



Analysis of ethical challenges in the field of artificial intelligence with an approach to Islamic ethics

Sayedh Mahshid Miri Balajorshari*, Amirreza Mahmoudi**

Abstract

Today, one of the controversial topics is artificial intelligence (AI), which is broadly defined as the capability of a machine to perform intelligent operations such as reasoning, learning, communication, and other intelligent behaviors. The primary goal of research in the field of AI is to create systems that operate similarly to humans and possess characteristics such as awareness, thinking, and ethics. Therefore, it is necessary for AI systems that are capable of thinking and awareness to incorporate artificial ethics to ensure they do not harm any living beings, directly or indirectly. Otherwise, the lack of an ethical foundation for these systems can create problems for humans and other creatures. In Islam, ethics concerning human behavior and actions within the social system is of high importance. Thus, adhering to Islamic ethical principles in the design and use of AI can enhance and ensure the quality of intelligent systems' performance. The main question in the present research is: What ethical challenges exist in the field of artificial intelligence? Therefore, the aim of this paper is to explain the ethical challenges in the field of artificial intelligence with a focus on Islamic ethics.

Keywords

Artificial Intelligence, Natural Intelligence, Ethics, Awareness, Islamic Ethics.

*PhD student in Criminal Law and Criminology, Department of Law, Islamic Azad University, Lahijan Branch, Lahijan, Iran (Responsible Author). mahshid.miri.id@gmail.com

** Assistant Professor of Law, Department of Law, Islamic Azad University, Lahijan Branch, Lahijan, Iran. Amirreza.mahmodi@liau.ac.ir

واکاوی مسائل اخلاقی در زمینه هوش مصنوعی با نگاهی به اخلاق اسلامی

سیده مهشید میری بالاجورشری*، امیررضا محمودی**

چکیده

امروزه یکی از موضوعات بحث برانگیز، هوش مصنوعی است که به طور گسترده به عنوان قابلیت یک ماشین برای انجام عملیات هوشمندانه مانند استدلال، یادگیری، ارتباط و دیگر رفتارهای هوشمندانه تعریف می شود. هدف اصلی تحقیقات در حوزه هوش مصنوعی، تولید سیستم هایی است که به شکلی شبیه به انسان عمل کنند و دارای ویژگی هایی مانند آگاهی، تفکر و اخلاق باشند. بنابراین ضرورت دارد در سیستم های هوش مصنوعی که قادر به تفکر و آگاهی هستند، اخلاق مصنوعی وجود داشته باشد تا آن ها به هیچ کدام از موجودات زنده به طور مستقیم یا غیرمستقیم آسیب نرسانند. در غیر این صورت، نبود پایه اخلاقی برای این سیستم ها می تواند مشکلاتی را برای انسان ها و موجودات دیگر پدید آورد. در اسلام، اخلاق در زمینه رفتار و اعمال انسان ها در نظام جامعه از اهمیت زیادی برخوردار است. بنابراین، هنگام طراحی و استفاده از هوش مصنوعی، رعایت اصول اخلاقی اسلامی می تواند

* دانشجوی دکتری حقوق کیفری و جرم شناسی، گروه حقوق، دانشگاه آزاد اسلامی واحد لاهیجان، لاهیجان، ایران (نویسنده مسئول).

mahshid.miri.id@gmail.com

** استادیار گروه حقوق، دانشگاه آزاد اسلامی واحد لاهیجان، لاهیجان، ایران.

Amirreza.mahmodi@liau.ac.ir

تاریخ تأیید: ۱۴۰۳/۰۲/۲۲

تاریخ دریافت: ۱۴۰۳/۰۱/۰۹



بهبود و کیفیت کارکرد سیستم‌های هوشمند را تضمین کند. پرسش اصلی در پژوهش حاضر این است که در زمینه هوش مصنوعی چه معضلات اخلاقی وجود دارد؟ لذا هدف این مقاله تبیین مسائل اخلاقی در زمینه هوش مصنوعی با نگاهی به اخلاق اسلامی است.

کلیدواژه‌ها

هوش مصنوعی، هوش طبیعی، اخلاق، آگاهی، اخلاق اسلامی.

مقدمه

دهه ۱۹۵۰ میلادی و پس از آن، دورانی است که هوش مصنوعی توسعه و پیشرفت‌های مهمی را تجربه کرده است. از گذشته تا به امروز، در زمینه توسعه هوش مصنوعی، شاهد ارائه انواع نوآوری‌های علمی برای بهبود زندگی خود بوده‌ایم؛ از جمله این نوآوری‌ها می‌توان به بازوهای رباتیک که به جراحان در عملیات جراحی کمک می‌کنند، خودروهای بدون راننده، تلفن‌های هوشمند، محصولات فناوری و تغییرات جدید در هر زمینه‌ای اشاره کرد. این نوآوری‌ها زندگی ما را آسان‌تر کرده‌اند (Başkurt, 2018, p. 7).

دانشمندان به این موارد بسنده نکرده و همچنان برای تولید سیستم‌ها و ربات‌های شبیه انسان که بتوانند آگاهانه فکر کنند، می‌کوشند.

زمانی که ربات‌ها قابلیت تفکر پیدا کنند، این سؤال مهم پیش می‌آید که اگر چه ما به دلیل توانایی تفکر، موفق به غلبه بر ناتوانی‌های جسمی خود و برتری نسبی نسبت به موجودات زنده دیگر هستیم، اما اگر این ویژگی اساسی که برتری ما را اثبات می‌کند، روزی به ربات‌ها منتقل شود، آیا مجبور خواهیم شد از آن‌ها اطاعت کنیم؟ آیا این سیستم‌های هوش مصنوعی تولیدشده می‌توانند به اندازه مغز انسان به طور طبیعی، تفکر، احساس و تجربه داشته باشند؟ زیرا مغز انسان هیچ برنامه یا



نرم‌افزاری ندارد و بیشتر به یک شبکه عصبی پیچیده که به‌طور مداوم خود را ساختاردهی می‌کند، شبیه است (Kaku, 2016, p. 84).

به همین دلیل، هدف اصلی در زمینه هوش مصنوعی این است که سیستم‌های هوش مصنوعی را که از طریق برنامه‌نویسی کار می‌کنند، به سطح مغز انسان نزدیک کند. فناوری هوش مصنوعی باید به‌صورت کنترل‌شده و با در نظر گرفتن مسائل اخلاقی پیشرفت کند. اصول اخلاقی متعددی با رویکرد اسلامی وجود دارند که می‌توانند در طراحی هوش مصنوعی استفاده شوند. هنگام طراحی هوش مصنوعی و استفاده از آن، رعایت اصول اخلاقی اسلامی می‌تواند بهبود عملکرد سیستم‌های هوشمند و کاهش آسیب‌های احتمالی را به همراه داشته باشد.

در بررسی پیشینه پژوهش مشخص شد تاکنون مطالعه‌ای درباره بررسی معضلات اخلاقی در زمینه هوش مصنوعی با رویکردی بر اخلاق اسلامی انجام نشده است. در مقاله‌ای با عنوان «اهمیت اخلاق در استفاده از هوش مصنوعی در آموزش پزشکی» (صراف‌زاده و ابوطالب، ۱۴۰۲)، نویسندگان درصدد بررسی مسائل اخلاقی که هوش مصنوعی در حوزه آموزش پزشکی ایجاد می‌کند، بوده‌اند. این مقاله صرفاً حوزه آموزش پزشکی را در زمینه هوش مصنوعی بررسی کرده است. در حالی که در مقاله حاضر، به‌طور کلی مسائل اخلاقی در زمینه هوش مصنوعی را با رویکردی بر اخلاق اسلامی مورد تحلیل و بررسی قرار دادیم.

همچنین در مقاله‌ای با عنوان «معضلات و رویکردهای مسائل اخلاقی در هوش مصنوعی و ابرهوش» (رمضانی و فیضی درخشی، ۱۳۹۲)، نویسندگان به معرفی ابرهوش و ویژگی‌های آن و ضرورت بررسی مسائل اخلاقی در این ماشین‌ها پرداخته‌اند. این در حالی است که مقاله حاضر با نگاهی به اخلاق اسلامی، مسائل اخلاقی را در زمینه هوش مصنوعی مطالعه کرده است.

در مقاله‌ای دیگر با عنوان «مقایسه ماهوی هوش انسانی با هوش مصنوعی از منظر فلسفه اسلامی» (شریعتی و اکبرزاده، ۱۴۰۱)، نویسندگان، اطلاعاتی درباره تفاوت‌ها





و شباهت‌های موجود میان هوش مصنوعی و هوش انسان ارائه کرده‌اند. از آنجا که این مقاله به مسائل اخلاقی موجود در این زمینه توجه نداشته، مقاله حاضر از این وجه دارای نوآوری است.

هدف مقاله حاضر تبیین مسائل اخلاقی در زمینه هوش مصنوعی با نگاهی به اخلاق اسلامی است. در این پژوهش، نوع تحقیق، نظری و روش تحقیق، توصیفی تحلیلی می‌باشد که با ابزار فیش و با استفاده از منابع کتابخانه‌ای تنظیم شده است.

