

Artificial Intelligence, Virtue Ethics, and Human Flourishing

Bahman Zahedi

Assistant Professor, Department of Religions and Islamic Mysticism and Political science, Faculty of Literature and Humanities, University of Mohaghegh Ardabili, Iran, (Corresponding Author): zahedi@uma.ac.ir

Rogheh Nazemfar

Roghayeh Nazemfar, Ph.D of econmic, Department of Social Sciences, University of Mohaghegh Ardabili, Iran. (r.nazemfar@uma.ac.ir)

۷۳

پژوهشنامه ادیان

Abstract:

In the networked world of the 21st century, artificial intelligence (AI) has increasingly penetrated many aspects of human existence and biology. Such emerging technologies have caused complex and unpredictable socio-technical changes. These developments have created practical uncertainties and opaque conditions, and have confronted humans with many ethical challenges. To address ethical concerns, a flexible ethical approach is needed that can stretch across disciplinary, national, cultural, and other boundaries, so that it can be applied to emerging technologies, including artificial intelligence, by covering global gaps and guiding human performance. Understanding the importance of the subject, in this paper, the goals of artificial intelligence and philosophical perspectives in technological ethics are explained. In addition, a comparison of virtue ethics with principled ethics and how to live well with this approach, as well as the ethical challenges of social robots and the ethics of these robots, have been examined. The results of the research suggest that in this fluid ethical landscape, what is needed is an ethics that relies on the presence of flexible, insightful, and practically wise human agents. Virtue ethics can illuminate the ethics of emerging technologies and guide the good life in the age of artificial intelligence.

Keywords: Artificial Intelligence, Virtue Ethics, Human Flourishing.

هوش مصنوعی، اخلاق فضیلت و شکوفایی انسان^۱

رقیه ناظم فر^۳

بهمن زاهدی^۲

چکیده

در جهان شبکه‌ای قرن بیست و یکم، هوش مصنوعی (AI) در بسیاری از ابعاد وجودی و زیستی انسان نفوذ فزاینده‌ای پیدا کرده است. چنین فن‌آوری‌های نوظهوری موجب تغییرات فنی-اجتماعی پیچیده و غیرقابل پیش‌بینی شده است. این تحولات، عدم قطعیت‌های عملی و شرایط ناشفافی را به وجود آورده و انسان را با چالش‌های اخلاقی زیادی مواجه ساخته است. برای رفع نگرانی‌های اخلاقی به رویکرد اخلاقی منعطفی نیاز است که بتواند در سراسر مرزهای انضباطی، ملی، فرهنگی و سایر مرزها کشش پیدا کند، طوری که با پوشش‌دهی شکاف‌های جهانی برای فن‌آوری نوظهوری از جمله هوش مصنوعی اعمال شود و عملکرد انسان را هدایت کند. با درک اهمیت موضوع، در نوشتار حاضر، اهداف هوش مصنوعی و دیدگاه‌های فلسفی در اخلاق فن‌آوری تبیین شده‌اند. علاوه بر این، مقایسه‌ی اخلاق فضیلت با اخلاق‌های اصولگرا و نحوه زندگی خوب با این رویکرد، همچنین، چالش‌های اخلاقی ربات‌های اجتماعی و اخلاق این ربات‌ها مورد بررسی قرار گرفته است. نتایج پژوهش حاکی از آن است که در این چشم انداز اخلاقی سیال، آنچه که نیاز هست، اخلاقی است که بر حضور عوامل انسانی انعطاف‌پذیر، فهیم و عملاً خردمند متکی باشد. اخلاق فضیلت می‌تواند اخلاق فن‌آوری‌های نوظهور را روشن کند و راهنمای زندگی خوب در عصر هوش مصنوعی باشد. در رباتیک نیز اخلاقی که بتواند

۷۴
پژوهشنامه ادیان

سال ۲۰، شماره ۳۸، تابستان ۱۴۰۴

۱. تاریخ دریافت: ۱۴۰۴/۰۸/۰۶؛ تاریخ پذیرش: ۱۴۰۴/۰۸/۲۷

۲. عضو هیئت علمی گروه ادیان، عرفان اسلامی و علوم سیاسی، دانشگاه محقق اردبیلی، اردبیل، ایران (نویسنده مسئول):

gh.babajani@urd.ac.ir

۳. دانش آموخته دکتری رشته علوم اقتصادی، دانشکده علوم اجتماعی، دانشگاه محقق اردبیلی، اردبیل، ایران

ویژگی‌های مورد نیاز برای قضاوت عاقلانه و خوب در چنین شرایط ناپایداری را پرورش دهد و منجر به تعالی و شکوفایی انسان شود، اخلاق فضیلت است.

کلیدواژه: هوش مصنوعی، اخلاق فضیلت، شکوفایی انسان

مقدمه

برای یونانیان باستان، فلسفه تلاشی برای یافتن راه‌هایی برای «زندگی خوب» بوده است. مکاتب اصلی فلسفی یونان باستان توافق داشتند که جهان هدفی دارد و زندگی خوب فردی که منجر به سعادت می‌شود بر اساس نقشی است که مردم در جامعه ایفا می‌کنند (ارسطو^۱، ۲۰۰۷، ص ۱۰۰). آنها به دنبال پاسخ به سوال «چه چیزی به انسان ارزش اخلاقی می‌دهد» یا «کدام انسان دارای ارزش‌های اخلاقی است» بودند. اما پیگیری اخلاقی زندگی خوب در قرن بیست و یکم فوق‌العاده چالش‌برانگیز و مملو از عدم اطمینان شده است. شرایط نوظهور عدم شفافیت حاد فنی-اجتماعی^۲ از جمله تغییرات سریع تکنولوژیکی، اجتماعی-سیاسی و محیطی، بشر را با چالش‌های اخلاقی جدید روبرو کرده است. پیشرفت در هوش مصنوعی در حال دگرگونی بسیاری از جنبه‌های زندگی انسان از جمله ارزش‌ها می‌باشد. ظهور ربات‌های اجتماعی که توسط هوش مصنوعی هدایت می‌شوند، به نظر می‌رسد که معنای ارزش‌های متعدد و طولانی‌مدت در روابط بین فردی مانند دوستی را تغییر می‌دهد. علاوه بر این، اتکای فزاینده بازیگران و مؤسسات قدرتمند به هوش مصنوعی برای تصمیم‌گیری‌های مؤثر بر نحوه زندگی مردم، ممکن است تأثیر قابل توجهی بر حریم خصوصی داشته باشد و مسائلی را در مورد شفافیت الگوریتمی و کنترل انسانی نیز مطرح کند. و البته این‌ها تنها برخی از چالش‌های اخلاقی معاصر هستند.

از اواخر قرن بیستم، رویکردهای اخلاقی فلسفی سنتی دیگر نتوانستند جوابگوی چالش‌های اخلاقی فراروی انسان باشند. به این دلیل از آن زمان اقبال معاصر به اخلاق فضیلت^۳ و ایده شکوفایی انسان آغاز گشت و تلاش‌هایی شد که اخلاق فضیلت را به زمینه‌های مدرن برگردانند (مک‌این‌تایر^۴،

1. Aristotle
2. technosocial
3. virtue ethics
4. MacIntyre

۲۰۰۷، ص ۲۵۱). شکوفایی اغلب به عنوان یک حالت رفاه و بهزیستی تعریف می شود که شامل احساسات و رفتارهای مثبت است. این حالت با توازن در رفاه عاطفی، روانی و اجتماعی مشخص می شود. شکوفایی انسان ریشه در فلسفه ارسطو دارد، جایی که به عنوان "Eudaimonia" شناخته می شود. "Eudaimonia" اغلب به "خوشبختی" ترجمه می شود، اما معنای دقیق تر آن "زندگی خوب" یا "شکوفایی" است. پژوهش ها در زمینه اخلاق فضیلت و شکوفایی انسان به دلیل نیاز امروز بشر هست به زبان اخلاقی منعطفی که بتواند در سراسر مرزهای انضباطی، ملی، فرهنگی و سایر مرزها کشش پیدا کند. فن آوری های نوظهور و هوش مصنوعی^۱ که جهانی و فراگیر هستند، چالش های اخلاقی جدیدی ایجاد می کنند که با رویکردهای اخلاقی سنتی به راحتی قابل حل نیستند. اخلاق فضیلت با تمرکز بر توسعه شخصیت و فضایل، می تواند به ما کمک کند تا این چالش ها را به شیوه ای متفکرانه و مسئولانه حل کنیم.

هدف مقاله حاضر بررسی این موضوع است که آیا اخلاق فضیلت و ایده شکوفایی انسان می تواند اخلاق فن آوری های نوظهور و هوش مصنوعی را روشن کند؟. به عبارتی دیگر آیا اخلاق فضیلت و ایده شکوفایی انسان می تواند برای هدایت و رویارویی با چالش های اخلاقی جدید در عصر هوش مصنوعی بکار گرفته شود؟

در این راستا، در ساختار مقاله حاضر، ابتدا اهداف هوش مصنوعی تشریح شده اند، سپس دیدگاه های فلسفی در اخلاق فن آوری، اقبال معاصر به اخلاق فضیلت و چرایی اخلاق فضیلت برای هدایت عملکرد انسان در عصر هوش مصنوعی تبیین شده است. در ادامه، مقایسه ای بین اخلاق فضیلت گرا با اخلاق های مبتنی بر قاعده در راستای موضوع مورد بحث صورت گرفته است. همچنین، ربات های اجتماعی و چالش های اخلاقی آن ها و رویکرد اخلاق فضیلت به عنوان رویکردی که بتواند منجر به تعالی و شکوفایی انسان در اخلاق ربات شود، مورد بررسی قرار گرفته است. در نهایت، نحوه زندگی خوب با هوش مصنوعی بر اساس اخلاق فضیلت تشریح شده و نتیجه گیری حاصل از پژوهش نیز در پایان مقاله آمده است.

هوش مصنوعی و اهداف آن

1. Artificial intelligence

خلق یک ذهن مصنوعی که قادر به خودآگاهی و اراده آزاد باشد، یکی از چالش‌برانگیزترین و جاه‌طلبانه‌ترین وظایف علم مدرن است. درک مکانیسم‌های عصبی که باعث ایجاد آگاهی در مغز انسان می‌شوند، راه را برای خلق سیستم‌های مصنوعی باز می‌کند، اما این سیستم‌ها هنوز از تکثیر کامل ذهن فاصله زیادی دارند. در این مسیر، بسیاری از سوالات اخلاقی و حقوقی در مورد حقوق و وظایف افراد مصنوعی، نقش آنها در جامعه و مسئولیت سازندگان مطرح می‌شوند. فن‌آوری‌های فعلی مانند محاسبات نورومورفیک و شبکه‌های عصبی عمیق، راه‌حل‌های امیدوارکننده‌ای ارائه می‌دهند، اما در توانایی خود برای مدل‌سازی تجربه انسانی و خودآگاهی محدود هستند. ادامه تحقیقات در این زمینه، ضمن توسعه استانداردهای اخلاقی و حقوقی برای ادغام هوش خودآگاه در سیستم‌های اجتماعی و حقوقی، مهم است (یاشچنکو، ۲۰۲۴).

