

Reflection on Artificial Intelligence and the Enhancement of Moral Judgment:

Exploring Challenges and Opportunities

Reza Jafari

PhD student of Islamic Studies, Department of Ethics, University of Islamic Studies, Qom,
Iran. rezajafari128@gmail.com

Abstract

Objective: This research delves deeply into the issues and challenges associated with artificial intelligence (AI) and moral judgment, seeking to answer the fundamental question of how AI can play a role in human moral decision-making.

Research Methodology: The study employs an analytical-descriptive approach, utilizing library research to examine existing literature and resources related to AI, ethics, and moral judgment.

Findings: AI, through complex algorithms and machine learning, is capable of processing information and making decisions across various domains. The application of AI in moral judgment offers benefits such as increased accuracy, speed, and impartiality. However, several challenges persist, including a lack of understanding of moral nuances, bias, and a deficiency in creativity and innovation.

Discussion and Conclusion: Currently, fully entrusting moral judgment to AI is unreliable. AI can serve as a tool to assist humans in the moral judgment process, but final decision-making should remain with humans.

Keywords: Artificial Intelligence, Moral Judgment, Applied Ethics, Moral Reasoning, Judgment Challenges, prejudice.

Cite this article: Jafari, R. (2024). Reflection on Artificial Intelligence and the Enhancement of Moral Judgment: Exploring Challenges and Opportunities. *Religious Studies in Media*, 6(24): 55-78. <https://doi.org/10.22034/jmrs.2024.210084>

Received: 2024/05/06 Revised: 2024/07/07 Accepted: 2024/07/29 Published online: 2024/12/30

© The Author(s).

Article type: Research Article

Published by: IRIB University



تأملی در باب هوش مصنوعی و ارتقای قضاوت اخلاقی؛ واکاوی چالش‌ها و فرصت‌ها

رضا جعفری

دانشجوی دکتری مدرسی معارف اسلامی، گروه اخلاق، دانشگاه معارف اسلامی، قم، ایران. rezajafari128@gmail.com

چکیده

هدف: این پژوهش به بررسی عمیق مسائل و چالش‌های مرتبط با هوش مصنوعی و قضاوت اخلاقی می‌پردازد و به دنبال پاسخ به این پرسش اساسی است که هوش مصنوعی چگونه می‌تواند در فرایند قضاوت اخلاقی انسان ایفای نقش کند.

روش‌شناسی پژوهش: این پژوهش با رویکردی تحلیلی-توصیفی و با استفاده از روش کتابخانه‌ای به بررسی منابع و ادبیات موجود در زمینه هوش مصنوعی، اخلاق و قضاوت اخلاقی می‌پردازد.

یافته‌ها: هوش مصنوعی با استفاده از الگوریتم‌های پیچیده و یادگیری ماشین قادر به پردازش اطلاعات و تصمیم‌گیری در زمینه‌های مختلف است. به کارگیری هوش مصنوعی در

استناد به این مقاله: جعفری، رضا (۱۴۰۳). تأملی در باب هوش مصنوعی و ارتقای قضاوت اخلاقی؛ واکاوی چالش‌ها و فرصت‌ها. مطالعات دینی رسانه، ۶(۲۴): ۵۵-۷۸.

<https://doi.org/10.22034/jmrs.2024.210084>

تاریخ دریافت: ۱۴۰۳/۰۲/۱۷ تاریخ اصلاح: ۱۴۰۳/۰۴/۱۷ تاریخ پذیرش: ۱۴۰۳/۰۵/۰۸ تاریخ انتشار آنلاین: ۱۴۰۳/۱۰/۱۰

ناشر: دانشگاه صدا و سیما

نوع مقاله: پژوهشی

© نویسندگان.



قضاوت اخلاقی مزایایی مانند افزایش دقت، سرعت و بی طرفی را به همراه دارد؛ با این وجود، چالش‌های متعددی در این زمینه وجود دارد، از جمله فقدان درک ظرافت‌های اخلاقی، تعصب و فقدان خلاقیت و ابتکار.

بحث و نتیجه‌گیری: واگذاری کامل قضاوت اخلاقی به هوش مصنوعی در حال حاضر امری غیرقابل اعتماد است. هوش مصنوعی می‌تواند به‌عنوان ابزاری برای کمک به انسان در فرایند قضاوت اخلاقی مورد استفاده قرار گیرد، اما اتخاذ تصمیمات نهایی باید بر عهده انسان باشد. **کلیدواژه‌ها:** هوش مصنوعی، اخلاق، قضاوت اخلاقی، اخلاق کاربردی، استدلال اخلاقی، شناخت اخلاقی.

مقدمه

هوش مصنوعی به‌طور فزاینده‌ای در حال تبدیل شدن به یک بخش جدایی‌ناپذیر از زندگی روزمره ماست. هوش مصنوعی در حال حاضر در طیف گسترده‌ای از زمینه‌ها از تشخیص چهره و ترجمه زبان گرفته تا رانندگی خودکار و تشخیص بیماری کاربرد دارد.

با افزایش نقش هوش مصنوعی در زندگی بشر، اهمیت بررسی مسائل اخلاقی مرتبط با این فناوری نیز افزایش می‌یابد. یکی از مهم‌ترین چالش‌های پیش‌روی هوش مصنوعی، قضاوت اخلاقی است.

قضاوت اخلاقی فرایندی پیچیده و چندوجهی است که در آن فرد با اتکا به معیارهای اخلاقی، ارزشی یا دینی، درستی یا نادرستی یک عمل یا تصمیم را تعیین می‌کند. این فرایند نقشی حیاتی در زندگی فردی و اجتماعی ایفا می‌کند و به ما در اتخاذ تصمیمات درست و اخلاقی یاری می‌رساند.

یکی از زمینه‌های نوظهور کاربرد هوش مصنوعی، کمک به حل چالش‌های اخلاقی و قضاوت اخلاقی در مورد اعمال و رفتارها بر اساس اصول اخلاقی است. با این حال، واگذاری قضاوت اخلاقی به هوش مصنوعی، چالش‌های متعددی را نیز به وجود می‌آورد.

چالش‌های پیش‌روی قضاوت اخلاقی هوش مصنوعی اغلب حول محور ارکان قضاوت

اخلاقی می‌چرخند. وجود این ارکان و شرایط قضاوت در هوش مصنوعی محل تردید است و این سؤال را مطرح می‌کند که آیا می‌توان به قضاوت اخلاقی ماشین‌ها اعتماد کرد؟ این پژوهش به بررسی عمیق و موشکافانه مسائل مربوط به هوش مصنوعی و قضاوت اخلاقی می‌پردازد و با اتخاذ رویکردی تحلیلی-توصیفی و بهره‌مندی از روش کتابخانه‌ای، به کاوش در منابع و ادبیات موجود در این حوزه پرداخته می‌شود. در گام نخست به تبیین مفهوم قضاوت اخلاقی و مؤلفه‌های بنیادین آن می‌پردازیم. سپس چگونگی عملکرد هوش مصنوعی و سامانه‌های هوشمند در زمینه قضاوت اخلاقی مورد واکاوی قرار می‌گیرد. در ادامه، چالش‌های پیش‌روی اتکا به هوش مصنوعی در قضاوت‌های اخلاقی مورد مذاقه قرار گرفته و میزان اعتمادپذیری آن در این عرصه سنجیده می‌شود.

