

حکم و مسئولیت اخلاقی در هوش مصنوعی بر مبنای آراء اخلاقی ملاصدرا

فروغ السادات رحیم پور^۱، مرجانه حاجیان^۲

چکیده

تحقیقات گذشته درباره پیوند بین هوش مصنوعی و اخلاق، تا حد زیادی بر برنامه‌ریزی یا استفاده انسان از هوش مصنوعی متمرکز بوده است. با پیشرفت هوش مصنوعی و امید به خودمختار شدن این فناوری، بنظر میرسد زمان آن رسیده که ماشینهای هوشمند، بهره‌مند از بُعد اخلاق شوند. گام اول برای این پروژه، عبارتست از ایجاد ماشینی که از یک اصل یا مجموعه‌یی از اصول اخلاقی پیروی کند. این مقاله با روش توصیفی-تحلیلی و بشیوه مسئله‌محور، درصدد است احکام و اصول اخلاقی در اندیشه ملاصدرا را بررسی کرده و امکان بکارگیری آن برای هوش مصنوعی را ارزیابی نماید. از نظر ملاصدرا، اگرچه اصول

اخلاقی ثابتند، ولی میتوان در شرایط جدید با محوریت همان اصول ثابت، احکامی جدید را که قبلاً انشاء نشده بود، استنباط نمود. پس باید واکاوی شود که آیا میتوان از هوش مصنوعی بعنوان مشاوره برای استنباط احکام اخلاقی جدید بهره گرفت یا خیر؟ برای تحقق چنین مطلوبی باید اصول اولیه اخلاقی و الگوریتمهایی که همان مراحل پیدایش احکام اخلاقی هستند، برای ماشین تعریف شوند تا ماشین با استفاده از آنها بتواند احکام اخلاقی جدید را صادر کند. از نگاه ملاصدرا، برای آنکه احکام اخلاقی اجرایی گردند، باید الزامی وجود داشته باشد و این الزام، باعث بوجود آمدن مسئولیت اخلاقی میشود. کسی دارای مسئولیت اخلاقی است که از شرایطی چون

۱. دانشیار گروه فلسفه و کلام اسلامی، دانشگاه اصفهان، اصفهان، ایران (نویسنده مسئول): f.rahimpour@ltr.ui.ac.ir

۲. دانش‌آموخته کارشناس ارشد فلسفه و کلام اسلامی، دانشگاه اصفهان، اصفهان، ایران؛ m.hajian987@gmail.com

تاریخ دریافت: ۱۴۰۳/۷/۳۰ تاریخ پذیرش: ۱۴۰۴/۳/۱۰ نوع مقاله: پژوهشی

DOR: 20.1001.1.15600874.1404.31.2.1.7

اختیار، قدرت و علم برخوردار باشد. بررسی امکان تحقق این شرایط در ماشینهای هوشمند و مسئولیت‌پذیری آنها، مسئله دیگر پژوهش پیش‌رو خواهد بود.

کلیدواژگان: هوش مصنوعی، احکام اخلاقی، مسئولیت اخلاقی، ملاصدرا.

مقدمه

با اختراع کامپیوتر در اواسط قرن بیستم میلادی، برخی دانشمندان در پی آن بودند که هوش انسان را با استفاده از ماشین شبیه‌سازی کنند. این رشته نوظهور، هوش مصنوعی نام گرفت. بعد از چند دهه تلاش و پیشرفت چشمگیر ماشینهای هوشمند، این پرسش ایجاد شد که آیا این ماشینها میتوانند در کنار تمام دستاوردهای مثبت خود، تهدیدی برای جامعه بشری باشند؟ بهمین دلیل، جستجوی راه‌حل برای پیشگیری از تهدیدهای هوش مصنوعی آغاز گردید. یکی از راه‌حلهایی که در سالهای اخیر مورد توجه قرار گرفت، افزودن بُعد اخلاق به نظام برنامه‌ریزی این ماشینها بود.

متخصصان، سه رویکرد متفاوت برای اجرای عاملهای اخلاقی مصنوعی اتخاذ کرده‌اند:

۱. رویکرد بالا به پایین: در این رویکرد، یک نظریه اخلاقی بعنوان بخشی از اطلاعات

اولیه عامل تعریف میشود و عامل در شرایط مختلف، با توجه به آن تصمیم‌گیری و عمل میکند. تحقیقاتی که تاکنون انجام شده، اثبات کرده‌اند که چنین رویکردهایی برای کار در دنیای واقعی به اندازه کافی قوی نیستند.

۲. رویکرد پایین به بالا: در این رویکرد، نظریه اخلاقی خاصی تحمیل نمیشود، بلکه ایجاد محیطی که در آن رفتار مناسب انتخاب گردد، دنبال میشود و عامل، تصمیمات اخلاقی خود را براساس فرایند یادگیری استوار میسازد. پس از فرایند یادگیری اولیه، سیستم باید بتواند معضلات اخلاقی جدید را بتنهایی حل کند.

۳. رویکرد ترکیبی: در این رویکرد سعی میشود با توسل به ترکیبی پیچیده از رویکرد بالا به پایین و پایین به بالا، تصمیم‌گیری انسانی شبیه‌سازی گردد (Allen & Wallach, 2005: pp. 151-153).

برای آنکه بتوانیم مشاوره اخلاقی و سپس ماشینی اخلاق‌مند داشته باشیم، باید حکم و مسئولیت اخلاقی را بررسی کنیم؛ چراکه چنین ماشینی نیازمند پایگاه داده‌ی (Database) از اصول اولیه اخلاقی است تا با استفاده از این اصول، بتواند اصول دیگری را در موقعیتهای مختلف استنتاج کند و از آنها پیروی نماید. برای رسیدن به حکم اخلاقی باید به شناخت خوب و بد اخلاقی پرداخت؛ اما ملاک خوب و بد بودن عمل در نظامهای اخلاقی مختلف، متفاوت است و بسته به اینکه چه نظام

اخلاقی‌بی ملاک کار هوش مصنوعی قرار میگیرد، عملکرد ماشین متفاوت خواهد شد. نوشتار حاضر درصدد است نظام اخلاقی مورد نظر ملاصدرا را بررسی و قابلیت آن را برای عرضه به پروژه اخلاقمند کردن هوش مصنوعی بیازماید. صدرالمتهلین مطابقت و عدم مطابقت با فضیلت را ملاک حسن و قبح عمل میدانند و معتقد است فضیلت یا رذیلت بودن یک صفت منوط به اینست که موجب ازدیاد یا باعث زوال کمالات شود (ملاصدرا، ۱۳۸۳ج: ۲۰۴).

یکی از توابع حکم اخلاقی، مسئولیت اخلاقی است. مسئولیت اخلاقی بمعنای الزام به انجام افعالی است که حکم اخلاقی بر آن واقع شده است. از آثار ملاصدرا چنین برمی آید که زمانی انسان واجد مسئولیت اخلاقی است که دارای شرایطی همچون مختار بودن، قادر بودن و عالم بودن باشد. پس برای اینکه یک عامل اخلاقی بتواند مسئولیت اخلاقی افعال خود را بپذیرد و به آن ملزم باشد، باید شرایطی را برای آن بوجود آورد که ملاصدرا آن شرایط را برای انسان الزامی میدانند. بهمین دلیل، عامل هوشمند، باید مختار، قادر و عالم باشد.

