سازگاری یا ناسازگاری توصیه‌نامه‌ی اخلاق هوش مصنوعی یونسکو با نظریه اخلاقی اسلام

حمید محسنی [۱]، کاظم فولادی قلعه [۲]

[۱] دانشجوی دکتری، مؤسسه آموزشی و پژوهشی امام خمینی، قم
h.mohseni1297@mailfa.com
[۲] استادیار، گروه مهندسی کامپیوتر، دانشکده مهندسی دانشکدگان فارابی دانشگاه تهران؛ سرپرست آزمایشگاه پژوهشی فضای سایبر دانشگاه تهران
kfouladi@ut.ac.ir

چکیده

یونسکو به تازگی سندی بین‌المللی با عنوان «توصیه‌نامه‌ی اخلاق هوش مصنوعی» تصویب کرده است. در این سند مجموعه ای از ارزش‌ها مشخص شده و از کشورهای عضو خواسته شده است مفاد آن را در جوامع خود به نحوی اجرا کنند. در واقع، نهادهای بین‌المللی سعی دارند با تصویب این‌گونه اسناد بالادستی، قوانین و ارزش‌های ملت‌ها را با خود همراه سازند. بنابراین، مسئله اصلی این پژوهش، بررسی سازگاری و یا عدم سازگاری مبانی ارزش‌های اخلاقی توصیه‌نامه‌ی مذکور با مبانی اخلاقی اسلام می‌باشد. بر اساس یافته‌های این تحقیق می‌توان گفت برخی مفاد این توصیه‌نامه مبتنی بر مکتب اخلاقی قراردادگرایی است. از این رو، این توصیه‌نامه به لحاظ موضوع و هدف اخلاقی، وظیفه‌گرا و عمل‌محور می‌باشد. اما نظریه‌ی اخلاق مکتب اسلام به لحاظ موضوع و هدف اخلاقی، فضیلت‌محور، غایت‌گرایانه و کمال‌جو می‌باشد. نتیجه اینکه در سند اخلاق هوش مصنوعی یونسکو تنها اصلاح رفتار فرد بدون معنویت و دین در نظر گرفته شده است. ولی در نظام اخلاقی اسلام، هدف، اصلاح رفتار و صفات نفسانی و ساخت شخص انسان و به پیرو آن نزدیک شدن به کمال مطلق می‌باشد. افزون براین، در پژوهش حاضر دو نظریه، یکی ناظر به امکان اخلاق‌مندی خود سیستم‌های هوشمند و دیگری ناظر به تعامل انسان با هوش مصنوعی ارائه شده است.

کلمات کلیدی: اخلاق، هوش مصنوعی، یونسکو، نظریه اخلاقی اسلام.

۱ مقدمه

هوش مصنوعی به‌عنوان یک تکنولوژی نوظهور، برای انسان معاصر فایده‌هایی را به ارمغان آورده است، ولی خطرهایی نیز از این ناحیه ارزش‌های اخلاقی انسان را تهدید می‌کند. بسیاری معتقدند که در آینده‌ای نه‌چندان دور، ظهور ابرهوش [۱] امری اجتناب‌ناپذیر است؛ هر چند در مدت زمان روی دادن این رخداد اختلاف

[1] Super Intelligence

نظر وجود دارد [۱۳] . اما این پدیده می‌تواند مخاطراتی بسیار جدی برای بشریت به‌همراه داشته باشد. از این رو دانشمندان می‌بایست مجموعه‌ای از ملاحظات را در نظر بگیرند. افزون بر این، همین هوش مصنوعی فعلی (هوش مصنوعی محدود[۲]) در زندگی بسیاری از انسان ها تأثیرگذار می‌باشد [۱۵]. این فناوری‌ها در کنار مزایای متعدد، خطرات و چالش‌هایی را نیز ایجاد می‌کنند که بخشی از آن ناشی از استفاده‌ی بدخواهانه از فناوری است. از این رو، سازمان بین‌المللی یونسکو به این فکر افتاده است که چارچوب‌های نظارت بین‌المللی و ملی را فراهم آورد تا این فناوری در جهت منفعت بشریت توسعه پیدا کند.