در مقاله حاضر ابتدا به تفاوت‌های هوش طبیعی و هوش مصنوعی پرداخته و با مقایسه تفصیلی، ارتباط بین آن‌ها بررسی شده است. سپس با توجه به این مسئله که آیا یک سیستم هوش مصنوعی می‌تواند همانند مغز انسان به آگاهی و تفکر دست یابد، با بررسی رابطه هوش مصنوعی و اخلاق بر اهمیت و دلایل نیاز به اخلاق مصنوعی تأکید و در نهایت، رویکرد اخلاق اسلامی در زمینه هوش مصنوعی بررسی می‌شود.

۱. تفاوت بین هوش مصنوعی و هوش طبیعی

هوش مصنوعی به سیستم‌هایی گفته می‌شود که توسط انسان، از طریق الگوریتم‌ها و برنامه‌های نحوی تولید شده‌اند و سعی دارند قابلیت‌هایی مانند فکر کردن، درک کردن، ارزیابی، تجزیه و تحلیل و کسب تجربه را به طور مشابه با انسان به دست آورند (صراف زاده و ابوطالب، ۱۴۰۲، ص ۲).

اصطلاح «هوش مصنوعی» را برای اولین بار جان مک‌کارتی در سال ۱۹۵۶ در کنفرانس دارتموث که کنفرانسی مربوط به هوش مصنوعی بود، به کار برد. همچنین برخی معتقدند هوش مصنوعی، فنون استفاده از برنامه‌های کامپیوتری برای درک اندیشه انسان است (Atrkinson & Boden, 1979, p. 130).



زمانی که مغز انسان ب رایانه‌ای مجهز به هوش مصنوعی مقایسه شود، آشکار می‌گردد که مغز انسان دارای سازکاری بسیار پیچیده و برتر است و بدون نیاز به برنامه‌ها خود را سازمان‌دهی می‌کند. مغز انسان از طریق تجربیات خود تکامل می‌یابد و از پویایی برخوردار است؛ اما سیستم رایانه‌ای و هوش مصنوعی ثابت هستند و تنها با الگوریتم‌ها و برنامه‌ها پیشرفت می‌کنند (Crick, 2005, p. 41-45)

در مفهوم هوش طبیعی، انسان موجودی است که می‌تواند رویدادها و مسائلی را که یاد گرفته و تجربه کرده است، به مرور زمان فراموش کند، مگر اینکه به‌طور مکرر با آن‌ها مواجه شود یا تجربه‌های مشابهی داشته باشد (قاسمی، ۱۴۰۲، ص ۱۴۱)؛ اما در مفهوم هوش مصنوعی، اطلاعاتی که در حافظه یک رایانه ذخیره شده است، تا زمانی که پاک نشوند، در حافظه باقی می‌مانند. بنابراین، انسان ممکن است برخی از مسائل را فراموش کند، اما کامپیوتر قابلیت فراموش کردن ندارد. تفاوت دیگر هوش طبیعی و هوش مصنوعی این است که انسان با هوش طبیعی خود قادر نیست دانش کلی خود را به دیگری منتقل کند؛ اما از طریق هوش مصنوعی، انتقال دانش جامع یک رایانه به رایانه دیگر به راحتی انجام می‌شود. افزایش سطح اطلاعات یک رایانه نسبت به افزایش سطح هوش یک انسان، آسان‌تر و ساده‌تر است. بنابراین، در حالی که اطلاعات و داده‌های هوش مصنوعی به آسانی دست‌یافتنی است، اطلاعات هوش طبیعی به سادگی قابل دستیابی نیست. گاهی دو انسان ممکن است به یک رویداد به شکل‌های مختلفی نظر داشته باشند که دلیل آن، تفاوت سطح هوش آنان است؛ اما در سیستم‌های هوش مصنوعی، واکنش به رویدادهای مشابه، یکسان است که دلیل آن، الگوریتم‌ها یا کدگذاری‌های یکسان، روی سیستم‌های هوش مصنوعی است (Adali, 2017, p. 9).

در مفهوم هوش طبیعی، انسان می‌تواند با ارزیابی تجربیات و رفتارهای خود و دیگران، هوش خود را بهبود بخشد، در حالی که هوش مصنوعی به اطلاعاتی محدود است که در رایانه بارگذاری می‌شود. بنابراین، می‌توان گفت هوش طبیعی نسبت به

هوش مصنوعی، از خلاقیت بیشتری برخوردار است (Penrose, 1997, p. 2). همچنین هوش طبیعی بر پایه تجربه استوار است؛ به این ترتیب که انسان می‌تواند از تجاربی که با کمک اعضای حسی خود یاد گرفته است، استفاده کند و از آنها بهره‌مند گردد، اما سیستم‌های هوش مصنوعی معمولاً با ورودی‌های نمادین کار می‌کنند (چناری و شبستانی، ۱۴۰۰، ص ۱۵۵).

تفاوت دیگر آن است که انسان با هوش طبیعی خود در مواجهه با هر موقعیت جدید می‌تواند با تکیه بر تجربیات و خلاقیت خود، به سرعت یک راه‌حل جدید ارائه دهد؛ زیرا انسان با توجه به اصول منطقی فکر می‌کند و می‌کوشد مسائل را حل کند. اما راه‌حلی که هوش مصنوعی می‌تواند ارائه دهد، به اطلاعات و روش‌های حل مسئله‌ای محدود است که به‌عنوان کد در هوش مصنوعی بارگذاری شده است (Adali, 2017, p. 10). بنابراین، هدف حوزه هوش مصنوعی، تقلید از هوش طبیعی است.

۲. آزمون تورینگ در مبحث قابلیت هوش مصنوعی

هدف تحقیقات هوش مصنوعی، تولید ماشین‌های هوشمند، آگاه، با قابلیت یادگیری و قادر به اندیشیدن است. به عبارت دیگر، هدف این است که ماشین‌های مذکور احساساتی مانند درد، خوشحالی، مهربانی و افتخار را واقعاً حس و درک کنند (Penrose, 1997, p. 19).

بهترین مثال برای این سؤال که «آیا کامپیوترها می‌توانند فکر کنند؟»، آزمون تورینگ است. این آزمون ابتدا در سال ۱۹۵۰ توسط آلن تورینگ ارائه شد. هدف اصلی آزمون تورینگ این است که بررسی کند آیا یک ماشین قادر است مشابه با انسان به سؤالات و وظایفی که برایش مطرح می‌شود، پاسخ دهد یا نه. در این آزمون، یک پرسشگر انسانی با یک رایانه و یک انسان در اتاق جداگانه در ارتباط است. پرسشگر فقط از طریق تبادل پیام می‌تواند با آنها ارتباط برقرار کند. اگر



آزمایشگر بتواند رایانه را از انسان تشخیص دهد، می‌توان گفت رایانه قدرت تفکر ندارد. در غیر این صورت، می‌توان گفت رایانه دارای قابلیت تفکر است. به‌طور معمول، آزمون تورینگ، ابزار سنجش هوش مصنوعی است. اگر یک سیستم رایانه‌ای بتواند به‌طور معقول به سؤالات پاسخ دهد و توانایی برقراری ارتباط مشابه با یک انسان را داشته باشد، یک هوش مصنوعی موفق در نظر گرفته می‌شود (Gödelek, 2013, p. 118). این آزمون به‌عنوان ابزاری برای سنجش پیشرفت هوش مصنوعی و تعیین میزان توانایی یک سیستم رایانه‌ای در تقلید رفتار انسانی استفاده می‌شود.

۲.۱. انتقادات وارد به آزمون تورینگ

به‌طور کلی، درباره آزمون تورینگ در زمینه هوش مصنوعی بسیار بحث و انتقاد شده است. برخی از انتقادات شامل مسائلی مانند بی‌توجهی به آگاهی و تجربه، نقدهای فلسفی، مسائل مذهبی و اخلاقی، استدلال‌های مرتبط با آگاهی و مفهوم و ماهیت ذهن، نادیده گرفتن جنبه‌هایی از هوش انسانی مانند هنر و خلاقیت و نبود معیارهای دقیق برای ارزیابی هوش و آگاهی مصنوعی است (Turing, 2008, p. 62-72).

یکی از مهم‌ترین انتقاداتی که به آزمون تورینگ وارد شد، ازسوی آدا لاولیس مطرح شده است. او بیان می‌کند پاسخی که یک رایانه در آزمون تورینگ ارائه می‌دهد، نتیجه کدگذاری‌ها و برنامه‌نویسی‌های انجام‌شده در آن رایانه است. طبق دیدگاه او، هیچ رایانه‌ای قادر به ارائه پاسخی که از قبل کدگذاری و برنامه‌نویسی نشده باشد، نیست. به عبارت دیگر، رایانه‌ها در واقع فقط قادر به ارائه پاسخی که از قبل کدگذاری شده و پیش‌فرض هستند و قدرت شناختی و ذهنی حقیقی ندارند (Gödelek, 2013, p. 119). انتقاد دیگری که به آزمون تورینگ مطرح شده به بعد احساسات مرتبط است. این انتقاد بر این اصل مبتنی است که احساسات و تجربیات



ذهنی، از طریق فرایندهای ذهنی و عصبی انسان به وجود می‌آید و بنابراین یک ماشین که این ویژگی‌ها را ندارد، نمی‌تواند به‌طور واقعی احساس و تجربه کند (Çelebi, 2019, p. 354).