پیشینه پژوهش

در زمینه حکم و مسئولیت اخلاقی از نظر ملاصدرا پژوهشهایی صورت گرفته که از میان آنها میتوان به کتاب درآمدی بر اخلاق فلسفی در

حکمت متعالیه (طوسی، ۱۳۹۲) اشاره کرد. در این کتاب، نویسنده اخلاق را با اتکا بر مبانی هستی‌شناسی و انسان‌شناسی ملاصدرا بررسی نموده، سپس به تبیین تأثیرات علم و عمل بر تکامل و سعادت انسان پرداخته است. در بخش دیگر، با توجه به فلسفه اخلاق، درباره اوصاف اخلاقی، باید و نبایدها در علم اخلاق و منشأ آنها سخن میگوید. مقاله «مسئولیت اخلاقی در نگاه ملاصدرا» (طاهریان دهکردی و دیگران، ۱۳۹۷)، با هدف یافتن دیدگاه ملاصدرا درباره انسان و مسئولیتهای اخلاقی او در برابر خود، خدا و دیگران، به بررسی مفهوم و عوامل احساس مسئولیت اخلاقی، لوازم مسئولیت اخلاقی، شرایط مسئولیت اخلاقی و تبیین مسئولیت اخلاقی از نگاه ملاصدرا پرداخته است. همچنین میتوان به مقاله دیگری با عنوان «موانع اخلاق‌ورزی از دیدگاه ملاصدرا» (دهقانی سیمکانی، ۱۳۹۱) اشاره کرد. این مقاله بدنبال پاسخ به این پرسش است که علت بی‌اخلاقیها چیست و چه عواملی باعث میشود عقل مغلوب جهل شده و در پی آن، بی‌اخلاقی بروز کند.

در منابع مربوط به هوش مصنوعی، یکی از مقالاتی که به نوعی با موضوع این پژوهش ارتباط دارد مقاله «The status of machine ethics» (Anderson, 2007) است. در این پژوهش،

نویسندگان پیشنهاد میکنند اصولی را به ماشین ارائه کنیم تا بر اساس آنها عمل اخلاقی را تشخیص دهد و حتی بتواند این اصول را گسترش دهد. بر همین اساس، ماشینهایی را نیز عرضه میکند که بعنوان مشاور اخلاقی عمل میکنند. در مقاله «Building ethics into artificial intelligence» (Yu et al., 2018) نیز نویسندگان پیشنهاد میدهند برای دستیابی به ماشین اخلاقمند، زمینه را به چهار قسمت تقسیم کرده و در هر زمینه معضلات اخلاقی بررسی و قوانینی اولیه صادر شود. از این احکام، ماشین میتواند برای موارد بعدی استدلال کرده و قوانینی جدید ارائه دهد.

ملاصدرا و حکم اخلاقی

نظریات اخلاقی به سه دسته وظیفه‌گرا، سودگرا و فضیلت‌گرا تقسیم میشود و ملاصدرا از این میان، نظریه فضیلت‌گرایی را برگزیده است. از نظر او، فضیلت بمعنای رعایت حد اعتدال است، زیرا هرکدام از خلقیات دارای دو طرف افراط و تفریط هستند. افراط آنها بدلیل حصول ضرر برای دیگران و تفریط آن بسبب فقدان یا کمبود خیر، رذیلت محسوب میشود، اما حد وسط آن فضیلت بشمار می‌آید. فقط درمورد حکمت بمعنای ادراک کلیات است که ملاصدرا معتقد است هرچه بیشتر باشد، فضیلتش بیشتر است و افراط آن رذیلت نیست (ملاصدرا، ۱۳۸۳ج: ۲۰۵).

خوب و بد اخلاقی از نظر ملاصدرا

برای رسیدن به حکم اخلاقی، ابتدا باید خوب و بد اخلاقی بررسی شوند. از نظر صدرالمتألهین ملاک خوب و بد اعمال، مطابق بودن یا نبودن آنها با فضیلت است. عملی خوب است که مطابق فضیلت باشد و عملی بد است که مطابق فضیلت نباشد، یا عبارتی، مطابق رذیلت باشد؛ اما اینکه چه چیزی فضیلت است و چه چیزی رذیلت، متناسب با هدف نفسی است که فعل از آن صادر میشود (همان: ۲۰۴). برای بیان تضاد بین فضیلت و رذیلت، ملاصدرا شجاعت و بزدلی را مثال میزند و میگوید:

شجاعت ذیل فضیلت و بیباکی که ضد آن است، ذیل رذیلت قرار میگیرد و این امر با این قاعده (که بیشتر به آن اشاره شد و مطابق آن تضاد در انواع اخیر تحت یک جنس - که در نهایت اختلاف هستند - روی میدهد) منافاتی ندارد، چراکه شجاعت و بزدلی، هر یک فی‌نفسه کیف نفسانیند و فضیلت و رذیلت بودنشان عرضی است، نه اینکه مقوم آنها باشد. پس فضیلت و رذیلت جنس فضایل و رذائل نفسانی نیستند. شجاعت نه ضد بیباکی است نه ضد بزدلی، زیرا واسطه بین آنهاست. اما دو طرف شجاعت (یعنی بیباکی و بزدلی) چون نهایت دوری از هم را دارند، متضاد یکدیگرند. تضاد شجاعت با هر یک از دو

طرف مذکور، تضاد بالعرض است، زیرا ماهیت شجاعت یعنی فضیلت بودن، عارض آن است و ماهیت بیباکی و بزدلی، یعنی رذیلت شمرده شدن، هم عارض آن دو است و تضاد در حقیقت بین دو امر عارضی واقع شده است (همو، ۱۳۸۰ الف: ۱۱۶-۱۱۵).

این قاعده را میتوان در مورد سایر ملکات نیز بکار برد و گفت تقابل فضایل و رذائل تضاد بالعرض است. همچنین بنظر میرسد برای تطابق فعل با فضیلت و رذیلت نمیتوان مبنایی هستی‌شناسانه یافت، زیرا تطابق فعل با فضیلت یا رذیلت، جزئی از ساختار هستی نیست. بعبارتی، نفس‌الامر و مطابق یک فعل برای خوب یا بد بودن، فضیلت و رذیلت است و این تطابق اگرچه امری حقیقی است (که در آن مطابقت داشتن با فضیلت، ارزشی مثبت و مطابقت نداشتن ارزشی منفی دارد)، اما مابازاء خارجی در هستی ندارد. اساساً، فضیلت وجود خارجی جداگانه‌یی از فرد فضیلت‌مند ندارد تا بتوان با تطابق آن با افعال، آن را بررسی هستی‌شناسانه کرد. این بدان معناست که چیز جدای از انسان در عالم خارج با عنوان شجاعت، حکمت و... وجود ندارد تا تطابق با آن مابازاء خارجی داشته باشد. بنابراین میتوان گفت این مسئله وارد حیطه معرفت‌شناسی میشود؛ اما میتوان بنیاد

هستی‌شناسانه این مسئله را در ارتباط خوب و بد با فضیلت و رذیلت جستجو کرد، نه تطابق میان فعل با فضیلت یا رذیلت.