هوش مصنوعی یک حوزه چند رشته‌ای است که تلاش می‌کند تا عامل‌های هوشمندی را توسعه دهد که قادر به انجام فعالیت‌هایی باشند که معمولاً به هوش انسانی نیاز دارند. استدلال، حل مسئله، یادگیری، ادراک و درک زبان نمونه‌هایی از این وظایف هستند. هوش مصنوعی با اجازه دادن به ربات‌ها برای تفسیر داده‌ها، تصمیم‌گیری و سازگاری با محیط‌های جدید، قصد دارد توانایی‌های شناختی انسان را تقلید کند. این ذهن مصنوعی طیف گسترده‌ای از کاربردها را دارد، از سیستم‌های ساده مبتنی بر قانون گرفته تا الگوریتم‌های قدرتمند یادگیری عمیق (سها رواس^۲، ۲۰۲۴). با توجه به کاربردهای زیاد هوش مصنوعی، می‌توان از آن برای اهداف بی‌شماری استفاده کرد. برخی از اهداف سازندگان هوش مصنوعی عبارتند از: پیشرفت علمی و فناوری، حل مسائل پیچیده، بهبود کارایی و بهره‌وری، سودآوری اقتصادی، قدرت نظامی و امنیتی. به طور کلی، مطالعات سه دسته از اهداف اصلی را برای هوش مصنوعی در نظر می‌گیرند: هوش مصنوعی برای کارایی، هوش مصنوعی برای کنترل اجتماعی و هوش مصنوعی برای شکوفایی انسان (استال^۳، ۲۰۲۳، ص ۲۳). در اولین هدف، هوش مصنوعی می‌تواند با صرفه‌جویی در هزینه‌ها، کارایی اقتصادی را بهبود بخشد و باعث افزایش منافع اقتصادی و رفاه جامعه شود (گراگلیا و همکاران^۴، ۲۰۱۸، ص ۸۰). همچنین، هوش مصنوعی منجر به توسعه مدل‌های تجاری جدید می‌شود و کسب سودهای هنگفت را تسهیل

1. Yashchenko
2. Soha Rawas
3. Stahl
4. Craglia

می‌کند (زوبف^۱، ۲۰۱۹، ص ۴۹۵). هدف دوم استفاده از هوش مصنوعی، کنترل اجتماعی است. با توانایی‌های رو به رشد سریع برای جمع‌آوری داده‌ها، همراه با توانایی هوش مصنوعی برای تشخیص الگوها و همبستگی‌های بین متغیرها، راه‌های جدید برای کنترل رفتار انسان فراهم شده است. با استفاده از ایده «تلنگر^۲» بر اساس اقتصاد رفتاری می‌توان این را انجام داد (کامرر^۳ و همکاران، ۲۰۰۴، ص ۳۹۳) یا آن را با شدت بیشتری در سیستم امتیازدهی اعتبار اجتماعی می‌توان دید (کامرز^۴، ۲۰۱۸، ص ۱۹، لیو^۵، ۲۰۱۹، ص ۲۸). هوش مصنوعی با استفاده از داده‌های رسانه‌های اجتماعی برای تأثیرگذاری بر نتیجه انتخابات مردم سالارانه در زمینه‌ی کنترل اجتماعی می‌تواند محدودیت‌های قانونی را نیز نقض کند (ایزاک و هانا^۶، ۲۰۱۸، ص ۵۸). بر اساس مطالعات انجام شده، کنترل اجتماعی یک نیروی محرکه و شرط لازم برای موفقیت «سرمایه‌داری نظارتی^۷» است. هدف سوم استفاده از هوش مصنوعی، شکوفایی انسان است. در توصیف هوش مصنوعی، ویرجینیا دیگنوم^۸ (۲۰۱۹، ص ۲۱۹) بیان می‌کند: «هوش مصنوعی مسئولیت‌پذیر در مورد مسئولیت انسان برای توسعه سیستم‌های هوشمند در راستای اصول و ارزش‌های بنیادین انسانی است تا شکوفایی و رفاه انسان در یک دنیای پایدار تضمین شود». در واقع هوش مصنوعی به گونه‌ای توسعه یافته و به کار گرفته شده است که شکوفایی انسان را ارتقا دهد. از آن می‌توان به عنوان ابزاری استفاده کرد که به افراد و گروه‌ها کمک می‌کند تا اهداف ارزشمند را شناسایی کنند و از آنها در تعالی برای دستیابی به این اهداف حمایت کنند. پیشنهاداتی برای اطمینان از اینکه هوش مصنوعی پیامدهای مثبتی برای افراد و جوامع دارد، وجود دارد که بخشی از این هدف سوم استفاده از هوش مصنوعی برای شکوفایی انسان است: به عنوان مثال، تلاش برای ایجاد یک «جامعه هوش مصنوعی خوب»^۹ یا «گفتمانی در زمینه هوش مصنوعی برای خوبی» (کات و همکاران^{۱۰}، ۲۰۱۶، ص ۲۲۳).

1. Zuboff
2. nudging
3. Camerer
4. Creemers
5. Liu
6. Isaak and Hanna
7. Surveillance capitalism
8. Dignum
9. good AI society
10. Cath

این سه هدف ممکن است با هم تضاد داشته باشند، اما لزوماً متناقض نیستند. یک اقتصاد کارآمد با تولید ثروت، راه‌هایی را برای شکوفایی انسان باز می‌کند که در غیر این صورت ناممکن خواهد بود. علاوه بر این، کارآیی و سود می‌تواند زمینه فعالیت مشروع برای تعالی انسان باشد و افراد می‌توانند در این فعالیت شکوفا شوند. کنترل اجتماعی اغلب مشکل‌ساز و در تضاد با آزادی‌های فردی دیده می‌شود. استفاده از فناوری‌های اطلاعات و ارتباطات از دیرباز با نقض حریم خصوصی و رشد نظارت همراه بوده است. امروزه که شرکت‌های بزرگ دنیا بسیاری از داده‌ها و زیرساخت‌های فنی مورد نیاز برای هوش مصنوعی را کنترل می‌کنند، نگرانی‌هایی از جمله استثمار افراد در اشکال جدید «سرمایه‌داری نظارتی» است (ژوبف، ۲۰۱۹، ص ۲۵). اما باز هم لازم نیست بین کنترل اجتماعی و شکوفایی انسان تناقضی وجود داشته باشد. انسان‌ها به عنوان موجودات اجتماعی نیاز به تعریف راه‌های همکاری دارند که شامل توافق بر سر قوانین اخلاقی است و اینها باید به نحوی کنترل و اجرا شوند. همزمان با همه‌گیری کووید ۱۹، بررسی‌هایی در مورد برنامه‌های کاربردی وجود داشت که برای ردیابی مبتلایان و پشتیبانی از ردیابی استفاده شود (کلار و لانزراث، ۲۰۲۰، ص ۵). بنابراین هوش مصنوعی می‌تواند برای ارتقای آگاهی از حریم خصوصی استفاده شود، که مسلماً شرط شکوفایی است (اکویستی، ۲۰۰۹، ص ۸۴). از این رو، حتی کنترل اجتماعی اجباری از طریق فناوری‌های دیجیتال ممکن است در برخی شرایط منجر به شکوفایی انسان شود.

دیدگاه‌های فلسفی در اخلاق فن‌آوری

اخلاق شکوفا بخشی از سنت اخلاق فضیلت و ریشه تاریخی آن در اخلاق ارسطویی است؛ با این حال برای پاسخ به سوال پژوهش حاضر، نگاهی به دیدگاه‌های دیگری که هدف مشترک ارتقای شکوفایی انسان را دارند، مفید است. هر کدام از این‌ها رویکرد نظری مشخصی را ارائه می‌دهند که با شکوفایی انسان سازگار است و هر کدام منجر به درک بیشتر در مورد چگونگی مشاهده و ترویج شکوفایی شده است. در مورد فن‌آوری و زندگی خوب، تأملاتی در افلاطون^۱، ارسطو^۲، سر فرانسیس

-
1. Klar and Lanzerath
 2. Acquisti
 3. Plato

بیکن^۱، جی. جی. روسو^۲، جی. دبلیو. هگل^۳، کارل مارکس^۴، جان دیویی^۵ و خوزه اورتگا ی گاست^۶ و دیگر متفکران یافت می‌شود، اما فلسفه فن‌آوری پس از جنگ جهانی دوم با نوشته‌های متفکرانی چون مارتین هایدگر^۷، ژاک ایلول^۸، هربرت مارکوزه^۹ و لوئیس مامفورد^{۱۰} مورد توجه قرار گرفت (هانکس^{۱۱}، ۲۰۱۰، ص ۲ و ۳). آثار این گروه به‌عنوان فلسفه‌های بدبینانه و حتی سرنوشت‌گرا توصیف می‌شوند که در آن فن‌آوری نیرویی یکپارچه است که همواره آزادی انسان را تضعیف می‌کند و ما را به زندگی‌های استثمار، مصرف‌گرا و با بهره‌وری بی‌معنی محکوم می‌کند (وربیک^{۱۲}، ۲۰۰۵، ص ۱۷۲). اگر این «جبرگرایی فن‌آورانه» را بپذیریم، مطالعه در اخلاق که آزادی انسان برای انتخاب یک زندگی خوب پیش فرض آن است، بیهوده خواهد بود.

با این حال، در پایان قرن بیستم، موج دومی از فیلسوفان، در حالی که هنوز به شدت تحت تأثیر موج اول بود، دلیل کافی برای امید به اخلاق قوی فن‌آوری پیدا کردند. مهمترین آنها هانس جونا^{۱۳} است که در دهه‌های ۱۹۷۰ و ۱۹۸۰ اولین فراخوان صریح را برای یک اخلاق فنی-اجتماعی جدید مطرح کرد. جونا با اینکه همچنان تکنولوژی را به عنوان یک «غول» به تصویر می‌کشد که پیشرفت آن «نابودی انسان»^{۱۴} را به دنبال دارد، ادعا کرد که ما برای بقا، به اخلاق جدید فن‌آوری نیاز داریم تا با چالش‌های عمیقی که این فن‌آوری‌های نوظهور برای تصویر انسان و تصورات ما از زندگی خوب ایجاد می‌کنند، مقابله کنیم (جونا، ۱۹۷۹، ص ۲۸). با این حال، او یادآوری می‌کند که «زمانی که به کمترین آن باور داشته باشیم، بیشتر به خرد نیاز داریم» - در عصری که پایه‌های همه نظریه‌های اخلاقی به چالش کشیده شده‌اند، زمانی که نسیت‌گرایی اخلاقی برای بسیاری یک

-
1. Sir Francis Bacon
 2. J.J. Rousseau
 3. G.W. Hegel
 4. Karl Marx
 5. John Dewey
 6. José Ortega y Gasset
 7. Martin Heidegger
 8. Jacques Ellul
 9. Herbert Marcuse
 10. Lewis Mumford
 11. Hanks
 12. Verbeek
 13. Hans Jonas
 14. obsolescence of man

موضع پیش فرض است، از کجا باید جستجوی خرد فنی - اجتماعی را آغاز کرد که شدیداً به آن نیاز است؟ به همین دلیل جوناس به طور مشخص اظهار می کند که فلسفه «متأسفانه برای این، یعنی اولین وظیفه کیهانی خود، آماده نیست» (جوناس، ۱۹۷۹، ص ۲۸). جوناس دریافت که سیستم ها و اصول اخلاقی سنتی کمک چندانی نمی کنند، در درجه اول به این دلیل که چنین سیستم هایی از نظر تاریخی برای حل معضلات اخلاقی که یک فرد با آنها مواجه می شود طراحی شده اند، جایی که پیامدهای اخلاقی مربوط به عمل او در یک زمان قابل پیش بینی آشکار می شود که برای سایر ذینفعان اخلاقی مشترک است (جوناس، ۱۹۸۴، ص ۹۰). اما فن آوری های نوظهور، مشکلات جدیدی را برای کنش اخلاقی جمعی ایجاد می کنند، احتمالاً افراد، گروه ها و سیستم های آینده را بیشتر یا بیشتر از ذینفعان فعلی تحت تأثیر قرار می دهند و پیامدهای غیر قابل پیش بینی دارند که در افق زمانی نامعلوم آشکار می شوند. شادی کلی برای فایده گرا غیر قابل محاسبه می شود، و ضرورت کانتی نمی تواند خطرات اخلاقی را که فراتر از «سازگاری انتزاعی»^۱ انتخاب خصوصی من با اراده عقلانی من قرار دارد، به دست آورد (جوناس، ۱۹۸۴، ص ۱۲). جوناس ادعا می کند که دیدگاه سنتی ارسطویی از فضایی مانند خرد و فروتنی، از آنچه ما نیاز داریم ناقص است. این به دو دلیل است؛ اول، به این دلیل که تصور ارسطو از آنچه برای زندگی خوب است، مبتنی بر «تغییرناپذیری های مفروض طبیعت و شرایط انسان» (مانند زیست شناسی اولیه انسان) است که به لطف فن آوری های جدید و هوش مصنوعی، دیگر تغییرناپذیر نیستند. دوم، ارسطو معتقد است که موضوع انتخاب هر آن چیزی است که تحت اراده و توانایی ما قرار بگیرد و ما توانایی دخل و تصرف در آن را داشته باشیم (صمدیه و ملایوسفی، ۱۳۹۶، ص ۱۶)، اما طبق نظر جوناس، انتخاب های صورت گرفته توسط انسان های قرن بیست و یکم بر شکوفایی کل گونه ها (و نه تنها گونه ما)، تا نسل های آینده زندگی در این سیاره تأثیر خواهد گذاشت (جوناس، ۱۹۸۴، ص ۶).