در سال‌های اخیر، رایانه‌ها در برخی از زمینه‌ها به همراه هوش طبیعی انسان، پیشرفت کرده‌اند. اما آیا در پیشرفت هوش مصنوعی و کارهایی که انجام می‌شود، ماشین‌ها می‌توانند جای انسان را بگیرند؟

در راستای انتقاد به آزمون تورینگ، جان سرل به این سؤال پاسخ منفی می‌دهد. جان سرل، عقیده دارد ذهن انسان و سیستم‌های دارای هوش مصنوعی با یکدیگر متفاوت‌اند. ماشین‌ها تنها زمانی می‌توانند مثل انسان فکر کنند و جایگزین او شوند که دارای آگاهی و وجدان باشند (Searle, 2005, p. 39).

هرچند سیستم‌های هوش مصنوعی در برخی از رفتارها به انسان شباهت دارند، نمی‌توانند به‌طور کاملاً طبیعی مانند انسان فکر کنند و آگاهی داشته باشند (هوشیار، ۱۳۸۴، ص ۵۰).

همان‌طور که جان سرل اشاره کرده است، اگر ماشین‌ها فقط ابزارهایی بدون تفکر و آگاهی باشند، هیچ بعد اخلاقی در رفتار انسان در برابر آن‌ها وجود ندارد (Dore, 2012, p. 36).

به عبارت دیگر، اگر ماشین‌ها فقط ابزارهایی باشند که نمی‌توانند فکر کنند و زنده باشند، رفتار انسان در برابر آن‌ها هیچ تأثیر اخلاقی ندارد؛ اما اگر ماشین‌ها ابزارهایی با تفکر و آگاهی باشند، رفتار انسان در برابر آن‌ها بعد اخلاقی خواهد داشت.

به این دلیل است که انسان به‌عنوان یک موجود آگاه، حقوق انسانی دارد؛ در حالی که ربات که نمی‌تواند فکر کند و از آگاهی بی‌بهره است، حقوقی ندارد. اگر یک ربات وابسته به سیستم هوش مصنوعی با تفکر، آگاهی و تجربه وجود داشته باشد، می‌توان به حقوق آن ربات اشاره کرد.



۳. رابطه هوش مصنوعی و اخلاق

در زمینه مسائل اخلاقی، برخی از فلاسفه به بررسی اخلاق در زمینه عقل یا آگاهی پرداخته‌اند، در حالی که برخی دیگر آن را در زمینه حس و ادراک پیگیری کرده‌اند. دسته دیگری از فلاسفه هم عقل و حس را هم‌زمان در نظر می‌گیرند (آهنگری، ۱۳۸۶، ص ۱۳). به همین دلیل، ویژگی‌های دقیقی که وضعیت اخلاقی را تعیین می‌کنند، کاملاً مشخص نیستند. زمانی که عقل یا آگاهی و حس را ارزیابی می‌کنیم، ذهن به عنوان مفهومی مرتبط با تفکر، ارزیابی شده و حس به عنوان توانایی تجربه‌های فیزیکی یا ظاهری توصیف می‌شود؛ برای مثال، آشکار است بسیاری از حیوانات از توانایی حسی برخوردارند و بنابراین این حیوانات دارای وضعیت اخلاقی نیز هستند. بنابراین، در رفتار با حیوانات باید وضعیت اخلاقی آن‌ها را در نظر گرفت و فراموش نکرد که آن‌ها نیز حقوقی دارند. در عین حال، نباید فراموش کنیم موجوداتی که بالاترین درجه عقل یا آگاهی را دارند، انسان‌ها هستند و بنابراین، انسان‌ها بالاترین وضعیت اخلاقی را دارند (Çelebi, 2017, p. 8).

اگر فرض کنیم ربات یا رایانه‌ای دارای هوش و بینش فوق‌العاده قوی باشد و از آگاهی و توانایی یافتن راه‌حل‌های مختلف برخوردار باشد، می‌توانیم بگوییم این ربات یا رایانه از لحاظ توانایی‌های ذهنی با انسان‌ها تفاوتی ندارد. اگر این دو موجود (انسان و ربات دارای هوش مصنوعی)، همان عملکرد، همان آگاهی و همان تجربه را داشته باشند و تنها از لحاظ جوهری از یکدیگر متمایز باشند، انتظار داریم در چنین شرایطی دارای همان وضعیت اخلاقی باشند. با این حال، اگر یک ربات در زمینه تجربه‌های ظاهری یا فیزیکی مثلاً احساس درد یا غم را تجربه کند، باید فرض کنیم که این ربات نیز احساساتی دارد و بنابراین، دارای یک وضعیت اخلاقی است و باید از حقوقی برخوردار باشد (Çelebi, 2017, p. 8).

اگر هوش مصنوعی یک ربات به سطح هوش یک انسان بالغ برسد، ممکن است ربات دارای هوش مصنوعی، شخصیت حقوقی به دست آورد. بنابراین، اگر ربات یا



رایانه دارای هوش مصنوعی از توانایی‌هایی مشابه انسان برخوردار باشد، انتظار داریم از لحاظ اخلاقی همان حقوق و وضعیتی که برای انسان‌ها قائل هستیم، به آن‌ها تعلق گیرد.

هرچند سیستم‌های هوش مصنوعی در بعضی از جوانب ممکن است شباهت‌هایی با موجودات زنده داشته باشند، اما همواره از طریق برنامه‌هایی که روی آن‌ها بارگذاری می‌شود، عمل می‌کنند. در این معنا، حس و شعور و آگاهی هوش مصنوعی همیشه یک حس و آگاهی برنامه‌ای است و به همین دلیل، به آن هوش مصنوعی گفته می‌شود. در نتیجه، آن‌ها وضعیت اخلاقی معادل با انسان‌ها ندارند. به‌طور طبیعی، برترین هوش، هوش انسان است. به همین دلیل است که برترین وضعیت اخلاقی به انسان تعلق دارد؛ برای مثال، ما می‌توانیم این برنامه‌های رایانه‌ای را به دلخواه تغییر دهیم، کپی کنیم، متوقف کنیم، حذف کنیم یا از آن‌ها استفاده کنیم. با اینکه سرعت پردازش و یادگیری هوش‌های مصنوعی بسیار زیاد است، در بیشتر موارد به تجربه انسان‌ها نیاز است. در برخی از موارد، یک فرد متخصص ممکن است از روش‌های مختلفی مانند استنباط، قیاس و گزینه استفاده کند (Allahverdi, 2002, p. 25).

به‌طور کلی، عملیات محاسباتی مانند حسابداری با هوش مصنوعی بهتر انجام می‌شود، در حالی که عملیاتی مانند استنباط و قیاس با هوش انسانی بهتر انجام می‌شود. هوش مصنوعی از طریق برنامه‌هایی که برایش بارگذاری می‌شود، عمل می‌کند؛ بنابراین، بیشتر فیلسوفان و دانشمندان به این نتیجه رسیده‌اند که سیستم‌های رباتیک یا رایانه‌ای که به هوش مصنوعی وابسته‌اند، در حال حاضر وضعیت اخلاقی ندارند (Celebi, 2017, p. 7).

از جنبه‌های بسیاری یک ربات یا سیستم رایانه‌ای با انسان تفاوت دارد؛ برای نمونه، از نظر اخلاقی یک ربات یا سیستم رایانه‌ای با برنامه‌ای که در آن بارگذاری شده است، می‌تواند اطلاعات اخلاقی را بداند (بختیاری، ۱۳۹۱، ص ۵۲)؛ مثلاً



می‌داند قتل، عملی نادرست است. اما اگر برنامه‌ای برای سیستم در نظر گرفته نشده باشد که «دزدی اشتباه است»، آن ربات یا سیستم رایانه‌ای نمی‌داند دزدی اشتباه است، درحالی که انسان با تجربه و یادگیری در زندگی، به‌طور طبیعی می‌تواند بفهمد که کشتن و دزدی، اعمالی نادرست هستند. این تفاوت نشان می‌دهد ربات یا سیستم رایانه‌ای هرچقدر هم برنامه‌های اخلاقی داشته باشد، وضعیت اخلاقی آن نمی‌تواند به‌اندازه وضعیت اخلاقی انسان باشد (Çelebi, 2017, p. 9). به همین دلیل، هوش طبیعی انسان در مقایسه با هوش مصنوعی، بسیار گسترده‌تر و کامل‌تر است و در نتیجه، وضعیت اخلاقی انسان برتر از وضعیت اخلاقی سیستم رایانه‌ای با هوش مصنوعی است.