نیک و بد یک عمل با تفاوت فاعل و قابل، متفاوت میشود. مثلاً تنبیه خطاکار نیکوست، ولی تنبیه فرد بیگناه ناپسند است. این تفاوت به اعتبار مراتب نفس انسان هم متباین است؛ بدین معنا که افعالی مانند شهوترانی برای مرتبه حیوانی نفس خیر است، اما در مقایسه با مرتبه عالی آن، یعنی ناطقه، قبیح است (همو، ۱۳۸۵ الف: ۳۴۲). از عبارات او میتوان دریافت که خوب و بد از نظر او عقلی است. اگر عقل سلیم به فعلی دعوت کند، آن فعل حتماً خوب است و اگر از فعلی نهی کند، قطعاً بد است (همو، ۱۳۸۹ ب: ۱۳۳۵). گاهی عقل بدون کسب و اکتساب به خوب و بد عملی علم پیدا میکند و گاهی هم نیازمند کسب و اکتساب است (همو، ۱۳۸۳ ب: ۴۵۲). بعنوان نمونه، ملاصدرا با اشاره به آثار شرب خمر میگوید: اگر شخصی عاقل باشد و از این آثار مطلع گردد، از این کار کناره‌گیری میکند، حتی اگر شارع آن را حرام نکند (همو، ۱۳۸۳ ج: ۲۷۴). برای آنکه انسان بتواند اوصاف اخلاقی را آنگونه که هست درک کند، باید به صفات عدالت و حکمت متصف باشد:

- حکمت بمعنای ادراک کلیات و صور عقلانی ثابت که کمال انسان بما هو انسان است و هرچه افزایش یابد، بهتر است (همان: ۲۰۴).
- عدالت عبارتست از اعتدال صفات باطنی نفس؛ یعنی غضب، شهوت و عقل. حاصل این اعتدال، باطن زیبا و نیک است، همانطورکه ظاهر زیبا و نیک، حاصل اعتدال اجزاء صورت است (همو، ۱۳۸۲ الف: ۱۷۱-۱۷۲).

طبع و عقل، مبدأ باید و نبایدهای اخلاقی

بر اساس دیدگاه ملاصدرا، انسان بواسطه عقل نظری میتواند خوب و بد افعال را درک کند و بعد از آن توسط عقل عملی، قوای مادون (نباتی و حیوانی) را برای رسیدن به اهداف انسانی بکار گیرد. در روند بکارگیری این قواست که باید و نبایدهای اخلاقی شکل میگیرد.

نفس انسان بمنظور رسیدن به کمال در هر مرتبه، الزاماتی را متناسب با آن مرتبه ایجاد میکند. افعال انسان برای نیل به کمال در مراتب گوناگون وجودی، متفاوت است؛ به این معنا که در یک مرتبه، افعال او اختیاری و مسبوق به تصور و تصدیق است، درحالیکه در مرتبه دیگر، مستند به علم حضوری و بینای از تصور و تصدیق و اعمال اختیار است. بگفته ملاصدرا، علم انسان به ذات خود، همچنین به قوای بدن و اعضا و آلات مربوط به حواس

ظاهری و باطنی، حضوری است و بکارگیری حواس و اعضای مختص به آنها به تصور «فعل بکارگیری اعضا» و تصدیق به «مفید بودن این فعل» نیاز ندارد. چنین فعلی، فعل اختیاری بشمار نمیرود، بلکه فعلی منبعث از ذات نفس است که از علم به ذات و حب ذات نشئت میگیرد (همو، ۱۳۸۱ ب: ۱۳۳-۱۳۴).^۲ در مقابل، انجام برخی از افعال نیازمند تفکر و تأمل و اعمال اختیار است. اینجاست که ملاصدرا میگوید: «هرگاه تصور چیزی قبل از موجود شدن آن حاصل شود و یکی از دو طرف امکانش رجحان یابد، اراده مختص آن طرف (رجحان یافته)، حاصل میشود. پس ترجیح یافتن (که حاصل تفکر است) مقدم بر اراده کردن است (همو، ۱۳۸۰ الف: ۲۸۳).^۳ این عملیات فکری، موجب ساخته شدن «باید و نباید اخلاقی» میشود.

بنابراین، میتوان گفت اراده انسان در مرتبه نباتی و حیوانی تابع طبیعت اوست و این طبیعت اوست که به باید و نباید فرمان میدهد (البته این بایدها و نبایدها جسمانی بوده و تأمین کننده اخلاق نیستند). در مرتبه عقلانی است که نفس با تشخیص ملاکهای عقلانی میتواند از لذایذ جسمانی فراتر برود و بایدها و نبایدهای اخلاقی را صادر کند (همو، ۱۳۸۳: ۱۴۹-۱۵۴).

در مسئله باید و نباید اخلاقی، الزام‌کننده و الزام‌شونده، نفس انسان است و این الزامات برای همه انسانها یکسان است، زیرا انسانها در نیازهای جسمی و روحی مشترکند (همو، ۱۳۸۰: ۲۵۷).

احکام اخلاقی و مراحل پیدایش آنها

احکام و قوانین اخلاقی قبل از آنکه صورت باید و نباید عملی بخود بگیرند، باید چهار مرحله را طی کنند:

۱. تحقق مصلحت حکم: اقتضای حکم بودن و نقطه شروع احکام اخلاقی، ملاحظه مصلحت آن حکم است؛ خواه این مصلحت فردی باشد یا جمعی، خواه مادی باشد یا معنوی. درباره چگونگی شکلگیری احکام برحسب مصلحت آنها، میتوان گفت: ملاک تفاضل اعمال، مقدار تأثیری است که هر عمل بر اکتساب اخلاق میگذارد. ملاصدرا در اینباره معتقد است: برخی افعال باعث تاریکی قلب میشوند که آنها را معصیت گویند و برخی باعث صفای دل هستند که آنها را طاعت مینامند. پس ارزش هر فعل به میزان تأثیری است که بر اخلاق میگذارد و این میزان تأثیر، مصلحت فعل مذکور را مشخص میکند (همو، ۱۳۸۱: ۷۹).

۲. قابلیت انشاء: بدون وجود مصلحت، حکمی بوجود نمی‌آید، اما صرف وجود

مصلحت هم موجب پدید آمدن حکم نیست، بلکه لازم است حکم قابلیت انشاء داشته باشد. از آنجاکه هدف از حیات دنیوی سیر بسوی خداوند است و نفس انسان باید در این سیر انسانیت خود را حفظ کند تا به مطلوب حقیقی برسد، پس نباید به حال خود واگذار گردد، بلکه با وضع و انشاء قوانین اخلاقی باید از جدال و انحراف او جلوگیری شود. مقصود انشاء قوانین و قواعد شریعت که همواره اخلاقی هستند، تأمین نفع عامه مردم و مصلحت آنهاست (همو، ۱۳۸۲: ۳۱۳).

۳. قابلیت عملی شدن: علاوه بر تحقق مصلحت و قابلیت انشاء، حکم باید قابلیت عملی شدن نیز داشته باشد و برای اجرایی شدن، ضروری است از صلاحیت اظهار و اعلان نیز برخوردار باشد. از نظر ملاصدرا مهمترین ویژگی حکم برای آنکه اظهار شود و منشأ مسئولیت قرار گیرد، اینست که آن حکم نفع عامه مردم را تأمین کند (همو، ۱۳۸۱: ج: ۲۲۷).

۴. فراهم بودن شرایط عمل به حکم: پس از اظهار احکام اخلاقی، باید در انسان مکلف شرایطی باشد تا بتواند به این احکام عمل کند و در مقابل آنها مسئولیت داشته باشد. ملاصدرا مهمترین شرایط را علم، قدرت، عقل و اختیار میداند. همچنین دعوت الهی، ارشاد، تهذیب، وعد و وعید را بعنوان شرایط خارج از وجود انسان برمی‌شمارد (همو، ۱۳۸۹: الف: ۲۰۵).