هدف نهایی این مقاله، ارائه پاسخی روشن به این پرسش است که آیا می‌توان مسئولیت قضاوت اخلاقی را به هوش مصنوعی واگذار کرد یا اینکه هوش مصنوعی تنها ابزاری برای کمک به انسان در این فرایند پیچیده خواهد بود؟

پیشینه پژوهش

هوش مصنوعی با پیشرفت‌های چشمگیر خود، در کنار مزایای فراوان، چالش‌های اخلاقی متعددی را نیز به وجود آورده است. از همان ابتدای ظهور ایده هوش مصنوعی دغدغه‌های اخلاقی در این حوزه توجه پژوهشگران را به خود جلب کرده است و تحقیقات بسیاری به بررسی این چالش‌ها معطوف شده‌اند.

بررسی‌ها نشان می‌دهد که تحقیقات غربی در این حوزه پیشروتر از پژوهش‌های داخلی بوده است. با این حال، با توجه به شتاب فزاینده پیشرفت سیستم‌های هوش مصنوعی و ظهور کاربردهای جدید برای آن‌ها، نیاز به بازنگری و به‌روزرسانی این مباحث و همچنین توجه به زمینه‌های نوظهور در این حوزه امری ضروری است.

یکی از زمینه‌های نوظهور که تاکنون به‌طور مستقل مورد توجه قرار نگرفته، نقش هوش مصنوعی در قضاوت اخلاقی است. اگرچه برخی تحقیقات به رابطه هوش مصنوعی و اخلاق

پرداخته‌اند، اما به‌طور خاص بر نقش هوش مصنوعی در فرایند قضاوت اخلاقی تمرکز نداشته‌اند. برخی از این تحقیقات مرتبط عبارت‌اند از هوش مصنوعی و اخلاق از کتاب کارآفرینی سلامت دیجیتال (روزن استراخ^۱ و دیگران، ۲۰۲۳)، هوش مصنوعی و چالش‌های اخلاقی فناوری‌ها (مارینو^۲ و کانانزی^۳) و مقاله «مروری نظام‌مند بر دلالت‌های اخلاقی استفاده از هوش مصنوعی در فناوری‌های دیجیتال و نسبت آن با اخلاق شکوفایی» (بلبلی و پارسانیا، ۱۴۰۲). در حوزه‌های خاص‌تر نیز پژوهش‌هایی که به موضوع چالش‌های به‌کارگیری هوش مصنوعی در برخی زمینه‌ها مانند قضاوت حقوقی یا مراقبت بهداشتی می‌پردازد را نیز می‌توان پژوهش‌های به این پژوهش نزدیک دانست، مانند مقاله «مشروعیت به‌کارگیری هوش مصنوعی در قضاوت» (ایزدی‌فرد و دیگران، ۱۴۰۲) و مراقبت‌های بهداشتی مقاله «اخلاق سیستم‌های هوش مصنوعی در مراقبت‌های بهداشتی و پزشکی» (واندمولبروک، ۲۰۲۴^۴).

با این وجود هیچ پژوهش مستقل به‌طور خاص بر موضوع قضاوت اخلاقی در هوش مصنوعی تمرکز نکرده است. با توجه به اهمیت فراینده هوش مصنوعی در تصمیم‌گیری‌ها و قضاوت‌های مختلف، این خلأ در پژوهش، مسئله محوری این تحقیق را تشکیل می‌دهد.

چیستی و جایگاه قضاوت اخلاقی

توافق‌زبانی صرف بر سر اصول اخلاقی، به‌مثابه نوشتن کلماتی زیبا بر روی کاغذی بی‌جان است که نه اهداف اخلاقی را محقق می‌کند و نه جهان را به جای بهتری بدل می‌کند (الدر؛ پل، ۱۳۹۷: ۱۲). اصول اخلاقی فقط زمانی معنا و مفهوم واقعی پیدا می‌کنند که در رفتار و کردار انسان‌ها متجلی شوند. در واقع این تعهد به عمل است که به اصول اخلاقی جان می‌بخشد و آن‌ها را از قلمرو انتزاع به دنیای واقعی می‌آورد (هینمن، ۱۴۰۰: ۷۲۳).

قضاوت اخلاقی، به‌عنوان پیش‌نیاز عمل اخلاقی، خود فرایندی پیچیده و گاه دشوار

1. Rosenstrauch
2. Marino
3. Cananzi
4. Vandemeulebroucke

است. ما همیشه با قضایای روشن اخلاقی سروکار نداریم. در بسیاری از موارد، مسائل اخلاقی با ابهام و پیچیدگی همراه هستند و پاسخ‌های قطعی و یگانه‌ای برای آن‌ها وجود ندارد. درگیر شدن در مسائل انتزاعی نظریه‌های اخلاق، می‌تواند ما را از ضرورت و دشواری مسائل اخلاقی زندگی روزمره غافل کند، درحالی‌که در دنیای واقعی، انسان‌ها در موقعیت‌های مختلف با چالش‌های اخلاقی متعددی روبه‌رو می‌شوند که قضاوت در مورد آن‌ها ساده نیست. به‌عنوان مثال در نظر بگیریید که یک پزشک با چه معضلات اخلاقی در انتخاب بین نجات جان دو بیمار با شرایط متفاوت مواجه است.

قضاوت اخلاقی عنصری محوری در اخلاق کاربردی است. اخلاق نظری به مبانی فلسفی و تئوریک اخلاق می‌پردازد، درحالی‌که اخلاق کاربردی به دنبال یافتن راه‌حل‌های عملی برای مسائل اخلاقی در دنیای واقعی است. قضاوت اخلاقی به‌عنوان پلی بین این دو حوزه عمل می‌کند و به ما کمک می‌کند تا اصول اخلاقی انتزاعی را در موقعیت‌های خاص به‌کار ببریم (شریفی، ۱۳۹۶: ۱۰۸/۱).

در دنیای پیچیده امروز، افراد به‌طور مداوم با انتخاب‌های اخلاقی در زمینه‌های مختلف زندگی مانند حرفه، روابط و محیط‌زیست روبه‌رو هستند. همچنین جوامع بشری با مسائل اخلاقی پیچیده‌ای مانند سقط جنین، مرگ با عزت و هوش مصنوعی مواجه هستند. قضاوت اخلاقی ابزاری برای بحث و گفت‌وگو در مورد این مسائل و یافتن راه‌حل‌های مناسب است. علاوه بر این، قضاوت اخلاقی به افراد کمک می‌کند تا وجدان اخلاقی خود را تقویت کنند و فضایل اخلاقی مانند عدالت، شجاعت و مهربانی را در خود پرورش دهند.

قضاوت اخلاقی عنصری ضروری برای زندگی اخلاقی در دنیای واقعی است. این توانایی به افراد کمک می‌کند تا تصمیمات درست بگیرند، مسائل اخلاقی را حل کنند و به ارتقای فضایل اخلاقی و پیشرفت جوامع خود کمک کنند.