هوش مصنوعی از طریق برنامه‌هایی که انسان‌ها ایجاد می‌کنند، به وجود می‌آید. بنابراین، هرگونه عملکرد، تصمیم‌گیری یا تجربه‌ای که یک سیستم هوش مصنوعی دارد (مهدیانی، ۱۴۰۰، ص ۴۰)، در واقع نتیجه‌ای از تلاش و تصمیمات انسان‌هایی است که آن برنامه را طراحی کرده‌اند. به عبارتی، وقتی یک ربات با استفاده از الگوریتم‌ها و برنامه‌هایی که برایش طراحی شده، وظیفه‌ای را انجام می‌دهد، این عملکرد به‌نوعی نشانه پیشرفت و دانش انسان است که آن برنامه را برای ربات ایجاد کرده است؛ برای مثال، پیروزی دیپ بلو در مقابل قهرمان شطرنج جهان در واقع موفقیت دیپ بلو نیست، بلکه موفقیت و پیشرفت برنامه‌نویس یا گروهی است که برنامه را برای دیپ بلو فراهم کرده است. همان‌طور که پروفیسور جفرسون می‌گوید، هیچ سیستمی نمی‌تواند از موفقیت لذت ببرد، ناراحت شود یا از تحسین خوشحال گردد.

به عبارت دیگر، هر موفقیت یا وضعیت اخلاقی در این سیستم‌ها، در واقع نتیجه تلاش و توانایی انسانی است که آن را برنامه‌ریزی و اجرا کرده است (Turing, 2008, p. 65). به همین ترتیب، وضعیت اخلاقی یک سیستم هوش مصنوعی وابسته به مقررات و برنامه‌هایی است که انسان در آن بارگذاری کرده



است. بنابراین، هوش مصنوعی تابع وضعیت اخلاقی انسان است؛ زیرا آن که وضعیت اخلاقی را در سیستم هوش مصنوعی بارگذاری می‌کند، انسان است؛ یعنی انسان تصمیماتی را درباره اخلاقیات و ارزش‌ها برای آن برنامه‌نویسی می‌کند. امروزه تحقیقات در زمینه هوش مصنوعی پیشرفت چشمگیری داشته است. با این حال، هنوز سیستم‌های هوش مصنوعی نمی‌توانند به‌طور کامل با مفهوم اخلاق سازگار باشند (رمضانی و فیضی درخشی، ۱۳۹۲، ص ۹). دانشمندان برای ساختن یک هوش مصنوعی که از هوش طبیعی تقلید کند، باید اخلاق مصنوعی را به گونه‌ای تنظیم کنند که از اخلاق طبیعی تقلید کند.

۴. اهمیت اخلاق در هوش مصنوعی

اصول اخلاقی، از گذشته تا امروز، در زندگی انسان‌ها جایگاه بسیار مهمی داشته است. به‌طور کلی، مفهوم اخلاق بر پایه مفاهیم خوب و بد بنا شده و از طریق مفاهیمی مانند اراده، مسئولیت، آزادی، وجدان، فضیلت، خوشبختی و وظیفه بررسی می‌شود. بنابراین، بسیاری از فلاسفه از طریق اخلاق، اندیشه‌های مختلفی را مطرح کرده‌اند؛ برای مثال، براساس افکار سقراط، پایه و اساس خوبی در فضیلت است، براساس افکار افلاطون، اساس خوبی، داشتن اراده خوب است (Platon, 2013, p. 87).

طبق افکار امانوئل کانت، خوبی، اراده یا نیت خوب ناشی از وظیفه است (Kant, 2013, p. 8) و براساس افکار جرمی بنتام و جان استوارت میل، اساس خوبی در منفعت و سودمندی است. به‌طور کلی، اخلاق در هر دوره براساس خوب و بد بنا شده است. هنگامی که سیستم‌های هوش مصنوعی به‌عنوان موجودی آگاه، متفکر و تجربه‌کننده مانند انسان توسعه یابند، انتظار داریم به اخلاق مصنوعی مشابه با اخلاق طبیعی انسان نیز مجهز باشند. این به معنای تعبیه مفاهیمی مانند خوبی، بدی، فضیلت، مسئولیت، آزادی، وجدان و وظیفه در سیستم هوش مصنوعی است. این سیستم



اخلاقی مصنوعی باید با سیستم قوانین و اصولی تطابق داشته باشد که مشابه اخلاق طبیعی انسان است و براساس آن اقدام کند؛ برای مثال، در مواقعی که یک سیستم هوش مصنوعی تصمیم‌گیری می‌کند، باید قادر باشد بین گزینه‌های خوب و بد تمایز قائل شود و تصمیمی را براساس خوبی و فضیلت بگیرد. همچنین، باید از مسئولیت تصمیمات خود آگاه باشد و عواقب آن‌ها را در نظر بگیرد. از طرف دیگر، باید به آزادی فرد در انتخاب و تصمیم‌گیری‌ها احترام بگذارد و به‌عنوان یک عامل اخلاقی، وظیفه‌ها را به‌درستی انجام دهد. هدف از اینکه یک سیستم هوش مصنوعی به سیستم اخلاقی مصنوعی مجهز شود، اطمینان‌یافتن از این است که رفتار آن سیستم با ارزش‌ها و قوانین مربوط به جامعه سازگاری داشته باشد.

برای ادغام ارزش‌های اخلاقی در هوش مصنوعی، رویکردهای متنوعی وجود دارد. اولین رویکرد، اعمال قوانین اخلاقی است؛ این رویکرد بر این اساس است که برخی از قوانین اخلاقی به‌صورت نرم‌افزاری روی هوش مصنوعی اعمال شود. این قوانین باید به‌گونه‌ای باشند که هیچ ابهامی درباره رفتار ربات باقی نماند. رویکرد دیگر این است که ربات باید بدون نیاز به مجموعه‌ای از قوانین خارجی، بتواند خودبه‌خود مفهوم درست و غلط را درک کند. برای این منظور، از الگوریتم‌های تکاملی مانند الگوریتم ژنتیک و انطباق نظریه بازی برای اقدامات اخلاقی استفاده می‌شود. رویکرد سوم توصیه می‌کند که دو رویکرد قبلی ترکیب شوند. این رویکرد پیشنهاد می‌دهد ربات ابتدا با یک مجموعه قوانین شروع کند و سپس طی زمان، با آموزش و تجربه این قوانین را تغییر داده، بهبود ببخشد و یاد بگیرد چگونه از آن‌ها استفاده کند (Ağırman, 2018, p. 16).

با انتقال عملکردها، استنتاج‌ها و بیانات انسانی به ربات‌ها از طریق برنامه‌نویسی و کدگذاری رباتیک، باز هم نواقصی وجود دارد. این نواقص به توانایی انسان در تصمیم‌گیری سریع و تأثیر وجدان بر نتیجه تصمیمات مربوط است. به عبارت دیگر، ربات‌ها هنوز نمی‌توانند به‌طور کامل توانایی‌ها و وجدان و ویژگی‌های اخلاقی



انسانی را دارا باشند (Ağırman, 2018, p. 17). به نظر می‌چو کا کو زمانی که سیستم‌های هوش مصنوعی و رباتیک از انسان‌ها هوشمندتر شوند، می‌توانند نسخه‌های هوشمندتر از خود بسازند؛ به طوری که هر نسخه جدید، هوشمندتر از نسخه قبلی خود باشد. این موضوع ممکن است به بهبود پیوسته هوشمندی ربات‌ها و احتمالاً پیشروی آن‌ها از انسان‌ها منجر شود (Kaku, 2016, p. 132).

در نهایت، این ربات‌ها ممکن است به اندازه‌ای پیشرفت کنند که از انسان پیشی بگیرند و جهان، زیر سلطه و حاکمیت آن‌ها قرار بگیرد. در این صورت، دیگر جهان تنها جهان انسان‌ها نخواهد بود، بلکه جهان ربات‌ها نیز خواهد بود. در این زمینه برای جلوگیری از آسیب‌ها و تهدیدات محتمل از سوی ربات‌ها، در ابتدا باید دانشمندان، محققان و برنامه‌نویسان اطمینان حاصل کنند که سیستم‌های هوش مصنوعی یا ربات‌ها خطرناک نیستند. همچنین باید تدابیر امنیتی و اخلاقی مناسبی درباره سیستم‌های هوش مصنوعی و رباتیک اتخاذ کنند (Kaku, 2016, p. 133-136)؛ زیرا در صورت تولید و استفاده نادرست از این فناوری‌ها، ممکن است انسانیت با خطرات جدی مواجه شود. بنابراین، برنامه‌نویسان و محققان در این حوزه باید در طراحی سیستم‌های هوش مصنوعی، اصول اخلاقی را در نظر بگیرند.