یونسکو متن اولیه سندی را با عنوان «توصیه‌نامه اخلاق هوش مصنوعی» زیر نظر یک گروه ۲۴ نفری از مختصصان ویژه (AHEG) تنظیم کرد. با بررسی رزومه‌ی این گروه ویژه مشاهده می‌شود که تخصص بیشتر آنها در حوزه‌ی فنی و مهندسی است، هرچند تعدادی حقوق‌دان و فیلسوف نیز از کشورهای مختلف –غیر از ایران– در بین آنها دیده می‌شود. همچنین حدود سه کارشناس با رویکرد اخلاق هوش مصنوعی از کشورهای آمریکا، بریتانیا و چین نیز در فرآیند تهیه‌ی این سند حضور داشته‌اند. به هر حال، پس از ماه‌ها گفت‌وگو و مذاکره‌ی نمایندگان کشورهای عضو و أخذ دیدگاه آنها و اعمال برخی نکات زیر نظر کارشناسان مزبور، اعضای یونسکو بر روی یک متن برای توسعه‌ی ارزش‌های اخلاقی در حوزه‌ی هوش مصنوعی به توافق رسیدند. سرانجام، این توصیه‌نامه در چهل‌ویکمین جلسه‌ی کنفرانس عمومی یونسکو در تاریخ ۲۴ نوامبر ۲۰۲۱، به تصویب رسید.

در این سند که متشکل از یک مقدمه و ۱۴۱ بند است، یک سری اهداف و باید و نبایدها آورده شده و از کشورهای عضو –از جمله ایران– خواسته شده است مفاد این سند را در کشورهای خودشان به‌نحوی اجرا کنند. نهادهای بین‌المللی با تصویب این‌گونه سندها –هر چند به‌ظاهر غیر الزام‌آور– سعی دارند آنها را به‌عنوان اسناد بالادستی قرارداده تا ارزش‌ها و قوانین جوامع دیگر را با خود همراه سازند. در بخشی از گزارش‌هایی که پیش از سند اصلی آمده است، بدین مطلب تصریح شده است که یونسکو سعی دارد رهبری اخلاقی خود را با تنظیم استانداردها و ایجاد ظرفیت‌هایی در حوزه علم و فناوری و هوش مصنوعی با رویکردی انسان‌محور تقویت کند [۱۶][۳][۴]. از این رو، این توصیه‌نامه سعی دارد هنجارهای اخلاقی را متناسب با اهداف خود در بین ملت‌ها مورد توجه قرار دهد.

بنابراین مفاد این توصیه‌نامه در قلمروی ارزش و هنجاری‌های اخلاقی بوده و هنجارهای اخلاقی نیز مبتنی بر یک‌سری مبانی و اصول هستند. در علم اخلاق، مکاتب متعدد اخلاقی وجود دارد که هرکدام مبتنی

[2] Artificial Narrow Intelligence (ANI)

[3] Reinforce UNESCO's leadership on global ethical reflection, standard setting, and building analytical capacities, to ensure human-centred progress in science and technologies, including in digital innovations and converging fields such as AI, gene-editing, and neuro-technologies; and build national capacities to maximize the benefits of these technologies and to address the associated risks. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

[۴] این متن در سند اصلی نیامده است ولی این عبارت در یکی از گزارش‌های پیش از متن نهایی تصویب شده، ذکر شده است. در این عبارت تصریح شده است که یونسکو سعی دارد رهبری خود درباره‌ی تنظیم استانداردها و ایجاد ظرفیت‌های اخلاقی را در جهان تقویت کند و زمینه‌ی پیشرفت انسان‌محور را در علم، فناوری، نوآوری‌های دیجیتال و زمینه‌های همگرا همچون هوش مصنوعی، ویرایش ژن‌ها و فناوری عصبی را به‌نفع بشریت فراهم آورد.

بر قواعد و اصولی می‌باشند. به‌طور کلی، مکاتب اخلاقی در برخی کتب فلسفه‌ی اخلاق به مکاتب واقع‌گرا و غیر واقع‌گرا تقسیم می‌شود.