متأسفانه، پیشنهاد سازنده ی هانس جوناس برای شکل گیری اخلاقیاتی نو، همچنان با لحنی بدبینانه و با نوعی ذات گرایی متافیزیکی پیوند خورده است که یادآور نگرش فیلسوفان موج نخست فناوری است. ذات گرایی در این زمینه به معنای واقع گرایی مابعدالطبیعی است و خود سه جریان عمده را دربر می گیرد: ذات گرایی قدیم، ضد ذات گرایی، و ذات گرایی جدید. در ذات گرایی قدیم، اشیا

دارای ذوات و خواص ذاتی جوهری تلقی می‌شوند؛ اجسام در عین انفعال، و نفس و عقل در عین فعال بودن، در نهایت اموری ایستا و ناپویا هستند. از این منظر، علم به اشیا همان علم به خواص ذاتی آنها از راه برهان است و معرفتی مطابق با واقع پنداشته می‌شود. در برابر آن، ضد ذات‌گرایی وجود ذات یا خواص ذاتی را برای اشیا منکر است. در این دیدگاه، شناخت اشیا براساس تعریف، قرارداد و بهترین تبیین ممکن از آن‌ها شکل می‌گیرد. اما ذات‌گرایی معاصر گذار از این دو قطب است؛ این نگرش اشیا را دارای خواص ذاتی جوهری، رویدادی، فرایندی، رابطه‌ای و پویا می‌داند. انسان نیز همچون دیگر موجودات، برخوردار از این خواص ذاتی است و همین ویژگی‌ها، مبنای اعتبار ذوات انسانی در چارچوب اعتبارات عام انسان‌ها محسوب می‌شوند. در این نگاه، علم به اشیا، بهترین تبیین از آنها است و تقریباً صادق و مطابق با واقع است (دادجو، ۱۳۹۹). هانس جونا س مانند هایدگر و رهروان او، فن‌آوری را به عنوان تهدیدی یکپارچه برای «جوهر انسان»^۱ در نظر گرفت. از این رو، برای مقابله با سناریوهای عمدتاً دیستوپیایی، آنچه جونا س «هیورستیک ترس»^۲ می‌نامد، پیشنهاد جونا س این است که به یک اخلاق فنی-اجتماعی جدید فکر کنیم. اما کسانی که فن‌آوری‌های نوظهور را فرصت‌های سازنده خوبی برای بشریت می‌دانند، و فکر می‌کنند که ترس آشکار از فن‌آوری بیش از آنکه حل کند، مشکلاتی را به وجود می‌آورد، پذیرای دیدگاه جونا س نیستند (جونا س، ۱۹۸۴، ص ۲۰۳). برخلاف جونا س، بورگمان، از فیلسوفان موج دوم فن‌آوری، در مورد فن‌آوری مستقیماً به مشکل یافتن یک ساختار نظری مناسب برای یک اخلاق فنی-اجتماعی نمی‌پردازد. تأملات بورگمان در مورد تکنولوژی مدرن و زندگی خوب بر اساس اندیشه ارسطویی است که علاقه گسترده‌تر به فضیلت و خرد عملی را با توجه به آزادی انسان در مواجهه با جبر فن‌آورانه ادغام می‌کند. بورگمان «جستجوی تعالی»^۳ را به عنوان پرورش ظرفیت‌ها یا فضایل اجتماعی، فیزیکی، و فکری، معیاری برای ارتباط ما با زندگی خوب در نظر می‌گیرد (بورگمان، ۱۹۸۴، ص ۱۰۲). با این حال، دیدگاه بورگمان در مورد رابطه بین اخلاق انسانی و فن‌آوری‌های نوظهور بیشتر بدبینانه است. بنا به نظر او، هدف غالب زندگی فن‌آورانه مدرن این است که پرورش

-
1. essence of Man
 2. heuristics of fear
 3. pursuit of excellence
 4. Borgmann

تعالی اخلاقی را به نفع یک زندگی مصرفی بدون فکر - چیزی که او «پارادایم دستگاه»^۱ می‌نامد - دگرگون کند. به نظر بورگمان نبایست فن آوری را کنار گذاشت. پیشنهاد او این است که بدانیم هدف نهایی برای انسان‌ها یک زندگی مشارکتی است، که بسیاری از کارها، علاوه بر پرورش برتری‌های شخصی، آن را تسهیل می‌کنند. بورگمان از ما می‌خواهد که فن آوری‌های خود را به گونه‌ای پیکربندی مجدد کنیم که به جای اینکه مانع زندگی پرشور شوند، آن را تسهیل کنند (بورگمان، ۱۹۸۴، فصل ۱۸). در حالی که مفهوم کلی «زندگی خوب» در دیدگاه بورگمان امیدوارکننده به نظر می‌رسد، اما روایت او مبهم و ناقص است. چگونه تعیین کنیم که چه چیزی به عنوان تسهیل‌کننده تعامل و چه چیزی به عنوان براندازی آن محسوب می‌شود؟ کدام برتری‌های خاص ارزش توسعه بیشتری دارند؟ مهم‌تر از آن برای اهداف ما، چگونه تمرکز محدود بر اصلاحات می‌تواند به ما در مورد فن آوری‌های نوظهوری که تأثیر آنها بر رویه‌های ما هنوز آشکار نشده است کمک کند؟ علاوه بر این، پیشنهاد بورگمان برای اصلاح، در حل مشکلات کنش اخلاقی جمعی مربوط به فن آوری‌های نوظهور کامل نیست. او نیاز به گفتمان عمومی در مورد زندگی خوب را به عنوان ابزاری برای مذاکره سیاسی و اعمال اصلاحات فنی اجتماعی معنادار مطرح می‌کند (بورگمان، ۱۹۸۴، فصل ۲۵). اما از نظر او، هدف اصلی چنین گفتمانی حفظ فرصت‌ها برای پیگیری زندگی خصوصی متعهدانه، شکوفایی جمعی و بلندمدت خانواده جهانی بشر است. در حالی که او متذکر می‌شود اصلاحات فنی - اجتماعی مستلزم تأیید «مشترک و عمومی» انواع خاصی از رفتار مدنی است که برای تسهیل گفتمان عمومی لازم است، محتوای چنین تأییداتی مبهم است و فقط به طور خلاصه به ارزش انتظارات اجتماعی «ادب، معاشرت یا مدنیت» اشاره می‌کند (بورگمان، ۱۹۸۴، ص ۲۳۳). از این رو به شرح کاملتر و صریح‌تری از فضایل خاص نیاز است که گفتمان عمومی بهبود یافته در مورد فن آوری و زندگی خوب را تسهیل می‌کند. در حالت ایده‌آل، همچنین باید ایده‌ای در مورد چگونگی تشویق پرورش جهانی این فضیلت‌ها به عنوان پیش‌شرطی برای ظهور خرد جمعی و نه فقط شخصی و خصوصی فنی - اجتماعی شکل بگیرد.

در حالی که تأثیرگذارترین فیلسوفان فن آوری قرن بیستم، هر کدام مفهوم‌پردازی منحصر به فردی برای مسائل اخلاقی فن آوری داشتند، همه آنها تمایل داشتند که اخلاق فن آوری را به عنوان پاسخی

به یک مشکل یا پدیده منحصر به فرد توصیف کنند، آیا به عنوان «قالب‌بندی» انسانی از واقعیت به عنوان منبعی که باید دستکاری شود (مارتین هایدگر) تصور شود، اثربخشی فن آوری خودمختار (ژاک ال^۱)؛ گسترش «انسان تک-ساحتی» اسیر منافع فنی - سیاسی (هربرت مارکوزه)؛ چرخه فن آورانه و بی‌امان نیازها و خواسته‌های ایجاد شده (جوناس)؛ یا قاعده خفقان «پارادایم دستگاه» بورگمان؟. همچنین بایستی توجه داشت که هر یک از آنها، انسان‌ها را به عنوان سوژه‌های تا حدی منفعل «مشکل» فن آورانه‌ی منحصر به فرد معرفی می‌کنند، که باید به نحوی آزادی خود را در رابطه با فن آوری بازیابند. در مقابل، اوایل قرن بیست و یکم یک تغییر اساسی در نحوه برخورد فیلسوفان به اخلاق فن آوری رخ داد. نسل جدیدی از متفکران آگاهانه از مفاهیم ذات‌گرایانه فن آوری در مقابل انسانیت فاصله گرفته و به تحلیل تک‌فن‌آوری‌ها به عنوان ویژگی‌های زمینه‌های خاص انسانی رفته‌اند. به دور از توصیف جهانی «مشکل فن آوری» به توصیف‌های خنثی‌تر و محلی‌تر از انبوهی از مسائل اخلاقی که توسط فن‌آوری‌های خاص و کاربردهای آن‌ها مطرح می‌شوند؛ و در نهایت، به دور از نگرانی‌های متافیزیکی در مورد آزادی و جوهر انسان به سمت گزارش‌های تجربی مبتنی بر روابط انسان و فن‌آوری روی آوردند. چرخش تجربی اول در دهه‌های ۱۹۸۰ و ۱۹۹۰ شکل گرفت، هنگامی که فلاسفه بیشتری که درون سنت کلاسیک فعالیت میکردند، از برخی از پیش فرض‌ها و روش‌های آن فاصله گرفتند. نوهایدگری‌ها، نظریه‌پردازان انتقادی نو و پس‌پدیدارشناسان شروع به تمرکز کردن بر تکنولوژی‌ها و موضوعات انضمامی کردند، سعی داشتند نظریه‌هایی سیاق‌محور و کمتر تعین‌گرایانه درباره تکنولوژی شکل دهند یا آنکه این نظریه‌ها را از مطالعات علم و تکنولوژی اخذ کردند و طرح‌نگرش‌هایی کمتر ویران‌شهرگرایانه، عمل‌گرایانه‌تر و متعادل‌تر درباره تکنولوژی مدرن را آغاز کردند. برای مثال، اندرو فینبرگ نظریه‌ای را درباره تکنولوژی بسط داد که درون سنت نظریه انتقادی جای می‌گرفت، اما تلقی‌اش درباره تکنولوژی را از مطالعات علم و تکنولوژی اخذ می‌کرد. این دیدگاه بر ماهیت سیاق‌مند تکنولوژی و امکان توسعه و به کارگیری آن به نحوی متفاوت تأکید داشت. دان‌آید پدیدارشناسی‌ای کمتر ارزشیابانه و بیشتر توصیفی را درباره تکنولوژی بسط داد که چندان به مطالعه تأثیر تکنولوژی بر تجربه انسانی نمی‌پرداخت، بلکه به این موضوع توجه داشت که تکنولوژی چگونه به انحائی متفاوت میان انسان‌ها و محیطشان وساطت می‌کند.

¹ Jacques Ellul

فیلسوف نوهایدگری، هوبرت دریفوس نیز با رویکردی پدیدارشناسانه و فیلسوفانه، امکان تحقق هوش مصنوعی به صورت موجودی با آگاهی مشابه آگاهی انسانی در آینده را رد کرده و آن را غیرقابل تحقق می‌داند. کرزویل منتقد این دیدگاه دریفوس است و باور دارد که نباید بیش از اندازه بر ویژگی‌های انسانی و تجربه‌های زیسته انسانی تاکید کرد زیرا هوش مصنوعی می‌تواند به روشی نوآورانه عمل کند. و این کار فقط باعث محدود شدن چشم‌اندازها نسبت به هوش مصنوعی خواهد شد. ری کرزویل یکی از آینده پژوهانی است که پیش‌بینی می‌کند در آینده‌ای نه چندان دور، توانایی هوش مصنوعی از هوش انسانی پیشی خواهد گرفت (رضایی اقچری و رعایت جهرمی، ۱۴۰۳). چرخش تجربی برای احیای مجدد فلسفه فن‌آوری و گسترش دامنه تحقیق در مفاهیم اخلاقی فن‌آوری کمک زیادی کرده است. اکنون مطالعات در زمینه‌هایی مانند اخلاق و فن‌آوری اطلاعات از انواع رویکردهای تجربی و فلسفی برای کشف طیف گسترده‌ای از پدیده‌های فنی-اجتماعی بومی شده استفاده می‌کنند، بدون اینکه به یک فراروایت در مورد فن‌آوری و ارتباط آن با زندگی خوب پایبند باشند (وربیک، ۲۰۰۵، ص ۵۰).

