

بررسی اخلاقی عملکرد دو سامانه هوش مصنوعی با تکیه بر حکمت عملی صدرایی

20.1001.1.24767220.1404.15.3.5.5

محمدسجاد زارع میرک‌آباد^۱

چکیده

تحقیق حاضر با هدف بررسی مسائل اخلاقی دو سامانه هوش مصنوعی تشخیص چهره و تصمیم‌گیری خودکار انجام شده است. همچنین، به منظور بررسی نحوه عملکرد این دو سامانه و نقض اصول اخلاقی به وسیله آن‌ها از روش تحقیق تحلیل محتوا استفاده شده است. بر پایه نتایج این تحقیق، اصول اخلاقی نقض شده عبارت‌اند از: انصاف، آزادی و اختیار، حریم خصوصی و حفظ اسرار، مسئولیت‌پذیری و پاسخ‌گویی و عمل به ظن معتبر. همچنین، این اصول در چهارچوب حکمت عملی صدرایی تحلیل و این نتایج دسته‌بندی شدند: ۱. در حکمت صدرایی انصاف از صفات ضروری حاکم است و رعایت بی‌طرفی الزامی است؛ ۲. آزادی و اختیار برای تکامل و سعادت انسان اساسی است؛ ۳. حریم خصوصی بسیار مهم و تجسس در آن ناپسند است؛ ۴. مسئولیت‌پذیری و پاسخ‌گویی از ویژگی‌های لازم رهبران است؛ ۵. عمل به ظن غیر معتبر قبیح است. بنابراین، براساس حکمت عملی صدرایی، به کارگیری این دو سامانه در شکل کنونی خود از لحاظ اخلاقی پذیرفتنی نیست. مگر آنکه طراحی یا اجرای آن‌ها به گونه‌ای تغییر کند که مسائل اخلاقی مطرح شده برطرف شود.

واژگان کلیدی: اخلاق هوش مصنوعی، حکمت عملی صدرایی، انصاف، آزادی، حریم خصوصی، پاسخ‌گویی، عمل به ظن معتبر

تاریخ پذیرش: ۲۷ آذر ۱۴۰۳

تاریخ بازنگری: ۱ آذر ۱۴۰۳

تاریخ دریافت: ۱۷ مرداد ۱۴۰۳

مقدمه

رایانه‌ها در پی تحقق این ایده بودند که مصنوعي کارهای انسان را به تنهایی و بدون نیاز به کمک یا با سرعت ودقتی بیشتر از انسان انجام دهد. بدون رایانه‌ها و شبکه‌های کامپیوتری، بیشتر فرایندهای تولید، صنعت، تجارت، حمل‌ونقل و توزیع، خدمات دولتی، نظامی، بهداشتی، آموزشی و تحقیقی کاملاً متوقف می‌شود. اکنون پیشرفت رایانه‌ها به خودکارسازی فرایندها و شبیه‌سازی عملکرد و توانایی‌های انسان رسیده است. محققان از طریق هوش مصنوعی که یکی از کاربردهای پیشرفته رایانه‌هاست، در پی شبیه‌سازی هوش انسان‌اند تا این مصنوع مانند انسان فکر کند و کارهای انسان‌ها را انجام دهد.

در شاخه هوش مصنوعی سامانه‌های متفاوتی طراحی و توسعه داده شده است، به گونه‌ای که هریک از آن‌ها در عرصه‌های مختلفی از زندگی بشروارد شده‌اند. برای نمونه، از طریق سامانه‌های تشخیص چهره، یکی از انواع سامانه‌های هوش مصنوعی، تشخیص چهره‌های افراد در تصاویر یا ویدئو امکان‌پذیر است. همچنین، این سامانه‌ها قابلیت ترکیب با ابزارهای دیگر را به منظور انجام شکل‌های پیچیده‌تر نظارت دارد. شکل پیشرفته این سامانه‌ها تشخیص و تحلیل زنده چهره به وسیله شبکه گسترده دوربین‌های شهری است. این دوربین‌ها داده‌های عظیمی از زندگی فردی و اجتماعی تک تک افراد جامعه و ارتباطاتشان را ذخیره می‌کنند. ذخیره چنین داده‌هایی و نحوه حفاظت و استفاده دولت‌ها از این داده‌ها مهم‌ترین مسئله اخلاقی در این حوزه است که در این نوشتار بررسی می‌شود.

یکی دیگر از انواع سامانه‌های هوش مصنوعی سامانه‌های تصمیم‌گیری خودکار است. سامانه هوش مصنوعی برای تصمیم درست باید اولویت و اهمیت هر کدام از مؤلفه‌ها را در شرایط مختلف تشخیص دهد. سامانه‌های تصمیم‌گیری خودکار برای بازسازی عملکرد تصمیم‌گیری در انسان طراحی شدند. امروزه، در حالی که بسیاری از این سامانه‌ها هنوز آزمایش نشده و برای اهداف خود به شکل مناسبی طراحی نشده‌اند به صورت گسترده به کار گرفته می‌شوند. برای نمونه، طبق گزارش کمیسیون اروپا^۱ (2020) حدود ۴۰ درصد از الگوریتم‌های به کاررفته در تصمیم‌گیری خودکار در حوزه‌های حساسی مانند استخدام یا دسترسی به وام، هنوز فاقد شفافیت و سازوکار مناسب برای رفع خطا هستند. این شواهد نشان می‌دهد استفاده گسترده از

سامانه‌های تصمیم‌گیری خودکار بدون ارزیابی‌های کافی، مسائل اخلاقی و فنی بسیاری پدید می‌آورد. همچنین در بعضی مواقع، برخلاف حقوق اساسی و به صورت نامشروع حقوق فردی و اجتماعی را نقض می‌کند. گاهی سامانه‌های تصمیم‌گیری خودکار، علی‌رغم طراحی برای بهبود کارایی و دقت، خطای می‌کنند و تصمیم‌های نادرستی می‌گیرند. در بیشتر مواقع، شک کردن، مقابله کردن و اصلاح این تصمیم‌ها به علت پیچیدگی و شفاف نبودن الگوریتم‌ها دشوار یا حتی غیرممکن است. برای نمونه، گزارشی از مؤسسه هوش مصنوعی اکنون^۲ (2018) نشان می‌دهد در حوزه‌های حساسی مانند استخدام و نظام‌های قضایی استفاده از این سامانه‌ها بدون ارزیابی‌های کافی به تبعیض و نابرابری منجر می‌شود. همچنین، تحقیقاتی در دانشگاه آکسفورد در سال ۲۰۲۰ نشان داده است که نبود شفافیت در الگوریتم‌های تصمیم‌گیری خودکار اعتماد عمومی را کاهش می‌دهد و پیامدهای منفی اجتماعی به همراه دارد. بنابراین، استفاده گسترده از این سامانه‌ها بدون ارزیابی‌های مناسب مشکلات اخلاقی و اجتماعی گوناگونی را به دنبال دارد.

هر پدیده اقتصادی، اجتماعی یا سیاسی براساس وسعت فراگیری و سطح نفوذی که دارد، در جامعه و زندگی مردم تأثیرات کم و بیش گسترده و عمیقی می‌گذارد. سامانه‌های هوش مصنوعی از این قاعده مستثنا نیستند، زیرا در جهان طبیعت نمی‌شود ادعا کرد که تأثیرات پدیده‌ای همگی مثبت یا همگی منفی‌اند. بنابراین، عقلانی است که حکومت‌ها قبل از فراگیری هر پدیده‌ای در جامعه، مشکلات و مزایای آن را بررسی کنند. در شرایط امروز که عصر فناوری و فضای مجازی نامیده می‌شود، فناوری‌های جدید تأثیرات گسترده و عمیقی در زندگی مردم می‌گذارند.

با توجه به اهمیت فلسفه اسلامی و حکمت عملی متکی بر حکمت نظری ضروری است که جوامع اسلامی با تکیه بر این میراث علمی و فرهنگی زمینه بهره‌گیری عملی از اصول اخلاقی و فلسفی را فراهم کنند. می‌شود از این امر مانند مبنایی برای حل مسائل معاصر، تسهیل حرکت جوامع به سوی کمال و ارتقای کیفیت زندگی در چهارچوب ارزش‌های اسلامی استفاده کرد. همچنین، توجه به عملی‌سازی این گنجینه فکری، به ویژه در حوزه‌هایی مانند اخلاق کاربردی و حکمرانی و در پیشبرد اهداف متعالی بسیار اثرگذار است.

حکمت مرتبط با رفتار و تصمیم‌گیری‌های عملی انسان است. ملاصدرا، یکی از برجسته‌ترین فیلسوفان صدرایی، حکمت عملی را یکی از ابعاد حکمت متعالی معرفی کرده است و آن را به هدایت عملی انسان در زندگی روزمره مرتبط می‌داند (Shirazi, 1981a). این نوع حکمت براساس اصول اخلاقی و ارزش‌های انسانی، راهنمایی برای انتخاب‌های صحیح در مواجهه عملی با مشکلات ارائه می‌دهد. در مقابل، حکمت نظری بیشتر به مسائل معرفتی و متافیزیکی می‌پردازد و به فهم عمیق واقعیت‌ها و اصول بنیادین جهان اختصاص دارد (Nasr, 1997). ارتباط میان حکمت عملی و حکمت نظری در این است که حکمت نظری بنیان‌گذار تفکرات و اصولی است که حکمت عملی براساس آن‌ها عمل می‌کند. به عبارت دیگر، حکمت نظری چهارچوب فلسفی و معرفتی را فراهم می‌کند که حکمت عملی در تصمیم‌گیری‌های اخلاقی و عملی به کار می‌برد. این تعامل میان دو نوع حکمت امکان توسعه رویکردی جامع و متوازن را در مواجهه با مسائل اخلاقی و عملی فراهم می‌کند.