از این رو برای بررسی سند مذکور نخست می‌بایست یک نگاه کلی به مکاتب اخلاقی داشته باشیم، سپس متن توصیه‌نامه را بر اساس آن ارزیابی کنیم. در این صورت، مشخص خواهد شد که منشأ ارزش‌های اخلاقی این سند، آیا برخاسته از دیدگاه‌های واقع‌گرایی و متناسب با نظریه اخلاقی مکتب اسلام است یا اینکه ارزش‌های اخلاقی مزبور متأثر از نظریه‌های غیرواقع‌گرایی و قراردادگرایی بوده و با ارزش‌های اخلاقی اسلام بسیار فاصله دارد. بنابراین، ضروری است مبانی و ارزش‌های اخلاقی این سند بین‌المللی با مبانی ارزش‌های اخلاق اسلامی مقایسه شود. پس مسئله‌ی اصلی این مقاله، بررسی سازگاری و یا ناسازگاری مبانی ارزشی سند اخلاق هوش مصنوعی یونسکو با نظریه‌ی اخلاقی مکتب اسلام می‌باشد. از این رو، عمده‌ی بحث ما ـ پس از تبیین اخلاق هوش مصنوعی و امکان اخلاق‌مندی ربات‌ها ـ تحلیل مبانی ارزش‌های سند مذکور و ارائه‌ی نظریه‌ی اخلاق هوش مصنوعی با استفاده از اندیشه‌ی برخی از فلیسوفان اسلامی می‌باشد. به همین دلیل با روش توصیفی و تحلیلی به کنکاش این سند بین‌المللی در حوزه‌ی اخلاق هوش مصنوعی می پردازیم.

۲ مروری بر کارهای دیگران

در خصوص کارهای مشابه، تنها پیشینه‌ای که به‌صورت رسمی و مدون یافت شد، گزارشی است که با عنوان «نقد و بررسی اولین پیش‌نویس توصیه‌نامه‌ی اخلاق هوش مصنوعی (یونسکو)» توسط بهزاد خداقلیزاده و همکاران و تحت نظر محمدمهدی نصرهرندی (مدیر گروه مطالعات اخلاقی) در پژوهشگاه مرکز ملی فضای مجازی چاپ شده است [۱]. در این گزارش نقدهایی به‌صورت بند به بند و نقدهایی کلی مطرح شده است، اما در آن هیچ اشاره‌ای به مبانی اخلاقی و سازگاری آن با نظریه‌ی اخلاقی اسلام نشده است. افزون بر این طبق فراخونی که دفتر فقه معاصر حوزه علمیه قم برای بررسی سند مزبور داد، تحقیقاتی در این راستا انجام شد که از میان آنها حدود ده مقاله برگزیده شده و در قالب کتاب به‌زودی انتشار خواهد یافت. در آن مقالات –به دلیل اینکه پیش از تصویب نهایی سند تدوین شده‌اند– با رویکردهای مختلف به نقد و بررسی نقاط قوت و ضعف سند پرداخته شده است و در ضمن آن پیشنهادهای تکمیلی نیز ارائه داده شده است. مقاله‌ای از نگارنده اول نیز در آن کتاب وجود دارد که در این پژوهش به برخی نتایج آن تحقیق اشاره خواهیم کرد. افزون برآن، در مقاله‌ی حاضر با زاویه دیدی متفاوت به بررسی سند اخلاق هوش مصنوعی خواهیم پرداخت.

۳ تبیین اخلاق هوش مصنوعی

به نظر می‌رسد برای تبیین و تحلیل بهتر، پیش از ورود به بررسی سند، بایسته و شایسته است به‌طور خلاصه مقصود از اخلاق هوش مصنوعی و رویکردهای مختلف در این زمینه تبیین شود. در یک رویکرد، خود هوش مصنوعی و ربات‌ها مورد خطاب باید و نباید قرار می‌گیرند. می‌توان «قوانین سه‌گانه‌ی رباتیک» که نخستین بار در سال ۱۹۴۲ توسط ایزاک آسیموف برای جلوگیری از آسیب‌رسانی ربات‌ها به انسان ارائه شد را از این قسم دانست [۱۵]. در رویکردی دیگر، اخلاق هوش مصنوعی بخشی از اخلاق ماشین می‌باشد. هدف اخلاق

ماشین این است که ماشینی می‌بایست ساخته شود که دست کم از یک اصل اخلاقی ایده‌آل و یا مجموعه‌ای از اصول اخلاقی پیروی کند [۱۲]. اما مراد از اخلاق هوش مصنوعی در رویکردی دیگر، این است که چگونه می‌توان کدهای اخلاقی را در سیستم‌های هوشمند تدوین کرد تا از خطرات احتمالی هوش مصنوعی علیه انسان‌ها جلوگیری شود [۱۳].