اقبال معاصر به اخلاق فضیلت

اقبال معاصر به اخلاق فضیلت به این معنا نیست که باید ثمرات «چرخش تجربی»^۱ در فلسفه فن‌آوری را کنار گذاشت. در واقع، چرخش تجربی می‌تواند با گشودن راه‌های بیشتری برای جمع‌آوری اطلاعات زمینه‌ای حیاتی درباره واقعیت‌های جدید و نوظهور فن‌آوری اجتماعی، خرد عملی را تغذیه کند. به لطف رویکرد انتقادی چرخش تجربی، مطالعه اخلاق و فن‌آوری سرانجام طیفی از دیدگاه‌ها و تعهدات نظری را توسعه داده است که به اندازه کافی متنوع هستند تا یک حوزه واقعاً فلسفی را تشکیل دهند. با این حال، درست است که ایده نیازمندی به چارچوب اخلاقی جدید در زندگی فن‌آورانه معاصر به طور کامل کنار گذاشته نشده است، اما این مشکل به طور فزاینده‌ای به دلیل تنوع و پویایی آشکار پدیده‌های فن‌آورانه اجتماعی به چالش کشیده می‌شود (فلوریدی^۲، ۲۰۱۳، ص ۳۰۰). این دشواری با واکنش انتقادی طولانی مدت علیه روایت‌های کلی و ذات‌گرایانه، که دعوت به اخلاق جدید فن‌آوری با آنها همراه است، بیشتر تشدید می‌شود. اما، اگر مایل نباشیم با اهمیت

1. empirical turn
2. Floridi

گسترده تر روابط فنی اجتماعی معاصر روبرو شویم، اخلاق فن آوری به فهرستی بی پایان از پدیده های به ظاهر نامرتب کاهش می یابد. یک بار دیگر، این معضل، قابلیت دوام اخلاق فضیلت را به عنوان چارچوبی برای اخلاق جدید فن آوری تقویت می کند. زیرا همانطور که بیان شد، این تنها ساختار نظری است که برای حداکثر انعطاف طراحی شده است تا ویژگی های متنوع و متغیر زندگی اخلاقی را در خود جای دهد. با وجود منتقدان نظریه های اخلاقی سنتی، از جمله اخلاق فضیلت ارسطویی، اما این امر اخلاق فضیلت را به عنوان چارچوب مناسبی رد نمی کند. در عوض، افق اخلاقی جدید ما از کنش جمعی انسانی در مواجهه با کدورت فزاینده فناوری اجتماعی، ما را دعوت می کند تا امکان پرورش جهانی فضایل را که به تازگی با نیاز کنونی ما سازگار شده اند، در نظر بگیریم. اخلاق فضیلتی ارسطویی، کنفوسیوس، و بودایی، هر کدام ویژگی هایی از افراد را بیان می کنند که حکمت اخلاقی را در زمان ها، مناطق یا فرهنگ های خودشان تجسم می دهند. پس آیا نتیجه نمی شود که اخلاق فضیلت معاصر می تواند ویژگی های مشخصه کسانی را توصیف کند که عاقلانه و خوب در محیط شبکه ای جهان قرن بیست و یکم زندگی می کنند؟ مسلماً، این گروه دارای اندازه و تنوعی است که در تاریخ جوامع اخلاقی بی سابقه است. با این حال، این گروهی است که نیاز به اتخاذ تصمیمات جمعی و از نظر اخلاقی عاقلانه ی فنی - اجتماعی دارد، که با افق های بازتر و پیچیده تر از آن چیزی که خردمندان دنیای کلاسیک تصور می کردند هماهنگ باشند.

اخلاق فضیلت، با رویکرد منحصر به فرد و جذاب خود، می تواند چارچوبی برای بررسی طیف گسترده ای از پیامدهای اخلاقی ناشی از فن آوری های نوظهور باشد. این رویکرد، با تمرکز بر پرورش فضایل اخلاقی، پیشنهادهایی سازنده برای بهبود سیستم های فن آورانه و افزایش مشارکت مثبت انسان در آن ها ارائه می دهد. به عبارت دیگر، اخلاق فضیلت نه تنها به ما کمک می کند تا چالش های اخلاقی فن آوری های جدید را درک کنیم، بلکه راهکارهایی عملی برای توسعه و استفاده مسئولانه از این فن آوری ها ارائه می دهد. همچنین به انسان اجازه می دهد از خطرات ذات گرای، تعمیم بیش از حد، و انتزاع که تلاش های موج اول و دوم برای چارچوب بندی اخلاق فن آوری را تضعیف می کنند، اجتناب کند. اخلاق فضیلت برای انطباق با رویارویی های باز و متنوع با فن آوری های خاص که شرایط بشر را در این قرن و قرن های آینده شکل می دهد، ایده آل است. سنت های اخلاقی فضیلت آمیز، خودانگیختگی و انعطاف پذیری خرد عملی را نسبت به پایبندی سخت به قوانین رفتار ثابت برتری می دهد - این یک مزیت بزرگ برای کسانی است که با

چالش‌های اخلاقی پیچیده، بدیع و دائماً در حال تکامل مانند چالش‌های جدید ناشی از استفاده از هوش مصنوعی مواجه می‌شوند. اخلاق فضیلت با افراد نه به عنوان افراد اتمیستی که با انتخاب‌های محدودی روبرو هستند، بلکه به عنوان موجوداتی می‌نگرد که اعمال آن‌ها همیشه توسط یک زمینه اجتماعی خاص از نقش‌های مرتبط، روابط و مسئولیت‌های عینی به دیگران اطلاع داده می‌شود. این رویکرد اجازه می‌دهد تا درک سنتی از راه‌هایی را که انتخاب‌ها و تعهدات اخلاقی را به یکدیگر پیوند می‌دهند و به یکدیگر متصل می‌کنند، گسترش یابد (اسکارف^۱، ۲۰۱۲، ص ۱۵۹).

روابطی که انسان‌ها برای شکوفایی شخصی و جمعی خود به آنها وابسته هستند همچنان در مقیاس و پیچیدگی رشد می‌کنند. در نهایت، در حالی که دیدگاه‌های زندگی خوب و عمل درست در سنت‌های فضیلتی به طور قابل ملاحظه‌ای متفاوت است، ساختار اساسی و ابزارهای خودپروری اخلاقی که آنها به کار می‌برند، شباهت بین فرهنگی و همپوشانی مفهومی زیادی را نشان می‌دهد. بنابراین اخلاق فضیلت فنی - اجتماعی این پتانسیل را دارد که به طور گسترده‌ای در یک شبکه متنوع فرهنگی و جهانی از افراد که با نیاز به خرد اخلاقی جمعی در برخورد با عدم قطعیت‌ها، خطرات، و فرصت‌های ایجاد شده توسط تکنولوژی‌های جدید مواجه هستند، طنین‌انداز شود.

چرا اخلاق فضیلت؟

در طول تاریخ مکاتب متعددی از سوی متفکران و فیلسوفان به وجود آمده است؛ از جمله مکتب اخلاقی کانت و پیامدگرایی اخلاقی. بر اساس اخلاق کانتی، ارزش‌های خاصی به صورت مطلق وجود دارند و تابع فرهنگ و زمان نیستند. بر اساس این فلسفه، عمل اخلاقی و وجدانی از مفهوم وظیفه‌گرایی ناشی می‌شود و اخلاق در مورد این که چه چیزی وظیفه است، سخن می‌گوید. از این رو عمل اخلاقی را انجام وظیفه یا مطابقت با ارزش‌های مطلق تعریف می‌نماید. از سوی دیگر، در خصوص پیامدگرایی اخلاقی می‌توان ابراز نمود که بر اساس فلسفه سودگرایی، که تفکر اقتصادی را به شدت متأثر ساخته است، کاری اخلاقی است اگر و تنها اگر مطابق با اصل مطلوبیت (یعنی مطابق با تولید رفاه بیشتر یا افزایش لذت و شادی یا ممانعت از رنج و درد) باشد. بنابراین برای اظهارنظر در مورد اخلاقی بودن عمل، باید به نتیجه و پیامدهای آن عمل توجه داشت (نعمت‌اللهی و همکاران، ۱۳۹۸). اما در مقایسه با این نظام‌های هنجاری، اخلاق فضیلت مزایایی در مواجهه با

موقعیت‌های نوظهور دارد. یک نظام اخلاقی که بتواند به طور انعطاف‌پذیر با موقعیت‌های جدید برخورد کند، در عصری با فناوری‌های به سرعت در حال تغییر، ضروری است. مزیت رویکرد اخلاق فضیلت این است که حداقل یک مبنا برای قضاوت وجود دارد که یک زندگی خوب با هوش مصنوعی چگونه به نظر می‌رسد. گزارش ارسطو از سعادت (Eudaimonia) تلاش می‌کند تا توضیح دهد که به طور کلی، یک انسان چگونه می‌تواند به خوبی زندگی کند، یعنی زندگی با فعالیت عقلانی مطابق با فضیلت. معنای زندگی خوب با هوش مصنوعی این است که در نظر بگیریم چگونه فناوری‌های هوش مصنوعی می‌توانند توانایی ما برای مشارکت در چنین فعالیت‌هایی را کاهش یا افزایش دهند و بر اساس آن تعیین کنیم که چگونه پیش برویم. به عبارت دیگر، فضیلت‌هایی که برای زندگی با هوش مصنوعی به آن‌ها نیاز داریم از قبل به خوبی شناخته شده‌اند، تنها سوال این است که چگونه به طور مناسب آن‌ها را در موقعیت‌های فعلی به کار ببریم. علاوه بر این، نحوه استفاده صحیح از چنین فناوری‌هایی بستگی به این دارد که یک فرد با توجه به ویژگی‌های برجسته یک موقعیت، چگونه از چنین فناوری‌هایی استفاده کند. بنابراین، اخلاق فضیلت می‌تواند طراحان و مهندسان چنین فناوری‌هایی را راهنمایی کند و به همه اعضای جامعه که با این فناوری‌ها تعامل دارند، نحوه تعامل مناسب با آن‌ها را آگاه کند.

چون هم آینده و هم سهم زیادی از زمان حال، مبهم هستند. اغلب تعیین اینکه چه چیزی در یک تصمیم اخلاقی در خطر است، دشوار یا غیرممکن است. ما قادر به شناخت ذهن و اراده‌ی سایر عوامل نیستیم. درک ما از آینده مبهم است و به اندازه‌ی روشن نیست که بتوانیم برنامه‌های خاصی را برای تضمین نتایج اخلاقی معین اجرا کنیم. نمی‌توانیم بدانیم چه فناوری‌های جدیدی توسعه داده می‌شوند یا به طور گسترده مورد استفاده قرار می‌گیرند. ما قادر به پیش‌بینی عواقب فناوری‌هایی که حتی نمی‌توانیم تصورش را بکنیم، نیستیم. ما نیز مانند والور معتقدیم که این عدم شفافیت، به ویژه در مورد تغییرات تکنولوژیکی نگران‌کننده است. والور، وضعیت کنونی ما را به صورت «ابهام حاد فنی-اجتماعی» توصیف می‌کند. در بخش‌های بزرگی از تاریخ، انسان‌ها این امکان را داشته‌اند که با اطمینان معقول و به درستی فرض کنند که فن‌آوری همچنان مانند گذشته به کار خود ادامه می‌دهد یا دست‌کم به روش‌هایی قابل پیش‌بینی گسترش خواهد یافت. در حالت اول، آینده به اندازه‌ی کافی شبیه حال خواهد بود، به طوری که هر هنجار و قاعده‌ی اخلاقی که داشتیم، برای اطمینان از ادامه‌ی زندگی خوب در مواجهه با تغییرات تکنولوژیک کافی باشد. یا در حالت دوم، دست‌کم می‌توانیم

تأثیرات برخی تغییرات تکنولوژیکی را پیش‌بینی کنیم و زمان کافی برای آماده کردن خود برای هنجارها و اخلاقیات جدید داشته باشیم (والور^۱، ۲۰۱۶).