مؤلفه‌های لازم در قضاوت اخلاقی

قضاوت اخلاقی فرایندی پیچیده و چندوجهی است که در آن، فرد با اتکا به معیارهای



اخلاقی، ارزشی یا دینی، درستی یا نادرستی یک عمل یا تصمیم را تعیین می‌کند. این فرایند نقشی حیاتی در زندگی فردی و اجتماعی ایفا می‌کند و به ما در اتخاذ تصمیمات درست و اخلاقی یاری می‌رساند.

۱. شناخت اخلاقی

شناخت اخلاقی، به‌عنوان اولین قدم در قضاوت اخلاقی، نقشی کلیدی در این فرایند ایفا می‌کند. تا زمانی که فرد موضوعی را به‌عنوان یک مسئله اخلاقی درک نکند، فرایند قضاوت اخلاقی آغاز نخواهد شد (دین؛ پرات، ۲۰۰۷).

تشخیص اینکه یک موقعیت یا عمل، دارای ابعاد اخلاقی است یا خیر، گامی اساسی در فرایند قضاوت اخلاقی به‌شمار می‌آید. این تشخیص، تحت‌تأثیر عوامل متعددی از جمله ارزش‌ها و باورهای فرد قرار می‌گیرد. این ارزش‌ها، تفسیر افراد از موقعیت‌ها را شکل می‌دهند و به آن‌ها در تشخیص مسائل اخلاقی کمک می‌کنند. هنجارهای اجتماعی حاکم بر جامعه نیز نقشی کلیدی در این زمینه ایفا می‌کنند و به افراد در تشخیص هنجارهای اخلاقی یاری می‌رسانند. تجربیات زندگی و آموزش اخلاقی نیز از دیگر عوامل مؤثر در شناخت موضوع اخلاقی هستند. تجربیات شخصی، آگاهی افراد را نسبت به مسائل اخلاقی افزایش می‌دهد و آموزش اخلاقی به آن‌ها به درک مفاهیم اخلاقی و تشخیص این مسائل کمک می‌کند (جیمز، ۲۰۱۹: ۳۷ - ۵۴).

با این وجود موانعی نیز می‌توانند در مسیر شناخت موضوع اخلاقی وجود داشته باشند. جهل به مفاهیم اخلاقی، تشخیص مسائل اخلاقی را دشوار می‌کند. تعصبات شخصی نیز می‌تواند مانع از درک صحیح موقعیت‌ها و اعمال به‌عنوان مسائل اخلاقی شود. خودخواهی و تمرکز بر منافع شخصی نیز می‌تواند مانع از توجه به ابعاد اخلاقی یک موقعیت شود.

۲. استدلال اخلاقی

پس از شناخت موضوع اخلاقی، گام بعدی در فرایند قضاوت اخلاقی، انجام استدلال اخلاقی است.

دو رویکرد متضاد و افراطی در حوزه تصمیم‌گیری اخلاقی وجود دارد. دیدگاه اول، موسوم به اصول‌گرایی، بر آگاهی از اصول اخلاقی به‌عنوان عنصری ضروری در تصمیم‌گیری تأکید می‌کند. این رویکرد، شیوه قیاسی استدلال اخلاقی را اتخاذ می‌کند و معتقد است که اصول اخلاقی، مبنای قضاوت‌های درست و عادلانه هستند. در مقابل، دیدگاه دوم اتکا به اصول را محل اشکال می‌داند و نقش جزئیات و شهودات را در تصمیم‌گیری اخلاقی اساسی می‌داند. این دیدگاه، رویکردی مبتنی بر تجربه و شهود را در پیش می‌گیرد و معتقد است که قضاوت‌های اخلاقی باید بر اساس درک عمیق موقعیت و شرایط خاص هر موقعیت اتخاذ شوند (داودآبادی و دیگران، ۱۳۸۹: ۵۲).

هیچ‌کدام از این دو رویکرد به‌تنهایی قادر به پاسخگویی به پیچیدگی‌های قضاوت اخلاقی نیستند. در عمل، افراد در قضاوت‌های اخلاقی خود از ترکیبی از اصول، حقایق و شهودات استفاده می‌کنند. اتکا به اصول اخلاقی، چارچوبی کلی برای تصمیم‌گیری ارائه می‌دهد، درحالی‌که توجه به جزئیات و شهودات، درک عمیق‌تر از موقعیت را فراهم می‌کند. بنابراین برای اتخاذ قضاوت‌های اخلاقی درست و عادلانه، باید از تعادلی بین این دو رویکرد استفاده کرد. پس استدلال اخلاقی، فرایندی پویا و چندوجهی تلقی می‌شود که در آن فرد با اتکا به معیارهای اخلاقی، ارزشی یا دینی، به تعیین درستی یا نادرستی یک عمل یا تصمیم می‌پردازد. این فرایند نقشی حیاتی در زندگی فردی و اجتماعی ایفا می‌کند و یاریگر انسان در اتخاذ تصمیمات درست و اخلاقی خواهد بود.

عوامل دخیل در استدلال اخلاقی

عوامل متعددی در فرایند استدلال اخلاقی نقش‌آفرینی می‌کنند. این عوامل را می‌توان به دو دسته کلی تسهیل‌کننده و بازدارنده تقسیم کرد:

۱. عوامل تسهیل‌کننده

الف) دانش اخلاقی: آشنایی عمیق و دقیق با مبانی نظری استدلال اخلاقی، شامل

نظریه‌های اخلاقی (فایده‌گرایی، وظیفه‌گرایی، فضیلت‌گرایی) و قواعد و هنجارهای اخلاقی رایج در جامعه، به تحلیل دقیق‌تر مسائل اخلاقی کمک می‌کند. این دانش به‌مثابه ابزاری قدرتمند، فرد را در واکاوی شفاف و بی‌طرفانه مسائل اخلاقی یاری می‌رساند.

(ب) مهارت‌های تفکر انتقادی: توانایی تحلیل مسائل از زوایای مختلف، شامل تفکر منطقی، استدلال دقیق، و ارزیابی بی‌طرفانه شواهد و اطلاعات، به انتخاب گزینه‌ای درست و عادلانه کمک می‌کند. این مهارت‌ها، قدرت تشخیص و تمایز بین گزینه‌های اخلاقی مختلف را به فرد می‌دهد و او را در انتخاب گزینه‌ای عادلانه و منصفانه یاری می‌رساند (بومر،^۱ ۱۹۹۰).

(ج) همدلی: درک عمیق و همدلانه احساسات و دیدگاه‌های دیگران، به انتخاب گزینه‌ای که به نفع همه باشد و حقوق و منافع همه افراد را در نظر بگیرد، کمک می‌کند. همدلی به‌مثابه پلی ارتباطی، فرد را قادر می‌سازد تا خود را جای دیگران بگذارد و پیامدهای اعمال خود را بر آن‌ها درک کند. این امر زمینه انتخاب گزینه‌ای را فراهم می‌کند که منافع و حقوق همه افراد در آن لحاظ شده باشد (لاریجانی؛ نیکو، ۱۴۰۱).