به نظر می‌رسد این نوع نگاه‌ها متأثر از دو گونه هدف‌گذاری مختلف برای مقوله‌ی هوش مصنوعی است. یک هدف‌گذاری، با اصطلاح هوش مصنوعی ضعیف[5] بیان می‌شود. در این هدف‌گذاری تنها معیار این است که کارکرد ایجاد شده در قالب هوش مصنوعی، به‌درستی ورضایت‌مندانه انجام شود. اما در هدف‌گذاری مقابل که با عنوان هوش مصنوعی قوی[6] از آن یاد می‌شود، رفتار هوش مصنوعی با رفتار انسان سنجیده می‌شود، در حالی که در هدف‌گذاری ضعیف با عملکرد سیستم سنجیده می‌شود. در هدف‌گذاری قوی به‌دنبال شبیه‌سازی سیستم ادراکی و آگاهی انسان در هوش مصنوعی هستند. افزون بر این، طرفداران هوش مصنوعی قوی بر این باورند که به‌دلیل دست‌یابی به اکتشافات، الگوریتم‌ها و دانش برنامه‌های نوین، هوش مصنوعی می‌تواند دارای حس هشیاری و آگاهی باشد [۱۴].

به نظر می‌رسد در توصیه‌نامه‌ی اخلاقی یونسکو، هدف‌گذاری ضعیف اتخاذ شده است. همان‌طور که در بند ۲، به‌خصوص بخش a تأکید شده است که نگاه این سند به سیستم‌های هوش مصنوعی به‌عنوان یک فناوری پردازش اطلاعات می‌باشد. افزون بر این، در بند ۲۶ اشاره می‌کند که هوش مصنوعی نباید تصمیم‌گیر نهایی در مسائل مهم همچون مرگ و زندگی باشد، بلکه اختیار آن باید به دست انسان‌ها باشد. افزون بر این، مراد از اخلاق هوش مصنوعی در توصیه‌نامه‌ی یونسکو بیشتر در ارتباط با رفتار انسان‌ها و بازیگران هوش مصنوعی می‌باشد، نه اینکه خود سیستم‌های هوشمند را مورد خطاب باید و نباید اخلاقی قرار دهد.

آیا بنا بر مبانی علم النفس اسلامی ربات‌ها می‌توانند به باید و نباید اخلاقی متصف شده و همانند انسان تصمیم‌های اخلاقی بگیرند؟ بررسی این مسئله نیازمند یک تحقیق و پژوهش مستقل و اساسی هست. از این رو، در اینجا به‌طور مختصر این مسئله را بررسی و نقد می‌کنیم. انجام رفتار اخلاقی متوقف بر آگاهی و درک معنای خوب و بد و مبتنی بر اختیار و اراده می‌باشد. اراده نیز مستلزم داشتن ذهن و یا نفس می‌باشد. در عین حال، بنابر مبانی علم النفس اسلامی، هوش مصنوعی نمی‌تواند ذهن و یا نفس خودآگاه داشته باشد. دلیل این امر به‌طور مختصر با تأمل در تعریف و حقیقت نفس در اندیشه‌ی حکمای اسلامی مشخص می‌شود. ابن سینا نفس را این‌گونه تعریف کرده است: «نفس کمال اول برای جسم طبیعی است» [۵]. بنا بر این تعریف، نفس یک حقیقت و کمال وجودی است که به جسم طبیعی تعلق پیدا می‌کند، نه به جسم مصنوعی. در اندیشه‌ی صدرالمتألهین بین ترکیب جسم مصنوعی و جسم طبیعی و آثار و احکام مترتب به آن تفاوت‌های بسیار عمیقی وجود دارد. در ترکیب صناعی، اجزاء به یکدیگر ضمیمه می‌شوند و با کنار هم قرار گرفتن آنها، شئ دارای یک حقیقت واحد نمی‌گردد. اما در ترکیب «طبیعی بالذات» شیء به‌تمام ذات تبدیل به شیء دیگری می‌شود، همچون جنین که تبدیل به انسان می‌شود [۶]. بنابراین، در هوش مصنوعی با ترکیب اجزاء آن، هر چند اجزاء موجود هستند و کارکردهای بسیاری نیز از ماشین پیچیده‌ی هوشمند بروز و ظهور پیدا

[5] Weak AI
[6] Strong AI

می‌کنند ولی بنابر علم النفس اسلامی، جسم مصنوع نمی‌تواند دارای ذهن (غیر فیزیکال) یا نفس مجرد شود تا بتواند از روی اختیار آگاهانه فعل اخلاقی را تشخیص داده و تصمیم به انجام آن بگیرد.