تغییرات تکنولوژیکی کنونی، سریع و غیرقابل پیش‌بینی هستند: فناوری‌ها تغییر می‌کنند، ابزارهای جدید و غیرمنتظره معرفی می‌شوند و زندگی اخلاقی و اجتماعی ما می‌تواند به سرعت به چیزی ناآشنا تبدیل شود. در مواجهه با این ابهام، نمی‌توان انتظار داشت که یک نظام اخلاقی با قواعد ثابت، به‌ویژه قوانینی که منصفانه ارائه می‌شوند، بتوانند راهنمایی دقیق یا متناسب با شرایط کنونی را برای زندگی خوب به ما ارائه دهند. اگر بر نوعی رویکرد مبتنی بر قاعده اصرار کنیم، ممکن است قوانین موجود چنان پیچیده و تحریف شوند که حتی ظاهری از تصمیم‌گیری اخلاقی خوب را نیز غیرممکن سازند. در مواجهه با این شرایط، باید در نظر بگیریم که آیا اخلاق فضیلت‌گرایانه، که هرگز قصد نداشته اخلاق را به فهرست محدودی از قوانینی که باید رعایت کنیم تقلیل دهد، اما با این وجود بینشی در مورد خوب زیستن ارائه می‌دهد، ما را قادر می‌سازد تا با آینده‌ای مبهم کنار بیاییم.

مقایسه اخلاق فضیلت‌گرا با اخلاق‌های مبتنی بر قاعده

جایگزین‌های اخلاق فضیلت‌گرا یا مبتنی بر قواعد یا مبتنی بر محاسبات اهداف هستند. قواعد، روشی ناکارآمد برای هدایت رفتار در موقعیت‌های بدیع یا غیرقابل پیش‌بینی هستند، زیرا بعید است که هر نوع فرمول‌بندی بتواند به سؤالاتی که از ویژگی‌های آن موقعیت برمی‌آید یعنی مسائل یا پیچیدگی‌های غیرمنتظره پاسخ دهد. مثال خوبی از این موضوع در زمینه ویرایش یا اصلاح ژن و به ویژه ویرایش ژن انسانی در اخلاق زیست‌پزشکی است. با وجود این واقعیت که قوانین کلی در زمینه زیست‌پزشکی وجود دارد (از جمله عدم ضرررسانی)، قوانین خاصی باید برای هر نوآوری جدید ایجاد شود. در واقع، زمانی که جنیفر دودنا^۲، راه انجام ویرایش هدفمند ژن با کریسپر - کس ۹ (CRISPR Cas-9) را کشف کرد، متوجه شد که بایستی مجموعه جدیدی از اجلاس‌ها را برای مدیریت استفاده از بیوتکنولوژی جدید تشکیل دهد. او در کتاب خود، "شکافی در آفرینش"، به جزئیات ملاقات خود با دانشمندان و افراد تجاری می‌پردازد که می‌خواستند از بیوتکنولوژی جدید برای اهداف مشکوک استفاده کنند. این موضوع او را متقاعد کرد که او باید اجلاس‌هایی برای

1. Vallor
2. Jennifer Doudna

ایجاد مجموعه‌ای از قوانین خاص برای کنترل آزمایش و نوآوری با استفاده از کریسپر_کس ۹ (CRISPR Cas-9) تشکیل دهد که فقط پس از مدتی می‌توان به آن‌ها دست یافت (دودنا و استمبرگ^۱، ۲۰۱۷، ۱۸۸-۱۸۴). نشست کریسپر در ژانویه ۲۰۱۵ همراه با اجلاس‌های بعدی، اشکالات رویکرد قانونمند را در زمینه زیست‌پزشکی و همچنین در حاکمیت پیشرفت‌های تکنولوژیکی در هوش مصنوعی نشان داد. اجلاس اولیه‌ای که دودنا^۲ ایده‌های خود را بر اساس آن‌ها پایه‌گذاری کرد- آسیلومار^۳، ۱۹۷۳، ۱۹۷۵ - برای تعیین قوانین مربوط به ویرایش ژن در طول انقلاب دی‌ان‌ای (DNA) نو ترکیب اولیه تشکیل شد. در حالی که کریسپر - کس ۹ بسیار قوی‌تر و دقیق‌تر از ویرایش ژن دی‌ان‌ای نو ترکیب است که در اجلاس آسیلومار مورد بحث قرار گرفت، این تکنولوژی‌ها تا جایی مشابه هستند که هر دو ژن را تغییر می‌دهند. قوانینی که از کنفرانس‌های آسیلومار بیرون آمدند، بسیار مختص دی‌ان‌ای نو ترکیب بودند و مبنای ناکافی برای تنظیم ویرایش ژن به‌طور کلی هستند. با توجه به اینکه جهش‌های رو به جلو در فناوری‌های هوش مصنوعی بسیار گسترده‌تر از دو رویکرد ویرایش ژن ذکر شده در بالا است، بعید به نظر می‌رسد که هر رویکرد مبتنی بر قانون بتواند همه موارد را پوشش دهد، در حالی که ویژگی کافی را ارائه دهد.

در ادبیات اخلاق هوش مصنوعی، اغلب با استناد صریح یا وارد کردن مستقیم اصولی که بوچمپ و چایلدرس در متن اصلی خود، اصول اخلاق زیست‌پزشکی، مطرح کرده‌اند، از رویکرد اصولگر استفاده شده است. مثال خوبی از این موضوع در فلوریدی و کالز^۴ (۲۰۲۱) آمده است که استدلال می‌کند اخلاق هوش مصنوعی می‌تواند از چهار اصل اخلاق زیست‌پزشکی بوچمپ و چایلدرس (احترام به اتونومی افراد، عدم ضرررسانی، سودرسانی و عدالت) استفاده کند، و اینکه اخلاق هوش مصنوعی باید یک اصل پنجم را در رابطه با شفافیت اضافه کند.

فلوریدی و همکاران (۲۰۲۱) شش کد اخلاقی تحقیق و توسعه هوش مصنوعی را ترکیب کردند و فهرستی از ۴۷ اصل فردی تهیه کردند. آن‌ها دریافتند که این ۴۷ اصل در واقع می‌تواند به‌طور موثر حول چهار اصل اخلاق زیست‌پزشکی به علاوه یک اصل پنجم "شفافیت" یا "توضیح پذیری" دسته‌بندی شوند. براساس این دیدگاه، کار سخت اخلاقی، اگرچه شاید کار سخت تجربی

1. Doudna & Sternberg
2. Doudna
3. Asilomar
4. Cows

یا سیاسی نباشد، یا انجام شده است یا چندان مرموز نیست. ما می‌دانیم چه کاری باید انجام دهیم، حتی اگر هنوز دقیقاً در حال کار بر روی چگونگی انجام آن یا تلاش برای تدوین این اصول و قوانین روشن در قوانین یا قوانین رسمی سازمانی هستیم. این از دیدگاه ارسطویی، در بهترین حالت، به شدت گمراه‌کننده است. اگر این قواعد یا اصول واضح، ملموس و قابل اجرا باشند، احتمالاً نسبت به انواع موقعیت‌های بدیع که ناگزیر به وجود می‌آیند، حساس نیستند. این یک نگرانی اصلی برای هر کسی است که به اخلاق هوش مصنوعی و یادگیری ماشین علاقه دارد. به هر حال، زندگی خوب با هوش مصنوعی نیازمند راهنمایی در مورد مسائل اخلاقی مرتبط با فناوری است که به سرعت و به طور غیرقابل پیش‌بینی تغییر می‌کند و اغلب کارهایی را انجام می‌دهد که خارج از درک ما هستند. ما باید انتظار داشته باشیم که با پیامدها و نتایج غیرمنتظره و فرآیندهایی مواجه شویم که به روش‌های ناآشنا عمل می‌کنند. حتی اگر بتوانیم قوانینی را تدوین کنیم و آنها را به تصویب برسانیم یا به طور مؤثر آموزش دهیم، باید نگران مواجهه با موقعیت‌هایی در آینده باشیم که قوانین و اصول در مورد آنها اعمال نمی‌شوند. به طور خلاصه، اصول غالباً به ویژه در شرایط جدید، راهنمایی مشخصی ارائه نمی‌دهند.

برخی از نظام‌های اصولگرا علاوه بر مجموعه‌ای از اصول راهنما، فضیلت‌هایی را نیز مورد استفاده قرار می‌دهند. به طور کلی، این سیستم‌های اصولگرا قواعد (معمولاً استثنایی یا مطلق) را پیش‌زمینه می‌کنند و از فضایل برای پر کردن شکاف‌ها، قضاوت در موارد تعارض، و راهنمایی استفاده می‌کنند (به عنوان مثال اخلاق کانت، همانطور که در متافیزیک اخلاق بیان شده است، از فضایل در ارتباط با امر مطلق استفاده می‌کند، بوچمپ و چایلدرس^۱ علاوه بر اصول خود، فضایل را نیز طلب می‌کند. بوچمپ و چایلدرس (۲۰۱۲) اشاره می‌کنند که فضایل در اصول اخلاقی شان جایگاهی دارند، اگرچه معتقدند که فضایل تنها یکی از چیزهایی هستند که در یک نظام هنجاری نیاز به توجه دارند (بوچمپ و چایلدرس، ۲۰۱۲، ۱۴ - ۳) و یک دیدگاه کاملاً غیر ارسطویی از فضایل دارند (به عنوان مثال، آن‌ها وحدت فضایل را انکار می‌کنند و معتقدند که فضایل می‌تواند افراد را از نظر اخلاقی به بیراهه بکشاند، ادعایی که در اخلاق فضیلت ارسطویی معنا ندارد. قواعد و اصولی که اغلب در اخلاق هوش مصنوعی مورد استناد قرار می‌گیرند، از مفاهیم اخلاقی مورد

مناقشه استفاده می‌کنند. بنابراین، اعمال این قوانین به معنای قضاوت خوب است. به نظر می‌رسد که بسیار شبیه به این است که بگوییم اعمال مناسب این قوانین موضوعه است، که در این صورت قوانین خوبی نیستند که به ما کمک کنند تا با هوش مصنوعی خوب زندگی کنیم، بلکه حکمت عملی است. ممکن است فرد فکر کند که توسل به این قوانین و اصول، تا زمانی که آن‌ها قوانین صحیح یا به اندازه کافی دقیق باشند، به معنای پشتیبانی از اخلاق فضیلت بدون تلاش است. با این حال، ما در خطر این هستیم که اصول خود را به خوبی مشخص کنیم بدون اینکه به عنوان یک راهنما به آن‌ها اشاره کنیم (البته رویکردهایی به اخلاق فضیلت وجود دارد که کم‌تر با اصول‌گرایی خصومت دارند، اگرچه بسیاری از این رویکردها شامل نوعی انحراف از موضع ارسطویی مطرح شده، یا برداشتی بسیار سست از آنچه که یک قانون را تشکیل می‌دهد، هستند. باخرز^۱ (۲۰۱۸) استدلال می‌کند که اصول را می‌توان در یک چشم انداز اخلاقی کلی، فضیلت یا اخلاق به عنوان شیوه‌ای برای بیان آن چشم‌انداز یا کمک به آموزش اخلاقی گنجانده (هر چند که باخورست همچنین اذعان می‌کند که پایداری به اصول باید توسط فرنیس هدایت شود. هایدون^۲ (۲۰۰۹) دیدگاه آر. اس. پیترز را بررسی می‌کند که آموزش اخلاقی، در حالی که در نهایت با هدف توسعه یک شناخت اخلاقی به طور مناسب عقل‌گرا و خدانا‌باورانه است، باید یک نوع فضیلت اومانستی (به جای ارسطویی) را در خود جای دهد که منجر به رفتار در راستای اخلاق عقل‌گرا می‌شود. گزارش معاصر و کاملاً ارسطویی هرستهاوس^۳ (۱۹۹۹) از اخلاق فضیلت شامل آن چیزی است که او آن را قوانین وی^۴ می‌نامد (اگرچه این قوانین شکل "کاری را انجام دهید که فرد با فضیلت انجام می‌دهد" را دارند و بنابراین تنها در سست‌ترین حالت قوانین هستند). خود ارسطو اشاره می‌کند که به نظر می‌رسد دست کم قوانین سخت و سریعی وجود دارد، به این معنا که برخی چیزها به خودی خود بد هستند (مانند قتل و زنا) و هیچ راه پرهیزگاری ممکن برای انجام آن‌ها وجود ندارد. این اقدامات، و بنابراین هر قانونی که آن‌ها را منع کند، بسیار کم و دور از آن است که یک اخلاق کامل را شامل شود. استاندارد خوبی اخلاقی به عنوان عامل فضیلت عمل می‌کند). اخلاق فضیلت ارسطویی رویکردی متضاد با رویکردهای اصولی دارد که فضایل را در بر می‌گیرد. از نظر ارسطو، فضایل، و به ویژه

1. Bakhurst

2. Haydon

3. Hursthouse

4. v-rules

فرونسپس، راهنمای نهایی عمل هستند. قوانین، هنگامی که وجود دارند، ثانویه هستند؛ یعنی قواعد و قوانین، در درجه دوم اهمیت قرار دارند. ارسطو بیشتر در مورد قوانین به عنوان ابزاری برای ایجاد یک جامعه خوب صحبت می‌کند. هدف اصلی قوانین، تشویق و ترویج فضیلت‌ها در جامعه است. قوانین نباید به صورت سفت و سخت اجرا شوند، بلکه باید با توجه به شرایط خاص هر موقعیت، تعدیل و تفسیر شوند. به همین دلیل، اخلاق فضیلت یک سیستم اخلاقی منعطف‌تر و سازگارتر با شرایط جدید و غیرمنتظره است.