آیزاک آسیموف درباره اینکه آیا یک هوش مصنوعی می‌تواند مفاهیمی مانند خوب و بد، و درست و غلط را درک کند یا نه، دیدگاهی مطرح کرده است. قوانین سه‌گانه رباتیک آیزاک آسیموف چنین است:

- یک ربات هیچ‌گونه آسیبی به انسان نمی‌رساند و نمی‌تواند با فعالیت‌نکردن باعث آسیب‌دیدن یک انسان شود؛

- یک ربات موظف است به دستورهای دریافتی از سوی انسان عمل کند، مگر در مواردی که این دستورها در تناقض با قانون شماره یک باشد؛

- یک ربات باید از وجود خویش مراقبت کند؛ البته مادامی که این محافظت در تضاد با قوانین شماره یک یا دو قرار نگیرد.



بعدها آسیموف با توجه به ناکارآمدی این سه قانون، قانون دیگری را به نام قانون صفر اضافه کرد. قانون صفر مقرر می‌دارد یک ربات نمی‌تواند به بشریت آسیب برساند یا با بی‌توجهی اجازه دهد بشریت آسیب ببیند.

این سه قانون طراحی شده‌اند تا ربات‌ها اخلاقیات و امنیت را در تعامل ربات‌ها با انسان‌ها و با دیگر ربات‌ها حفظ کنند. آیزاک آسیموف معتقد بود با این سه قانون، می‌توان اثبات کرد که هوش مصنوعی قادر است خوب و بد را از هم تشخیص دهد و به انسان‌ها آسیب نرساند. اما به نظر می‌رسد این قوانین به‌طور کامل این ادعا را تأیید نمی‌کنند. برای برنامه‌نویسی یک هوش مصنوعی عمومی براساس این مفاهیم، باید به‌طور جزئی برای هر موقعیت مشخص گردد که آن موقعیت خوب یا بد است. با این حال، به نظر می‌رسد عمل به این شرط نیز به‌طور کامل ممکن نیست و به بروز مشکلاتی می‌انجامد.

بعضی از فیلسوفان معتقدند با بارگذاری برنامه‌های مشخصی در سیستم‌های هوش مصنوعی، ربات‌ها قادر به آسیب‌رساندن به انسان‌ها نخواهند بود. آن‌ها فرض می‌کنند ربات‌ها از ابتدا با نیت خوب طراحی شده و انسان به‌عنوان خالق آن‌ها، آن‌ها را فقط برای انجام وظایف مفید و مثبت برنامه‌ریزی می‌کند.

به‌طور مثال، در زمینه آزادی، فرض کنید ربات‌ها به‌طور کامل در ارتکاب قتل آزاد هستند؛ این یک وضعیت بسیار ترسناک برای انسان‌هاست. بنابراین، برای جلوگیری از وقوع چنین وضعیتی، دانشمندان باید اصول اخلاقی را در سیستم‌های هوش مصنوعی بارگذاری کنند. پس، ربات‌ها از ابتدا باید به‌گونه‌ای طراحی شوند که به‌جای آسیب‌رسانی به انسان‌ها، به کمک و همکاری با آن‌ها تمایل داشته باشند. به عبارت دیگر، ربات‌ها باید به‌شکلی کارآمد و خیرخواهانه طراحی شوند (Asimov, 2006, p. 24).

درباره مسئولیت یک سیستم هوش مصنوعی نیز باید گفت شخص یا نهادی که طراحی، توسعه و استفاده از سیستم را برعهده دارد، باید مسئولیت‌های اخلاقی و



عملکرد سیستم را به دقت بررسی کند و در صورت لزوم، تدابیر پیشگیرانه و مناسبی را اتخاذ کند تا مسئولیت اجرای صحیح و اخلاقی سیستم تضمین شود.

۵. دلایل نیاز به وجود اخلاق در هوش مصنوعی

فرض کنیم یک هوش مصنوعی ساخته شود که قادر باشد تصمیم بگیرد و در زندگی روزمره با انسان‌ها تعامل داشته باشد. سؤال مهمی که وجود دارد این است که چگونه می‌توانیم اخلاقیات را برایش برنامه‌ریزی کنیم؛ زیرا اگر تصمیمات حیاتی این هوش مصنوعی به زندگی انسان‌ها مرتبط باشد، باید به مسئله اخلاق نیز توجه گردد. در غیر این صورت، نادیده گرفتن اصول و ارزش‌های اخلاقی به نادیده گرفتن ارزش‌های انسانی منجر می‌شود (Uyar, 2017, p. 18).

اما آیا می‌توان قدرت اراده و تصمیم‌گیری را در یک سیستم هوش مصنوعی قرار داد؟ در اینجا باید به پرسش مشهور فیلسوف اخلاق، فیلیپا فوت، اشاره کنیم: فرض کنید یک قطار در حال حرکت روی ریل است و شما در حاشیه راه آهن، در کنار یک اهرم ایستاده‌اید. با کشیدن این اهرم مسیر حرکت قطار تغییر خواهد کرد. در همین حین، یک قطار با سرعت به شما نزدیک می‌شود و از طرفی دیگر، در مسیر اصلی قطار پنج نفر ایستاده‌اند و در مسیری دیگر یک نفر روی ریل است. شما توانایی متوقف کردن قطار و همچنین اخطار دادن به افراد روی ریل را ندارید. تنها کاری که از دست شما برمی‌آید انتخاب بین کشته شدن پنج نفر یا فشردن اهرم و کشته شدن یک نفر برای نجات افراد دیگر است. شما کدام را انتخاب می‌کنید؟ (Uyar, 2017, p. 19).

در چنین شرایطی بیشتر افراد ترجیح می‌دهند با کشیدن اهرم، مسیر حرکت قطار را تغییر دهند و جان پنج نفر را نجات دهند. طبق منطق اخلاقی و رفتار مناسب انسانی، این تصمیم، منطقی بودن را نشان می‌دهد. اما پرسش این است که آیا یک ربات وابسته به هوش مصنوعی هم ترجیح می‌دهد مسیر حرکت قطار را تغییر دهد؟



در واقع برای اینکه یک هوش مصنوعی بتواند تصمیمات حیاتی بگیرد، نیازمند یک اخلاق مصنوعی (شامل اراده و تصمیم‌گیری) است.

اگر توانایی تصمیم‌گیری و اخلاق انسانی را از طریق برنامه‌نویسی در یک ربات بارگذاری کنیم، چرا نتوانیم به این نتیجه برسیم؟ با این حال، حتی اگر بتوان اخلاق مصنوعی را به‌طور کامل از طریق برنامه‌نویسی در یک ربات بارگذاری کرد، آن ربات باز هم نمی‌تواند طبق اخلاق طبیعی انسان‌ها عمل کند. همچنین، ما نمی‌توانیم به‌طور کامل توجیه کنیم که چه چیزی تصمیمات اخلاقی ما را تعیین می‌کند و براساس چه معیاری تصمیم می‌گیریم.

سیستم هوش مصنوعی، یک سیستم مداوم و پایدار است. در زمینه وضعیت اخلاقی، سیستم هوش مصنوعی قادر است با استفاده از برنامه‌هایی که در آن بارگذاری می‌شود، میان خوب و بد تمایز قائل شود. به عبارت دیگر، یک سیستم هوش مصنوعی می‌تواند درک کند که انجام کارهای اخلاقی یا شاد بودن، امری مطلوب است، در حالی که دزدی یا قتل، امری ناپسند است؛ اما نمی‌تواند ناگهان تصمیم خود را تغییر دهد و وقتی قصد دارد کار بدی انجام دهد، از کار بد به کار خوب بپردازد.

دلیل این است که سیستم هوش مصنوعی، یک سیستم برنامه‌ای است؛ برای مثال، اگر برنامه‌ای که حاوی دستور دزدی است بارگذاری شود، سیستم هوش مصنوعی قادر نخواهد بود، تصمیم خود را تغییر داده و از دزدی خودداری کند؛ چون سیستم هوش مصنوعی یک سیستم پیوسته و پایدار است که عملکرد آن در حوزه اخلاق از طریق برنامه‌هایی که در آن بارگذاری می‌شود، تعیین می‌گردد. بنابراین، وضعیت اخلاقی سیستم هوش مصنوعی به وضعیت اخلاقی برنامه‌ای که روی آن اجرا می‌شود، وابسته است.

مبحث دیگری که در خصوص وضعیت اخلاقی وجود دارد به کار و شغل مربوط است؛ فرض کنید چندین بیمار، نیازمند درمان فوری هستند و یک پزشک مبتنی بر



هوش مصنوعی باید تصمیم بگیرد به کدام بیمار اولویت دهد. این موقعیت اخلاقی باعث پیچیدگی اوضاع و ورود به یک تناقض اخلاقی می‌شود؛ به این معنا که هر تصمیمی که گرفته شود، ممکن است به زیان یکی از بیماران باشد. این تناقض اخلاقی نشان می‌دهد در مواجهه با تصمیم‌گیری‌های پیچیده، حتی سیستم‌های هوش مصنوعی نیز ممکن است در موقعیت‌هایی قرار گیرند که تصمیم‌گیری صحیح اخلاقی برایشان دشوار باشد (Köse, 2018, p. 95)؛ برای مثال، فرض کنید فردی از طریق ارتکاب جرم به فقیران کمک می‌کند. اگر قرار باشد یک ربات قاضی یا سیستم قضایی مبتنی بر هوش مصنوعی درباره این فرد حکم صادر کند، این موقعیت نیز دارای تناقض اخلاقی است؛ زیرا محکومیت این فرد می‌تواند به نوعی تناقض اخلاقی منجر شود. به طور کلی، این موارد نشان می‌دهند نیاز به تعبیه تصمیمات اخلاقی در سیستم‌های هوش مصنوعی از اهمیت زیادی برخوردار است. باید تلاش شود در طراحی ربات‌ها و سیستم‌های هوش مصنوعی به تصمیم‌گیری‌های اخلاقی صحیح و بحران‌زدایی از تناقضات اخلاقی توجه شود.