بنابراین، ربات و ماشین‌های هوشمند، نفس یا ذهن ندارند تا آگاهی و اراده‌ی آزاد در انجام افعال داشته باشند. از این رو، امکان اخلاق‌مندی ربات‌ها بدین معنا منتفی است. البته همان‌طور که در مقدمه و برخی بندهای توصیه‌نامه‌ی یونسکو آمده است باید هوش مصنوعی را به سمت ارزش‌های اخلاقی هدایت کرد. همچنین بعضی دانشگاه‌ها، مراکز و مؤسسات –همچون مرکز آینده‌ی اطلاعات در کمبریج یا مطالعه صدساله هوش مصنوعی AI۱۰۰ در دانشگاه استنفورد– به‌دنبال ابتکارات در زمینه‌ی به‌کارگیری کدهای اخلاقی در هوش مصنوعی هستند [۱۳]. بنابراین، می‌توان این ایده را درباره‌ی امکان اخلاق‌مندی ربات‌ها مطرح کرد: «می‌توان با به‌کاربردن الگوریتم‌ها و سیستم‌ها، داده‌ها و برنامه‌های مبتنی بر ارزش‌های اخلاقی اسلامی، از این سیستم‌های هوشمند به‌گونه‌ای در ترویج ارزش‌های اخلاقی و مقابله با مقوله‌های ضداخلاقی در فضای سایبر و فضای فیزیکی استفاده کرد».

۴ تحلیل منشأ ارزش اخلاقی توصیه‌نامه‌ی یونسکو و سازگاری آن با نظریه‌ی اخلاقی اسلام

اکنون نوبت آن است که ببینیم منشأ ارزش و لزوم اخلاقی به‌کاربرده شده در این توصیه‌نامه به کدام یک از مکاتب اخلاقی نزدیک است. به نظر می‌رسد منشأ ارزش و لزوم اخلاقی و هنجارهایی که دراین توصیه‌نامه به آن اشاره می‌شود، از نوع قراردادگرایی باشد. مکتب قراردادگرایی نیز ذیل مکاتب غیرواقع‌گرایی قرار می‌گیرد [۱۱]. با تأمل در برخی مفاد مقدمه‌ی سند مشخص می‌شود این سند حقوق بشر و آزادی‌ها را به تأیید و پذیرش مردم مقید کرده است. همچنین در بند ۶ بیان می‌شود: «هدف این توصیه‌نامه رسیدن به یک ابزار هنجاری پذیرفته شده در سطح جهانی است»؛ در نتیجه این مفاهیم را ملاک و منشأ هنجارهای اخلاقی معرفی می‌کند و براساس آنها «بایدها و نبایدهای اخلاقی» را در حوزه‌ی هوش مصنوعی معین می‌سازد. عبارت‌های مذکور ما را به این سمت سوق می‌دهد که بپذیریم برخی از مبانی به‌کاررفته در این متن برخاسته از مکاتب غیرواقع‌گرایی و از نوع قراردادگرایی هست.

اما قراردادگرایی چه اشکال و ایرادی دارد؟ مسئله نسبیت‌گرایی و کثرت‌گرایی برخی لوازم مهم قراردادگرایی هستند [۱۱]. با این وجود، مفاد بندهای ۱۳۱ تا ۱۳۴ که در حوزه‌ی نظارت و ارزیابی تأثیرات اخلاقی فناوری‌های هوش مصنوعی می‌باشد و سعی دارد روشی را برای این ارزیابی ارائه دهد، دچار مشکل اساسی می‌شوند. زیرا، بنابر کثرت‌گرایی، نمی‌توان روش واحدی را برای بررسی تأثیرات اخلاقی در کشورهای عضو ارائه داد.