علاوه بر این، با توجه به هنجارهای فرهنگی و اجتماعی موجود در تحقیقات هوش مصنوعی، به نظر می‌رسد سیستم‌های مبتنی بر اصول، به‌طور خاص برای اخلاق هوش مصنوعی مناسب نباشند. برای اینکه سیستم‌های اصول‌محور بتوانند رفتارها را محدود کنند، طرف‌های ذینفع باید با قوانین موافقت کنند، فرهنگی داشته باشند که پیروی از قوانین را تشویق کند و سازوکاری برای اجرای آن قوانین وجود داشته باشد. در غیر این صورت، قوانین صرفاً یک پیشنهاد خواهند بود. یک نمونه خاص از این موضوع را می‌توان در تلاش برای وضع برخی قوانین اخلاقی برای توسعه هوش مصنوعی مشاهده کرد. در یک نامه سرگشاده، گروهی از اینفلوئنسرها و توسعه‌دهندگان هوش مصنوعی خواستار توقف ۶ ماهه توسعه ابزارهای هوش مصنوعی در سطح یا قوی‌تر از ChatGPT-4 شدند. این تلاش برای ایجاد یک مهلت قانونی، مشابه مهلت قانونی بود که قبل از اولین کنفرانس آسیلومار در دهه ۱۹۷۰ وضع شد. کنفرانس‌های آسیلومار مجموعه‌ای از قوانین را وضع کردند که امکان ادامه تحقیقات را فراهم می‌کرد (برگ^۱، ۲۰۰۸).

در حوزه پزشکی، به دلیل اهمیت حفظ سلامت و ایمنی بیماران، قوانین و مقررات سختگیرانه‌ای وجود دارد که تمام جنبه‌های تحقیقات و درمان را پوشش می‌دهد. اما این فرهنگ پیروی از قوانین در زمینه‌های فناوری رایانه وجود ندارد. توقف موقت (Moratorium)، دوره‌ی وقفه موقت در انجام یک فعالیت خاص (مانند تحقیقات) است که معمولاً به دلیل نگرانی‌های اخلاقی یا ایمنی اعمال می‌شود. در گذشته، به دلیل نگرانی‌های اخلاقی در مورد دستکاری ژنتیکی، یک دوره توقف موقت در تحقیقات DNA اعمال شد. اما، تلاش برای اعمال یک توقف موقت در تحقیقات هوش مصنوعی (به دلیل نگرانی‌های اخلاقی و ایمنی) موفقیت‌آمیز نبود. این امر ممکن است تا حدی به این دلیل

باشد که بسیاری از افراد فعال در زمینه هوش مصنوعی خود را به عنوان کسانی معرفی می کنند که به دنبال شکستن قوانین، ایجاد نوآوری و ایجاد تحول در صنایع مختلف هستند. از نظر فرهنگی، این بخش از صنعت فن آوری از ذهنیت «سریع پیش برو و خراب کن»^۱ سود برده است، که با پیروی از قوانین مخالف است یا بعید است که آن را تشویق کند. بنابراین، رویکردهای مبتنی بر اصول با دو مانع روبرو هستند: تغییر فرهنگ جامعه فن آوری و ایجاد و اجرای مجموعه ای از قوانین. اخلاق فضیلت ارسطویی، تغییرات فرهنگی را در خود جای داده است. این امر مستلزم آموزش نسل بعدی، به ویژه کسانی که بر آینده فن آوری تأثیرگذار خواهند بود - کسانی که در طراحی، حکمرانی و غیره هستند - به منظور شکل دادن به شخصیت کسانی که تصمیمات اخلاقی می گیرند، است. این رویکرد احتمالاً پایه قوی تری برای تجسم راهی برای اداره و زندگی خوب با هوش مصنوعی در مقایسه با مجموعه ای از قوانین فراهم می کند (بنیامین^۲، ۲۰۱۹، ۱۷-۱۱).

ربات های اجتماعی و چالش های اخلاقی

در دهه گذشته، تلاش های قابل توجهی برای مطالعه استفاده های احتمالی از ربات های اجتماعی صورت گرفته است (جو هال^۳، ۲۰۲۰، ص ۷۷).

اینها ربات هایی هستند که به شکل انسان نما یا حیوان ساخته شده اند و برای تعاملات طبیعی، عاطفی و شبه بین فردی انسان و ربات ساخته شده اند. برای اینکه چنین تعاملاتی شبیه به تعاملات بین فردی مشاهده شده در بین انسان ها باشد، ربات های اجتماعی باید به احساسات انسان مانند مجهز شوند که معمولاً با استفاده از روش های هوش مصنوعی، معروف به هوش هیجانی مصنوعی (AEI) توسعه یافته است. چنین روش هایی معمولاً به ربات های اجتماعی در انجام وظایف زیر اجازه می دهند یا به آنها کمک می کنند. (۱) تشخیص احساسات، (۲) تولید احساسات، و (۳) تقویت احساسات (اسکالر و اسکالر^۴، ۲۰۱۸، ص ۳۹).

در حالی که دو مورد اول از این وظایف مورد مطالعه قرار گرفته و حتی تا حدی در رباتیک اجتماعی اجرا شده است. وظیفه تقویت احساسات تا حدی ناشناخته باقی مانده است. با این حال، باید توجه داشت که حتی ساده ترین کار؛ تشخیص احساسات ممکن است - تا به امروز - نشان دهنده یک

-
1. Move fast and break things
 2. Benjamin
 3. Johal
 4. Schuller and Schuller

چالش باشد، به خصوص زمانی که برای مثال سیگنال‌هایی را روی چهره انسان (حالت، حجم صدا، و غیره) در نظر بگیرند (کراکوسکی^۱، ۲۰۱۸، ص ۱۹). علاوه بر این، با توجه به ماهیت چند وجهی که بسیاری از ارتباطات انسانی را مشخص می‌کند، برای هوش مصنوعی کاملاً حیاتی است که به درستی روانشناسی انسان را تشخیص داده و درک کند. در این زمینه، هوش مصنوعی مشاوره‌ای - نوعی از هوش مصنوعی که به طور گسترده در خدمات کمک اطلاعاتی استفاده می‌شود - به ندرت می‌تواند اطلاعاتی را برای آرام کردن در اختیار فرد قرار دهد اگر علائمی از تحریک یا استرس را نشان دهد. به طور مشابه، بیشتر ربات‌های اجتماعی موجود تا به امروز می‌توانند نقش‌های خاصی مثلاً شامل مراقبت را ایفا کنند. با این حال، همه این نقش‌ها معمولاً در یک سناریوی ساختاریافته و قابل پیش‌بینی انجام می‌شوند که در آن متغیرهای روان‌شناختی در نظر گرفته نمی‌شوند. حتی اگر ربات‌های اجتماعی ممکن است فواید خاصی را ارائه دهند؛ به عنوان مثال، آنها می‌توانند استرس یا سایر احساسات منفی را کاهش دهند یا حتی به ایجاد همدلی در کودکان کمک کنند. آنها ممکن است نتوانند برخی از اثرات مثبت مشابه را در بزرگسالان تکرار کنند، مگر اینکه انسان‌های بالغ چنین ربات‌هایی را به اندازه کافی باهوش برای همدردی و تعامل با آنها در نظر بگیرند (به عبارت دیگر، چنین ربات‌هایی توانایی تشخیص احساسات و عمل مطابق با هنجارهای اجتماعی را ندارند). با این حال، از آنجایی که علاقه به ربات‌های اجتماعی احتمالاً افزایش می‌یابد، ممکن است افکاری در مورد امکان استفاده از ربات‌های اجتماعی به عنوان ابزاری برای پرورش دوستی وجود داشته باشد. بسیاری از قبل در مورد امکان ایجاد دوستان ربات حدس می‌زدند (گابریل^۲، ۲۰۲۰، ص ۴۱۷). به عنوان مثال، پروموبوت^۳ اخیراً یک ربات انسان‌نمای جدید را معرفی کرده است که قادر به ژست دادن و برقراری ارتباط واقعی با مردم تقریباً مانند یک دوست است.

با وجود امیدواری زیاد، انتقادات شدیدی نیز مطرح است. یکی از نگرانی‌ها این است که ربات‌ها در بهترین حالت فقط می‌توانند واکنش‌های احساسی و الگوهای رفتاری انسان‌ها را در رابطه دوستی، تقلید کنند (داناها^۴، ۲۰۱۹، ص ۱۳). حتی اگر به نظر برسد ربات‌های تعامل‌گرا بیشتر شبیه انسان شده‌اند و با معرفی هوش مصنوعی می‌توانند به شیوه‌ای شخصی‌تر ارتباط برقرار کنند، توانایی‌های

1. Krakovsky
2. Gabriel
3. Promobot
4. Danaher

اجتماعی آنها همچنان از توانایی‌های یک انسان عقب می‌باشد. در نتیجه، می‌توان نتیجه گرفت که تجربه انسان و ربات هنوز به طور قابل توجهی با نوع دوستی مشاهده شده در بین انسان‌ها متفاوت است. در این زمینه، نگرانی دومی نیز مطرح است که در ادبیات به طور گسترده مورد بحث قرار نگرفته است و به این ترتیب کاملاً بدیع است. نگرانی در مورد ناتوانی ربات‌ها در تقلید از سبک‌های دوستی مشاهده شده در بین دوستان واقعی است. یک سالمند تنها یا شخصی که در یک مکان دورافتاده زندگی می‌کند (مثلاً در کوهستان) را در نظر بگیرید، دوست ربات انسان‌نما مطمئناً ممکن است راه حل مطلوب برای چنین شخصی (مسلماً راه حلی بهتر از جایگزین آن؛ یعنی اصلاً دوستی ندارد). باشد. با این حال، جنبه اخلاقی حساسی که باید مورد بحث قرار گیرد، مربوط به سبک دوستی است که ربات باید برای دنبال کردن یا تمرین کردن آن آموزش دیده/از قبل برنامه‌ریزی شده باشد. آیا باید «تأیید کننده و مراقبت کننده» باشد. یعنی آیا همیشه باید از انسان لذت ببرد و از او حمایت کند؟ یا باید «اصلاح کننده» نیز باشد. یعنی به دلیل صلاحیت «خشتی» و منطقی خود باید سعی کند انسان را متقاعد کند تا در صورت نیاز به منطق، نگرش‌ها، رفتارها و/یا سبک‌های زندگی خاصی آن را اصلاح یا بهبود بخشد؟ آیا ربات باید «مخالف» باشد. یعنی دیدگاه خاص خود را داشته باشد (در صورت وجود کدام یک؟) و از آن در تعامل با انسان حمایت کند، حتی اگر منجر به درگیری شود؟ به طور خلاصه، نکته‌ای که در اینجا دنبال می‌شود این است که دوست انسان مقداری از تمام موارد بالا را انجام می‌دهد و احتمالاً حتی بیشتر از آن را انجام می‌دهد. به عنوان مثال، دوست انسانی نه تنها قادر است رفتار/موضع خود را تغییر دهد بلکه با توجه به شرایط محیطی و/یا سناریوهای روانی در حال تغییر می‌تواند تنظیم کند. برعکس، دوست رباتیک نه قادر است موضع خود را تغییر دهد (مگر اینکه دوباره برنامه‌ریزی شود) و نه قادر است رفتار خود را در رابطه با شرایط متغیر (چه محیطی یا روانی) تنظیم کند. به سادگی کاری را انجام می‌دهد که برای آن آموزش دیده یا از قبل برنامه‌ریزی شده است. در این زمینه، ممکن است این اتفاق بیفتد که دوست رباتیکی که در همه شرایط به دوست انسانی‌اش دست و پا می‌زند و لذت می‌دهد و حمایت می‌کند، ممکن است بر مثلاً یک انسان بیش از حد منتقد ترجیح داده شود. با این حال، این وضعیت، مسائل اخلاقی مهمی را ایجاد می‌کند، زیرا ممکن است به تدریج به از دست دادن ارزش‌های اجتماعی مهم (مانند صداقت، احترام، درست‌کاری و سایر ارزش‌های هنجاری) شود که دوستی واقعی را

مشخص می‌کند و برای درستی و روابط گروهی هماهنگ نیز از اهمیت بالایی برخوردار است (فارینا^۱ و همکاران، ۲۰۲۲، ص ۵).