هوش انسان و هوش مصنوعی از جنبه‌های مختلف، تفاوت‌های قابل توجهی دارند. انسان‌ها برخلاف ماشین‌ها، موجودات زیستی هستند؛ از این رو، ممکن است برخی از فعالیت‌هایی که ماشین‌ها انجام می‌دهند، مانند تصمیم‌گیری یا انجام کارهای مشابه انسان، از نظر اخلاقی مسئله‌ساز باشند؛ برای مثال اینکه انسان به عنوان یک موجود زیست‌شناختی، تصمیمی را که یک ماشین می‌گیرد، پذیرفته یا مسئولیت انجام برخی از کارها را به ماشین‌ها واگذار کند و با ماشین‌ها همکاری کند، چقدر اخلاقی است؟ (İnce, 2017, p. 17)

جامعه به مرور به ماشین‌ها وابسته شده است؛ اما اگر ما انسان‌ها این سیستم‌های هوش مصنوعی را ایجاد کرده‌ایم، به همان اندازه که می‌توانیم آن‌ها را کنترل کنیم، قادر به از بین بردن آن‌ها هم هستیم.



۶. رویکرد اخلاق اسلامی در هوش مصنوعی

رویکرد اخلاق اسلامی در هوش مصنوعی می‌تواند براساس اصول و ارزش‌های اخلاقی اسلامی و قوانین شرعی تعیین شود. در اسلام، اخلاق در زمینه رفتار و اعمال انسان‌ها در نظام جامعه از اهمیت زیادی برخوردار است. بنابراین، هنگام طراحی و استفاده از هوش مصنوعی، رعایت اصول اخلاقی اسلامی می‌تواند بهبود و کیفیت کارکرد سیستم‌های هوشمند را تضمین کند.

اصول اخلاقی متعددی با رویکرد اسلامی وجود دارند که می‌توانند در طراحی هوش مصنوعی به کار گرفته شوند؛ برای مثال، اصل عدل و انصاف در اسلام، یکی از اصول اخلاقی اساسی است. در طراحی هوش مصنوعی نیز باید تلاش کرد از هرگونه تبعیض ناعادلانه در پردازش اطلاعات و تصمیم‌گیری‌های سیستم هوشمند خودداری شود. عدالت و انصاف از اصول بسیار مهم اسلام است. براساس قوانین شرعی از مسلمانان خواسته شده سیستم‌های هوش مصنوعی باز، مسئولانه و بی‌طرفانه را ایجاد کرده و به کار گیرند.

این شامل حذف سوگیری‌ها در الگوریتم‌های هوش مصنوعی و اطمینان از دسترسی بودن فناوری‌های هوش مصنوعی برای همه افراد، صرف‌نظر از پیشینه یا جایگاه اجتماعی اقتصادی آن‌هاست. علاوه بر این، در اخلاق اسلامی، حفظ حریم خصوصی فرد و احترام به حقوق او بسیار مهم است. در هوش مصنوعی نیز باید سازکارهایی وجود داشته باشد که از نقض حریم خصوصی افراد جلوگیری کنند و از اطلاعات شخصی محافظت کنند. در اسلام، پرهیز از ضرر رساندن به دیگران و جامعه به‌عنوان اصلی اساسی مطرح است. در هوش مصنوعی نیز باید از ایجاد آسیب و ضرر برای افراد و جامعه پرهیز کرد و سیستم‌ها باید با حداقل ریسک و خطر عمل کنند (شریعتی و اکبرزاده، ۱۴۰۱، ص ۹۵-۹۸).

مهم است برای توجه به این اصول اخلاقی، تفسیری درست و کاربردی از مفاهیم دینی و اخلاقی صورت گیرد. همچنین باید توجه شود که هوش مصنوعی به‌طور



معمول براساس الگوریتم‌ها و قوانینی کار می‌کند که انسان‌ها تعیین می‌کنند. بنابراین، بر عهده طراحان و برنامه‌نویسان هوش مصنوعی است که از اصول اخلاقی اسلامی پیروی کنند و سیستم‌ها را به گونه‌ای طراحی کنند که با ارزش‌های اخلاقی اسلامی سازگار باشند.

رابطه اسلام و هوش مصنوعی، چندوجهی است و ابعاد مختلفی را در بر می‌گیرد. اسلام به عنوان یک نظام دینی و اخلاقی جامع، چارچوبی را فراهم می‌کند که از طریق آن می‌توان توسعه و کاربرد هوش مصنوعی را ارزیابی و درک کرد. ملاحظات اخلاقی، نقش مهمی در دیدگاه اسلام درباره هوش مصنوعی دارد. مسلمانان تشویق می‌شوند فناوری‌های هوش مصنوعی را به شیوه‌هایی توسعه دهند که با اصول اخلاقی اسلامی هم‌سو باشد. این امر تضمین می‌کند که سیستم‌های هوش مصنوعی، حقوق بشر را نقض نمی‌کنند، از حریم خصوصی محافظت می‌کنند، از تبعیض جلوگیری می‌کنند و عدالت و انصاف را ترویج می‌دهند. دانشمندان، محققان و فناوران اسلامی همچنان در بحث‌ها و گفت‌وگوها در این زمینه شرکت می‌کنند تا درک جامعی از پیامدهای اخلاقی و عملی هوش مصنوعی با توجه به اصول اسلامی ایجاد کنند (تشکری صالح و رجبی، ۱۳۹۷، ص ۱۴-۱۷).

اخلاق اسلامی می‌تواند درباره مشکلات ناشی از هوش مصنوعی در جامعه مسلمان، افراد را راهنمایی کند. به همین دلیل ضروری است اطمینان حاصل شود که پیشرفت فناوری به هیچ وجه حقوق بشر را نقض نمی‌کند، با ارزش‌های اسلامی مغایرت ندارد و موجب آسیب‌رسیدن به دیگران از جمله نسل‌های آینده نمی‌شود.

برای تشویق به استفاده از هوش مصنوعی اخلاقی و توسعه فناوری در جوامع اسلامی، هنجارهای اخلاقی اسلامی باید در طراحی و اجرای هوش مصنوعی و فناوری گنجانده شود. این امر ممکن است با القای ارزش‌های اخلاقی اسلامی در



فرایند توسعه، وضع استانداردهای اخلاقی برای استفاده از هوش مصنوعی و فناوری و تشویق به آموزش و گفت‌وگوی عمومی درباره هوش مصنوعی اخلاقی و شیوه‌های فناوری محقق شود (ربیع‌زاده، ۱۴۰۰، ص ۳۳).

آموزش و گفت‌وگو عمومی در ترویج هوش مصنوعی اخلاقی و شیوه‌های فناوری در جوامع اسلامی از مؤلفه‌هایی بسیار مهم در این زمینه است. افراد و سازمان‌ها می‌توانند از طریق افزایش آگاهی و دانش درباره اخلاق اسلامی و نتایج کاربرد آن در حوزه هوش مصنوعی و فناوری، درباره ایجاد و استفاده از این پیشرفت‌ها قضاوت‌های آگاهانه داشته باشند. علاوه بر این، توسعه قواعد اخلاقی برای استفاده از هوش مصنوعی و فناوری می‌تواند به ما اطمینان دهد که از این پیشرفت‌ها با روش‌هایی سازگار با اصول اسلامی، استفاده می‌شود.

مسلمانان باید اطمینان حاصل کنند که سیستم‌های هوش مصنوعی، حقوق بشر و حریم خصوصی را حفظ می‌کنند و در عین حال، عدالت و انصاف را ترویج می‌دهند. در جوامع اسلامی، القای اخلاق اسلامی در توسعه هوش مصنوعی، ایجاد هنجارهای اخلاقی و تقویت آموزش و گفت‌وگو برای توسعه هوش مصنوعی و فناوری اخلاقی، بسیار مهم است.

رویکرد اخلاق اسلامی در هوش مصنوعی می‌تواند بهبود عملکرد سیستم‌های هوشمند و کاهش آسیب‌های احتمالی را به همراه داشته باشد. این رویکرد می‌تواند به ایجاد سیستم‌های هوشمندی که با ارزش‌های اخلاقی اسلامی سازگار هستند و به انسان‌ها و جامعه خدمت می‌کنند، کمک کند.