مطلب دیگری که شایسته است در این کاوش به آن پرداخته شود، تبیین موضوع و هدف علم اخلاق می‌باشد. در فلسفه‌ی اخلاق دیدگاه‌های متفاوتی درباره‌ی موضوع و هدف اخلاقی وجود دارد. با تبیین برخی دیدگاه‌ها درباره‌ی موضوع و هدف علم اخلاق، رویکرد این توصیه در این مسئله نیز روشن می‌شود.

همان‌طور که گذشت ارزش‌های اخلاقی در این سند مبتنی بر نظریه‌ی قراردادگرایی است. نظریه‌ی

قراردادگرایی ذیل دیدگاه وظیفه‌گرایی اخلاقی[۷] می‌گنجد. در این دیدگاه، معیار ارزیابی اخلاقی، عمل به تکلیف و وظیفه است [۲]. نظریه‌ی «اخلاق وظیفه» دیدگاهی است که از سوی برخی اندیشمندان علم اخلاق مغرب زمین مطرح شده است و در آن ارزش‌های اخلاقی را صرفاً ناظر به رفتارها می‌دانند و علم اخلاق را بیان‌کننده‌ی ارزش و لزوم رفتارها معرفی می‌کند [۷] و هدف اصلی اخلاقی را نیز تصحیح خود رفتار دانسته و ورای آن هدف دیگری را نشان نمی‌دهند [۱۱].

اما در دیدگاه اخلاق فضیلت[۸] صفات درونی و نفسانی انسان متصف به ارزش‌های اخلاقی شده، و علم اخلاق عهده‌دار بیان ارزش و لزوم این صفات هست. هدف علم اخلاق در این رویکرد آراستن نفس به کمالات نفسانی پسندیده و زدودن صفات ناپسند از آن است. افزون براین، رفتارها و اعمال انسان که وسیله‌ای برای تحقق صفات نفسانی هستند، نیز در این رویکرد ارزش‌گذاری می‌شوند. به عبارت دیگر، هدف در این رویکرد دست‌یابی به کمال نفس است. برخی علمای شیعه براساس انسان‌شناسی خاص خود، همین دیدگاه را اختیار کرده‌اند. از نظر ایشان حقیقت وجود انسان، نفس مجرد اوست و کمال و نقص واقعی انسان نیز ناظر به همین نفس یا روح است. کمال و نقصی که با افعال اختیاری انسان در این دنیا به دست می‌آید در نفس او باقی می‌ماند و انسان پس از مرگ با همان‌ها محشور می‌شود [۸].

در رویکرد فضیلت‌محور، ارزش‌گذاری مربوط به اشخاص است نه به رفتار. اما این تأکید در فلسفه‌ی اخلاق نوین غربی جای خود را در رویکردی به‌شدت قانون‌پرستانه به کردار و رفتار داده است. از این رو، قواعد و اصول بر صدر نشسته‌اند [۳]. در واقع، وجه تمایز اصلی آنها این است که اخلاق فضلیت ناظر به ارزش‌های اخلاقی است و دیدگاه وظیفه‌گرا ناظر به الزام اخلاقی است. از این رو، مسئله‌ی اصلی در اخلاق فضیلت این است که «چگونه فردی باشیم» و در نظریه وظیفه‌محور دغدغه اصلی این است که «چه فعلی را باید انجام دهم» [۴]. بنابراین، یک فرد ممکن است از فضایل اخلاقی بی‌بهره باشد ولی به‌قصد انجام وظیفه یک عمل اخلاقی را انجام دهد.

از این رو، نقد مبنایی و اساسی به سند توصیه‌نامه‌ی اخلاق هوش مصنوعی یونسکو این است که مفاد این توصیه‌نامه به‌دلیل اینکه از قسم مکاتب قراردادگرایی هست، تابع یک مکتب اخلاقی غیرواقع‌گرا و وظیفه‌محور می‌باشد. پس ارزش‌های اخلاقی آن تنها متوجه کنش و رفتار ظاهری انسان است. در نتیجه، در این سند اخلاقی اصلاً به مباحثی همچون کمالات نفسانی، رابطه‌ی انسان با خدا، تأثیر واقعی ارزش‌های اخلاقی در روح و نفس انسان و مواردی از این دست، پرداخته نشده است. اما نظریه‌ی اخلاقی مکتب اسلام با رویکرد متعالیه در دیدگاه برخی فلیسوفان معاصر مسلمان یافت می‌شود. این دیدگاه نکات مثبت دیدگاه فضلیت‌گرایی اسلامی را در خود داشته و افزون برآن یک نوع دیدگاه غایت‌انگارانه و کمال‌جویانه می‌باشد. در این اندیشه، رفتار نیز موضوع علم اخلاق می‌باشد [۱۰]. زیرا برخی صفات نفسانی با تکرار رفتار اختیاری به‌دست می‌آید [۹]. در این اندیشه، هم رفتار اختیاری و هم صفات نفسانی مورد توجه است.