۳-۱. مسائل اخلاقی سامانه‌های تشخیص چهره

به منظور تشخیص چهره افراد در تصویرها یا ویدئوها از سامانه‌های تشخیص چهره استفاده می‌کنند که قابلیت ترکیب با ابزارهای دیگر را برای نظارت‌های پیچیده‌تر دارند. یکی از کاربردهای پیچیده این سامانه با استفاده از شبکه گسترده دوربین‌های زنده در سطح شهر مطرح می‌شود. این دوربین‌ها داده‌های زیادی را از زندگی فردی و اجتماعی افراد جامعه و ارتباطاتشان ذخیره می‌کنند. ذخیره، حفاظت و استفاده از این داده‌ها از سوی دولت‌ها از مهم‌ترین مسائل اخلاقی در این حوزه‌اند. در ادامه، به برخی از کاربردهای این سامانه‌ها اشاره می‌شود.

۴-۱. تشخیص تمایلات^۵

سازوکار تشخیص تمایلات زیررده‌ای از تشخیص چهره است که تصویرها و چهره‌ها را در ویدئوها تفسیر می‌کند و درباره شخصیت، احساسات درونی، سلامت ذهنی، مشارکت فرد در کار یا حتی مقاصد پنهانی افراد حاضر در ویدئوها اظهار نظر می‌کند. ارتباط یافتن فناوری تشخیص تمایلات با مسائلی مانند استخدام، دسترسی به بیمه، آموزش و امنیت عمومی در هر دو سطح فردی و اجتماعی خطراتی را در پی دارد (Khodagholizadeh, 2020).

در این نوشتار عملکرد سامانه‌های هوش مصنوعی را با تکیه بر حکمت عملی صدرایی از لحاظ اخلاقی نقد می‌کنیم. فرضیه ما این است که حکمت متعالیه دربردارنده اصول و ارزش‌های اخلاقی موردنیاز زندگی متعالی انسان است. از این رو، با پیدایش پدیده‌های جدید به وسیله تحلیل آن پدیده‌ها و عرضه به اصول حکمت متعالیه می‌توان چهارچوب و ضوابط کلی مواجهه با پدیده‌ها را استخراج یا استنباط کرد.

۱. مبانی نظری

هوش مصنوعی: هوش مصنوعی شاخه‌ای از علوم رایانه است که به توسعه سیستم‌ها و ماشین‌هایی می‌پردازد که قادر به انجام وظایفی هستند که معمولاً نیاز به هوش انسانی دارند. این وظایف شامل یادگیری، استدلال، حل مسئله و درک زبان طبیعی است. (Russell and Norvig, 2020)

۱-۱. تعریف‌ها

یادگیری ماشین^۱: زیرشاخه‌ای از هوش مصنوعی که به سیستم‌ها امکان یادگیری از داده‌ها را بدون برنامه‌ریزی صریح می‌دهد (Mitchell, 1997).

یادگیری عمیق^۲: زیرمجموعه‌ای از یادگیری ماشین که از شبکه‌های عصبی چندلایه برای تحلیل داده‌های پیچیده استفاده می‌کند (Goodfellow et al., 2016).

سامانه تشخیص چهره^۳: این سامانه‌ها با استفاده از الگوریتم‌های یادگیری عمیق، هویت افراد را از طریق تصاویر یا ویدئوها تشخیص می‌دهد (Jain et al., 2011). سامانه‌های تصمیم‌گیری خودکار^۴: سیستم‌هایی که بدون دخالت انسانی تصمیم‌گیری می‌کنند، مانند الگوریتم‌های اعطای وام یا تشخیص بیماری (OECD, 2021).

تعریف اخلاق و جایگاه آن در تصمیم‌گیری‌های انسانی: اخلاق مطالعه ارزش‌ها و اصولی است که رفتار درست و نادرست را تعیین می‌کنند. در تصمیم‌گیری‌های انسانی اخلاق اثری مهم در تعیین مسئولیت‌ها و حقوق دارد (Beauchamp and Childress, 2019).

۲-۱. مفهوم حکمت عملی و ارتباط آن با حکمت نظری

حکمت عملی در فلسفه اسلامی به معنای دانش و

1. Machine Learning
2. Deep Learning
3. Face Recognition System
4. Automated Decision-Making Systems

5. Emotion Recognition and Affect Recognition Systems

۵-۱. احراز هویت^۱ و اعتبار اجتماعی^۲

احراز هویت افراد یکی از نیازهای ضروری در تعاملات اجتماعی است. با توسعه هوش مصنوعی شکل سنتی احراز هویت جای خود را به دوربین‌های زنده تشخیص چهره دادند. از آنجایی که اطلاعات هویتی افراد در حافظه یک پارچه سامانه احراز هویت وجود دارد، با اسکن چهره به وسیله دوربین‌ها چهره فرد با اطلاعات موجود تطبیق داده و هویت فرد شناسایی می‌شود. همچنین، سامانه‌های اعتبار اجتماعی جای کارت‌های بانکی را گرفتند، به گونه‌ای که با تشخیص چهره فرد و اتصال به مرکز اعتبار اجتماعی، خرید کالا و خدمات تسهیل می‌شود. علاوه بر این، در برخی جوامع از اعتبار اجتماعی برای تشویق و تنبیه رفتار اجتماعی استفاده می‌شود، به گونه‌ای که تخلف از قوانین اعتبار اجتماعی را کاهش و رفتارهای مثبت آن را افزایش می‌دهد (Kostka, 2019).

۶-۱. امنیت^۳ و شناسایی مظنونین^۴

پلیس مدت‌هاست از فناوری تشخیص چهره برای شناسایی افراد در تصاویر بایگانی شده استفاده می‌کند. اما استفاده از سامانه‌های شناسایی در دوربین‌های زنده کاربرد جدیدی است. پلیس تمام شهر را با نصب دوربین‌های زنده تشخیص چهره در خیابان‌ها به ویژه مناطق پر ازدحامی مانند مراکز گردشگری و خرید زیر نظر می‌گیرد و جرائم خشن و جدی را شناسایی می‌کند. یکی دیگر از کاربردهای امنیتی سامانه‌های تشخیص چهره تشخیص رفتارهای خشنونت‌آمیز به صورت زنده است. این سامانه‌ها به وسیله هوش مصنوعی درگیری و رفتارهای خشن را شناسایی و به پلیس اعلام می‌کنند (Li and Zhang, 2017).

۷-۱. مسائل اخلاقی سامانه‌های تصمیم‌گیری خودکار^۵

یکی از مهم‌ترین مزیت‌های سامانه‌های تصمیم‌گیری خودکار افزایش بازدهی و کاهش هزینه‌ها به واسطه شبیه‌سازی تصمیم‌گیری انسان است. این سامانه‌ها در موقعیت‌های مختلف قرار می‌گیرند و بر اساس آموزش‌های ماشینی‌ای که دیده‌اند، انتخاب می‌کنند

و تصمیم می‌گیرند. آن‌ها تأثیرات زیادی بر عرصه‌های مختلف زندگی از جمله فرصت‌ها، دسترسی‌ها، آزادی، امنیت، حقوق، نیازها و رفتار می‌گذارند و اگر این موارد بررسی نشوند، تأثیرات سوء گسترده‌ای به همراه دارند (Richardson, 2021).

امروزه در سراسر جهان، سازمان‌های دولتی با عناوینی مانند بهره‌وری و صرفه‌جویی در هزینه در حال فراهم کردن و به کارگیری سامانه‌های تصمیم‌گیری خودکارند. در حالی که بسیاری از این سامانه‌ها هنوز آزمایش نشده و برای اهداف خود به شکل مناسبی طراحی نشده‌اند. اگر این سامانه‌ها اشتباه کنند، یعنی تصمیم نادرستی بگیرند، در بیشتر مواقع، زیرسؤال بردن این تصمیمات و اصلاحشان دشوار یا غیرممکن است (Khodagholizadeh, 2020).

از سامانه‌های تصمیم‌گیری خودکار در بخش‌های دولتی و خصوصی استفاده می‌شود. برای نمونه، در بخش دولتی می‌توان به کاربردهایی در نظام کیفری، رفاه، آموزش، پزشکی و مهاجرت اشاره کرد. همچنین، در بخش خصوصی می‌توان از پیشنهاد فیلم یا موسیقی مورد علاقه یا پیشنهاد کار به رانندگان تاکسی‌های اینترنتی نام برد. علاوه بر این، بهینه‌سازی مسیر خطوط هوایی، سامانه‌های مدیریت عملکرد شرکت‌ها، پیشنهاد ادغام یا تفکیک شرکت‌ها مبتنی بر سود، سامانه‌های تأیید یا رد صلاحیت، سامانه‌های کنترل مالی و تشخیص تخلف و نمونه‌هایی مشابه کاربردهایی از این سامانه‌ها هستند (Harris and Davenport, 2005).

۲. پیشینه

اخلاق در حوزه هوش مصنوعی همواره یکی از موضوعات کلیدی و مورد توجه محققان و متخصصان بوده است. با پیشرفت سریع فناوری‌های هوش مصنوعی نیاز به بررسی دقیق‌تر ابعاد اخلاقی این سامانه‌ها بیش از پیش احساس می‌شود. در مطالعات بسیاری مشکلات و راهکارهای اخلاقی مرتبط با هوش مصنوعی بررسی شده است. در ادامه، برخی از مهم‌ترین آن‌ها بازبینی می‌شود.