۲. عوامل بازدارنده

(الف) تعصبات شخصی: تعصبات، دیدگاه فرد را نسبت به مسائل اخلاقی محدود می‌کند و مانع از درک صحیح موقعیت و انتخاب گزینه‌ای درست و عادلانه می‌شود (جیمز، ۲۰۱۹: ۵۵). تعصبات، همچون غباری بر دیدگاه فرد، مانع از واکاوی شفاف و بی‌طرفانه مسائل اخلاقی می‌شود.

(ب) احساسات: احساسات قوی مانند خشم، ترس و اضطراب می‌توانند قضاوت اخلاقی فرد را تحت تأثیر قرار داده و او را به انتخاب گزینه‌ای نادرست بر اساس احساسات آنی، و نه عقل و منطق سوق دهد. برای مثال، ترس از طرد شدن توسط گروهی خاص، ممکن است فرد را به انتخاب گزینه‌ای وادار کند که با ارزش‌ها و باورهای او مغایرت دارد (همان: ۶۱).

ج) فشارهای اجتماعی: فشارهای اجتماعی از سوی خانواده، دوستان یا جامعه می‌توانند فرد را به انتخاب گزینه‌ای نادرست و مغایر با ارزش‌ها و باورهای خود وادار کنند. این فشارها می‌توانند به صورت مستقیم یا غیرمستقیم اعمال شوند و فرد را در موقعیتی دشوار قرار دهند (همان: ۶۹).

انواع استدلال اخلاقی

۱. استدلال مبتنی بر وظیفه

این نوع استدلال بر قوانین و مقررات اخلاقی تمرکز دارد. در این روش، فرد با بررسی قوانین و مقررات اخلاقی، به دنبال تعیین درستی یا غلطی عمل خود است (راس، ۱۹۳۰: ۳۵). به عنوان مثال، فرد ممکن است با اتکا به قانون «دروغ گفتن ممنوع است» به این نتیجه برسد که دروغ گفتن در هر موقعیت، غلط است.

۲. استدلال مبتنی بر پیامد

این نوع استدلال بر پیامدهای یک عمل اخلاقی تمرکز دارد. در این روش فرد با بررسی پیامدهای احتمالی عمل خود، به دنبال انتخاب گزینه‌ای است که بیشترین خیر و کمترین ضرر را به همراه داشته باشد. به عنوان مثال، فرد ممکن است با در نظر گرفتن پیامدهای احتمالی دروغ گفتن، به این نتیجه برسد که در یک موقعیت خاص، دروغ گفتن مجاز است (جیمز، ۲۰۱۹: ۱۰۷).

۳. استدلال مبتنی بر فضیلت

این نوع استدلال بر ویژگی‌های اخلاقی فرد تمرکز دارد. در این روش فرد با بررسی ویژگی‌های اخلاقی خود و دیگران، به دنبال انتخاب گزینه‌ای است که با فضایل اخلاقی همخوانی داشته باشد (ریچلز، ۱۳۹۶: ۲۸۳). به عنوان مثال، فرد ممکن است با اتکا به فضیلت «صداقت» به این نتیجه برسد که در یک موقعیت خاص، باید راست بگوید، حتی

اگر این کار پیامدهای منفی برای خودش داشته باشد، اما چون فضیلت و ویژگی برون‌نگری است، اگر دروغ برای دیگران منفعت حقیقی داشته باشد و آتشی را خاموش کند، پس با فضیلت برخوردی ندارد.

احساسات اخلاقی

علاوه بر شناخت موضوع اخلاقی و تسلط بر مهارت استدلال اخلاقی، احساسات اخلاقی نیز نقشی اساسی در قضاوت اخلاقی ایفا می‌کنند. این احساسات، واکنش‌های عاطفی به موقعیت‌ها و اعمال اخلاقی هستند که می‌توانند طیف وسیعی از احساسات مثبت مانند شادی، رضایت، و افتخار تا احساسات منفی مانند خشم، غم، و شرم را شامل شوند (پرینز^۱ و نیکولز^۲، ۲۰۱۰).

احساسات اخلاقی، نقشی حیاتی در هدایت مسیر تفکر ما در مورد مسائل اخلاقی ایفا می‌کنند (مرنیگان^۳ و وانگ^۴، ۲۰۱۸). به عنوان مثال، احساس خشم نسبت به یک عمل ظالمانه، می‌تواند توجه ما را به پیامدهای آن عمل جلب کند و انگیزه ما را برای مقابله با آن افزایش دهد. به طور مشابه، احساس همدلی با فردی که در حال رنج است، می‌تواند ما را به یاری رساندن به او ترغیب کند.

نظریه‌های مختلفی در مورد منشأ و ماهیت احساسات اخلاقی وجود دارد. برخی از نظریه‌ها، مانند «نظریه شناختی-تکوینی» بر نقش یادگیری و تجربه در شکل‌گیری احساسات اخلاقی تأکید دارند. این نظریه‌ها معتقدند که احساسات اخلاقی در طول زندگی فرد و در تعامل با محیط اجتماعی شکل می‌گیرند (کولبرگ^۵، ۱۳۴۸: ۳۴-۷۱).

نظریه‌های دیگر مانند «نظریه تکاملی» معتقدند که احساسات اخلاقی ریشه تکاملی

1. Prinz
2. Nichols
3. Murnighan
4. Wang
5. Kohlberg

دارند و به بقا و پیشرفت انسان کمک می‌کنند. این نظریه‌ها بر این باورند که احساسات اخلاقی، سازگاری‌هایی هستند که در طول فرایند تکامل به وجود آمده‌اند و به انسان‌ها کمک می‌کنند تا در جوامع پیچیده همکاری کنند و به‌طور مؤثر با یکدیگر تعامل داشته باشند (گرین،^۱ ۱۳۸۳).

احساسات اخلاقی و استدلال اخلاقی به‌طور جدایی‌ناپذیری با یکدیگر مرتبط هستند. احساسات اخلاقی می‌توانند به‌عنوان منبع اطلاعاتی برای استدلال اخلاقی عمل کنند و به ما در درک و ارزیابی مسائل اخلاقی کمک کنند. به‌طور مشابه، استدلال اخلاقی می‌تواند به ما در درک و تنظیم احساسات اخلاقی خود کمک کند.

شخصیت اخلاقی

«شخصیت اخلاقی» به‌عنوان یک مفهوم چندوجهی و پیچیده، نقشی اساسی در قضاوت و عمل اخلاقی ایفا می‌کند (ویتمر،^۲ ۱۹۹۲: ۴۳). این مفهوم مجموعه‌ای از ویژگی‌های شخصیتی را شامل می‌شود که بر رفتار اخلاقی فرد، شیوه استدلال او در مورد مسائل اخلاقی، و درک او از موقعیت‌های اخلاقی اثر می‌گذارد (بومر^۳ و دیگران، ۱۹۸۷: ۲۶۵-۲۶۹).