مطلق یا نسبی بودن احکام اخلاقی

یکی از مسائل مهم در مبحث احکام اخلاقی، پاسخ به این پرسش است که این احکام در تمامی شرایط و برای همه انسانها، یکسان و ثابتند یا خیر؟ بتعبیر دیگر، احکام اخلاقی مطلق هستند یا نسبی؟ نسبی بودن اخلاق، نظریه‌ی است که برخی نظیر سوفیستها به آن معتقدند. مطابق این باور، اخلاق قراردادی است برای آسانتر کردن زندگی اجتماعی که متناسب با شرایط مختلف، تغییرپذیر است. در مقابل، برخی مانند ملاصدرا اصول اخلاقی را مطلق دانسته و معتقدند اگر برخی اصول اخلاق را ثابت نمی‌یابیم، بدلیل نسبی بودن آنها نیست، بلکه به این خاطر است که تعیین یک مصلحت از میان مصالح گوناگون و گاه متعارض، باعث میشود مصلحتی نادیده گرفته شود.

ملاصدرا با اتکا به سه اصل زیر، مطلق بودن احکام اخلاقی را پذیرفته است:

۱. واحد بودن کمال همه انسانها: از منظر ملاصدرا کمال هر انسان در ظهور آثار خاص اوست و از آنجاکه انسان بواسطه قوه عاقله از موجودات دیگر متمایز میشود، کمال خاصش در بروز و ظهور آثار قوه عاقله است. آثار این قوه از حیث اندیشه و نظر، ادراک معقولات و از حیث عمل، تشخیص خوب از بد است

(همو، ۱۳۸۱ج: ۱۶۱-۱۶۰). بر این اساس، از آنجاکه کمال همه انسانها واحد است، پس احکام و قوانین اخلاقی نیز باید واحد باشند تا انسان را به کمال برسانند.

۲. واحد بودن مدرکات عقلی انسانها: ادراک انسان دارای سه مرتبه است: گاهی در مرتبه حسی است، گاه به مرتبه بالاتر، یعنی مرتبه مثالی و خیالی ارتقا می‌یابد و بالاترین مرتبه آن، ادراک عقلی است. در دو مرحله اول، تفاوت مدرکات مقبول و مشهود است، اما هنگام درک حقایق معقول (بسبب ارتباط با عقل فعال و صور عقلانی آن) همگی به حقیقتی واحد میرسند و در این مرحله میتوان گفت صاحب مدرکاتی واحد و یکسانند (همو، ۱۳۸۰ب: ۵).

۳. رعایت صلاح نظام کلی: از نظر صدرالمتألهین، مقصود خداوند متعال از اظهار دین و قواعد شریعت (که زیربنای احکام اخلاقی هستند) نفع عمومی و مصلحت کلی انسانهایی است که تا روز قیامت به دنیا می‌آیند و تعداد آنها و سعادت و خیری که برایشان حاصل میشود، بشمار است (همان: ۱۳۳).

امکان‌پذیری اعتبار و استنباط احکام اخلاقی جدید

چنانکه گذشت، ملاصدرا اصول اخلاقی را اموری مطلق میداند. حال به این پرسش

میپردازیم که آیا از آن اصول مطلق میتوان احکامی جدید و پویا استنباط کرد؟ بیان دیگر، آیا میتوان حکم اخلاقی جدیدی اعتبار نمود یا خیر؟

در اینباره دو احتمال وجود دارد:

۱. احکام اخلاقی منحصر به همان اصول و قواعد ثابت هستند و امکان استنباط احکام جدید وجود ندارد.

۲. از نظر ملاصدرا، پس از ختم نبوت و انقطاع وحی، پس از ائمه معصومین (علیهم‌السلام)، مجتهدین (دربارۀ احکام دین) از طریق اجتهاد فتوا میدهند (همو، ۱۳۸۲ب: ۴۳۸). اصول اخلاق هم مانند احکام دین، میتوانند سرچشمۀ اعتبار و استنباط احکامی جدید باشند و احکام متناسب با شرایط زمانی و مکانی مختلف را بوجود آورند.

ملاصدرا معتقد است شکلگیری امور اعتباری، بخاطر ضرورتی است که نسبت به وجود آنها احساس میشود. آنچه اعتبار میگردد گرچه در عالم خارج یافت نمیشود، اما باید بگونه‌یی باشد که بر حقیقت خارجی دلالت کند و اگرچه خود، حقیقتی عینی نیست، ولی نباید چیزی جدا از آن باشد، بلکه لازم است بنحوی حکایتگر عالم عین باشد (همو، ۱۳۸۵ب: ۷۵۸).

درمورد دیدگاه ملاصدرا درباره اعتبار احکام اخلاقی، باید به نظریۀ او درباره دلیل بوجود

آمدن اخلاق توجه کرد. مطابق نظر وی، انسان از مراحل گوناگون هیولانی، جسمی، جمادی، نباتی، شهوی، غضبی، احساسی، تخیلی و توهمی‌گذر میکند و به مرتبۀ انسانی میرسد (همو، ۱۳۸۲ب: ۴۲۲). همه انسانها در طی این مسیر مشترکند. امر مشترک دیگر بین انسانها مقصدی است که همه در پی رسیدن به آن هستند. ملاصدرا میگوید: «غرض از خلقت و آفرینش، سوق دادن همه افراد بشر به جوار قرب الهی و دار کرامت اوست تا مشمول رحمت گردند» (همو: ۴۲۳). اعتبارکننده اخلاق باید احکامی را اعتبار کند که اسباب رسیدن به این مقصد را فراهم آورد و بتواند انسانها را به کمال مطلوب برساند.

حکم اخلاقی در هوش مصنوعی

پس از آنکه مراحل پیدایش احکام اخلاقی از نظر ملاصدرا روشن شد، باید بسراغ هوش مصنوعی برویم و ببینیم تا چه حد میتوان این مراحل را در ماشین اجرا کرد و بعنوان یک مشاور اخلاقی از آن کمک گرفت. اندرسون هدف نهایی اخلاق ماشین را ایجاد ماشینی میداند که از یک اصل اخلاقی ایده‌آل یا مجموعه‌یی از اصول پیروی میکند (در اقداماتی که ممکن است انجام دهد توسط این اصول هدایت شود). او دستیابی به این هدف را فراتر از ایجاد محدودیت در رفتار ماشین بر اساس

اصول اخلاقی در نظر می‌گیرد. وی این هدف را شامل ایجاد ماشینی میدانند که میتواند با استفاده از اصول اخلاقی دربارهٔ مسائلی که قبلاً برایش پیش نیامده است، استدلال کند. او در جایی دیگر به سه قانونی که آیزاک آسیموف برای ایجاد اخلاق در ماشین پیشنهاد کرده بود، اشاره میکند. سه قانون آسیموف عبارتند از:

۱. ربات نباید به انسان آسیب برساند، یا با عدم اقدام، اجازه آسیب رسیدن به انسان را بدهد.

۲. ربات باید از دستورات انسان پیروی کند، مگر در مواردی که این دستورات با قانون اول در تضاد باشد.

۳. ربات باید از وجود خود محافظت کند، تا زمانی که این حفاظت با قانون اول و دوم مغایرت نداشته باشد.