خداقلی‌زاده^۶ (2020) در ترجمه گزارش سال ۲۰۱۸ مؤسسه هوش مصنوعی اکنون تأثیرات اجتماعی و اخلاقی هوش مصنوعی را بررسی و اثر مهم شفافیت و مسئولیت‌پذیری را در توسعه و پیاده‌سازی سامانه‌های هوش مصنوعی برجسته می‌کند. همچنین، خلیلی^۷

6. Khodagholizadeh

7. Khalili

1. Facial Recognition Authentication Systems

2. Social Credit Systems

3. Security

4. Suspect Identification Systems

5. Automated Decision Systems (ADS)

کوستکا^{۱۱} (2019) تأثیر سیستم‌های اعتبار اجتماعی چین را از منظر عمومی بررسی کرد و نشان داد این سیستم چگونه پذیرش فراوانی در میان مردم دارد. بیوچمپ^{۱۲} و چیلدرس^{۱۳} (2019) در کتاب *اصول اخلاق زیست‌پزشکی*^{۱۴} اصول اخلاقی‌ای را تشریح کرده‌اند که می‌توان آن‌ها را در توسعه سامانه‌های هوش مصنوعی نیز به کار برد. همچنین، وینفیلد^{۱۵} و جیروتکا^{۱۶} (2018) بر اهمیت حاکمیت اخلاقی به منظور اعتمادسازی به سامانه‌های رباتیک و هوش مصنوعی تأکید می‌کنند. گیلپین^{۱۷} (2018) نیز به اهمیت توضیح‌پذیری در ماشین‌های یادگیرنده پرداخته است. علاوه بر این، مطالعاتی مانند دُشی-ولز^{۱۸} و کیم^{۱۹} (2017) بر ضرورت تفسیرپذیری الگوریتم‌های یادگیری ماشین تأکید کرده‌اند.

گودفلو^{۲۰} و همکاران (2016) در کتاب *یادگیری عمیق*^{۲۱} اصول و روش‌های یادگیری عمیق را مطرح می‌کنند که پایه‌ای برای بسیاری از سامانه‌های هوش مصنوعی مدرن است. اونیل^{۲۲} (2016) در کتاب *سلاح‌های نابودگر ریاضی*^{۲۳} خطرات افزایش نابرابری و تهدید دموکراسی به وسیله داده‌های بزرگ و هوش مصنوعی را مطرح می‌کند. همچنین، مطالعات میتل‌اشتات^{۲۴} و همکاران (2016) اخلاق الگوریتم‌ها و ایجاد نقشه‌ای از مباحث مختلف این حوزه را بررسی می‌کنند. هریس^{۲۵} و داوِنپورت^{۲۶} (2005) نیز به بلوغ تصمیم‌گیری خودکار و تأثیر آن در جوامع پرداخته‌اند. درنهایت، باویر^{۲۷} (2004) نیز در مقاله خود تعارض بین امنیت و حفظ حریم خصوصی را در فناوری تشخیص چهره مطرح کرده است.

این مطالعات به صورت مستقیم به ترکیب دیدگاه‌های سنتی مانند حکمت عملی صدرایی و اصول اخلاقی مدرن با فناوری‌های جدید نپرداخته‌اند، اما نشانه‌هایی

(2019) در ترجمه گزارش اخلاق هوش مصنوعی در گوگل اخلاق هوش مصنوعی در شرکت‌های بزرگ مانند گوگل را بررسی و درباره مسائل مرتبط با تبعیض الگوریتمی و حفظ حریم خصوصی بحث می‌کند.

متوادی^۱ (2018) در مقاله خود مفهوم حریم خصوصی انسان را از منظر قرآن و حدیث مطالعه می‌کند و بر اهمیت حفظ اسرار و اطلاعات شخصی در مقابل فناوری‌های نوین تأکید می‌کند. طوسی^۲ (2014) در کتاب خود با عنوان *درآمدی بر فلسفه سیاسی صدرالمآلهین اصول حکمت عملی صدرایی و کاربرد آن‌ها در حوزه سیاست و حکمرانی* بررسی می‌کند. لکزایی^۳ (2005) نیز در تحقیقی با عنوان *سیاست اخلاقی در اندیشه صدرالمآلهین مفاهیم اخلاقی مطرح شده در حکمت عملی صدرایی و ارتباط آن‌ها با مسائل روزمره اجتماعی و سیاسی دقیق مطالعه می‌کند.*

در حوزه بین‌المللی، سازمان همکاری و توسعه اقتصادی^۴ (2021) در توصیه‌های خود درباره هوش مصنوعی بر اصول اخلاقی و حقوقی تأکید کرده است. این توصیه‌ها شامل اصولی برای توسعه و استفاده مسئولانه از فناوری‌های هوش مصنوعی است و بر اهمیت تضمین شفافیت، عدالت، مسئولیت‌پذیری، احترام به حقوق بشر در طراحی و پیاده‌سازی سامانه‌های هوش مصنوعی تأکید می‌کند. کمیسیون اروپا^۵ (2020) در سند رسمی خود درباره هوش مصنوعی^۶ چهارچوبی برای توسعه و استفاده اخلاقی از هوش مصنوعی در اتحادیه اروپا ارائه می‌دهد. در این سند بر اهمیت همسویی نوآوری‌های فناورانه و اصول اخلاقی، از جمله شفافیت، عدالت، حفظ حریم خصوصی و امنیت تأکید شده است. همچنین، توصیه‌هایی برای تنظیم‌گری هوش مصنوعی در زمینه‌های پرخطر و حساس ارائه شده است. راسل^۷ و نورویگ^۸ (2020) در کتاب *هوش مصنوعی: رویکردی نوین*^۹ مروری جامع بر مفاهیم و کاربردهای هوش مصنوعی ارائه داده‌اند. همچنین، بوکن^{۱۰} (2020) در مقاله خود سیستم اعتبار اجتماعی چین را معرفی و تأثیرات آن را در آزادی و اختیار افراد بررسی می‌کند.

11. Kostka
12. Beauchamp
13. Childress
14. Principles of Biomedical Ethics
15. Winfield
16. Jirotko
17. Gilpin
18. Doshi-Velez
19. Kim
20. Goodfellow
21. Deep Learning
22. O'Neil
23. Weapons of Math Destruction
24. Mittelstadt
25. Harris
26. Davenport
27. Bowyer

1. Metvadi
2. Tusi
3. Lakzai
4. Organisation for Economic Co-operation and Development
5. European Commission
6. White Paper on Artificial Intelligence
7. Russell
8. Norvig
9. Artificial Intelligence: A Modern Approach
10. Boquen

اخلاقی ناسازگار است. در ادامه، طراحی و کاربری‌های فنی-مهندسی و مشکلات اجتماعی ناشی از آن‌ها در قالب ادبیات اخلاقی صورت‌بندی می‌شوند تا در بخش بعد با تکیه بر حکمت عملی صدرایی نقد و بررسی شوند.

۴-۲. انصاف (عدالت و بی‌طرفی)

با توجه به مسائل مطرح‌شده، یکی از اصول اخلاقی موردتعرض این سامانه‌ها انصاف^۱ و بی‌طرفی است. اگر در برنامه‌ریزی‌های مرتبط با مردم دیدگاه جامع و درستی از انسان نباشد و تفاوت‌های فردی و اجتماعی در نظر گرفته نشود، به قضاوت نادرست، ایجاد مشکل برای مردم و در نتیجه تضییع حقوق آن‌ها منجر می‌شوند. سامانه‌های هوش مصنوعی براساس مجموعه محدودی از داده‌ها آموزش می‌بینند و مبتنی بر همان محدودیت برای همه گزینه‌های آینده تصمیم‌گیری می‌کنند (Khalili, 2019). بنابراین، خطر بروز سوگیری یا نبود بی‌طرفی در سامانه‌های هوش مصنوعی تأثیراتی گسترده و جدی در جوامع انسانی می‌گذارد (O'Neil, 2016). ازاین‌رو، طراحی سامانه‌هایی که بی‌طرف و دربرگیری بیشتری داشته باشند، امری ضروری است تا از تبعیض‌های ناعادلانه جلوگیری شود (Noble, 2018). بااین‌حال، حذف کامل سوگیری در الگوریتم‌ها غیرممکن است و تحقیقات نشان می‌دهد در عمل توجه کافی از سوی دولت‌ها به این مسئله نشده است (Mittelstadt et al., 2016). ازاین‌رو، کمی توجه به تضییع حقوق افراد منجر شده و اعتماد عمومی را به سامانه‌های هوش مصنوعی کاهش داده است (IEEE, 2019; Jobin et al., 2019). برای نمونه، در سیستم‌های تصمیم‌گیری خودکار سوگیری‌های الگوریتمی ممکن است به تبعیض علیه گروه‌های خاص منجر شود که این موضوع نقض حقوق بشر است (Winfield and Jirotko, 2018). بنابراین، ضروری است سیاست‌گذاران و طراحان سامانه‌های هوش مصنوعی با رعایت استانداردهای اخلاقی، افزایش شفافیت و پاسخ‌گویی به کاهش سوگیری‌ها و افزایش عدالت در کاربرد این فناوری‌ها کمک کنند (Doshi-Velez and Kim, 2017).

۴-۲. آزادی و اختیار

سامانه‌های اعتبار اجتماعی که یکی از انواع سامانه‌های تشخیص چهره‌اند، مجموعه‌ای از اطلاعات رفتاری

برای توسعه و استفاده اخلاقی از هوش مصنوعی را مطرح کرده‌اند. ازاین‌رو، با توجه به تنوع و گستردگی مطالعات انجام‌شده، پژوهش حاضر قصد دارد با تلفیق این دیدگاه‌ها و ارائه راهکارهای عملی به بررسی اخلاقی سامانه‌های هوش مصنوعی پردازد و نقشی میان‌افزار بین فلسفه سنتی و فناوری‌های جدید ایفا کند.

۳. روش تحقیق

روش تحقیق این پژوهش کیفی و مبتنی بر تحلیل محتوا بوده است. از آنجا که تحلیل محتوا روشی نظام‌مند برای بررسی و تفسیر متن‌ها و داده‌هاست، در این تحقیق از آن برای شناسایی مسائل اخلاقی دو سامانه هوش مصنوعی (سامانه تشخیص چهره و سامانه تصمیم‌گیری خودکار) و بررسی اصول اخلاقی نقض شده به وسیله آن‌ها به کار گرفته شده است.

تحلیل محتوا شامل بررسی محتوای مقاله‌های علمی، گزارش‌های فنی، کتاب‌های معتبر در زمینه هوش مصنوعی و اخلاق و متن‌های فلسفی مربوط به حکمت عملی صدرایی است. این متن‌ها براساس مفاهیم اصلی مانند اخلاق هوش مصنوعی، سامانه‌های تصمیم‌گیری خودکار، سامانه‌های تشخیص چهره، انصاف، آزادی، حریم خصوصی، مسئولیت‌پذیری و عمل به ظن معتبر کدگذاری شدند. داده‌های کدگذاری شده به منظور شناسایی الگوها و مسائل اخلاقی مرتبط با سامانه‌ها تحلیل و سپس نتایج در چهارچوب حکمت عملی صدرایی طبقه‌بندی و تفسیر شدند.