شخصیت اخلاقی نه‌تنها بر نحوه استدلال و انگیزه عمل اخلاقی فرد، بلکه بر درک موقعیت‌های اخلاقی نیز اثرگذار است. (ویتمر، ۱۹۹۲). فردی با شخصیت اخلاقی قوی، مسائل اخلاقی را با دقت و ظرافت بیشتری تجزیه و تحلیل می‌کند، به ابعاد مختلف آن توجه می‌نماید و با در نظر گرفتن عواقب احتمالی هر انتخاب، به گزینه‌ای عادلانه و منصفانه دست پیدا می‌کند.

همچنین شخصیت اخلاقی می‌تواند مقاومت فرد در برابر فشارهای اجتماعی برای عمل

1. Greene
2. Wittmer
3. Bommer

غیراخلاقی را افزایش دهد. به عنوان مثال، فردی که دارای شخصیت اخلاقی قوی است، احتمال کمتری دارد که به خاطر ترس از طرد شدن یا مورد انتقاد قرار گرفتن، از ارزش‌های اخلاقی خود دست بکشد و در برابر هنجارهای غلط اجتماعی تسلیم شود.

شخصیت اخلاقی با هوش اخلاقی که به معنای توانایی درک و حل مسائل اخلاقی است، ارتباط تنگاتنگی دارد. شخصیت اخلاقی قوی، زمینه‌ای را برای ارتقای هوش اخلاقی فراهم می‌کند و به فرد کمک می‌کند تا در موقعیت‌های چالش‌برانگیز، تصمیمات درست و اخلاقی اتخاذ کند.

عملکرد هوش مصنوعی و ماشین‌ها

هوش مصنوعی به عنوان یکی از تحولات پیش‌گام عصر حاضر، به سرعت در حال دگرگونی ابعاد مختلف زندگی بشر است. هوش مصنوعی با ارائه راه‌حل‌های نوین و ارتقای کارایی در زمینه‌های گوناگون از حوزه‌های تخصصی و صنعتی گرفته تا زندگی روزمره نقشی حیاتی ایفا می‌کند.

۱. اصول کلی عملکرد هوش مصنوعی

الف) یادگیری: هوش مصنوعی از طریق الگوریتم‌های یادگیری ماشینی، یادگیری عمیق و شبکه‌های عصبی مصنوعی، قادر به استخراج دانش از داده‌ها و شناسایی الگوها است. این امر به هوش مصنوعی امکان می‌دهد تا بدون برنامه‌ریزی صریح، وظایف مختلف را یاد گرفته و به صورت خودکار انجام دهد.

ب) استدلال: هوش مصنوعی با تجزیه و تحلیل داده‌ها و استنتاج منطقی، قادر به پیش‌بینی و تصمیم‌گیری است. این توانایی، هوش مصنوعی را به ابزاری قدرتمند برای حل مسائل پیچیده در حوزه‌های مختلف تبدیل می‌کند.

ج) حل مسئله: هوش مصنوعی با شناسایی و درک مسائل، ارائه راه‌حل‌ها و انتخاب بهترین گزینه، به انسان در حل مسائل یاری می‌رساند.

۲. انواع هوش مصنوعی

الف) هوش مصنوعی محدود (Narrow AI): این نوع هوش مصنوعی برای انجام یک وظیفه خاص طراحی شده است. به عنوان مثال، سیستم‌های تشخیص چهره، ربات‌های صنعتی و موتورهای جستجو نمونه‌هایی از هوش مصنوعی محدود هستند.

ب) هوش مصنوعی عمومی (Artificial General Intelligence - AGI): این نوع هوش مصنوعی قادر به انجام طیف گسترده‌ای از وظایفی است که انسان می‌تواند انجام دهد. AGI هنوز در مراحل توسعه اولیه قرار دارد و دستیابی به آن به‌عنوان یکی از بزرگترین چالش‌های علمی شناخته می‌شود.

ج) هوش مصنوعی فوق‌العاده (Superintelligence): این نوع هوش مصنوعی از هوش انسان فراتر می‌رود و در حال حاضر یک مفهوم تخیلی است.

هوش مصنوعی در حال حاضر در طیف گسترده‌ای از زمینه‌ها، از جمله پزشکی (تشخیص بیماری‌ها، تجویز دارو، توسعه درمان‌های جدید، جراحی رباتیک)، مهندسی (طراحی و ساخت محصولات، بهینه‌سازی فرایندهای تولید، پیش‌بینی خرابی تجهیزات)، امور مالی (تشخیص تقلب، مدیریت ریسک، پیش‌بینی بازارهای مالی)، حمل‌ونقل (توسعه خودروهای خودران، بهینه‌سازی سیستم‌های حمل‌ونقل، مدیریت ترافیک) و دانش‌های نظری (اخلاق، فلسفه، حقوق) کاربرد دارد.

در حوزه اخلاق، هوش مصنوعی می‌تواند برای شناسایی و حل مسائل اخلاقی مانند تبعیض، بی‌عدالتی و نقض حریم خصوصی، توسعه چارچوب‌های اخلاقی برای استفاده از هوش مصنوعی و ارائه آموزش اخلاقی به توسعه‌دهندگان و کاربران هوش مصنوعی استفاده شود.

چالش‌های قضاوت اخلاقی هوش مصنوعی

در این بخش به بررسی چالش‌های قضاوت اخلاقی هوش مصنوعی می‌پردازیم. هدف از این بررسی ارائه دیدگاهی جامع در مورد این موضوع مهم و بیان راهکارهایی برای رفع این چالش‌ها است.

۱. فقدان درک ظرافت‌های اخلاقی

هوش مصنوعی به‌عنوان یک سیستم مبتنی بر داده، قادر به یادگیری الگوها و استنتاج از داده‌ها است. با این حال، این سیستم‌ها در درک مفاهیم پیچیده‌ای مانند عدالت، انصاف و شفقت که در قضاوت اخلاقی نقشی اساسی دارند، ناتوان هستند. این نقصان می‌تواند منجر به تصمیم‌گیری‌های نادرست یا غیر منصفانه توسط هوش مصنوعی شود. به عنوان مثال، هوش مصنوعی ممکن است قادر به تشخیص موارد احتمالی تبعیض در استخدام باشد، اما نتواند ظرافت‌های تبعیض را درک کند و در نهایت یک فرد را به دلیل نژاد یا جنسیت او رد صلاحیت کند.

عواقب منفی فقدان درک ظرافت‌های اخلاقی در هوش مصنوعی می‌تواند گسترده و مخرب باشد. این عواقب شامل تبعیض، ازدست‌رفتن اعتماد، آسیب به افراد و جوامع، تهدید حقوق بشر، افزایش نابرابری، ازدست‌رفتن شغل و خطرات امنیتی می‌شود.

برای رفع چالش فقدان درک ظرافت‌های اخلاقی در هوش مصنوعی لازم است که توانایی هوش مصنوعی در درک مفاهیم اخلاقی ارتقا یابد. این امر باید با آموزش هوش مصنوعی با داده‌های اخلاقی متنوع و باکیفیت، توسعه الگوریتم‌های اخلاقی، و همکاری بین متخصصان هوش مصنوعی و متخصصان اخلاق انجام شود.