اخلاق ربات

اگرچه رباتیک و هوش مصنوعی فن‌آوری‌های متمایز هستند، همگرایی آنها در بازار امروز به این معنی است که رباتیک یک حوزه التقاطی فزاینده است و به سؤالات اخلاقی ناشی از تحقیقات هوش مصنوعی و توسعه و همچنین مطالعه تعامل انسان- ربات (HRI)^۲ می‌پردازد. ادامه نوشتار حاضر به سمت نگرانی در مورد ربات‌ها و پرورش فضیلت هدایت می‌شود. زیرا بسیاری از کارهای رباتیک در چارچوب‌های اخلاقی مبتنی بر قانون یا فایده‌گرا عمل می‌کنند. چنین چارچوب‌هایی می‌توانند برای تطبیق با جریان تغییرات مداوم، تنوع زمینه‌ای، و افق مبهم فزاینده فن‌آوری‌های در حال ظهور و کاربردهای آنها تلاش کنند. این بدان معنا نیست که چنین رویکردهایی هرگز در اخلاق ربات‌ها مفید نیستند؛ بلکه تنها به این معنا است که آنها ناقص می‌باشند و می‌مانند.

موضوع با اخلاق فایده‌گرا توضیح داده می‌شود. فایده‌گراها فقط یک اصل اخلاقی ثابت و جهانی را تشخیص می‌دهند: وظیفه ترویج بیشترین شادی کلی یا لذت خالص. بنابراین، هنگام قضاوت در مورد اصول اخلاقی مراقبت از کودکان رباتی، یک فایده‌گرا باید تعیین کند که آیا مراقبت از کودکان رباتی در طولانی‌مدت، بیش‌ترین شادی کلی را برای والدین، کودکان و سایر ذینفعان تحت تاثیر ایجاد خواهد کرد یا خیر. بایستی پاسخ داده شود که آیا ساختن یا خرید یک پرستار بچه ربات، اخلاقی است. به طور ساده، تصور کنید که فرد فایده‌گرا فقط باید اخلاق استفاده از یک پرستار ربات را خودش قضاوت کند، نه مراقبت از کودکان رباتی به عنوان یک عمل اجتماعی گسترده. با این حال، همانطور که شری ترکل و سایر محققان کشف کرده‌اند، تعامل انسان و ربات به‌طور آشکاری احساس انسان را نسبت به ربات‌ها به شیوه‌هایی تغییر می‌دهد که به راحتی قابل پیش‌بینی نیستند، و در طول زمان به روش‌های پیچیده تکامل می‌یابد (شوترز^۳، ۲۰۱۱، ص ۲۱۲).

1. Farina
2. human- robot interaction
3. Scheutz

بنابراین فردی که ایده استفاده از یک پرستار ربّاتی را عمیقاً نفرت انگیز یا بیگانه می بیند، نمی تواند بداند که آیا پس از یک ماه یا یک سال استفاده از آن هنوز هم آن احساسات ناخوشایند را خواهد داشت یا خیر. مجموع لذتی که او و فرزندانش از حضور یک پرستار ربّات خواهند برد، به گونه ای خواهد بود که او نمی تواند به راحتی آن را پیش بینی کند. نکته این است که هر محاسبه سودگرایی که فرد در یک برهه زمانی معین انجام می دهد، با قدرت ربّات های اجتماعی برای تغییر علایق و نفرت های ما به گونه ای که دائماً ما را شگفت زده می کند، به شدت بی ثبات می شود. و چگونه قضاوت کنیم که به کدام یک از احساسات خود اعتماد کنیم؟ آن هایی که وقتی یک ربّات اجتماعی برای اولین بار وارد زندگی ما می شود، داریم؟ یا احساسات بسیار متفاوتی که ممکن است بعداً داشته باشیم؟ علاوه بر این، توانایی ها و رفتارهای خود ربّات ها به طور مداوم در حال تکامل است؛ اینکه ما و فرزندانمان چقدر از نسخه یک نوعی ربّات خوشحال شویم، معلوم نمی کند که این ربّات تا چه حد می تواند ما را خوشحال کند، زمانی که به نسخه بالاتر ارتقا پیدا کند. البته این ممکن است به این معنی باشد که ما دائماً باید محاسبات خود را در مورد کاربرد مورد انتظار ربّات اصلاح کنیم. پس چگونه می توان کاربرد مورد انتظار این ربّات را برای خانواده ارزیابی کرد؟ در این چشم انداز اخلاقی سیال، آنچه ما نیاز داریم، اخلاقی است که بر حضور عوامل انسانی انعطاف پذیر، فهم و عملاً خردمند متکی باشد، که فعالانه ویژگی های مورد نیاز برای قضاوت عاقلانه و خوب در چنین شرایط ناپایداری را پرورش می دهند. اینجا جایی است که اخلاق فضیلت آغاز می شود. در حالی که فایده گرایان و کانتی ها یک فرمول یا دستورالعمل ثابت برای استفاده در هر قضاوت اخلاقی به ما ارائه می دهند، اخلاق گرایان فضیلت ترجیح می دهند یاد بگیریم که چگونه قاضی خوبی شویم. علاوه بر این، رویکردهای کاملاً ریشه شناختی و فایده گرایانه به اخلاق ربّاتیک تمایل به تغییر یا حتی دور زدن این سؤال دارند که در واقع چه نوع کالاهای و فعالیت ها و زمینه های دریافت آن ها برای انسان مفید است. این را می توان به وضوح در مورد سوء استفاده یا شکنجه ربّات ها نشان داد، موردی که نمونه های قابل توجهی از آن در دنیای واقعی داریم (ویتبی^۱، ۲۰۱۱، ص ۲۴۰). خرده فرهنگ آینده انسان ها را در نظر بگیرید که عادت به شکنجه ربّات های اجتماعی دارند، لذت بردن از فریادهای ربّات ها، پیچ خوردن، و التماس های ناامیدانه برای رحمت، بسیار زیاد می شوند. می توان

فرض کرد که این ربات‌ها حساس نیستند، آنها فقط درد را تقلید می‌کنند و واقعا رنج نمی‌برند. با این حال، از آنجایی که انسان‌هایی که به طور گسترده با ربات‌های اجتماعی تعامل دارند، معمولاً حالت‌های ذهنی قوی را بر روی آن‌ها ایجاد می‌کنند، فرض می‌کنند که شکنجه‌گران به وضوح درک می‌کنند که ربات‌های آن‌ها به شدت رنج می‌برند (ترکل^۱، ۲۰۱۱). اجازه دهید صرفاً برای استدلال، تصور کنیم که این عادت سادیستی باعث سرایت به رفتار صاحبان ربات با هموعان خود نخواهد شد- نتیجه‌ای که در غیر این صورت تقریباً هر نظریه اخلاقی را آسان می‌کند که این عادت را غیراخلاقی بداند. اگر از منظر کانتی به آن بپردازیم، به نظر نمی‌رسد سناریویی که در بالا توضیح داده شد ناقض الزام قاطع باشد. آن قاعده اخلاقی فقط مستلزم آن است که یک عمل عمومیت پذیر باشد و یک عامل عقلانی را صرفاً وسیله‌ای برای رسیدن به هدف تلقی نکند (کانت^۲، ۱۹۹۷). تصور یک عمل جهانی از شکنجه ربات‌ها غیرممکن نیست. از طرفی دیگر، این ربات‌های خاص واقعاً موجوداتی منطقی مستقل و با وقار نیستند و سزاوار این نیستند که به عنوان «هدف فی نفسه^۳» با آنها رفتار شود. بنابراین، از دیدگاه فایده‌گرایانه، شکنجه این ربات‌ها نه تنها مجاز به نظر می‌رسد، بلکه فعالیت‌های اخلاقی مثبتی برای سادیست‌ها به نظر می‌رسد، زیرا بدون ایجاد درد برای کسی لذت می‌برد. این نتایج نشان می‌دهد که چرا چنین نظریه‌های اخلاقی ناقص هستند. هیچ یک از این چارچوب‌ها منابعی ندارند برای پاسخ به این که چه اعمال یا ویژگی‌های انسانی نجیب، قابل تحسین و شایسته آرزوهای ما هستند. از منظر اخلاق فضیلتی، افرادی که بیشتر اوقات فراغت خود را صرف کسب لذت بی‌شمار از شکنجه ربات‌ها می‌کنند، حتی اگر با لذت بی‌ضرر خود، لذت خالص را در جهان افزایش دهند، خوب زندگی نمی‌کنند یا شکوفا نمی‌شوند، چرا که این فعالیت، هیچ کدام یک از ویژگی‌ها، مهارت‌ها یا انگیزه‌های شخصیتی را که باعث تعالی و شکوفایی انسان می‌شود، پرورش نمی‌دهد. هیچ چیز در مورد فعالیت مورد علاقه شکنجه‌گران آنها را انسان‌های بهتری نمی‌سازد، و از آنجا که یک انسان خوب دارای ویژگی‌های شخصیتی و الگوهای فعالیتی است که مستقیماً بر خلاف سادیسم است، آزاردهندگان ربات‌ها بخاطر عادات سادیسمی خود کمتر نجیب هستند، حتی اگر هیچ فرد زنده‌ای توسط عادات آن‌ها آسیب نبیند.

1. Turkle
2. Kant
3. ends in themselves

اخلاق فضیلت و زندگی خوب با هوش مصنوعی

بحث در مورد چگونگی استفاده یا زندگی با هوش مصنوعی به گونه‌ای که در واقع راهنمایی عینی را در شرایط جدید فراهم کند، بسیار مهم است. با توجه به دیدگاه ارسطویی که تشریح شد، پاسخ سوالاتی از این دست: چه باید بکنیم یا چگونه باید با هوش مصنوعی زندگی کنیم؟، به این صورت است: هر آنچه حکیم با هوش مصنوعی انجام می‌دهد. در حالی که این پاسخی است که یک نظریه پرداز فضیلت به نوعی باید بدهد، این به طور قابل درک برای کسانی که می‌خواهند یک اخلاق کاربردی از هوش مصنوعی برای ارائه راهنمایی عینی یا خاص فراهم کند، ناخوشایند است. زمانی که حکمت‌های اخلاقی با درک کافی از تکنولوژی‌های هوش مصنوعی وجود داشته باشند، راهنمایی‌های عینی و خاص از تعیین اینکه حکیم در یک موقعیت خاص چه کاری انجام خواهد داد، حاصل می‌شود. اما، پیاده سازی با فقدان قوانین عینی یا راهنمایی ارائه شده توسط یک چارچوب اخلاقی فضیلت و فقدان نمونه‌های اخلاقی به راحتی قابل شناسایی، پیچیده می‌شود. در بخش‌های قبلی مقاله، توجیهات نظری برای رویکرد ارسطویی به اخلاق فضیلت را مورد بررسی قرار گرفت؛ در ادامه عملی بودن این مسیر از نظر نیکلاس اسمیت و داربی ویکرز بررسی می‌شود.

نیکلاس اسمیت استدلال می‌کند که پیاده‌سازی رویکرد ارسطویی به هوش مصنوعی دشوار است، زیرا یافتن افراد با فضیلت که نمونه‌های حیاتی برای یادگیری اخلاقی هستند، یک چالش جدی است. این افراد باید دانش فنی مرتبط را نیز داشته باشند و در توسعه، استقرار و سیاست‌گذاری هوش مصنوعی دخیل باشند. مشکل این است که افراد با فضیلت باید هم عادت به فضیلت داشته باشند و هم آموزش ببینند که هر دو فرایندهایی زمان‌بر و پرهزینه هستند. برای شروع این برنامه، حداقل به یک فرد با فضیلت و دانش فنی نیاز است که به عنوان الگو و راهنما عمل کند. می‌توان آموزش فنی و پرورش شخصیت را به طور همزمان دنبال کرد، اما دیدگاه ارسطویی این است که شخصیت پس از مدتی نسبتاً ثابت می‌شود و تغییر آن دشوار است. با این حال، تلاش برای آموزش افراد به گونه‌ای که هم با فضیلت باشند و هم دانش فنی داشته باشند، بهترین کاری است که می‌توان انجام داد، حتی اگر با شکست مواجه شود. زیرا هدف باید دستیابی به فضیلت باشد، نه پذیرش شکست و یافتن

روشی برای زندگی با فناوری که نسبت به مشکلات رویکردهای فعلی بی تفاوت است (کروچه و پریچارد^۱، ۲۰۲۲، اورونا و پریچارد^۲، ۲۰۲۲).