نتیجه‌گیری

هدف اصلی سیستم‌های هوش مصنوعی، تولید سیستم‌هایی است که بتوانند شبیه انسان عمل کرده، دارای قابلیت آگاهی، تفکر و تجربه باشند. درحالی که برخی از فیلسوفان معتقدند سیستم‌های هوش مصنوعی قادر به تقلید از ویژگی‌های بشر هستند، برخی دیگر همچنان این موضوع را ناممکن می‌دانند.

اگر چه تصور می‌شود سیستم‌های هوش مصنوعی با ویژگی‌های انسانی و دارای آگاهی و تفکر تولید شوند، باید به یاد داشت این سیستم‌ها ساخت انسان و مشتق از هوش طبیعی خواهند بود. همچنین، اگر چه با تحقیقات هوش مصنوعی ممکن است اخلاق مصنوعی را هم بتوان تولید کرد، اما واضح است که این اخلاق نیز از هوش و اخلاق طبیعی انسان‌ها نشئت می‌گیرد. بنابراین، خالق سیستم‌های هوش مصنوعی انسان است. در این زمینه، با در نظر گرفتن موقعیت مرکزی انسان در مسئولیت، با حفظ تعادل صحیح، باید اجازه داد هوش مصنوعی توسعه یابد. به عبارت دیگر، باید در جهت پیشرفت هوش مصنوعی حرکت کنیم و در عین حال، مسئولیت و اخلاق را در نظر داشته باشیم.

اگر سیستم‌های هوش مصنوعی به قدرت یادگیری عمیق و نرم‌افزار خود دسترسی داشته باشند، آیا امکان اتخاذ تدابیری برای محدود کردن این دسترسی وجود خواهد داشت؟ همچنین اگر ربات‌ها مخالفت کنند و در مقابل اطاعت از دستورات انسان یا کنترل شدن اعتراض کنند، چگونه می‌توان از مسائل اخلاقی ناشی از این وضعیت جلوگیری کرد؟

اگر تلاش‌های علمی برای رساندن این سیستم‌های هوش مصنوعی به سطح مغز انسان یا هوش طبیعی، موفقیت‌آمیز باشد و آن‌ها را به مرحله تقلید و حتی دستیابی به سطحی از آگاهی برساند، نیاز به اخلاق مصنوعی برای این سیستم‌ها نیز وجود خواهد داشت. برای اینکه انسان بتواند سیستم‌های هوش مصنوعی را که خود تولید کرده است، در کنترل خود نگه دارد و از آسیب‌رسیدن به بشر جلوگیری کند، لازم



است عناصر اخلاقی انتزاعی مانند مسئولیت و اراده در چارچوب مفاهیم خوب و بد به‌طور مصنوعی در این سیستم‌ها بارگذاری شود. در غیر این صورت، سیستم‌های هوش مصنوعی ممکن است از کنترل سازنده خود خارج شوند و شرایط آشوب‌زا به وجود آورند.

بنابراین، مسئولیت متقابلی لازم است که یک سیستم هوش مصنوعی بتواند مسئولیت در قبال سازنده خود را بر عهده بگیرد. به عبارت دیگر، در صورتی که اخلاق مصنوعی در یک سیستم هوش مصنوعی تعبیه نشود، مشکلات جدی برای بشر و دیگر موجودات زنده در طبیعت ایجاد می‌شود.

بنابراین، ماشین‌ها و سیستم‌های هوش مصنوعی به تصمیمات و اقدامات انسان‌ها وابسته هستند و وضعیت اخلاقی سیستم‌های آن‌ها در دست انسان‌ها قرار دارد. با توجه به این موضوع، ابتدا باید موارد مثبت و منفی مربوط به پژوهش‌های هوش مصنوعی مشخص شود و با در نظر گرفتن هدف کمک به تسهیل و بهبود زندگی بشریت، باید از اینکه سیستم‌های هوش مصنوعی قوی‌تر و مقاوم‌تر از انسان شوند و به خطری برای ادامه زندگی بشر تبدیل گردند، جلوگیری شود. برای جلوگیری از این مشکلات، فناوری هوش مصنوعی باید به‌صورت کنترل‌شده و با در نظر گرفتن مسائل اخلاقی پیشرفت کند.





کتاب‌نامه

۱. آهنگری، فرشته (۱۳۸۶)، «پیشینه و بنیادهای اخلاق در ایران و جهان»، فصلنامه اخلاق در علوم و فناوری، ش ۳، ص ۱۱-۲۲.
۲. بختیاری، آمنه (۱۳۹۱)، «بررسی هنجارهای اخلاقی در فضای مجازی»، مطالعات رسانه‌ای، سال ۷، ش ۱۲، ص ۴۶-۵۳.
۳. تشکری صالح، ابوذر و محمود رجبی (۱۳۹۷)، «بررسی توان رقابت هوش مصنوعی با ذهن انسان از منظر قرآن»، قرآن‌شناخت، دوره ۱۱، ش ۲، ص ۵-۲۰.
۴. چناری، مهین و اعظم شبستانی (۱۴۰۰)، «پیش‌بینی میزان آشنایی با اخلاق در فضای مجازی با استفاده از سواد رسانه‌ای و شایستگی‌های فناورانه»، اخلاق در علوم و فناوری، سال ۱۶، ش ۱، ص ۱۵۴-۱۵۸.
۵. ربیعی‌زاده، احمد (۱۴۰۰)، «کاربرد هوش مصنوعی در پژوهش‌های علوم اسلامی»، ره‌آورد نور، دوره ۲۰، ش ۷۴، ص ۲۸-۳۷.
۶. رضانی، مجید و محمدرضا فیضی درخشی (۱۳۹۲)، «اخلاق ماشین، چالش‌ها و رویکردهای مسائل اخلاقی در هوش مصنوعی و ابرهوش»، اخلاق در علوم و فناوری، سال ۸، ش ۴، ص ۱-۹.
۷. شریعتی، فهیمه و محمدرضا اکبرزاده (۱۴۰۱)، «مقایسه ماهوی هوش انسانی با هوش مصنوعی از منظر فلسفه اسلامی»، حکمت معاصر، دوره ۱۳، ش ۲، ص ۸۹-۱۱۷.
۸. صراف‌زاده، شقایق و احسان ابوطالب (۱۴۰۲)، «اهمیت اخلاق در استفاده از هوش مصنوعی در آموزش پزشکی»، پژوهش در آموزش علوم پزشکی، سال ۱۵، ش ۲، ص ۱-۴.
۹. قاسمی، محمدرضا (۱۴۰۲)، «الهیات و هوش مصنوعی»، نشریه پاسخ، ش ۳۱، ص ۱۳۹-۱۶۰.



۱۰. مهدیانی، محمدرضا (۱۴۰۰)، «جایگاه اخلاق در فضای مجازی»، پژوهش مطالعات علوم اسلامی، دوره ۳، ش ۳۰، ص ۳۹-۵۱.
۱۱. هوشیار، مسعود (۱۳۸۴)، «هوش مصنوعی و سیستم‌های اخلاق»، علوم خلاقیت شناسی و حل مسئله ابداعی، ش ۹، ص ۴۴-۵۴.
12. Adalı, Eşref (2017). *Yapay Zekâ. İnsanlaşan Makineler ve Yapay Zekâ içinde, haz. Mehmet Karaca*, 8-13, İstanbul: İstanbul Teknik Üniversitesi Vakfı Yayınları.
13. Ağırman, Handan (2018). “Etik Sürücüsüz Araç Mümkün müdür?”. III. *Ulusal Uygulamalı Etik Kongresi*, ODTÜ.
14. Allahverdi, Novruz (2002). *Uzman Sistemler: Bir Yapay Zekâ Uygulaması*. İstanbul: Atlas Yayınları.
15. Asimov, Isaac (2006). *Ben, Robot*. (Çev. Ekin Odabaş), İstanbul: İthaki Yayınları.
16. Atrkinson M ve Boden A. (1979). “Artificial Intelligence and Natural Man”. *Philosophical Quarterly*, s. 130-132.
17. Başkurt, Batuhan (2018).. “Keskin Kılıç: Yapay Zekâ”. *Ankara Teknoloji Geliştirme Bölgesi kurucu ve İşleticisi A. Ş.* sayı: 13, s. 7.
18. Çelebi, Ömer, Faruk (2017). “Yapay Zekâ ve Etik”. *İstanbul Medeniyet Üniversitesi Sosyal Bilimler Enstitüsü*, s. 1-20.
19. Çelebi, Vedat (2019). “Searle ve Nagel’in Bilinç Kuramlarında Fizikalizm Eleştirisi Bağlamında Yapay Zekânın Sınırı Tartışması”. *Beyhtülhikme Felsefe Dergisi*, 9, 2, s. 351-376.
20. Crick, Francis (2005). *Şaşkırtan Varsayım (İnsan Varlığının Temel Sorularına Yanıt Arayışı)*. (Çev. Sabit. Say), Ankara: Tübitak Yayınları.