جامعه بررسی شده شامل ۵۰ مقاله علمی از پایگاه‌های معتبر، گزارش‌های سازمان‌های بین‌المللی مرتبط با هوش مصنوعی و اخلاق و همه آثار مرتبط با حکمت عملی ملاصدرا بوده است. این منابع با هدف شناسایی مسائل اخلاقی و تحلیل آن‌ها از منظر حکمت عملی صدرایی انتخاب شدند.

۴. بحث و بررسی یافته‌ها

۴-۱. اصول اخلاقی مورد تعرض به وسیله دو سامانه هوش

مصنوعی

بررسی کاربردهای گوناگون دو سامانه تشخیص چهره و تصمیم‌گیری خودکار و تأثیراتشان در جوامع نشان می‌دهد نحوه طراحی و به‌کارگیری آن‌ها به مشکلات اجتماعی منجر می‌شود، به‌گونه‌ای که کاربرد آن‌ها با برخی از اصول

۴-۴. تبیین‌پذیری^۱ و پاسخ‌گویی

پیش‌بینی و تصمیم‌گیری خودکار موجب ارتقای ابعاد گوناگون زندگی می‌شود. اما اگر سامانه‌ها نتوانند تصمیمات خود را به‌گونه‌ای پذیرفتنی تبیین کنند، نمی‌شود درباره صحت یا غلط بودن آن تصمیمات داوری کرد. بنابراین، تبیین‌پذیری امری اساسی است تا بتوان از چرایی تصمیم پرسید، آن را فهمید و به آن اعتماد کرد.

تبیین‌پذیری برای کاربران و طراحان این امکان را فراهم می‌کند تا شفاف‌تر باشند و درباره تصمیمات خود توضیحاتی با سطح مناسبی از جزئیات ارائه دهند. این توضیحات برای تضمین انصاف، شناسایی سوگیری بالقوه در داده‌های آموزشی و اطمینان از عملکرد مناسب الگوریتم‌ها مهم‌اند (Gilpin, 2018).

تبیین‌پذیری مقدمه‌ای برای پاسخ‌گویی به مردم درباره عملکرد سامانه‌های تصمیم‌گیری خودکار است. همچنین، بهره‌برداران این سامانه‌ها باید شرایط اعتراض افراد تحت تأثیر آن‌ها را فراهم کنند، به‌گونه‌ای که آمادگی پاسخ‌گویی و جبران تصمیم‌های غلط و ظلم‌های احتمالی را داشته باشند (Kuziemski and Misuraca, 2020).

۴-۵. عمل به ظن معتبر

از آنجایی‌که سامانه‌های تشخیص تمایلات براساس داده‌های محدود آموزش درباره احساسات درونی، مقاصد و شخصیت افراد قضاوت می‌کنند، براساس نشانه‌های کمی حکم کلی صادر می‌کنند. درواقع، چون درباره شخصیت و احساسات فرد از بیرون قضاوت می‌شود، ممکن است داوری آن‌ها با درصد زیاد از خطا همراه باشد. در نتیجه، تصمیماتی غلط گرفته و به افراد ظلم شود. برای نمونه، اگر سامانه فقط با دیدن حالت خاص چهره‌ای، که افرادی در داده‌های آموزش قبل از ارتکاب جرم داشته‌اند، قضاوت کند و بگوید این فرد تمایل به ارتکاب جرم دارد، در حالی که او واقعاً چنین قصدی ندارد، حتی اگر این قضاوت پیامد قضایی هم نداشته باشد، باعث هدر رفتن وقت و ضایع شدن حق او می‌شود. بنابراین، می‌توان گفت ظن حاصل از مطالعه داده‌های آموزش از نوعی تمثیل یا استقراء ناقص نتیجه می‌شود و به‌هیچ‌وجه ظنی معتبر نزد عقلانیست.

افراد، شرکت‌ها و نهادها را به‌وسیله شبکه‌های گسترده دوربین‌های زنده به‌طور دائم تحلیل و به‌روزرسانی می‌کنند. از این سامانه‌ها به آقابالاسر هم تعبیر می‌شود که با قضاوت درباره رفتار شهروندان ایده امتیازدهی را در همه جنبه‌های زندگی گسترش می‌دهند. نظارت بر رفتار مردم مرز مشخصی ندارد. بر پایه مطالعات کشورهای اجراکننده آن، از جمله چین با این استدلال که نظم و امنیت عمومی مقدم بر سایر ارزش‌هاست حریمی خصوصی برای شهروندان قائل نیستند (Boquen, 2020).

اعمال تشویق و تنبیه گسترده و عمیق باعث می‌شود مردم حتی در چهارچوب قوانین و هنجارهای اجتماعی نتوانند آزادانه تصمیم بگیرند و رفتارهای خود را با اختیار انجام دهند؛ بلکه همواره در اجرای کارهایشان نوعی اکراه احساس کنند. زیرا نظام تشویق و تنبیه به صورت خودکار، قطعی و دائمی آزادی و اختیار انسان را در انجام افعالش محدود می‌کند. درواقع، انسان پس از گذشت مدت کمی مانند ماشین به ناچار رفتارش را با سیاست‌ها و ارزش‌های حاکم بر سامانه اعتبار اجتماعی منطبق می‌کند.

البته مرز میان مجازات مشروع و معقول در مقابل سلب اختیار و آزادی‌های مشروع و مرز میان عرصه اجتماعی در مقابل حریم خصوصی در جوامع مختلف با فرهنگ‌های گوناگون ممکن است متفاوت باشد، اما این محدوده با توجه به ویژگی‌های فطری و ذاتی انسان و هدف خلقت محدوده‌ای مشخص است که افراط و تفریط از آن به نظم اجتماعی و کارکردهای انسانی آسیب‌هایی می‌زند.

۴-۳. حریم خصوصی و حفظ اسرار

شبکه‌های گسترده دوربین‌های زنده تشخیص چهره با جمع‌آوری عظیم داده‌های مردم و امکان ترکیب آن‌ها حریم خصوصی را با مشکل مواجه می‌کنند. سامانه‌های تشخیص چهره تمام داده‌های زیستی مردم در مکان‌های مختلف را ذخیره و پردازش می‌کنند و از این طریق اطلاعات زیادی از نحوه زندگی و ارتباط افراد جامعه جمع‌آوری می‌کنند که به‌آسانی در دسترس دولت‌ها و شرکت‌هاست. بنابراین، اگر محدودیت و ضوابطی برای استفاده از این سامانه‌ها وضع نشود، حریم خصوصی افراد تقریباً از بین می‌رود. درست است این سامانه‌ها امنیت را تقویت می‌کنند، اما در مقابل ممکن است حریم خصوصی و حفظ اسرار را خدشه‌دار کنند (Bowyer, 2004).

۱. Interpretability: تبیین‌پذیری، توضیح‌پذیری یا فهم‌پذیری الگوی است که از یادگیری ماشین حاصل می‌شود و این قابلیت را دارد تا بوسیله اصطلاحات فهم‌پذیری برای انسان توضیح داده شود. از آنجایی‌که اعتماد به سامانه‌های غیرقابل فهم مسئله‌برانگیز است، تبیین‌پذیری به مباحث اخلاق ارتباط پیدا می‌کند.

۴-۶. نقد اخلاقی عملکرد دو سامانه هوش مصنوعی

در این بخش عملکرد دو سامانه تشخیص چهره و تصمیم‌گیری خودکار با تکیه بر حکمت عملی صدرایی نقد و بررسی می‌شود. در گفتمان فلسفه اسلامی حکمت عملی عهده‌دار تبیین بایدها و نبایدها و حسن و قبح افعال در ساحت‌های گوناگون است. اخلاق فردی، تدبیر منزل و سیاست مدن سه ساحت حکمت عملی‌اند (Shirazi, 1981a). آنچه موضوع بحث این نوشتار است در حیطه اخلاق فردی و تدبیر منزل نیست، بلکه در حیطه سیاست، یعنی نحوه مواجهه حکومت‌ها با مسائل اجتماعی، قرار می‌گیرد.

۴-۷. جایگاه اخلاق و سیاست در حکمت متعالیه

تمامی فیلسوفان اسلامی تقسیم‌بندی حکمت را به دو قسم حکمت نظری و حکمت عملی که ارسطو ارائه کرد قبول دارند. این تقسیم‌بندی سیاست را ذیل حکمت عملی قرار می‌دهد. معرفت سیاسی نزد ارسطو و فیلسوفان اسلامی یکی از سه شاخه حکمت عملی یعنی اخلاق، تدبیر منزل و سیاست مدن است. این دسته از فیلسوفان سیاست را این‌گونه تعریف کرده‌اند: علم به مشارکت انسان‌ها در جامعه برای تأمین مصالح اجتماعی و بقای نوع بشر. همچنین، تعریف برخی دیگر از سیاست این‌گونه است: شناخت اموری که به وسیله آن‌ها زمینه انجام اعمال نیک برای مردم شهرها فراهم می‌شود (Najm al-Din Ali Katebi, 1874).