علاوه بر این، تدوین اصول اخلاقی برای توسعه و استفاده از هوش مصنوعی، و افزایش نظارت و کنترل بر استفاده از هوش مصنوعی، از دیگر راهکارهای مقابله با چالش‌های قضاوت اخلاقی هوش مصنوعی است. اصول اخلاقی باید بر اساس احترام به حقوق بشر، عدالت، انصاف، و شفقت تدوین شوند. نظارت و کنترل بر استفاده از هوش مصنوعی نیز باید توسط نهادهای مستقل و متخصص انجام شود.

بنابراین با ارتقای درک هوش مصنوعی از مفاهیم اخلاقی، تدوین اصول اخلاقی، و افزایش نظارت و کنترل، می‌توان از هوش مصنوعی به‌طور مسئولانه و اخلاقی برای رفاه بشر استفاده کرد.

۲. فقدان خلاقیت و ابتکار

«خلاقیت» به معنای توانایی تولید ایده‌ها و مفاهیم جدید و ابتکار به معنای توانایی عمل مستقل و بدون نیاز به دستورالعمل‌های دقیق است.

هوش مصنوعی در حال حاضر قادر به انجام وظایف پیچیده طبق دستورالعمل‌ها است، اما در تولید ایده‌های جدید یا اقدام مستقل بدون دخالت انسان، عملکرد ضعیفی دارد. این نقصان می‌تواند مانع استفاده کامل از هوش مصنوعی برای حل مشکلات پیچیده و ارتقای رفاه بشر شود.

دلایل متعددی برای فقدان خلاقیت و ابتکار در هوش مصنوعی وجود دارد. یکی از این دلایل، فقدان درک عمیق هوش مصنوعی از جهان است. هوش مصنوعی برای یادگیری و تصمیم‌گیری به داده‌ها متکی است. با این حال، داده‌ها فقط بازتابی از دنیای واقعی هستند و خلاقیت و ابتکار را به‌طور کامل نشان نمی‌دهند. به عبارت دیگر، هوش مصنوعی فقط می‌تواند از آنچه که در داده‌ها آموزش دیده است، یاد بگیرد و فراتر از آن عمل کند.

دومین دلیل، الگوریتم‌های هوش مصنوعی هستند. الگوریتم‌های هوش مصنوعی برای انجام وظایف خاص طراحی شده‌اند و توانایی تعمیم به وظایف جدید را ندارند. این امر مانع از خلاقیت و ابتکار هوش مصنوعی می‌شود. هوش مصنوعی با الگوریتم‌های فعلی خود قادر به تفکر خارج از چارچوب و ارائه ایده‌های جدید نیست.

سومین دلیل، فقدان تعامل هوش مصنوعی با انسان است. انسان‌ها می‌توانند به هوش مصنوعی در درک بهتر جهان و توسعه خلاقیت و ابتکار کمک کنند. با این وجود، هوش مصنوعی در حال حاضر به‌طور کامل از تعامل با انسان برای یادگیری و رشد استفاده نمی‌کند. هوش مصنوعی از طریق تعامل بیشتر با انسان می‌تواند از تجربیات و خلاقیت انسان‌ها یاد بگیرد و ایده‌های جدیدی را پرورش دهد.

فقدان خلاقیت و ابتکار در هوش مصنوعی می‌تواند پیامدهای منفی متعددی به دنبال داشته باشد. نخست، کاربرد هوش مصنوعی محدود می‌شود. هوش مصنوعی بدون خلاقیت

و ابتکار قادر به حل مشکلات جدید و پیچیده نیست و در نتیجه کارایی آن در بسیاری از زمینه‌ها کاهش می‌یابد. دومین ثمره، افزایش وابستگی به انسان است. هوش مصنوعی فاقد خلاقیت و ابتکار به‌طور کامل به انسان برای انجام وظایف متکی است و بدون دخالت انسان قادر به پیشرفت و یادگیری نیست. سومین و مهم‌ترین نتیجه فقدان خلاقیت، خطرات اخلاقی است. هوش مصنوعی بدون خلاقیت و ابتکار ممکن است قادر به درک ظرافت‌های اخلاقی و قضاوت اخلاقی صحیح نباشد و در نهایت در موقعیت‌های مختلف، تصمیمات غیراخلاقی و نادرستی اتخاذ کند.

بنابراین پرورش خلاقیت و ابتکار در هوش مصنوعی ضرورتی اجتناب‌ناپذیر است. هوش مصنوعی خلاق و مبتکر می‌تواند در حل مشکلات پیچیده، کاهش وابستگی به انسان و ارتقای سطح اخلاقی در تصمیم‌گیری‌ها نقشی کلیدی ایفا کند. آموزش هوش مصنوعی با داده‌های متنوع و توسعه الگوریتم‌های آن و افزایش تعامل هوش مصنوعی با انسان می‌تواند در رفع این نقصان مؤثر باشد. آموزش هوش مصنوعی با داده‌های متنوع و باکیفیت، به این سیستم‌ها کمک می‌کند تا درک عمیق‌تری از جهان پیدا کنند و از ظرافت‌ها و پیچیدگی‌های دنیای واقعی آگاه شوند. این امر به‌نوبه خود، خلاقیت و ابتکار هوش مصنوعی را در حل مسائل مختلف ارتقا می‌دهد و در نتیجه به استفاده کامل از هوش مصنوعی برای حل مشکلات پیچیده، ارتقای رفاه بشر، و دستیابی به اهداف توسعه پایدار کمک خواهد کرد.

۳. تعصب

یکی از مهم‌ترین چالش‌های قضاوت اخلاقی هوش مصنوعی، وجود تعصب در هوش مصنوعی است. «تعصب» به معنای قضاوت یا باور پیشینی است که بدون در نظر گرفتن شواهد و واقعیات شکل می‌گیرد. تعصبات به ندرت بیان می‌شوند و معمولاً در قالب اشکالی مبهم و ظاهرالصلاح به لحاظ اجتماعی معتبر و کاربردی نمود می‌یابند که باعث می‌شوند فرد متعصب بتواند حتی هنگامی که آشکارا حقوق دیگران را زیر پا می‌گذارد، بدون عذاب وجدان

به خواسته‌هایش برسد. تعصبات می‌توانند آگاهانه یا ناآگاهانه باشند و می‌توانند بر اساس نژاد، جنسیت، مذهب، گرایش جنسی، وضعیت اقتصادی-اجتماعی یا سایر عوامل باشند (لیندا الدر؛ ریچارد پل، ۱۳۹۷: ۷۶).

تعصبات می‌توانند به‌طور آگاهانه یا ناآگاهانه در هوش مصنوعی رخنه کنند و پیامدهای منفی متعددی به دنبال داشته باشند. تبعیض، ازدست‌رفتن اعتماد عمومی، و آسیب به افراد و جوامع، از جمله عواقب مخرب تعصبات در هوش مصنوعی هستند.