اندرسون معتقد است وضع این سه قانون برای ایجاد اخلاق ماشینها، وظایف غیرمستقیم انسان را نقض میکند، زیرا از آنجاییکه هدف هوش مصنوعی ایجاد موجوداتی است که بتوانند مانند انسان رفتار کنند (اگرچه کاملاً مانند آنها نباشند)، این احتمال نیز وجود دارد که ربات اخلاقی خودمختار، تا حد قابل توجهی به انسان شباهت داشته باشد؛ پس اگر انسانها قوانینی را وضع کنند که به آنها اجازه بدر رفتاری با ماشینهایی که شبیه انسانها هستند را بدهد، احتمال بدر رفتاری با

انسانها توسط همان ماشینها نیز افزایش می‌یابد. بنابراین، چنین قوانینی باید رد شوند، نه به این دلیل که ما در برابر ماشینها وظایفی داریم، بلکه بدلیل وظایف غیرمستقیم ما در قبال انسانها.

پس از آنکه اندرسون قوانین سه‌گانه آسیموف را رد میکند، برای آنکه به یک اصل اخلاقی برسد، روشی را پیشنهاد میکند و معتقد است باید ابتدا وظایف اولیه‌ی به ماشین داده شود، سپس ماشین با استفاده از منطق استقرائی (ILP) روابطی بین این وظایف ایجاد نماید و آنها را توسعه بخشد و در نهایت، به یک اصل تصمیم‌گیری اخلاقی یا نظریهٔ اخلاقی دست یابد. نمونهٔ این روش در ماشین هوش مصنوعی «MedEthEx» که یک مشاور اخلاقی در حوزهٔ اخلاق پزشکی است، پیاده شده است. در معماری این برنامه، در ماژول^۴ تعلیم، از یک دانشمند اخلاق اطلاعاتی دربارهٔ نام عمل و مقدار تأمین شدن هر کدام از قواعد اخلاقی مربوط به آن عمل گرفته میشود. ماژول واسطه پایگاه دانش، راهنماییهایی برای انتخاب وظیفه برای حالات خاص را فراهم میسازد و ماژول پیشنهاددهنده، بر اساس حالت فعلی و دانشی که توسط دانشمندان اخلاق ارائه شده است، سعی میکند مناسبترین را انتخاب کند. در این بخش از منطق استقرائی استفاده میشود (Anderson, 2007).

دانشمندان بررسیهای خود را برای توانمندی‌سازی ماشینها برای پیروی از اصول و احکام اخلاقی، به چهار حوزه تقسیم میکنند:

اول) «بررسی معضلات اخلاقی»: بمنظور ساختن سیستمهای هوش مصنوعی که مایلیم رفتار اخلاقی داشته باشند، اولین گام بررسی معضلات اخلاقی است. اندرسون برای این منظور «GenEth» را که یک ماشین مشاور اخلاقی است پیشنهاد میدهد. این ماشین بگونه‌ی طراحی شده که به اخلاقدانان در فرایند مدون کردن اصول اخلاقی در حوزه‌های کاربردی معین کمک میکند.

دوم) «چارچوب تصمیم‌گیری اخلاق فردی»: دانشمندان به این نتیجه رسیدند که تصمیم‌گیری اخلاقی توسط انسانها نه تنها ملاحظات سودگرایانه، بلکه قوانین اخلاقی را نیز شامل میشود. چنین قوانینی اغلب شامل ارزشهایی هستند که از نظر اخلاقی انجام برخی اقدامات را بدون توجه به عواقب آن، منع میکنند. بعنوان مثال، قتل، صرف‌نظر از عواقب آن، اخلاقاً هم غیرقابل‌قبول است و تقبیح میشود. دانشمندان برای این حوزه، ماشین هوشمند «MoralDM» را پیشنهاد کرده‌اند که یک عامل را قادر می‌سازد تا معضلات اخلاقی را با استفاده از دو سازوکار حل نماید: (۱) استدلال اصول اولیه، که بر اساس قوانین اخلاقی تثبیت شده کار میکند.

(۲) استدلال قیاسی، که سناریوی داده شده را با موارد مشابه قبل مقایسه میکند تا به تصمیم‌گیری کمک نماید.

سوم) «چارچوب تصمیم‌گیری جمعی»: با افزودن سازوکارهای تصمیم‌گیری اخلاقی همانند آنهایی که در بخش قبل ذکر شد. به عوامل فردی، عده‌یی از عواملان میتوانند نقشهایی مختلف را هنگام ارزیابی انتخابهای عملی اخلاقی بعهده بگیرند. بعنوان مثال، برخی عواملان، اخلاق وظیفه‌گرا را ارزیابی میکنند، برخی اخلاق نتیجه‌گرا را و گروهی اخلاق فضیلت‌گرا؛ سپس بر اساس مجموعه‌یی از قوانین اخلاقی اولیه، قوانینی پیچیده‌تر را از طریق یادگیری بدست می‌آورند.

چهارم) «اخلاق در تعاملات انسان و هوش مصنوعی»: در این حوزه، دانشمندان چارچوبهای اخلاق را در عوامل تأثیرگذار بر رفتارهای انسانی ادغام میکنند. مطالعات نشان میدهد که باید تفاوت‌های شخصیتی، نگرش اخلاقی و تخصص فردی انسانها در نظر گرفته شود.

در زمینه اخلاق هوش مصنوعی، غالب فعالیتها بر توسعه چارچوبهای فردی متمرکز شده و آثاری هم درباره تصمیم‌گیری اخلاق جمعی بوجود آمده است، ولی هنوز کار زیادی نیاز است (Yu et al., 2018: p. 1). «بررسی معضلات اخلاقی» و «چارچوب تصمیم‌گیری

اخلاق فردی» مهمتر از دو زمینه دیگرند، چراکه زیرساخت دو زمین، دیگرند. در این دو زمینه باید معضلات اخلاقی بررسی و قوانین اولیه‌یی صادر شود تا برای موارد بعدی، ماشین بتواند استدلال کرده و قوانینی جدید ارائه دهد.

تسری دیدگاه ملاصدرا در مورد حکم اخلاقی

به هوش مصنوعی

همانطور که در شرح آراء ملاصدرا گذشت، انسان برای صدور احکام اخلاقی، نخست خوب و بد اعمال را تشخیص می‌دهد و پس از شناسایی فعل خوب، در پی انجام آن برمی‌آید و از افعال ناپسند دوری میکند؛ اینجاست که باید و نباید اخلاقی شکل می‌گیرد. از طرفی، احکام اخلاقی مطلق هستند و فرقی نمی‌کند که برای انسان تدوین شده باشند یا برای ماشین، زیرا هدف از ایجاد ماشینهای هوشمند، داشتن ماشینهایی است که مانند انسان عمل کنند؛ اما از آنجاکه این احکام از اصول اخلاقی مطلق بدست می‌آیند، باید گفت ماشینها هم میتوانند مانند انسانها احکام اخلاقی را با استفاده از داده‌های اولیه و استدلال، اعتبار کنند و حتی بعنوان مشاور انسان عمل نمایند.