در نظریه‌ی اخلاقی اسلام، مفاهیم ارزش‌های اخلاقی مبتنی بر واقعیت هستند [۱۱]. از این رو این مکتب ذیل نظریه واقع‌گرا قرار می گیرد و همچنین این دیدگاه به‌نوعی دیدگاه غایت‌انگارانه و کمال‌جویانه

[7] Ethics of Duty
[8] Ethics of Virtue

به‌معنای خاص خود می‌باشد. هدف از فعل اختیاری، قرب وجودی به کمال مطلق و خداوند متعال می‌باشد. در این نظریه‌ی مبتکرانه، میان فعل اختیاری انسان و نتیجه‌ی آن رابطه‌ی ضرورت بالقیاس برقرار هست. از این رو، ارزش اخلاقی فعل اختیاری انسان تابع تأثیر آن فعل در رسیدن انسان به کمال حقیقی است. پس هرکاری به اندازه‌ای که در رسیدن به کمال حقیقی انسان مؤثر باشد، ارزنده خواهد بود و اگر تأثیر منفی داشته باشد ارزش آن نیز منفی خواهد بود و اگر هیچ ارزش مثبت و منفی نداشته باشد، خنثی خواهد بود [۱۰].

بنابراین، نظریه اخلاقی مکتب اسلام درباره ارزش‌های اخلاقی هوش مصنوعی را می‌توان این گونه تبیین کرد: «اگر چرخه‌ی حیات سیستم‌های هوش مصنوعی از اختراع و کشف گرفته تا کاربرد و استفاده از آن، در راستای این باشد که موجب کمال انسانی و تقرب وی به خداوند متعال باشد، این امر ارزشی اخلاقی خواهد داشت و بر عکس اگر فناوری‌های هوش مصنوعی موجب دوری انسان از کمال حقیقی خود باشد، ضد ارزش به حساب می‌آیند و در غیر این صورت، آن تکنولوژی خنثی بوده و دارای ارزش و یا ضد ارزش اخلاقی نمی‌باشد».

۵ نتیجه‌گیری

در گزارش [۱] به توصیه‌نامه‌ی یونسکو این ایراد گرفته می‌شود که در بیان نسبت ارزش‌ها و اصول باید در متن روشن‌تر عمل می‌کرده است و از کلی‌گویی‌های نابجا پرهیز می‌نموده است؛ از این رو منتقدان پیشنهاد داده‌اند که بند ۹۱ به عنوان الگویی برای نگارش سایر بندها باشد [۱]. در حالی که مفاد این بند با توجه به محتوای سایر بندها در راستای ترویج فرهنگ برابری جنسیتی می‌باشد. حتی پیشنهاد می شود در این سند تکثر ارزشی و فرهنگی گنجانده شود [۱]. اگر منظور از این عبارت این است که یونسکو به ارزشهای فرهنگ اسلام نیز بها بدهد، این نکته تا حدی قابل پذیرش است؛ اما قبول تکثر ارزشی منجر به نسبی‌گرایی و کثرت‌گرایی در ارزش‌های اخلاقی می‌شود و لازمه‌ی این امر عدم امکان ارزیابی و داوری میان مکاتب مختلف اخلاقی بوده و همه به یک اندازه معتبر می‌باشند. در حالی که، تحقیق حاضر به لحاظ مبانی این سند اخلاقی را مورد بررسی قرار داد. و مشخص شد که این سند مبتنی بر قراردادگرایی بوده و در این مکتب، هدف عمل به وظیفه بدون توجه به صفات کمالی و حقیقی انسان است. اما نظریه‌ی اخلاقی اسلام جزو مکاتب واقع‌گرایی و در اندیشه‌ی برخی علما فضلیت‌محور و از منظر برخی غایت‌گرایانه و کمال‌جویانه می‌باشد. در این رویکرد افزون بر اصلاح رفتار و صفات نفسانی، تقرب به کمال مطلق و خداوند متعال مد نظر است.