برایند جامع دیدگاه صدرایی را می‌توان بدین‌گونه خلاصه کرد که حقیقت سیاست از نگاه وی در مرتبه‌ای از مراتب خلقت و ساحتی از افعال الهی است و از حیث شرف وجودی در مرتبه بالایی در نظام وجودی قرار ندارد. بنابراین، فقط مقدمه و ابزاری است برای رسیدن به مراتب بالاتر وجود و غایات برتر. ملاصدرا در بخش سوم از ملحقات کتاب المظاهر الالهیه آنجا که درباره مقاصد شش‌گانه کتاب و محتوای مباحث آن‌ها توضیح می‌دهد، دیدگاه خود را راجع به علم سیاست و تدبیر حیات اجتماعی بشر این‌گونه تبیین می‌کند (Lakzai, 2004):

یاددادن روش آبادسازی منازل و مراحل سلوک به سوی خدای تعالی و طریقه بندگی وی و چگونگی برداشت توشه و مرکب برای سفر آخرت و مهیاشدن به منظور ریاضت‌دادن و تعلیف کردن مرکب تن. مقصود از این قسمت علم به چگونگی رفتار و برخورد انسان با عینیات این‌جهانی است که برخی از این عینیات در درون خود

انسان است، مانند صفات نفسی از قبیل قوای شهوانی و غضبی و علم به این عینیات را تهذیب الاخلاق می‌نامند و برخی از این عینیات (امور) در برون انسان است (خارج از نفسانیات شخصی است)، چه در یک منزل، مانند خانواده و خدمتکاران و پدر و فرزند، که علم به این عینیات را تدبیر المنزل می‌نامند و چه در یک شهر که علم به این عینیات را علم سیاست و احکام شریعت، از قبیل دیات و قصاص و احکام نام می‌نهند (Shirazi, 1999).

۴-۸. جایگاه شریعت و سیاست در حکمت متعالیه

ملاصدرا در عبارتی در مقام بیان جایگاه شریعت می‌گوید مبدأ حرکت شریعت نهایت حرکت سیاست است که همانا حرکت دادن انسان از عالم جسمانی به سوی عالم الوهی است؛ مسیری که تدابیر سیاست بشری از آن عاجز است: «شریعت حرکتی است که آغاز آن پایان سیاست است، زیرا شریعت است که نفس و قوای او را از عالم ترکیب به ساحت عالم وحدت واصل می‌کند. شریعت انسان را به عالم الوهی تذکر می‌دهد و او را از انحطاط به سوی شهور و غضب و عواقب آن‌ها دور می‌سازد» (Shirazi, 1981b).

حکیم ملاصدرا در موضوعی دیگر شریعت را همچون روح می‌داند در جسم سیاست (Shirazi, 1981b). سیاست ساخته بشر هرچند نیازهای جسمانی و دنیوی مردان و زنان جامعه را تأمین می‌کند، هرگز قادر به تأمین نیازهای روحی و معنوی آنان نیست. اهمیت برنامه‌های ارائه شده در شریعت نه از منظر حیثیات مادی و تدبیر امور دنیوی، بلکه به علت حیثیات معنوی و روحانی آن‌هاست. شریعت تأمین‌کننده اهداف خلقت و سوق دادن بشر به جوار الهی و دار کرامت و رحمت اوست (Lakzai, 2004).

۴-۹. انصاف در حکمت متعالیه

از منظر این نوشته انصاف و عدالت مفهومی نزدیک با بی‌طرفی در قضاوت یا تصمیم‌گیری دارد. اما معنایی که ملاصدرا با لفظ عدالت اراده می‌کند، اعم از بی‌طرفی است. جست‌وجو در آثار صدر المتألهین نشان می‌دهد که او از کلمه النصف و الانصاف معنایی نزدیک به معنای بی‌طرفی اراده کرده است.

ملاصدرا در کتاب المبدأ و المعاد (1972) در بیان اوصاف رئیس اول مدینه به انصاف و عدل اشاره می‌کند

از دیدگاه ملاصدرا راه رسیدن به سعادت و خوشبختی مبتنی بر عمل اختیاری انسان است تا انسان با حرکت اختیاری خود به آن کمال که دست قدرت آفریدگار در باطن و نهاد او قرار داده است برسد. زیرا انسان فطرتاً طالب کمال و سعادت است، ولی رسیدن به آن بر پایه اختیار و اراده خودآیین انسان نهاده شده است. از نظر ملاصدرا نفس انسان این سعادت و بهجت را از راه تلاش و تکرار عمل و کارهایی که موجب طهارت نفس از کدورات است به دست می‌آورد. بنابراین، می‌توان چنین نتیجه گرفت که وضع قوانین و اعمال مجازات به وسیله انسان‌ها بدون داشتن منشأ الهی و غیبی ممکن است به زندگی دنیوی جوامع انسانی نظم دهد، اما انسان را محصور و تنیده در همین عالم طبیعی نگه می‌دارد و به مرتبه وجود مجرد و کمال در مرتبه عقلانی نمی‌رساند، زیرا راه رسیدن به آن‌ها استخدام ظاهر برای باطن است.

۴-۱۱. حریم خصوصی و حفظ اسرار در حکمت متعالیه

به منظور جست‌وجو و ریشه‌یابی مفهوم حریم خصوصی در متون ملاصدرا باید مفاهیم معادلی که در گذشته کاربرد داشته است شناسایی و بررسی کرد. رایج‌ترین مفاهیم مرتبط با حریم خصوصی در قرآن و روایات کلماتی همچون تجسس و دنبال کردن عورات اند. لغت‌شناسان در معنای تجسس گفته‌اند: «تفتیش از لایه‌های پنهان امور تجسس است که بیشتر در امور شر به کار می‌رود».

حریم خصوصی در فرهنگ اسلامی را می‌توان این‌گونه تعریف کرد: «حریم خصوصی خلوت و امور پنهانی افراد است که ورود یا هرگونه دخالت و تعرض به آن برای دیگران ممنوع است». در این تعریف امور پنهانی لزوماً عیوب افراد نیستند، بلکه ممکن است امور باشند که افراد به آشکار شدن آن‌ها راضی نیستند. همچنین، صرفاً رضایت فرد مجوز نقض حریم خصوصی و آشکار کردن عیب فرد نیست، زیرا در موقعیت‌هایی ممکن است فرد رضایت داشته باشد، اما این آشکار کردن عیب با حریم عمومی و امنیت اجتماعی از لحاظ شریعت اسلامی در تعارض باشد (Metvadi, 2018).

ملاصدرا در کتاب *الشواهد الربوبیه* (1981b) به مسئله حریم خصوصی و تجسس اشاره می‌کند. او در بخش صفات رئیس اول مدینه می‌فرماید: «هشتمین صفت رئیس اول مدینه این است که درباره همه خلق خدا مهربان و باعاطفه باشد. نه هنگام مشاهده منکر شدیداً غضبناک شود و نه حدود الهی را تعطیل کند؛ بدون اینکه به تجسس

و می‌فرماید: «رئیس اول مدینه ... باید ذاتاً دوست‌دار عدل باشد و کسانی را که اهل عدالت‌اند دوست بدارد. از ظلم منجر و نسبت به اهل آن خشمگین باشد. اهل انصاف باشد و حق مظلوم را بدهد و بر حکم عدل گردن نهد و بر رأی خود هنگامی که به عدل فراخوانده می‌شود لجاجت نرزد و هرگز دعوت به سوی جور و ظلم را نپذیرد».

صدرالمتألهین در کتاب *شرح اصول الکافی* (2004) نیز در شرح حدیث امام صادق (ع) معروف به جنود عقل و جهل در ادامه بخش صفات صاحبان عقل و ویژگی‌های رفتاری آن‌ها به عبارت «الْإِنصَافُ وَ ضِدُّهُ الْخَمِيَّةُ» اشاره می‌کند و در شرح آن می‌فرماید: «انصاف و ضد آن حمیت (طرف‌داری برخلاف حق) است: انصاف انتخاب یکی از دو جزء مساوی یک شیء است. انصاف از النصف گرفته شده و اسم انصاف است. النصف همان عدل است. برای نمونه، شایسته است که قاضی در برابر شاکی و متشاکی با انصاف رفتار کند، یعنی آن دو نزد قاضی مساوی باشند... منظور از حمیت نیز این است که انسان به واسطه غیرت برآمده از نفسانیات و تعصب روی مذهب یا چیز دیگری از عدل خارج شود و از حق تعدی کند و این از صفات جهل است، چون ضد انصاف و عدل است». ملاصدرا در اینجا دقیقاً به معنای بی‌طرفی اشاره می‌کند و حاکم را ملزم به رعایت انصاف می‌داند و نداشتن انصاف را خروج از عدالت و تعدی از حق معرفی می‌کند.

۴-۱۰. آزادی و اختیار در حکمت متعالیه

صدرالمتألهین (1981b) هنگامی که از آزادی سخن می‌گوید، آزادی از هوای نفس را مهم‌ترین دغدغه خود معرفی می‌کند و معتقد است تا زمانی که انسان خود را از قیدوبند تمایلات مراتب نباتی و حیوانی خلاص نکرده است، به هیچ فضیلت دیگری حتی عدالت دست نخواهد یافت. همچنین، صدرالمتألهین (1972) در موضعی دیگر می‌گوید: «غرض از وضع شرایع و ایجاب طاعات آن است که شهادت غیب را خدمت نماید و شهوات عقول را خدمت کنند و جز به کل و دنیا به آخرت برگردد و محسوس معقول شود و از عکس این امور منجر گردد».

باید توجه کرد به رسمیت‌شناختن قانون شریعت با این عنوان که کامل‌ترین قانون برای سعادت بشر است، این ارتکاز ذهنی را برای متقدمان پدید آورده است که موضوعاتی از قبیل حقوق فردی و اجتماعی، از جمله آزادی باید به متن قانون شریعت ارجاع داده شود و گویی تعیین تکلیف در آن‌ها از عهده فیلسوف خارج است (Tusi, 2014).

به گونه ای که وقتی دیگران از شما سؤال می کنند، انتظار توضیح آن عمل و تصمیم را دارند. انسان کامل و عالم حقیقی در مکتب صدرایی انسان عارف زاهدی نیست که تمام توجهش به عالم مجرد است و از عالم جسمانی غافل شده و گوشه انزوا برگزیده است.

