایانه در طیف وسیعی از تحقیقات است که علم را تشکیل می‌دهد. نگرانی در اینجا مقابله با بحران‌های هویت اجتناب‌ناپذیر است. با تجسم شناخت و احساسات ناخودآگاه. تمایز دقیقی را بین دستاوردهای مستدل فناوری دیجیتال و انتظارات بیش‌ازحد و بی‌اساس بروز می‌کند. با توجه به خطرات احتمالی مرتبط با سیستم‌های هوش مصنوعی، نظارت فعال در تمام مراحل «از طراحی و توسعه برای استقرار و استفاده» ضروری است. این رویکرد فعال، حکمرانی مؤثر هوش مصنوعی و امنیت و ترویج استفاده مسئولانه از فناوری‌های هوش مصنوعی از دیدگاه اخلاقی را تسهیل می‌کند. رابطه هوش مصنوعی و فلسفه شامل مفاهیم بسیاری است که هر دو موضوع شامل عمل می‌شوند، به‌عنوان مثال، عمل، اهداف، دانش، باور و آگاهی. با این حال، هوش مصنوعی در مورد این‌ها موضعی را اتخاذ می‌کند که ما می‌توانیم آن را طراح بدانیم. مفاهیم؛ می‌پرسد که چه نوع دانش، باور، آگاهی و غیره انجام می‌دهد یک سیستم کامپیوتری باید به‌منظور رفتار هوشمندانه و نحوه ساخت آن‌ها را به یک برنامه هوش مصنوعی تبدیل کنند. فیلسوفان عموماً چیزهای بیشتری گرفته‌اند دیدگاه انتزاعی و پرسیده شد که دانش چیست و غیره. موضع طراح مشابه است. سؤالات فلسفی به‌ویژه در سطح انسان به هوش مصنوعی مرتبط هستند به دنبال هوش مصنوعی است. با این حال، بیشتر تحقیقات هوش مصنوعی در سطح انسانی را هدف قرار نداده‌اند، بلکه با استفاده از تئوری‌ها و تکنیک‌های هوش مصنوعی به مشکلات خاص هستند. در نتیجه فلسفه هوش مصنوعی شامل تکنیک‌های بسیاری است که از مطالعه فلسفه به‌دست آمده‌اند. این‌ها شامل منطق استقرایی، نظریه احتمالات و استدلال عملی است.

### منابع

- عباسی، محمود، تیموری، مهرداد، (۱۴۰۳)، چالش‌های اخلاق رقومی هوش مصنوعی و امکان‌سنجی بیمه مسئولیت برای سامانه مبتنی بر هوش مصنوعی، فصلنامه دولت و حقوق، ۵(۱۵): ۳۱-۴.
- عباسی، محمود، تیموری، مهرداد، (۱۴۰۲)، مروری بر چالش‌های اخلاقی و حقوقی کاربرد هوش مصنوعی در نظام سلامت، مجله اخلاق پزشکی، ۱۷(۴۸): ۱۱-۱.
- طهماسبی، محمدرضا، (۱۳۸۵)، رهیافت‌های بنیادین فلسفی در هوش مصنوعی، فصلنامه حکمت و فلسفه، ۲(۶): ۲۵-۴۷.
- مطلبی کربکندی، حسین، مینایی، بهروز، دیرباز، عسگر، (۱۳۹۳)، بررسی فلسفی امکان تحقق هوش مصنوعی قوی با توجه به دیدگاه‌های مختلف در مسأله ذهن و بدن، مجله فلسفه دین، ۱۱(۱): ۱۹۶-۱۷۳.
- قاضوی، سید محمد، واعظی، احمد، (۱۴۰۲)، نسبت رویکردهای مختلف به ذهن با امکان هوش مصنوعی قوی، مجله نقد و نظر، ۲۸(۱۱۱): ۱۵۸-۱۸۷.
- شریعتی، فهیمه، اکبر زاده توتونچی، محمدرضا، (۱۴۰۱)، مقایسه ماهوی هوش انسانی با هوش مصنوعی از منظر فلسفه اسلامی با تأکید بر حکمت متعالیه ملاصدرا، راهگشایی در فهم جایگاه عقول برتر، فصلنامه حکمت معاصر، ۱۳(۳۵): ۱۱۷-۸۹.

### Reference

- Tripathi RL, (2024), Mind and Machine: A Philosophical Examination of Matt Carter's "Minds & Computers: An Introduction to the Philosophy of Artificial Intelligence", Journal of Data Science and Artificial Intelligence, 1(2): 1-3.
- Prabhu, M. & Premraj, J. Anil, (2024), Artificial consciousness in AI: a posthuman fallacy, AI and Society, Journal of AI & Society, 1(3): 1-14.
- Tubali, Shai, (2024), Will humans ever become conscious? Jiddu Krishnamurti's thought about AI as a fresh perspective on current debates, Journal of AI & Society, 1(3): 1-13.
- Krämer, Sybille, (2022), The artificiality of the human mind: a reflection on natural and artificial intelligence. In: Nagl-Docekal H, Zacharasiewicz W (eds) Artificial intelligence and human enhancement. De Gruyter, Berlin, 17-32
- Bar-Gil, Oshri, (2024), Redefining Human-Centered AI: The Human Impact of AI-based Recommendation Engines, In book: Human-Centered AI A Multidisciplinary Perspective for Policy-Makers, Auditors, and Users. pp. 34-45. Publisher: Taylor & Francis.

- Jedličková, Anetta, (2024), Ethical approaches in designing autonomous and intelligent systems: a comprehensive survey towards responsible development, *Journal of AI & Society*, 2(1): 1-14.
- Chalmers, David, (2011), A computational foundation for the study of cognition, *Journal of Cognitive Science*. 12(4): 323-357.
- Edelman, Shimon, (2008), *Computing the Mind: How the Mind Really Works*, 1st edition, Oxford: Publisher Oxford University Press.
- Stone, Peter, Rodney Brooks, Erik Brynjolfsson, Ryan Calo, Oren Etzioni, Greg Hager, Julia Hirschberg, Shivaram Kalyanakrishnan, Ece Kamar, Sarit Kraus, Kevin Leyton-Brown, David Parkes, William Press, AnnaLee Saxenian, Julie Shah, Milind Tambe, and Astro Teller, 2016, "Artificial Intelligence and Life in 2030", One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University, Stanford, CA, September 2016.
- Silver, David, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharrshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, and Demis Hassabis, 2018, "A General Reinforcement Learning Algorithm That Masters Chess, Shogi, and Go through Self-Play", *Science*, 362(6419): 1140-1144.
- Müller, Vincent C, (2021), Ethics of Artificial Intelligence and Robotics, *Stanford Encyclopedia of Philosophy*, Stanford University, 1-31.
- Bringsjord, Selmer, Arkoudas, Konstantine, Bello, Paul, (2006), Toward a General Logicist Methodology for Engineering Ethically Correct Robots, Published in: *IEEE Intelligent Systems* 21(4): 38-44.
- Colburn, Timothy R, (2000), *Philosophy and computer science*, 1st Edition, Publisher in Routledge.
- Hailperin, Max, Kaiser, Barbara, Knight, Karl, (1999), *Concrete Abstractions An Introduction to Computer Science Using Scheme*, 1st Edition, An Imprint of Brooks/Cole Publishing Company.
- McCarthy, John, (2006), The Philosophy Of Ai And The Ai Of Philosophy, *Philosophy of Information*, 8(1): 711-740.