برای رسیدن به مطلوب فوق، ماشین باید مرحله‌یی را که برای پیدایش احکام اخلاقی توسط انسان پیگیری میشود، دنبال کند؛ یعنی ابتدا مصلحت انسان و جامعه را بررسی کرده،

مشخص کند که هر عمل تا چه اندازه به مصلحت نزدیک است و میتواند آن را برآورده کند. پس از آن، اگر ماشین تشخیص داد که عمل مورد نظر مصلحتی را در پی دارد، نوبت به مرحله بعدی، یعنی قابلیت انشاء میرسد. در اینجا، ماشین بررسی میکند که آیا آن عمل که قرار است به حکم کلی تبدیل شود، قابلیت انشاء دارد یا خیر. سپس ماشین باید بررسی کند که آن حکم، نفع همگانی را تأمین میکند یا خیر؛ اگر نفع عامه توسط آن حکم تأمین میشود، ماشین آن را اظهار و اعلان مینماید. در واقع روند پیدایش احکام اخلاقی توسط ماشین هوشمند با دقت بیشتر از انسان انجام میشود، هرچند قوانین اولیه از انسان گرفته میشود و ماشین تنها توسعه‌دهنده قوانین است.

مسئولیت اخلاقی از نظر ملاصدرا

برای آنکه احکام اخلاقی به هدف خود که همان اجرایی شدن اخلاق است دست یابند باید همه انسانها ملزم به اجرای آن شوند و این الزام، مسئولیت اخلاقی را پدید می‌آورد. در اینجا ابتدا به مسئولیت اخلاقی از نظر ملاصدرا می‌پردازیم و سپس کاربرد آن در هوش مصنوعی و امکان تحقق یافتنش در ماشین اخلاقمند را ارزیابی میکنیم.

از آنجاییکه انسانها زندگی اجتماعی دارند، برای آنکه این نظام اجتماعی حفظ شده و به بیراهه کشیده نشود، باید مصلحت‌هایی رعایت شوند. از نظر ملاصدرا تک‌تک انسانها ملزم به رعایت مصلحت اجتماعی هستند. همچنین انسانها باید بتوانند عواقب افعالی را که از آنها سر میزند بعهده بگیرند و این همان معنای مسئولیت اخلاقی انسان است (ملاصدرا، ۱۳۸۲ الف: ۱۶۳). منشأ مسئولیتهای اخلاقی، احکام اخلاقی است و درواقع، اگر مسئولیت اخلاقی وجود نداشته باشد، احکام اخلاقی عبث و بیفایده خواهند بود. مسئولیت اخلاقی دارای مؤلفه‌ها و لوازمی است که درک آنها به درک این مسئولیت کمک میکند.

یکی از مؤلفه‌های مهم مسئولیت اخلاقی، فاعلی است که مسئول شناخته میشود و مورد بازخواست قرار میگیرد. این فاعل، باید بتواند فعل انجام دهد و این فعل چیزی جدا از او نیست. ملاصدرا معتقد است انسان برای آنکه به کمال برسد، باید فاعل باشد و فعلش از او جدا نمیشود (همو، ۱۳۸۳ ب: ۴۵۲). عمل انسان برگرفته از علم اوست و این علم باعث میشود انسان بسوی کمال برود و عملی را انجام دهد که باعث استکمال وجودی فاعل خود گردد. پس عمل و عامل (که همان انسان است) با هم متحدند (همو، ۱۳۹۰: ۷۴).

شرایط فاعل مسئول برای عمل اخلاقی

اختیار: دربارهٔ اختیار انسان، دیدگاه‌هایی گوناگون وجود دارد. معتزله برای آنکه خداوند را از قبايح منزه کنند و همچنین تکلیف انسان را توجیه نمایند، انسان را کاملاً مختار میدانند؛ درحالی‌که این نظر موجب شرک میگردد. از سویی، اشاعره با مطلق دانستن قدرت و ارادهٔ الهی، معتقدند تمام افعال صرفاً با اراده و فاعلیت خداوند محقق میشود و بر افعال انسان، جبر حاکم است. اما ملاصدرا نظر متکلمان امامیه را برگزیده و معتقد است حق آنست که بین جبر و تفویض راه میانه را بگیریم. او همچنین بر این باور است که انسان در تحقق افعال خود دارای اراده و اختیار است ولی این اختیار در طول ارادهٔ الهی است، پس فعل انسان همانطور که به او منسوب میشود به خداوند نیز منسوب میگردد. درواقع فعل انسان در عین اختیاری بودن، اجباری است و در عین اجبار، اختیاری است؛ مانند زمانی‌که فعل هرچند به قوای نفس نسبت داده شود، در عین حال به نفس انسان هم منتسب میگردد. برای مثال، فعل دیدن به قوهٔ باصره نسبت داده میشود و در عین حال به نفس انسان هم منسوب میگردد (همو، ۱۳۸۱ الف: ۳۴۲-۳۳۴).

قدرت: قدرت، کیفیتی نفسانی است؛ به این معنا که فاعل، فعل مورد نظر خود را «اگر بخواهد، انجام میدهد و اگر نخواهد، انجام

نمیدهد». با توجه به این بیان، قادر کسی است که بخواست خود کاری انجام می‌دهد و اراده او سبب انجام فعلش میشود، فارغ از اینکه اراده‌اش عین ذات باشد یا زائد بر ذات (همو، ۱۳۸۲ج: ۲ / ۷۲۶). قدرت به اشتراک معنوی، شامل قدرت مجردات، انسان و حیوان است و عالیت‌ترین مرتبه آن که منزله از امکان است، به خداوند اختصاص دارد. قدرت در انسان، تام نیست، زیرا مبادی افعال اختیاری مثل تصدیق، شوق و اراده از خارج به ما میرسد و بهمین دلیل میتوان گفت قدرت در انسان، امری بالقوه است (که با جمع بودن شرایط، به فعلیت میرسد) (همو، ۱۳۸۱ب: ۲۰۸-۲۰۷).

علم: حقیقت علم، حضور نزد عالم است (همان: ۱۱۱). علم را به حضوری و حصولی تقسیم میکنند و به اعتقاد ملاصدرا، علم حصولی به همان علم حضوری باز میگردد (همان: ۱۲۴). شرط حضور معلوم نزد عالم، مجرد است و نفس انسان بدلیل مجرد بودن، دارای ویژگیهای لازم عالم شدن هست. این علم میتواند علم نفس به خود، یا به قوای خود، یا به صور ذهنی خود باشد (همو، ۱۳۸۳ب: ۳۲۳). ملاصدرا در یک تقسیم‌بندی، علم را سه قسم میکند: (۱) علم فعلی، مانند علم خداوند به همه موجودات غیر از ذات خود و مانند علم سایر علل به معلولات خود؛ (۲) علم انفعالی، مانند علم ماسوی الله به

چیزهایی که معلول خود آنها نیست و آفریده خداوند هستند. میتوان گفت این علم همان صورتهایی است که در نفس حاصل میشود. (۳) علمی که نه فعلی است و نه انفعالی، مانند علم انسان به ذات خود و علم او به اموری که صورشان هنوز از او غایب نشده است (ملاصدرا، ۱۳۸۳ب: ۴۱۷).

از آنجاییکه انسان اختیار، قدرت و علم دارد، باید گفت شرایط لازم و کافی مسئولیت اخلاقی را دارد، اما آیا ماشینهای هوشمند هم دارای این شرایط هستند؟ و اگر ماشین دارای چنین شرایطی باشد، آیا بتنهایی دارای مسئولیت است؟ از طرف دیگر، باید دید مسئولیت اخلاقی در برابر چه کسانی معنا پیدا میکند و اینکه ماشینها هم مانند انسانها در برابر آنها مسئول هستند؟

مسئولیت اخلاقی در هوش مصنوعی

برای آنکه بدانیم مسئولیت اخلاقی تا چه حد میتواند متوجه ماشینهای هوشمند باشد، باید بررسی کنیم که این ماشینها چقدر از شرایط مسئول بودن برخوردارند؛ یعنی تا چه حد دارای اختیار، قدرت و علم هستند. دانشمندان، هوش مصنوعی را به سه نوع تقسیم کرده، برای هر کدام ویژگیهایی را ذکر میکنند. هر کدام از این سه نوع ماشین، حدودی از اختیار و قدرت را دارند؛ پس لازم

است هر کدام را جداگانه بررسی کرده و ببینیم کدامیک شرایط مسئولیت اخلاقی را بطور کامل دارد و مسئول اعمال خود است و کدامیک دارای محدودیت است و آن را نمیتوان مسئول اخلاقی دانست.