انسان کامل ناگزیر است هم به اسرار علوم الهی آگاه باشد و هم به علوم مربوط به معاش، یعنی همان علم به احکام حلال و حرام امور دنیوی. صدرالمتألهین (۱۹۸۴) در این زمینه می فرماید: «علما بر سه دسته اند: ۱. عالم بالله غیر عالم به امر الله؛ عالم به امر الله و غیر عالم بالله؛ ۳. عالم بالله و به امر الله. دسته اول آنانی اند که معرفت الهی بر قلب آنان مستولی شده و غرق در مشاهدۀ نور جلال و صفات کبریایی حق اند و از احکام (حلال و حرام) جز به مقدار ضرورت نمی دانند. دسته دوم عالمانی اند که عالم به احکام الهی اند و غیر آگاه از دقایق احکام و اسرار جلالت الهیه اند. دسته سوم عالمانی اند که همواره در مرز میان عالم معقولات و عالم محسوسات اند. این گروه به حب الهی همواره با حق اند و در عین حال با شفقت و رحمت با خلق رفتار می کنند. هرگاه که از سوی (توجه) پروردگار خود به سوی خلق مراجعت می کنند، یکی از افراد آنان می شوند، گویی که اصلاً خدا را نمی شناسند و هنگامی که مشغول به ذکر و بندگی خدایند، گویا که اصلاً خلق را نمی شناسند و این روش سلوک مرسلین و صدیقین است.

از خصلت های انسان کامل آن است که به دور از هرگونه ریا و نفاق و ابراز تقدس در جامعه و در میان مردم به خدمتگزاری مشغول است. در حالی که کثرات عالم جسمانی (ناشی از اشتغالات جسمانی و اجتماعی) مانع و مزاحم حضور قلب و عبادت و تهجد او با حق تعالی نیست (Tusi, 2014).

با این تبیین، در حکمت عملی صدرایی رسیدگی به مشکلات مردم انسان کامل را از سلوک باز نمی دارد، بلکه کمال نفس است. بنابراین، پاسخ گویی به مردم و رسیدگی به مشکلات آن ها با شفقت و رحمت، خود سلوک الی الله است. از نظر صدرالمتألهین (۱۹۷۲) رئیس اول مدینه باید پاسخ گو باشد، حق مظلوم را بدهد، بر حکم عدل گردن نهد، بر رأی خود هنگامی که به عدل فراخوانده می شود لجاجت نرزد و هرگز دعوت به سوی جور و ظلم را نپذیرد.

۴-۱۳. عمل به ظن معتبر در حکمت متعالیه

صدرالمتألهین در کتاب تفسیر القرآن کریم (۱۹۸۲) در چندین موضع به آیه «وَمَا يَتَّبِعْ أَكْثُرُهُمْ إِلَّا ظَنًّا إِنَّ الظَّنَّ

همت گمارد. پس چگونه این طور نباشد! در حالی که او آگاه به اسرار الهی در لوازم مقدرات است.» ملاصدرا قید مهمی را در این فرایند ذکر می کند و آن این است که رئیس مدینه نباید تجسس کند. به عبارت دیگر، حریم خصوصی مردم باید در نظر گرفته شود.

حفظ اسرار و اطلاعات نیز یکی از مصادیق حفظ حریم خصوصی است. این موضوع به صورت مستقیم و متمایز در آثار ملاصدرا در حیطه برون دینی بحث نشده است. اما چنان که گفته شد، از نظر ملاصدرا قانون جامع و کاملی که انسان ها را به بالاترین مقاصد انسانی رهنمون می کند، قانون شریعت اسلام است. صدرالمتألهین به واسطه اشاره به آیات و روایات مرتبط به این موضوعات اشاره کرده است.

صدرالمتألهین در کتاب شرح اصول الکافی (۲۰۰۴) ذیل حدیث معروف جنود عقل و جهل به این فرمایش حضرت امام صادق (ع) اشاره می کند: «وَالْكَيْتُمَانُ وَضِدُّهُ الْإِفْشَاءُ» (حفظ اسرار و ضد آن، افشای اسرار و پرده دری است) و در شرح آن می نویسد: «همانا از محسنات اخلاق یک فرد عاقل کامل این است که عیوب برادران مسلمانش را می پوشاند و از افشای سری که به او سپرده شده است، اجتناب می کند. بر شخص واجب است که عیوب برادرش را انکار کند، اگرچه به کذب. زیرا راست گویی در همه حال واجب نیست. همچنان که برای شخص جایز است عیوب و اسرار خود را پنهان کند، حتی اگر نیاز به کذب باشد. همین حالت برای برادر مسلمانش هم مطرح است، زیرا برادرش هم به منزله خود آن شخص است و مانند یک شخص واحدند. پس همانا مؤمنین مانند یک نفس واحدند و اختلافی بین آن ها نیست. مگر بدنشان ... به کسی گفته شد چگونه حفظ اسرار می کنی؟ گفت: خبردهنده را انکار می کنم و برای خبرگیرنده قسم می خورم. همچنین، کسی به دیگری سری را بیان کرد، سپس گفت آن را حفظ می کنی؟ پاسخ داد بلکه فراموش می کنم. و به ابی یزید گفته شد: با چه کسی از مردم هم نشین می شوی؟ پاسخ داد با کسی که از تو همانی را که خدا می داند بداند. سپس مانند پوشاندن خدا برای تو بپوشاند.» بنابراین، بنا به نظر صدرالمتألهین نه تنها افشای سیر کار قبیحی است، بلکه کتمان آن واجب است. حتی اگر کتمان سیر نیازمند گفتن دروغ باشد.

۴-۱۲. مسئولیت پذیری و پاسخ گویی در حکمت متعالیه

مراد از پاسخ گویی چنان که در فرهنگ لغت آکسفورد نیز آمده، مسئول بودن در برابر تصمیمات و اعمال است،

یعنی همان مرتبه علم و یقین است. هیچ شکی نیست که تمثیل صرفاً موجب ظن می‌شود. بنابراین، از ناحیه عقلا باطل شمرده می‌شود، زیرا تمثیل طریق عقلا برای تحصیل علم عقلایی به واقعیت نیست. منظور این است که تمثیل دارای ملاک عام مشترک نیست و عقلا برای کشف واقع به آن تکیه نمی‌کنند. راز بطلان تمثیل این است که عقلا در امور خود صرفاً به ظن و تخمین تکیه نمی‌کنند، مگر در آنجا که موردی را مسامحه کنند.»

ملاصدرا در اینجا تمثیل را نزد عقلا معتبر نمی‌داند و بیان می‌کند که تمثیل طریق عقلا حتی برای تحصیل علم عقلایی هم نیست. عقلا در امور خود به تمثیل تکیه نمی‌کنند، مگر در مواردی که مسامحه کنند. به کارگیری لفظ عقلا به وسیله ملاصدرا مهم است و نشان می‌دهد او در اینجا در مقام بیان ملاکات عقلی نیست و عرف عقلا را در نظر گرفته است. البته استثنایی را ذکر می‌کند و آن در موقعیتی است که عقلا مسامحه می‌کنند. با کمی تأمل روشن می‌شود که اقدام حکومت در قبال مردم، آن هم در جایی که امکان خطا زیاد است، و تصمیم حکومت براساس تمثیل به ظلم به مردم منجر می‌شود. همچنین، جایگاه مسامحه در میان عقلا نیست.

نتیجه‌گیری

سامانه‌های تشخیص چهره اموری مانند احراز هویت، شناسایی مظنونان و امنیت را ظاهراً بهبود داده‌اند، اما در مقابل، اگر ملاحظات اخلاقی در دامنه و شیوه به کارگیری آن‌ها به‌ویژه در زمینه کاربری تشخیص تمایلات رعایت نشود، آسیب‌های زیادی به جامعه وارد می‌کنند. آن‌ها اگر منصفانه عمل نکنند و بی‌طرف نباشند، به سوگیری، جانب‌داری و در نتیجه تصمیم و قضاوت نادرست و تضییع حقوق مردم منجر می‌شوند. صدرالمتألهین در آثار خود به «انصاف» و بی‌طرفی توجه کرده است و به آن در بحث صفت رئیس اول مدینه و رفتار قاضی اشاره می‌کند و انصاف را از مصادیق عدل می‌شمارد. صدرالمتألهین در اثر دیگر خود حاکم را ملزم به رعایت انصاف می‌داند. بنابراین، در نظام حکمت عملی صدارایی تصمیم‌گیری‌ها و اقدامات نباید به‌گونه‌ای باشد که به تعدی به حقوق دیگران و ظلم به آن‌ها منجر شود و حکومت باید کاملاً بی‌طرفانه در برابر آحاد مردم رفتار کند.

سامانه اعتبار اجتماعی که به آقابالاسر هم تعبیر می‌شود، با قضاوت درباره رفتار شهروندان ایده امتیازدهی را در همه جنبه‌های زندگی گسترش می‌دهد. این

لا يُغْنِي مِنَ الْحَقِّ شَيْئًا^۱ (و اکثر این مردم جز از خیال و گمان باطل خود از چیزی پیروی نمی‌کنند، در صورتی که گمان و خیالات موهوم هیچ از حق بی‌نیاز نمی‌گرداند.) اشاره می‌کند و ظن باطل و غیرمعتبر را مذمت می‌کند و می‌فرماید:

«درک حقایق کلی و علم حقیقی از طریق نشانه‌های لفظی امکان‌پذیر نیست. هر حقیقتی علت خاصی دارد و ذات هر حقیقتی فقط با وجه خاصش محقق می‌شود. علم به آن حقیقت نیز صرفاً به وسیله علم به مبادی و مقدمات آن حاصل می‌شود و این همان راه حصول یقین و شناخت است. اما ظنون و سایر ادراکات حاصل وجهی غیر از این است. شناخت تام و کشف صریح با دلالت‌های لفظی محقق نمی‌شود، چراکه دلالت آن‌ها قطعی نیست. البته در امور عملی که احکامی خاص‌اند و مقصود از آن‌ها عملی خاص، ریاضت نفسانی، مصلحت نوعی و نظام جمله‌هاست، ظن و رجحان، برای عمل کافی است. چراکه علم در اینجا وسیله‌ای برای عمل است و چیزی بیشتر از آن نیست. در بیان خداوند متعال که می‌فرماید «إِنَّ الظَّنَّ لَا يُغْنِي مِنَ الْحَقِّ شَيْئًا» اشاره لطیفی به معنایی که ذکر شد وجود دارد. منظور از حق در اینجا اعتقاد صحیح مطابق با واقع است.»