21. Dore, Fatma (2012). “Güçlü Yapay Zekâya Karşı Çin Odası Argümanı”. *Sosyal Bilimler Dergisi*, XIV, 1, s. 23-38.
22. Gödelek, Kamuran (2013). *Zihin Felsefesi. Eskişehir. Anadolu Üniversitesi Yayınları.*
23. Gökhan, İnce (2017). *İnsanlığın Yapay Zekâyla İmtihanı. İnsanlaşan Makineler ve Yapay Zekâ içinde, haz. Mehmet Karaca, 26-30, Sayı 75. İstanbul Teknik Üniversitesi Vakfı Yayınları.*
24. Kaku, Michio (2016). *Geleceğin Fiziği. (Çev. Yasemin Saraç Oymak ve Hüseyin Oymak), Ankara: ODTÜ Geliştirme Vakfı Yayıncılık.*
25. Kant, Immanuel (2013). *Ahlâk Metafiziğinin Temellendirilmesi. (Çev. İoanna Kuçuradi), Ankara: Türkiye Felsefe Kurumu Yayınları.*
26. Köse, Utku (2018). “Güvenli Yapay Zekâ Sistemleri İçin İnsan Denetimli Bir Model Geliştirilmesi”. *Mühendislik Bilimleri ve Tasarım Dergisi* 6, 1, s. 93- 107.
27. Penrose, Roger (1997). *Kral'ın Yeni Usu (Bilgisayar ve Zekâ). (Çev. Tekin Dereli), Ankara: Tübitak Popüler Bilim Kitapları.*
28. Platon (2013). *Menon. (Çev. Furkan Akderin), İstanbul: Say Yayınları.*
29. Searle, John (2005). *Akıllar, Beyinler ve Bilim. (Çev. Kemal Bek), İstanbul: Say Yayınları.*
30. Turing, A., M. (2008). *Bilgi İşlem Makineleri ve Zekâ. Aklın Gözü içinde, haz. Douglas R. Hofstadter ve Daniel C. Dennet, (Çev. Füsun Doruker), 59- 73. İstanbul: Boğaziçi Üniversitesi Yayınevi.*



31. Uyar, Tefvik (2017). *Ya Yapay Ahlâk. İnsanlaşan Makineler ve Yapay Zekâ içinde*, haz. Mehmet Karaca, Sayı 75, 18-21. İstanbul: İstanbul Teknik Üniversitesi Vakfı Yayınları.



References

۱. Adali Esref (۲۰۱۷). *Artificial intelligence. In Humanizing Machines and Artificial Intelligence*, ed. Mehmet Karaca, ۸-۱۲, Istanbul: Istanbul Technical University Foundation Publications.
۲. Agirman, Handan (۲۰۱۸). Is an Ethical Driverless Vehicle Possible? III. *National Applied Ethics Congress*, METU.
۳. Ahangari, Fereshteh (۲۰۱۶). "Background and Foundations of Ethics in Iran and the World". *Ethics in science and technology*, No. ۳, p. ۱۱-۲۲.
۴. Allahverdi, Novruz (۲۰۰۲). *Expert Systems: An Application of Artificial Intelligence*. Istanbul: Atlas Publications.
۵. Asimov, Isaac (۲۰۰۶). *I'm Robot.* (Trans. Ekin Odabas), Istanbul: İthaki Publications.
۶. Atkinson M and Boden A. (۱۹۷۹). Artificial Intelligence and Natural Man. *Philosophical Quarterly*, p. ۱۳۰-۱۳۲.
۷. Bakhtiari, Amaneh (۲۰۱۷). "Ethical norms in cyber space". *Media Studies*, Year ۷, No. ۱۲, p. ۴۶-۵۳.
۸. Baskurt, Batuhan (۲۰۱۸). Sharp Sword: Artificial Intelligence. *Founder and Operator of Ankara Technology Development Zone Inc.* issue: ۱۲, p. ۷.
۹. Celebi, Omer, Faruk (۲۰۱۷). Artificial Intelligence and Ethics. *Istanbul Medeniyet University Social Sciences Institute*, p. ۱-۲۰.
۱۰. Celebi, Vedat (۲۰۱۹). Discussion on the Limits of Artificial Intelligence in the Context of Searle and Nagel's Critique of Physicalism in Consciousness Theories. *Beytülhikme Philosophy Journal*, ۹, ۲, p. ۳۵۱-۳۷۶.
۱۱. Chenari, Mahin & Shabestani, Azam (۲۰۲۰). "predicting the level of familiarity with ethics in virtual space, using media literacy and technological skills". *Ethics in Science and Technology Quarterly*, year ۱۶, No. ۱, p. ۱۰۴-۱۰۸.
۱۲. Crick, Francis (۲۰۰۵). *The Puzzling Hypothesis* (The Search for Answers to the Fundamental Questions of Human Existence). (Trans. Sabit Say), Ankara: Tübitak Publications.
۱۳. Dore, Fatma (۲۰۱۲). The Chinese Room Argument Against Strong Artificial Intelligence. *Journal of Social Sciences*, XIV, ۱, p. ۲۳-۳۸.
۱۴. Ghasemi, Mahmoudreza (۲۰۲۲). "Theology and Artificial Intelligence". *Answer magazine*, No. ۳۱, p. ۱۳۹-۱۶۰.
۱۵. Godelek, Kamuran (۲۰۱۳). *Philosophy of Mind*. Anadolu University Publications.
۱۶. Gokhan, Ince (۲۰۱۷). *Humanity's Test with Artificial Intelligence. In Humanizing Machines and Artificial Intelligence pleasure*. Mehmet Karaca, ۲۶-۳۰, Issue ۷۵. Istanbul Technical University Foundation Publications.
۱۷. Hoshiar, Masoud (۲۰۱۴). "Artificial Intelligence and Moral Systems". *Science Journal of Creativity and Innovative Problem Solving*, No. ۹, p. ۴۴-۵۴.
۱۸. Kaku, Michio (۲۰۱۶). *Physics of the Future*. (Trans. Yasemin Sarac Oymak and Huseyin Oymak), Ankara: METU Development Foundation Publishing.
۱۹. Kant, Immanuel (۲۰۱۳). *Foundation of the Metaphysics of Morals*. (Trans. İoanna Kucuradi), Ankara: Turkish Philosophical Society Publications.
۲۰. Kose, Utku (۲۰۱۸). Developing a Human-Supervised Model for Safe Artificial Intelligence Systems. *Journal of Engineering Sciences and Design* ۶, ۱, p. ۹۳- ۱۰۷.
۲۱. Mehdiani, Mohammadreza (۲۰۲۰). "the position of ethics in cyber space". *Islamic Studies Research*, Vol. ۳, No. ۳۰, p. ۳۹-۵۱.
۲۲. Penrose, Roger (۱۹۹۷). *The King's New Intelligence (Computers and Intelligence)*. (Trans. Tekin Dereli), Ankara: Tubitak Popular Science Books.
۲۳. Plato (۲۰۱۳). *Menon*. (Trans. Furkan Akderin), Istanbul: Say Publications.
۲۴. Rabieizadeh, Ahmad (۲۰۲۰). "Application of Artificial Intelligence in Islamic Science Research". *Rahevard Noor*. Vol. ۲۰, No. ۷۴, p. ۲۸-۳۷.
۲۵. Ramezani, Majid & Faizi Derakhshi, Mohammad Reza (۲۰۱۲). "Machine ethics, challenges and approaches to ethical issues in artificial intelligence and super-intelligence". *Ethics in Science and Technology Quarterly*, Year ۸, No. ۴, p. ۱-۹.
۲۶. Sarafzadeh, shaghayegh & Abutaleb, Ehsan (۲۰۲۲). "The Importance of Ethics in the Use of Artificial Intelligence in Medical Education". *Journal of Research in Medical Science Education*. Year ۱۵, No. ۲, p. ۱-۴.
۲۷. Searle, John (۲۰۰۵). *Minds, Brains and Science*. (Trans. Kemal Bek), Istanbul: Say Publications.
۲۸. Shariati, Fahimeh & Akbarzadeh, Mohammad Reza (۲۰۲۱). "The substantive comparison of human intelligence with artificial intelligence from the perspective of Islamic philosophy". *Contemporary Wisdom Quarterly*, Vol. ۱۳, No. ۲, p. ۸۹-۱۱۷.

۲۹. Tashakori Saleh, Abuzar & Rajabi, Mahmoud (۲۰۱۷). "examining the ability of artificial intelligence to compete with the human mind from the perspective of the Qur'an". *Qur'an Knowledge*, Vol. ۱۱, No. ۲, p. ۵-۲۰.
۳۰. Turing, A., M. (۲۰۰۸). *Computing Machines and Intelligence*. In *the Mind's Eye, pleasure*. Douglas R. Hofstadter and Daniel C. Dennet, (Trans. Fusun Doruker), ۵۹- ۷۳. Istanbul: Bogazici University Publishing House.
۳۱. Uyar, Tevfik (۲۰۱۷). *What about Artificial Morality? In Humanizing Machines and Artificial Intelligence, pleasure*. Mehmet Karaca, Issue ۷۵, ۱۸-۲۱. Istanbul: Istanbul Technical University Foundation Publications.