هوش مصنوعی ضعیف: هوش مصنوعی ضعیف یا محدود، برای انجام وظایفی خاص طراحی شده و قادر به درک و یادگیری چیزی فراتر از برنامه‌ها و الگوریتمهایی که توسط برنامه‌نویسان به آن ارائه شده، نیست. این نوع هوش مصنوعی فقط کار خاصی را که به آن سپرده شده است مانند تشخیص صدا یا تصویر یا جستجو در وب-انجام میدهد و به ورودی و برنامه‌نویسی انسان وابسته است. این سیستمها هوشیاری و خودآگاهی نداشته و توانایی کسب تجربه را ندارند.

هوش مصنوعی عمومی یا قوی: هوش مصنوعی قوی به سیستمهایی اطلاق میشود که توانایی انجام هر کار فکری را که یک انسان میتواند انجام دهد، دارند. این سیستمها میتوانند درک کنند، یاد بگیرند و کارهایی را که انسان انجام میدهد، انجام دهند. این نوع هوش مصنوعی با هوش انسانی مطابقت دارد و حتی فراتر از آن میرود. این ماشینها دارای آگاهی هستند و بدون دخالت انسان، خودمختار و مستقل عمل میکنند. در حال حاضر هوش

مصنوعی قوی فقط در حد یک فرضیه مانده و به واقعیت نرسیده است.

هوش مصنوعی فوق هوشمند: این نوع هوش مصنوعی که فعلاً بعنوان یک فرضیه مطرح میشود، دارای ویژگیهایی از جمله توانایی حل مسئله، یادگیری، خودمختاری و خلاقیت بوده، و تأثیری عمیق بر جوامع انسانی خواهد داشت. هوش مصنوعی مذکور حتی میتواند در هر زمینه‌یی از انسان توانمندتر باشد و تصور میشود توان حل مشکلات پیچیده جهان را دارد و میتواند فناوریهای جدید ابداع کند و بطور کلی، زندگی انسان را بهبود بخشد.

تسری دیدگاه ملاصدرا دربارهٔ مسئولیت اخلاقی به هوش مصنوعی

حال اگر بخواهیم به ارزیابی فاکتورهای اصلی مرتبط با مسئولیت‌پذیری -یعنی اختیار، قدرت و علم- در انواع هوش مصنوعی بپردازیم، باید بگوییم آنچه امروز عملاً از انواع سه‌گانه هوش مصنوعی در دسترس است، همان هوش مصنوعی ضعیف است. این ماشین چنانکه گذشت، کاملاً وابسته به سازندهٔ خود است و بنابراین داشتن اختیار و قدرت مستقل برای آن منتفی است. از طرفی، سیستمهای هوشمند، برخلاف انسانها، همیشه در معرض هک، ویروس و اختلالات شبکه هستند که هر

کدام از اینها میتواند منجر به خطا در عملکرد ماشین شود و ماشین نتواند عمل مورد نظرش را انجام دهد.

درباره علم نیز باید گفت: اگرچه بر اساس نظر ملاصدرا، همه موجودات از جماد گرفته تا انسان دارای علم و شعور هستند (ملاصدرا، ۱۳۸۳ الف: ۳)، اما ماشینهای هوشمند، علمی را که منشأ مسئولیت اخلاقی باشد ندارند. برای داشتن علمی که ایجاد مسئولیت کند، باید عالم، مرتبه تجردی (یا نفس مجرد) هم داشته باشد، اما در حال حاضر ماشینهای هوشمند کاملاً مادی هستند. البته ممکن است در آینده ساخت هوش مصنوعی قوی یا فوق هوشمند محقق گردد و ممکن است در آنها بخشی با عملکرد مشابه نفس مجرد انسانی تعبیه شود که بتواند واکنشهایی معادل اختیار کردن، قدرت و علم داشتن از خود نشان دهد.

با توجه به مطالبی که گذشت، ماشینهای هوشمند در شرایط کنونی نمیتوانند بتنهایی دارای مسئولیت اخلاقی باشند، زیرا اختیار، قدرت و علم آنها محدود و قابل کنترل توسط انسانهاست. بنابراین، چون ماشینها بطور مطلق مسئولیت اخلاقی را نمیپذیرند، پس در برابر خدا، انسانها و حیوانها هم مسئولیت چندانی متوجه آنها نمیشود؛ مگر آنکه روزی دانشمندان به هوش مصنوعی

قوی یا فوق هوشمند با شرایطی که در بالا فرض شد، دست پیدا کنند.

جمعبندی و نتیجه گیری

۱. در این پژوهش، حکم اخلاقی از آن جهت مورد توجه قرار گرفت که ماشینهای هوشمند، نیازمند مجموعه‌یی از اصول و احکام اولیه اخلاقی هستند تا با استفاده از آنها بتوانند در موقعیتهای مختلف، احکامی جدید را استنباط کنند. برای صدور احکام، ابتدا باید حُسن و قبح اعمال تشخیص داده شود. پس از تشخیص حُسن و قبح، باید و نباید اخلاقی بوجود می‌آید، تا به عمل حَسَن، امر و از قبیح، نهی کند. احکام اخلاقی برآمده از همین باید و نبایدهای اخلاقی هستند.

۲. از نظر ملاصدرا، اصول و مبانی احکام اخلاقی، مطلق و ثابتند، ولی برخی از احکامی که مورد نیازند و هنوز انشاء نشده‌اند، میتوانند توسط انسان اعتبار شوند. مشخص شد که ماشینهای هوشمند میتوانند بعنوان دستیار در صدور این قبیل احکام به انسان کمک کنند. بزرگترین مزیت دستیاری ماشینهای هوشمند اینست که مراحل پیدایش احکام اخلاقی را با دقتی بیشتر طی کرده و زمانی کمتر را صرف میکنند.

۳. غایت مورد نظر در احکام اخلاقی، اجرایی شدن اخلاق است. برای اینکه این غایت محقق شود، باید انسانها ملزم به انجام احکام اخلاقی باشند. این الزام، مسئولیت اخلاقی را پدید می‌آورد.

۴. مسئولیت اخلاقی، متوجه همه افراد جامعه نیست؛ فقط کسانی مسئولیت اخلاقی دارند که شرایط لازم برای مسئول بودن را داشته باشند. مهمترین این شرایط عبارتند از: اختیار، قدرت و علم. ماشینهای هوشمند نیز تنها زمانی مسئول مطلق اعمال خود هستند که مختار، قادر و عالم باشند. آنچه امروز بعنوان هوش مصنوعی در دسترس است، هوش مصنوعی ضعیف است که اختیار و قدرت چندانی ندارد و هنوز وابسته به سازنده خود است. همچنین ماشینهای هوشمند نمیتوانند دارای علم باشند، زیرا شرط عالم بودن، مجرد بودن است که درباره ماشینهای هوشمند فعلی صدق نمیکند. پس میتوان نتیجه گرفت که تمام مسئولیت اخلاقی متوجه سازنده ماشینهای هوشمند است.