صدرالمتألهین البته در ساحت عمل ظن و رجحان را کافی می‌داند، چراکه علم در این ساحت ابزاری برای عمل است. نکته‌ای که باید به آن توجه کرد این است که ملاصدرا ظن و رجحان را معتبر می‌داند و نه هرگونه وهمی که از طریق تمثیل یا سایر راه‌های سست به دست آمده است و نزد عقلا معتبر نیست. او این معنا را به صراحت در کتاب شرح اصول کافی (۲۰۰۴) بیان می‌کند: «همانا خداوند اهل ظن و تخمین را در این آیه مذمت فرموده است: وَمَا يَتَّبِعْ أَكْثَرُهُمْ إِلَّا ظَنًّا إِنَّ الظَّنَّ لَا يُغْنِي مِنَ الْحَقِّ شَيْئًا». همچنین، او در این کتاب در مواضع مختلف استدلال، صاحبان دیدگاه سست را با عبارت اهل الظن و التخمین معرفی می‌کند.

صدرالمتألهین در کتاب شرح اصول کافی (۲۰۰۴) در بخش برهان بر عدم حجیت قیاس حجیت نبودن قیاس فقهی (تمثیل) را بررسی می‌کند و در دلیل دوم خود با عنوان «التدليل العقلي» بیان می‌کند: «قیاس (فقهی) از آنجایی که همان تمثیل منطقی است، در مقابل قیاس و استقراء منطقی قرار دارد. لذا، فاقد پست‌ترین مرتبه برهان عقلی،

تصمیم‌گیری‌ای قرار دارند که پیش‌تر انسان قرار می‌گرفت. این سامانه‌ها براساس آموزش‌های ماشینی‌ای که دیده‌اند، انتخاب می‌کنند و تصمیم می‌گیرند، به‌گونه‌ای که تصمیم آن‌ها در زندگی مردم و سرنوشتشان اثرگذار است. از این‌رو، لازم است دولت‌ها و شرکت‌ها در قبال تصمیمات سامانه‌ها مسئولیت‌پذیر و پاسخ‌گو باشند. در حکمت عملی صدرایی رسیدگی به مشکلات مردم نه تنها انسان را از سلوک باز نمی‌دارد، بلکه کمال نفس است. بنابراین، پاسخ‌گویی به مردم و رسیدگی به مشکلات آن‌ها با شفقت و رحمت سلوک الی‌الله است. از نظر صدرالمتألهین رئیس اول مدینه باید پاسخ‌گو باشد، حق مظلوم را بدهد و براساس عدل رفتار کند.

سامانه تشخیص تمایلات براساس داده‌های محدود آموزش درباره احساسات درونی، مقاصد و شخصیت افراد قضاوت می‌کند. بنابراین، در عمل از تعداد اندکی از نشانه‌ها حکمی کلی صادر می‌کند. در واقع، چون کسی در خارج از فرد درباره شخصیت و احساسات او قضاوت می‌کند، این دآوری با درصد بسیار زیاد خطا به تصمیمات غلط و ظلم به افراد منجر می‌شود. صدرالمتألهین تمثیل را حتی فاقد پست‌ترین مرتبه برهان می‌داند و معتقد است عقلا برای کشف واقع به تمثیل تکیه نمی‌کنند، زیرا ظن و تخمین تنها دستاورد تمثیل است و عقلا آن را مگر در موضع مسامحه معتبر نمی‌دانند. استثنای تمثیل در موضع مسامحه شامل موضوع ما نمی‌شود، چراکه با کمی تأمل مشخص می‌شود اقدام حکومت‌ها یا شرکت‌ها در قبال مردم، آن‌هم در جایی که امکان خطا زیاد است و تصمیمشان به ظلم منجر می‌شود، موضعی معتبر برای مسامحه نیست.

مبتنی بر مطالب گفته‌شده عملکرد دو سامانه تشخیص چهره و تصمیم‌گیری خودکار حداقل پنج اصل اخلاقی را نقض می‌کنند. در بررسی متون حکمت عملی صدرایی مشخص می‌شود صدرالمتألهین به این پنج اصل توجه کرده است و رئیس اول مدینه را موظف به اجرای آن‌ها می‌داند. بنابراین، براساس حکمت عملی صدرایی این دو سامانه صلاحیت به‌کارگیری با این ویژگی‌ها را ندارند، مگر اینکه نحوه طراحی یا به‌کارگیری آن‌ها به‌گونه‌ای تغییر کند که مسائل اخلاقی و تعرض به اصول اخلاقی منتفی شوند. به‌منظور مواجهه با مسائل اخلاقی سامانه‌های هوش مصنوعی و همسو کردن آن‌ها با اصول حکمت عملی صدرایی لازم است سیاست‌گذاران و ذی‌نفعان اقدامات عملی بسیاری انجام دهند. اولین گام مهم، تدوین قوانین

سامانه با اعمال تشویق و تنبیه بر کوچک‌ترین رفتارهای فردی و اجتماعی آزادی و اختیار را محدود می‌کند و با ایجاد اکراه در مردم به رفتار آنان جهت می‌دهد. در حکمت صدرایی سیر تکامل انسان به‌گونه‌ای است که او با به‌خدمت‌درآوردن قوای جسمانی برای قوای عقلانی، استخدام ظاهر برای باطن، ملک برای ملکوت و دنیا برای آخرت به تکامل وجودی و سعادت می‌رسد. صدرالمتألهین غرض از وضع شرایع و ایجاب طاعات آن را همین می‌داند. بنابراین، به رسمیت شناختن قانون شریعت با این عنوان که کامل‌ترین قانون برای سعادت بشر است، موضوعات حقوق فردی و اجتماعی، مانند آزادی را به متن قانون شریعت ارجاع داده است. همچنین، از دیدگاه ملاصدرا راه رسیدن به سعادت و کمال وجودی مبتنی بر عمل اختیاری انسان است. زیرا انسان فطرتاً طالب کمال و سعادت است، ولی رسیدن به آن بر پایه اختیار و اراده انسان نهاده شده است. از این‌رو، وضع قوانین و اعمال مجازات به‌وسیله انسان‌ها، بدون داشتن منشأ الهی و غیبی، ممکن است به زندگی دنیوی جوامع انسانی نظم دهد، اما انسان را محصور و تنیده در همین عالم طبیعی نگه می‌دارد و به مرتبه وجود مجرد و کمال در مرتبه عقلانی نمی‌رساند.

سامانه‌های تشخیص چهره از طریق شبکه‌های گسترده دوربین‌های تحلیل زنده تمام داده‌های زیستی مردم در مکان‌های مختلف را ذخیره و پردازش می‌کنند. این‌گونه اطلاعات زیادی از نحوه زندگی و ارتباطات افراد جامعه به دست می‌آورند که به‌آسانی در دسترس دولت‌ها و شرکت‌هاست. بنابراین، حفظ حریم خصوصی در به‌کارگیری آن‌ها بسیار مهم است. در حکمت عملی صدرایی رئیس اول مدینه نباید تجسس کند، یعنی نباید حریم خصوصی مردم را نقض کند.

سامانه‌های هوش مصنوعی در فرایند یادگیری ماشین از داده‌های آموزش یاد می‌گیرند و براساس داده‌های ورودی پیش‌بینی می‌کنند. بنابراین، باید الزامات استفاده از داده‌های حساس از نظر حریم خصوصی، تضمین حفظ اسرار و اطلاعات مردم در نظر گرفته شود. صدرالمتألهین در شرح حدیث معروف جنود عقل و جهل به این فرمایش حضرت امام صادق (ع) اشاره می‌کند: «وَأَلِكْتُمَانُ وَضِدُّهُ الْإِفْشَاءُ» و بیان می‌کند که از محسنات اخلاق فرد عاقل کامل این است که عیوب برادران مسلمانانش را پنهاناند و از افشای سری که به او سپرده شده است اجتناب کند.