از این رو پیشنهاد می‌شود در قراردادهای بین‌المللی، بررسی اسناد، اعلام نظرها و توافق‌ها و به‌طور خاص در اجرای آنها در ایران اسلامی دقت لازم به‌کار رود و سازگاری آن با مبانی اندیشه‌ی اسلامی لحاظ شود؛ همچنان‌که مقام معظم رهبری (مدظله) می‌فرمایند ما در بحث فناوری باید پیشرفت کنیم و در این بخش شاگردی و علم‌آموزی کنیم، اما نباید به توصیه‌ها و برنامه‌های آنها [غربی‌ها] خوش‌بین بوده و عمل کنیم[۹]. افزون براین، پیشنهاد می‌شود در تدوین «سند ملی هوش مصنوعی کشور» متناسب با ارزش‌ها و چارچوب‌های تمدن نوین اسلامی برای حوزه‌ی اخلاق هوش مصنوعی ریل‌گذاری شود.

[۹] بیانات در دیدار مسئولان و محققان ستاد توسعه علوم شناختی با رهبر معظم انقلاب، https://khl.ink/f/41471

مراجع

[۱] ب، خداقلیزاده. نقد و بررسی اولین پیش‌نویس توصیه‌نامه اخلاق هوش مصنوعی (یونسکو). پژوهشگاه فضای مجازی، تهران، ۱۳۹۹.

[۲] ح، حمدی، ح. و محیطی. ظرفیت‌سنجی اندیشه‌های اخلاقی در ایجاد نظام ارزشی تمدن نوین اسلامی، مطالعات بنیادین تمدن نوین اسلامی، ۲(۱)، ۱۳۹۸.

[۳] ر، هولمز. مبانی فلسفه اخلاق هولمز، ترجمه‌ی مسعود علیا. ققنوس، تهران، ۱۳۸۹.

[۴] ز، خزاعی. اخلاق فضیلت. حکمت، تهران، ۱۳۸۹.

[۵] ابن سینا. الشفاء (الطبیعیات)، مکتبة آیت‌الله المرعشی، قم، ۱۴۰۴ق.

[۶] صدرالمتألهین. المتعالیه فی الاسفار العقلیه الاربعه، ویرایش ۳. دار احیاء،بیروت، ۱۹۸۱م.

[۷] ژکس. فلسفه اخلاق (حمکت عملی)، ترجمه‌ی ابوالقاسم پورحسینی، ویرایش ۲. امیرکبیر، تهران، ۱۳۶۲.

[۸] م، نراقی. جامع السعادات، تحقیق محمد کلانتر. موسسه الاعلمی للمطبوعات، بیروت، ۱۲۰۹ق.

[۹] م.ت، مصباح یزدی. دروس فلسفه اخلاق. اطلاعات، تهران، ۱۳۷۳.

[۱۰] م.ت، مصباح یزدی. نقد و بررسی مکاتب اخلاقی، تحقیق و نگارش احمدحسین شریفی. مؤسسه آموزشی و پژوهشی امام خمینی، قم، ۱۳۸۴.

[۱۱] م، مصباح. بنیاد اخلاق (روشی نو در آموزش فلسفه اخلاق)، ویرایش ۷. مؤسسه آموزشی و پژوهشی امام خمینی، قم، ۱۳۹۰.

[12] Anderson, M., and Anderson, S. L. Machine ethics: Creating an ethical intelligent agent. AI magazine, 28(4), 2007.

[13] Boddington, P. *Towards a code of ethics for artificial intelligence*. Springer, 2017.

[14] Lucci, S., and Kopec, D. *Artificial intelligence in the 21st century: A living introduction, mercury learning and information*. 2016.

[15] Russell, S., and Norvig, P. *Artificial Intelligence: A Modern Approach*, 4th ed., Prentice Hall, 2020.

[16] UNESCO. Recommendation on the ethics of artificial intelligence. https://en.unesco.org/artificial-intelligence/ethics, Ret: 6 Feb 2022.