الأجناس للفضائل و الرذائل النفسانية. ثم إن الشجاعة ليست ضداً لشيء من التهور و الجبن، لكونها واسطة بينهما. و أما الطرفان فلكونهما في غاية التباعد كانا متضادين، بل تضاد الشجاعة مع كل منهما تضاد بالعرض، إذ الشجاعة ماهية لها عارض وهو كونها فضيلة، و كل من التهور و الجبن ماهية، لها عارض هو كونها رذيلة؛ و التضاد بالحقيقة بين العارضين».

۲. «فاول علوم النفس هو علمها بذاتها، ثم علمها بقوى البدن و الآلات التي هي الحواس الظاهرة و الباطنة، و هذان العلمان إنما هما علمان من العلوم الحضورية. ثم بعد هذين العلمين ينبعث عن ذات النفس لذاتها استعمال الآلات، بدون تصور هذا الفعل الذي هو استعمال الآلات و التصديق بفائدته؛ فإن هذا الاستعمال ليس فعلاً اختيارياً بمعنى كونه حاصلًا بالقصد و الروية و إن كانت النفس عالمة به مريدة له، لأن إرادة ذلك الفعل إنما تنبعث عن ذاتها لا عن رويتها؛ فذاتها بذاتها موجبة لاستعمال الآلات لا بإرادة اختيارية زائدة عليها قائمة بها، بل لما كانت ذاتها في آن وجودها عالمة بذاتها و عاشقة لها و لفعالها، عشقاً ناشئاً عن الذات لذاتها، اضطرت إلى استعمال الآلات التي لا قدرة لها إلا عليه».

۳. «إذا حصل تصور شيء قبل وجوده و ترجح أحد جانبي إمكانه، تحصل إرادة متخصصة بأحدهما. فالترجيح متقدم على الإرادة كما مر».

۴. ماژول یک بخش مجزا از برنامه کامپیوتر است که وظیفه خاصی را انجام میدهد.

منابع

دهقانی سیمکانی، رحیم (۱۳۹۱) «موانع اخلاق و ورزی از دیدگاه ملاصدرا»، پژوهشنامه اخلاق، شماره ۱۸، ص ۱۴۳-۱۲۵.

طاهریان دهکردی، بتول؛ شیخ الاسلامی، محمود؛ اکبرزاده، رجب (۱۳۹۷) «مسئولیت اخلاقی در نگاه

پی نوشتها

۱. «الشجاعة تحت الفضيلة و التهور المضاد لها- تحت الرذيلة لا یرد نقضاً على هذه القاعدة، لأن كل واحد منهما في نفسه كيفية نفسانية، كونه فضيلة أو رذيلة إنما هو صفة عارضة له لا أنها مقومة؛ فالفضيلة و الرذيلة ليستا من

- ملاصدرا»، حکمت صدرایی، شماره ۱۳، ص ۹۲-۸۱؛
DOI:10.30473/PMS.2018.4691
طوسی، سیدخلیل‌الرحمان (۱۳۹۲) درآمدی بر اخلاق
فلسفی در حکمت متعالیه، قم: پژوهشگاه علوم و
فرهنگ اسلامی.
- ملاصدرا (۱۳۸۰ الف) حکمة المتعالیة فی الأسفار الأربعة،
ج ۲، تصحیح و تحقیق مقصود محمدی، تهران: بنیاد
حکمت اسلامی صدرا.
- (۱۳۸۰ ب) حکمة المتعالیة فی الأسفار الأربعة،
ج ۷، تصحیح و تحقیق مقصود محمدی، تهران: بنیاد
حکمت اسلامی صدرا.
- (۱۳۸۱ الف) حکمة المتعالیة فی الأسفار الأربعة،
ج ۶، تصحیح و تحقیق احمد احمدی، تهران: بنیاد
حکمت اسلامی صدرا.
- (۱۳۸۱ ب) المبدأ و المعاد، تصحیح و تحقیق
محمد ذیحی و جعفر شاه‌نظری، ج ۱، تهران: بنیاد
حکمت اسلامی صدرا.
- (۱۳۸۱ ج) کسر أصنام الجاهلیة، تصحیح و
تحقیق محسن جهانگیری، تهران: بنیاد حکمت
اسلامی صدرا.
- (۱۳۸۲ الف) حکمة المتعالیة فی الأسفار الأربعة،
ج ۹، تصحیح و تحقیق رضا اکبریان، تهران: بنیاد
حکمت اسلامی صدرا.
- (۱۳۸۲ ب) الشواهد الربوبیة فی المناهج السلوکیة،
تصحیح و تحقیق سیدمصطفی محقق داماد، تهران:
بنیاد حکمت اسلامی صدرا.
- (۱۳۸۲ ج) شرح و تعلیقه صدرالمتألهین بر الهیات
شفا، تصحیح و تحقیق نجفقلی حبیبی، تهران: بنیاد
حکمت اسلامی صدرا.
- (۱۳۸۳ الف) حکمة المتعالیة فی الأسفار الأربعة،
ج ۱، تصحیح و تحقیق غلامرضا اعوانی، تهران: بنیاد
حکمت اسلامی صدرا.
- (۱۳۸۳ ب) حکمة المتعالیة فی الأسفار الأربعة،
ج ۳، تصحیح و تحقیق مقصود محمدی، تهران: بنیاد
حکمت اسلامی صدرا.
- (۱۳۸۳ ج) حکمة المتعالیة فی الأسفار الأربعة،
ج ۴، تصحیح و تحقیق مقصود محمدی، تهران: بنیاد
حکمت اسلامی صدرا.
- (۱۳۸۳ د) حکمة المتعالیة فی الأسفار الأربعة،
ج ۸، تصحیح و تحقیق علی‌اکبر رشاد، تهران: بنیاد
حکمت اسلامی صدرا.
- (۱۳۸۵ الف) شرح اصول الکافی، ج ۲، تصحیح و
تحقیق سیدمهدی رجایی، تهران: بنیاد حکمت
اسلامی صدرا.
- (۱۳۸۵ ب) شرح اصول الکافی، ج ۴، تصحیح و
تحقیق محمود فاضل یزدی، تهران: بنیاد حکمت
اسلامی صدرا.
- (۱۳۸۹ الف) تفسیر القرآن الکریم، ج ۲، تصحیح و
تحقیق محسن بیدارفر، تهران: بنیاد حکمت اسلامی
صدرا.
- (۱۳۸۹ ب) تفسیر القرآن الکریم، ج ۴، تصحیح و تحقیق
محسن بیدارفر، تهران: بنیاد حکمت اسلامی صدرا.
- (۱۳۹۰) رساله سه اصل، تصحیح و تحقیق
حسین نصر، تهران: بنیاد حکمت اسلامی صدرا.
- Allen, C. & Wallach, W. (2005). Artificial
morality: Top-down, bottom-up, and hybrid
approaches. *Ethics and Information
Technology*.
DOI: 10.1007/s10676-006-0004-4.
- Anderson, M. & Leigh A., S. (2007). The status
of machine ethics; a report from the AAAI
Symposium. *Minds and machines*.
DOI: 10.1007/s11023-007-9053-7.
- Yu, H., Shen, Z., Miao, C., Leung, C., Lesser, V.
R., & Yang, Q. (2018). Building ethics into
artificial intelligence. ArXiv preprint arXiv:
1812. DOI: 1812.02953.