سامانه‌های تصمیم‌گیری خودکار در موقعیت‌های

منابع

- Beauchamp, T. L., and Childress, J. F. (2019). *Principles of Biomedical Ethics*. Oxford University Press.
- Boquen, A. (2020). "An Introduction to China's Social Corporate Credit System". Retrieved from *New Horizons*. <https://nhglobalpartners.com/chinas-social-credit-system-explained/>
- Bowyer, K. (2004). "Face Recognition Technology: Security Versus Privacy". *IEEE*.
- Doshi-Velez, F., and Kim, B. (2017). "Towards a Rigorous Science of Interpretable Machine Learning".
- European Commission. (2020). "White Paper on Artificial Intelligence: A European Approach to Excellence and Trust". *European Union*.
- Gilpin, L. H. (2018). "Explaining Explanations: An Overview of Interpretability of Machine Learning". *IEEE*.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MI
- Harris, J., and Davenport, T. (2005). "Automated Decision Making Comes of Age". *Accenture*.
- IEEE. (2019). "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems".
- Jain, A. K., Ross, A., and Nandakumar, K. (2011). *Introduction to Biometrics*. Springer Science and Business Media.
- Jobin, A., Ienca, M., and Vayena, E. (2019). "The Global Landscape of AI Ethics Guidelines". *Nature Machine Intelligence*, 1(9), pp. 389–399.
- Katebi, A. N. (1874). *Hikmat al-'Ain and its commentary*. Mashhad: Ferdowsi University.
- Khalili, M. (2019). "Ethics of Artificial Intelligence و مقررات جامع اخلاقی است که به صراحت به مسائل سوگیری، حریم خصوصی و مسئولیت‌پذیری در سامانه‌های هوش مصنوعی می‌پردازد. این قوانین باید براساس اصول انصاف و بی‌طرفی تنظیم شوند تا از تضییع حقوق افراد جلوگیری شود. همچنین، افزایش شفافیت در الگوریتم‌ها و فرایندهای تصمیم‌گیری خودکار امری ضروری است تا کاربران عملکرد و معیارهای تصمیم‌گیری را درک کنند و اعتماد عمومی به سامانه‌های هوش مصنوعی افزایش یابد. آموزش و ارتقای آگاهی عمومی در زمینه اهمیت اخلاق در هوش مصنوعی نیز باید یکی از اولویت‌ها در نظر گرفته شود تا توسعه‌دهندگان، سیاست‌گذاران و عموم مردم به خوبی با سوگیری‌ها و تبعیض‌های احتمالی مقابله کنند. علاوه‌براین، با سرمایه‌گذاری در توسعه فناوری‌های حفظ حریم خصوصی، مانند یادگیری ماشین خصوصی و رمزنگاری پیشرفته می‌توان به حفاظت از داده‌های حساس کمک کرد. همچنین، نهادهای نظارتی مستقل به منظور بررسی و ارزیابی سامانه‌های هوش مصنوعی از منظر اخلاقی، به گونه‌ای طراحی شود که تضمین کنند این سامانه‌ها با اصول عدالت و بی‌طرفی هماهنگ‌اند. درنهایت، تشویق به طراحی اخلاق‌محور در فرایند توسعه سامانه‌های هوش مصنوعی باید اجباری شود تا مسائل اخلاقی از ابتدا و در مراحل اولیه طراحی رعایت شوند.
- به منظور پیشبرد مطالعات در حوزه اخلاق هوش مصنوعی و تطبیق آن با حکمت عملی صدرايي بهتر است پژوهش‌های آتی بر جنبه‌های مختلفی متمرکز شوند. یکی از مسیرهای پیشنهادی انجام مطالعات تطبیقی بین اصول اخلاقی سنتی، مانند حکمت عملی صدرايي و اصول اخلاقی مدرن در هوش مصنوعی است تا نقاط اشتراک و تفاوت‌ها شناسایی و راهکارهای همسوسازی ارائه شود. همچنین، باید پژوهش‌هایی انجام شود که تأثیرات اجتماعی و فرهنگی استفاده از سامانه‌های هوش مصنوعی را در جامعه بررسی و نحوه تأثیرگذاری این سامانه‌ها را در ارزش‌ها و رفتارهای انسانی تحلیل کنند، به گونه‌ای که به درک بهتری از پیامدهای اجتماعی این فناوری‌ها منجر شوند. افزون‌براین، توسعه معیارهای مشخص و قابل اندازه‌گیری برای ارزیابی اخلاقی سامانه‌های هوش مصنوعی در مراحل مختلف طراحی و پیاده‌سازی نیز یکی دیگر از حوزه‌های مهم پژوهشی است.

- at Google". *National Cyberspace Center*. {In Persian}
- Khodaghalizadeh, B. (2020). "Translation of the 2018 Report of the AI Now Institute". *National Cyberspace Center*. {In Persian}
- Kolayni, M. (2028). *AI-Kafi*. Tehran: Dar al-Kutub Al-Islamiyyah.
- Kostka, G. (2019). "China's social Credit Systems and Public Opinion: Explaining High Levels of Approval". *Sage*.
- Kuziemski, M., and Misuraca, G. (2020). "AI Governance in the Public Sector: Three Tales from the Frontiers of Automated Decision-Making in Democratic Settings". *Elsevier - Telecommunications Policy*.
- Lakzai, N. (2004). "Ethical Politics in the Thought of Mulla Sadra". *Political Science Studies*, 1(1), pp. 45-67. {In Persian}
- Li, X., and Zhang, T. (2017). "An exploration on Artificial Intelligence Application: From Security, Privacy and Ethic Perspective". *IEEE*.
- Metvadi, H. (2018). "Conceptualizing Human Privacy in the Quran and Hadith". *Quranic Sciences and Teachings*, pp. 9-23. {In Persian}
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., and Floridi, L. (2016). "The Ethics of Algorithms: Mapping the Debate". *Big Data and Society*, 3(2).
- Nasr, S. H. (1997). *Three Muslim Sages: Avicenna, Suhrawardi, Ibn Arabi*. Oneworld Publications.
- Noble, S. U. (2018). "Algorithms of Oppression: How Search Engines Reinforce Racism". *NYU Press*.
- OECD. (2021). "OECD Recommendation on AI". [Online]. Available: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- O'Neil, C. (2016). "Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy". *Journal cover image*.
- Richardson, R. (2021). "Defining and Demystifying Automated Decision Systems". *SSRN*.
- Russell, S., and Norvig, P. (2020). "Artificial Intelligence: A Modern Approach". *Pearson*.
- Shirazi, S. M. (1972). *Al-Mabda' wa Al-Ma'ad*. Tehran: Iranian Institute of Wisdom and Philosophy.
- Shirazi, S. M. (1981a). *Al-Hikmah al-Muta'Aliyah Fi Al-Asfar al-'Aqliyyah al-Arba'ah*. Beirut: Dar Ihya al-Turath.
- Shirazi, S. M. (1981b). *Al-Shawahid al-Rububiyyah fi Al-Manahij Al-Sulukiyyah*. Mashhad: Al-Markaz Al-Jami'i li Al-Nashr.
- Shirazi, S. M. (1982). *Tafseer Al-Quran Al-Kareem*. Qom: Bidar.
- Shirazi, S. M. (1984). *Mafatih Al-Ghayb*. Tehran: Institute for Cultural Research.
- Shirazi, S. M. (1999). *Al-Mazahir Al-Ilahiyyah Fi Asrar Al-'Ulum Al-Kamaliyyah*. Tehran: Sadra Islamic Philosophy Foundation.
- Shirazi, S. M. (2004). *Sharh Usul Al-Kafi*. Tehran: Institute for Cultural Studies and Research.
- The Holy Quran.
- Tusi, K. R. (2014). *Introduction to the Political Philosophy of Mulla Sadra*. Qom: Research Institute of Islamic Sciences and Culture. {In Persian}
- Winfield, A. F., and Jirotko, M. (2018). "Ethical Governance is Essential to Building Trust in Robotics and AI Systems". *Philosophical Transactions of the Royal Society A*, 376(2133).



Ethical Review of the Performance of Two Artificial Intelligence Systems Based on Sadrai's ractical Wisdom

Mohammad Sajjad Zare Mirakabad¹

Abstract

The present study aims to investigate the ethical challenges associated with two artificial intelligence systems: facial recognition and automated decision-making. Furthermore, the content analysis method was employed to examine the operational mechanisms of these systems and their consequent violation of ethical principles. The findings indicate that the violated ethical principles include fairness, freedom and autonomy, privacy and confidentiality, responsibility and accountability, and acting upon valid conjecture (zann-e motabar). Moreover, these principles were analyzed within the framework of Sadraean practical wisdom, yielding the following conclusions: 1) In Sadraean philosophy, fairness is an essential attribute of the ruler, and the observance of impartiality is mandatory; 2) Freedom and autonomy are fundamental to human perfection and felicity; 3) Privacy is of paramount importance, and intrusion into it is reprehensible; 4) Responsibility and accountability are necessary characteristics of leadership; and 5) Acting upon invalid conjecture is morally reprehensible. Consequently, based on Sadraean practical wisdom, the deployment of these two systems in their current form is ethically unacceptable, unless their design or implementation is modified to address the aforementioned ethical concerns.

Keywords: AI Ethics, Sadraean Practical Wisdom, Fairness, Freedom, Privacy, Accountability, Acting upon Valid Conjecture

1. Ph.D. Candidate in Islamic Philosophy and Theology, Allameh Tabataba'i University, Tehran, Iran; sajad.zare93@gmail.com

نقش نامه و فرم تعارض منافع

الف) نقش نامه

پدیدآورندگان	محمد سجاد زارع میرک آباد
نقش	نویسنده
نگارش متن	همه
ویرایش متن و ...	همه
طراحی / مفهوم پردازی	همه
گردآوری داده	همه
تحلیل / تفسیر داده	همه

ب) اعلام تعارض منافع

عضویت سازمانی، افتخاری یا غیررسمی، اشتغال، مالکیت سهام، و دریافت حق اختراع، و البته محدود به این موارد نیست. منظور از رابطه و انتفاع غیرمالی عبارت است از روابط شخصی، خانوادگی یا حرفه‌ای، اندیشه‌ای یا باورمندانه، و غیره.

چنانچه هریک از نویسندگان تعارض منافی داشته باشد (و یا نداشته باشد) در فرم زیر تصریح و اعلام خواهد کرد:

مثال: نویسنده الف هیچ‌گونه تعارض منافی ندارد. نویسنده ب از شرکت فلان که موضوع تحقیق بوده است گزند دریافت کرده است. نویسندگان ج و د در سازمان فلان که موضوع تحقیق بوده است سخنرانی افتخاری داشته‌اند و در شرکت فلان که موضوع تحقیق بوده است سهام دارند.

در جریان انتشار مقالات علمی تعارض منافع به این معنی است که نویسنده یا نویسندگان، داوران و یا حتی سردبیران مجلات دارای ارتباطات شخصی و یا اقتصادی می‌باشند که ممکن است به طور نا عادلانه‌ای بر تصمیم‌گیری آن‌ها در چاپ یک مقاله تأثیرگذار باشد. تعارض منافع به خودی خود مشکلی ندارد بلکه عدم اظهار آن است که مسئله‌ساز می‌شود.

بدین وسیله نویسندگان اعلام می‌کنند که رابطه مالی یا غیرمالی با سازمان، نهاد یا اشخاصی که موضوع یا مفاد این تحقیق هستند ندارند، اعم از رابطه و انتساب رسمی یا غیررسمی. منظور از رابطه و انتفاع مالی از جمله عبارت است از دریافت پژوهانه، گزند آموزشی، ایراد سخنرانی،

اظهار (عدم) تعارض منافع: با سلام و احترام؛ به استحضار می‌رساند نویسندگان مقاله هیچ‌گونه تعارض منافی ندارد.

نویسنده مسئول: محمد سجاد زارع میرک آباد

تاریخ: ۱۴۰۴/۰۹/۰۱