واژه دیگر نزدیک به تعصب «سوگیری» است که دارای دو معناست که در یک معنا خنثی است و در دیگری منفی. در حالت خنثی، فرد به دلیل دیدگاهش به برخی مسائل بیشتر توجه می‌کند، و بر برخی مسائل بیشتر تأکید می‌کند. این امر به‌خودی‌خود مورد انتقاد نیست. در معنای منفی، ما نقاط ضعف دیدگاه خود و نقاط قوت یا بینش‌های دیدگاه مخالف را نادیده می‌گیریم یا به‌صورت غیرمنطقی در برابر آن مقاومت می‌کنیم (همان: ۸۰).

دلایل متعددی برای وجود تعصب در هوش مصنوعی وجود دارد. اولین دلیل، تعصبات موجود در داده‌هایی است که هوش مصنوعی با آن‌ها آموزش دیده است. داده‌ها بازتابی از جامعه هستند و تعصبات موجود در جامعه را نیز منعکس می‌کنند.

الگوریتم‌های هوش مصنوعی نیز می‌توانند به‌گونه‌ای طراحی شوند که تعصبات موجود در داده‌ها را تشدید کنند. فقدان تنوع در تیم‌های توسعه هوش مصنوعی، سومین دلیل وجود تعصب در این سیستم‌ها است.

برای مقابله با چالش تعصب در هوش مصنوعی، لازم است اقداماتی در سطوح مختلف انجام شود. در سطح داده‌ها، جمع‌آوری و استفاده از داده‌های باکیفیت و بدون تعصب، گامی اساسی است. طراحی الگوریتم‌های هوش مصنوعی برای شناسایی و از بین بردن تعصبات، نقشی حیاتی در این زمینه دارد. افزایش تنوع در تیم‌های توسعه هوش مصنوعی نیز گامی مهم در راستای ریشه‌کن کردن تعصب از این سیستم‌ها است.

با انجام اقداماتی در سطوح مختلف می‌توان تعصب در هوش مصنوعی را کاهش داد و از هوش مصنوعی به‌طور عادلانه و منصفانه برای رفاه بشر استفاده کرد. این امر مستلزم تعهد و همکاری بین ذی‌نفعان مختلف، از جمله دولت‌ها، سازمان‌های بین‌المللی، جامعه مدنی، و متخصصان هوش مصنوعی است.

قابلیت‌های ویژه هوش مصنوعی در قضاوت اخلاقی

با وجود چالش‌های متعددی مانند فقدان درک ظرافت‌های اخلاقی، فقدان خلاقیت و ابتکار، تعصب و تبعیض، هوش مصنوعی می‌تواند یاریگر انسان در قضاوت اخلاقی باشد. هوش مصنوعی قادر به تجزیه و تحلیل حجم عظیمی از داده‌ها در زمان کوتاه است. این امر به انسان کمک می‌کند تا در مورد مسائل اخلاقی، اطلاعات و بینش‌های جدیدی به دست آورد. برای مثال، هوش مصنوعی می‌تواند با بررسی میلیون‌ها مقاله علمی، موارد مشابه مخصوصه و تراحمات اخلاقی، راه‌حلهایی را برای انسان ارائه دهد.

هوش مصنوعی با شناسایی الگوها و روندها در داده‌ها، می‌تواند به انسان در پیش‌بینی تراحمات اخلاقی احتمالی کمک کند. برای مثال، هوش مصنوعی می‌تواند با بررسی داده‌های مربوط به یک موقعیت اجتماعی، شرایطی را که احتمال وجود تراحمات اخلاقی در آن بیشتر است، شناسایی کند و به انسان هشدار دهد.

هوش مصنوعی می‌تواند با شبیه‌سازی تراحمات و مسائل اخلاقی مختلف، به انسان در درک بهتر پیچیدگی‌های مسائل اخلاقی و ارتقای مهارت‌های قضاوت اخلاقی خود کمک کند. این امر به انسان کمک می‌کند تا در هنگام مواجهه با تراحمات اخلاقی در دنیای واقعی، تصمیمات بهتر و عادلانه‌تری اتخاذ کند.

هوش مصنوعی می‌تواند با ارائه اطلاعات بی‌طرفانه و بدون تعصب، به انسان در تصمیم‌گیری اخلاقی عادلانه‌تر و منصفانه‌تر کمک کند. برای مثال، هوش مصنوعی می‌تواند با بررسی داده‌های مربوط به استخدام، موارد تبعیض را شناسایی کند و به انسان در استخدام عادلانه‌تر افراد کمک کند.

با استفاده مسئولانه از هوش مصنوعی می‌توان از این فناوری نوظهور برای ارتقای قضاوت اخلاقی و ساختن جامعه‌ای عادلانه‌تر و منصفانه‌تر استفاده کرد. این امر مستلزم همکاری بین متخصصان هوش مصنوعی، متخصصان اخلاق، و دیگر ذی‌نفعان برای توسعه و استفاده از هوش مصنوعی در راستای قضاوت اخلاقی است.

هوش مصنوعی پتانسیل بالایی برای ارتقای قضاوت اخلاقی انسان دارد. با این حال، استفاده از این ابزار نوظهور مستلزم احتیاط و توجه به نقش حیاتی انسان در قضاوت اخلاقی است. با ایجاد تعادل بین هوش مصنوعی و قوه تشخیص انسانی می‌توان از مزایای این فناوری نوظهور برای ارتقای اخلاق و ساختن جامعه‌ای عادلانه‌تر و منصفانه‌تر استفاده کرد.

بحث و نتیجه‌گیری

هوش مصنوعی به‌عنوان یک فناوری نوظهور، پتانسیل‌های فراوانی برای ارتقای قضاوت اخلاقی در زمینه‌های مختلف دارد. این فناوری می‌تواند با سرعت و دقت بالا، اطلاعات و داده‌های مربوط به مسائل اخلاقی را جمع‌آوری و تحلیل کند و به انسان در تصمیم‌گیری‌های اخلاقی یاری رساند. با این حال، هوش مصنوعی در مسیر ارتقای قضاوت اخلاقی با چالش‌های متعددی نیز روبه‌رو است.

یکی از مهم‌ترین چالش‌ها، ناتوانی هوش مصنوعی در درک ظرافت‌های اخلاقی و موقعیت‌های پیچیده اخلاقی است. هوش مصنوعی در حال حاضر قادر به درک کامل مفاهیم اخلاقی مانند عدالت، انصاف و شفقت نیست و ممکن است در قضاوت‌های خود دچار خطا شود.

چالش دیگر، فقدان خلاقیت و ابتکار در هوش مصنوعی است. این فناوری در حل مسائل اخلاقی جدید و نوظهور به کمک انسان نیاز دارد و قادر به ارائه راه‌حل‌های خلاقانه و ابتکاری نیست.

علاوه بر این، الگوریتم‌های هوش مصنوعی ممکن است متأثر از تعصبات موجود در

داده‌های آموزشی خود باشند و این موضوع، منجر به تبعیض در قضاوت‌های اخلاقی آن‌ها شود. برای مثال اگر هوش مصنوعی با داده‌هایی آموزش داده شود که در آن، زنان به‌طور کلی از مردان ضعیف‌تر تلقی شده باشند، ممکن است در قضاوت‌های اخلاقی خود، زنان را در موقعیت‌های فرودست قرار دهد.

با وجود چالش‌های متعددی که ذکر شد، هوش مصنوعی ابزاری قدرتمند برای ارتقای قضاوت اخلاقی است. هوش مصنوعی با تجزیه و تحلیل داده‌ها، اطلاعات و بینش‌های جدیدی در مورد مسائل اخلاقی به انسان ارائه می‌دهد. همچنین با شناسایی الگوها و شبیه‌سازی تزااحات اخلاقی، به انسان در درک بهتر پیچیدگی‌های مسائل اخلاقی و ارتقای مهارت‌های قضاوت اخلاقی خود کمک می‌کند. همچنین با ارائه اطلاعات بی‌طرفانه و بدون تعصب، به انسان در تصمیم‌گیری اخلاقی عادلانه‌تر و منصفانه‌تر کمک می‌کند.

با استفاده مسئولانه از هوش مصنوعی و همکاری متخصصان هوش مصنوعی، متخصصان اخلاق و دیگر ذی‌نفعان، می‌توان از این فناوری نوظهور برای ارتقای قضاوت اخلاقی و ساختن جامعه‌ای عادلانه‌تر و منصفانه‌تر استفاده کرد.

پیشنهاد‌های تحقیقاتی

با توجه به نتایج پژوهش، پیشنهاد می‌شود برخی رویکردهای پژوهشی کلان و خرد در زمینه هوش مصنوعی و ارتقای قضاوت اخلاقی مورد بررسی قرار گیرند. این امر شامل انجام مطالعات تطبیقی و تحلیلی عمیق در زمینه رویکردهای اخلاقی مختلف به هوش مصنوعی از جمله فلسفه اخلاق، نظریه عدالت، حقوق بشر و فقه است. همچنین شناسایی و تبیین اصول اخلاقی بنیادین که بر هوش مصنوعی حاکم هستند، مانند عدالت، انصاف، شفافیت، پاسخگویی، احترام به حریم خصوصی و عدم تبعیض، ضروری است.

برای تدوین چارچوب‌های اخلاقی جامع و کارآمد، مشارکت فعال متخصصان از

زمینه‌های مختلف اخلاق، فلسفه، حقوق، هوش مصنوعی، علوم انسانی و فناوری ضروری است. همچنین تدوین سازوکارهای لازم برای به‌روزرسانی و بازنگری مستمر چارچوب‌های اخلاقی به‌منظور تطابق آن با پیشرفت‌های فناوری هوش مصنوعی و تغییرات اجتماعی توصیه می‌شود.

منابع

۱. اردشیر لاریجانی، مریم؛ نیکو، سید جابر (۱۴۰۱). همدمی در اخلاق؛ بررسی تأثیر همدمی بر قضاوت اخلاقی. نقد و نظر، ۲۷ (۴): ۳۷-۵۸. DOI:10.22081/jpt.2022.63931.1945.58
۲. ایزدی فرد، علی اکبر؛ رجبی، فرشید؛ زارعی، محمدمهدی؛ جهانی، علی اکبر (۱۴۰۲). مشروعیت به کارگیری هوش مصنوعی در قضاوت. مطالعات فقه اقتصادی، ۵: ۳۷۳-۳۸۶. DOI: 10.22034/ejs.2023.412639.1521
۳. بلبل قادیکلایی، سمیه؛ پارسانی، حمید (۱۴۰۲). مروری نظام‌مند بر دلالت‌های اخلاقی استفاده از هوش مصنوعی در فناوری‌های دیجیتال و نسبت آن با اخلاق شکوفایی. راهبرد اجتماعی فرهنگی، ۱۲ (۳): ۷۷۱-۷۹۸. DOI: 10.22034/scs.2022.160772.798-771
۴. داودآبادی، مهدیه و دیگران (۱۳۸۹). جستارهایی در اخلاق کاربردی. قم: دانشگاه قم.
۵. ریچلز، جیمز (۱۳۹۶). عناصر فلسفه اخلاق. ترجمه محمود فتحعلی و علیرضا آل‌بویه. قم: پژوهشگاه علوم و فرهنگ اسلامی.
۶. شریفی، احمدحسین و دیگران (۱۳۹۶). دانشنامه اخلاق کاربردی. قم: موسسه آموزشی و پژوهشی امام خمینی.
۷. الدر، لیندا؛ پل، ریچارد (۱۳۹۷). آشنایی با مبانی استدلال اخلاقی بر اساس مفاهیم و اصول تفکر انتقادی. ترجمه پیام یزدانی. تهران: نشر اختران.
۸. هینمن، لارنس (۱۴۰۰). اخلاق: رویکردی کثرت‌گرایانه به نظریه اخلاق. تهران: نشر کرگدن.
9. Boomer, M. (1990). Critical Thinking Skills in Ethics. *Business Ethics Journal*, 9(2): 201-221.
10. Boomer, M., Gratto, C., Gravander, J., & Tuttle, M. (1987). A behavioral model of ethical and unethical decision making. *Journal of Business Ethics*, 6: 265-280.
11. Dane, E., & Pratt, M. G. (2007). Exploring Intuition and its Role in Managerial Decision Making. *Academy of Management Review*, 32(1): 33-54. DOI: 10.5465/amr.2007.23463682
12. Greene, J. (2004). The Nature and Origins of Moral Emotions. *Journal of Experimental Philosophy*, 15: 33-70.
13. Kohlberg, L. (1969). Stage and Sequence: The Cognitive-Developmental Approach to Moralization. In M. L. Hoffman (Ed.), *The Psychology of Moral Development* (pp. 34-71). New York: Holt, Rinehart and Winston.
14. Marino, D., Cananzi, D., & Aragona, F. (2024). Artificial Intelligence and Ethics Challenges of Technologies. In *Ethics and Artificial Intelligence. Studies in Systems, Decision and Control* (Vol. 519). Cham: Springer.

15. Murnighan, K., & Wang, L. (2018). Emotions, Intuition, and Reasoning in Moral Decision Making. *Academy of Management Annual Meeting Proceedings*,(1): 13413-13413. DOI:10.5465/AMBPP.2013.13413abstract
16. Prinz, J. J., & Nichols, S. (2010). Moral emotions. In J. M. Doris (Ed.), *The Moral Psychology Handbook* (pp. 111–146). Oxford: Oxford University Press.
17. Rachels, J. (2019). *Ethics: Theories and Issues* (12th ed.). New York: McGraw-Hill Education.
18. Rosenstrauch, D., Mangla, U., Gupta, A., & Masau, C.T. (2023). Artificial Intelligence and Ethics. In A. Meyers (Ed.), *Digital Health Entrepreneurship. Health Informatics*. Cham: Springer.
19. Ross, D. (1930). *The Right and the Good*. Oxford: Clarendon Press.
20. Vandemeulebroucke, T. (2024). The ethics of artificial intelligence systems in healthcare and medicine: from a local to a global perspective, and back. *Pflugers Arch - European Journal of Physiology*.
21. Wittmer, D. P. (1992). Ethical sensitivity and managerial decision-making: An experiment. *Public Administration - Dissertations*, 43. Retrieved from https://surface.syr.edu/ppa_etd/43 (accessed Oct. 9, 2022).