ویکرز نیز با اسمیت هم عقیده است که آموزش فضیلت چالش‌های جدی دارد، اما نسبت به تاثیرگذاری مثبت آن در سطح دانشگاه خوش‌بین‌تر است. این خوش‌بینی بر سه پایه استوار است: مفهوم "انکراتیا"^۳ (خوشتنداری) در اندیشه ارسطو، تحقیقات جدید درباره انعطاف‌پذیری مغز در بزرگسالی، و تجربه شخصی نویسنده در تدریس دروس اخلاق. ارسطو معتقد بود که دستیابی به فضیلت بسیار دشوار است و تنها افراد بزرگ‌منش می‌توانند به آن برسند. با این حال، او "انکراتیا" را به عنوان گامی میانی به سوی فضیلت یا نقطه‌ای پایانی معرفی می‌کند که افراد را به اعضای محترم جامعه تبدیل می‌کند. افراد دارای انکراتیا مهارت‌های لازم برای تشخیص درست‌ترین عمل در موقعیت‌های مختلف را دارند، حتی اگر احساساتشان به درستی تنظیم نشده باشد. ویکرز معتقد است که آموزش انکراتیا چالش‌برانگیز است و نیازمند توسعه مهارت‌های تفکر اخلاقی در طول زمان است. پیشرفت سریع هوش مصنوعی علاقه به اخلاق را افزایش داده و امید است که این امر منجر به آموزش اخلاق و توسعه مهارت‌های تفکر اخلاقی در برنامه درسی، به ویژه برای دانشجویان رشته‌های فنی شود.

اگرچه ارسطو معتقد بود که شخصیت پس از مدتی ثابت می‌شود و آموزش نمی‌تواند آن را تغییر دهد، ویکرز با توجه به تحقیقات جدید درباره انعطاف‌پذیری مغز، معتقد است که ایجاد فضیلت یا حداقل انکراتیا در سنین بالاتر امکان‌پذیر است. مغز انسان انعطاف‌پذیر است و می‌تواند تغییر کند، حتی پس از دوره‌های رشد عصبی در کودکی و بزرگسالی. مطالعات نشان داده‌اند که دوره‌های اخلاق به طور موثری رفتار افراد را تغییر نمی‌دهند. اما ویکرز معتقد است که این دوره‌ها می‌توانند کمک کنند تا مهارت‌های لازم برای شناسایی ویژگی‌های اخلاقی یک موقعیت، درک فضیلت مورد نیاز، نحوه عمل، و تنظیم احساسات خود را به دست آورند. به عبارت دیگر، یادگیری اخلاقی نوعی یادگیری کارآموزی است که در آن دانشجویان باید به محیط یادگیری دسترسی داشته باشند تا بتوانند تصمیم‌گیری اخلاقی را تمرین کنند. او همچنین معتقد است که علاقه به اخلاق به دلیل

1. Croce and Pritchard
2. Orona
3. enkratia

پیشرفت سریع هوش مصنوعی افزایش یافته است. در نتیجه، امیدوار است که با ایجاد یک برنامه درسی اخلاقی، دانشجویان را به سمت انکراتیا سوق دهد. به طور خلاصه، هر دوی ویکرز و اسمیت معتقدند که آموزش اخلاق راه درستی برای کمک به نسل بعدی برای زندگی خوب با هوش مصنوعی است، هر چند که این مسیر دشواری است (شویتگیل^۱ و همکاران، ۲۰۲۰، اریکسون و پول^۲، ۲۰۱۶، ص ۴۹-۲۶).

نتیجه‌گیری

در حال حاضر فن‌آوری‌های گسسته با هم‌افزایی، دامنه و قدرت خود را برای تغییر زندگی‌ها تا حد زیادی بزرگ کرده‌اند. برای مقابله با چالش‌های اخلاقی عصر هوش مصنوعی و فن‌آوری‌های دیجیتال، منابع اخلاقی جدید را می‌توان در سنت فلسفی اخلاق فضیلت یافت. چون که بشر به یک اخلاق مثبت و سازنده در فن‌آوری‌های نوظهور و هوش مصنوعی نیاز دارد، که علاوه بر مقررات، به مسئله زندگی خوب و شکوفایی اجتماعی انسان نیز پردازد. اخلاق‌گرایان فضیلت ترجیح می‌دهند یاد دهند که چگونه قاضی خوبی شویم. به بیانی دیگر، در دنیای پرتحول امروز که فناوری‌های جدید به سرعت در حال ظهور هستند، به اخلاقی نیاز هست که به جای تکیه بر قوانین خشک و غیرقابل انعطاف، بر انسان‌های آگاه، فهمیده و خردمند تکیه کند. اخلاق فضیلت، با تاکید بر پرورش صفات نیکو در افراد، می‌تواند به آنها کمک کند تا در عصر هوش مصنوعی، زندگی خوب و سعادت‌مندی داشته باشند. بنابراین، یک اخلاق فضیلت فنی-اجتماعی این پتانسیل را دارد که با حکمت در برخورد با عدم قطعیت‌ها، ریسک‌ها و فرصت‌های ایجاد شده توسط فن‌آوری‌های جدید، برای شبکه‌ای جهانی مملو از تنوع فرهنگی و افرادی که نیاز به اخلاق جمعی دارند، اعمال شود. در حوزه رباتیک نیز، اخلاق فضیلت با پرورش ویژگی‌های لازم برای قضاوت درست و عاقلانه در شرایط پیچیده، می‌تواند به شکوفایی و تعالی انسان کمک کند. پس در این دنیای متغیر، به جای دستورالعمل‌های ثابت، به انسان‌هایی نیاز است که بتوانند با خردمندی و درایت، بهترین تصمیم‌ها را بگیرند و راه درست را پیدا کنند.

1. Schwitzgebel
 2. Ericsson & Pool

ملاحظه‌های اخلاقی

در پژوهش حاضر با معرفی منابع مورد استفاده، اصل اخلاق امانت داری علمی رعایت و حق معنوی مولفین آثار محترم شمرده شده است.

منابع

- دادجو، ابراهیم. (۱۳۹۹). ذات گرایی جدید مابعدالطبیعه جدید. ذهن. ۲۱(۸۴): ۷۷-۱۱۲.
- رضایی اقچری، نیلوفر و رعایت جهرمی، محمد (۱۴۰۳). بررسی و مقایسه خوانش دریفوس و کرزویل از هوش مصنوعی. فلسفه علم. ۱۴(۱)، ۱۹۷-۲۲۶.
- صمدیه، و ملایوسفی. (۱۳۹۶). ارسطو و فرونیسیس (حکمت عملی). آینه معرفت، ۱۷(۴)، ۱-۲۴.
- نعمت الهی، شرلی، آزادی، کیهان، و وطن پرست، محمدرضا. (۱۳۹۸). بررسی پیامدگرایی و وظیفه گرایی اخلاقی با تاکید بر نظریه کانت در جامعه حسابداران رسمی ایران و تاثیر آن بر میزان پابندی بر رفتار حرفه‌ای حساب‌رسان. پژوهش‌های اخلاقی (انجمن معارف اسلامی)، ۱۰(۱)، ۳۱-۴۸.

References

1. Acquisti A (2009) Nudging privacy: the behavioral economics of personal information. IEEE Secur Priv 7:82-85.
2. Aristotle (2007) The Nicomachean ethics. Filiquarian Publishing, Minneapolis.
3. Bakhurst, D. (2018). Practice, Sensibility and Moral Education. Journal of Philosophy of Education, 52(4).
4. Benjamin, R. (2019). Race After Technology. Polity Press.
5. Berg, P. (2008). Asilomar 1975: DNA modification secured. Nature 455: 290-291.
6. Borgmann, Albert. 1984. Technology and the Character of Contemporary Life. Chicago: University of Chicago Press.
7. Beauchamp, T., & Childress, J. (2012). Principles of Biomedical Ethics (7th ed.). Oxford University Press.
8. Camerer CF, Loewenstein G, Rabin M (2004) Advances in behavioral economics. Princeton University Press, Princeton, NJ
9. Cath CJN, Wachter S, Mittelstadt B, Taddeo M, Floridi L (2016) Artificial intelligence and the 'good society': the US, EU, and UK approach. Social Science Research Network, Rochester, NY
10. Craglia M, Annoni A, Benczur P et al (2018) Artificial intelligence: a European perspective. Publications Office of the European Union, Luxembourg
11. Creemers R (2018) China's social credit system: an evolving practice of control. Social Science Research Network, Rochester, NY

12. Croce, M., & Pritchard, D. (2022). Education as the Social Cultivation of Intellectual Virtue. In M. Alfano, C. Klein, & J. de Ridder (Eds.), *Social Virtue Epistemology*. Routledge, pp. 583–601.
13. Danaher J (2019) The philosophical case for robot friendship. *J Posthum Stud* 3(1):5–24.
14. Dignum V (2019) Responsible artificial intelligence: how to develop and use AI in a responsible way. Springer Nature Switzerland AG, Cham, Switzerland.
15. Doudna, J. & Sternberg, S. (2017) *A Crack in Creation: Gene Editing and the Unthinkable Power to Control Evolution*. Houghton Mifflin Harcourt.
16. Ericsson, A. & Pool, R. (2016) *Peak: The New Science of Expertise*. Houghton Mifflin Harcourt.
17. Farina, M., Zhdanov, P., Karimov, A., & Lavazza, A. (2022). AI and society: a virtue ethics approach. *AI & SOCIETY*, 1-14.
18. Floridi, L. (2013). *The ethics of information*. Oxford University Press, USA.
19. Floridi, L. et al. (2021). "An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." In: Floridi, L. (eds) *Ethics, Governance, and Policies in Artificial Intelligence*. Philosophical Studies Series, vol 144. Springer, Cham.
20. Floridi, L., Cows, J. (2021). "A Unified Framework of Five Principles for AI in Society." In: Floridi, L. (eds) *Ethics, Governance, and Policies in Artificial Intelligence*. Philosophical Studies Series, vol 144. Springer, Cham.
21. Gabriel I (2020) Artificial intelligence, values, and alignment. *Minds Mach* 30(3):411–437.
22. Hanks, Craig, ed. 2010. *Technology and Values*. Malden, MA: Wiley Blackwell.
23. Haydon, G. (2010). Reason and Virtues: The Paradox of R. S. Peters on Moral Education. *Journal of Philosophy of Education*, 43: 173–188.
24. Hursthouse, R. (1999). *On Virtue Ethics*. Oxford University Press.
25. Isaak J, Hanna MJ (2018) User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer* 51:56–59.
26. Johal W (2020) Research trends in social robots for learning. *Curr Robot Rep* 1:75–83.
27. Jonas, H. (1979). *Toward a philosophy of technology*. Hastings Center Report, 34-43.
28. Jonas, H. (1984). *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. University of Chicago, 202.
29. Kant, Immanuel. [1785] 1997. *Groundwork of the Metaphysics of Morals*. Translated by Mary Gregor. Cambridge, UK: Cambridge University Press.
30. Klar R, Lanzerath D (2020) The ethics of COVID-19 tracking apps—challenges and voluntariness. *Res Ethics* 16:1–9.
31. Krakovsky M (2018) Artificial (emotional) intelligence. *Commun ACM* 61(4):18–19.
32. Liu C (2019) Multiple social credit systems in China. *Social Science Research Network*, Rochester, NY.
33. MacIntyre AC (2007) *After virtue: a study in moral theory*. University of Notre Dame Press, Notre Dame, IN.

34. MacIntyre, A. (1984). *After Virtue*, 2nd edn (Notre Dame, IN. IN: University of Notre Dame Press.
35. Orona, G. A., & Pritchard, D. (2022). Inculcating curiosity: pilot results of an online module to enhance undergraduate intellectual virtue. *Assessment and Evaluation in Higher Education*, 47(3): 375–389.
36. Scharff, R. C. (2012). Empirical technoscience studies in a Comtean world: too much concreteness?. *Philosophy & Technology*, 25, 153-177.
37. Scheutz, M. (2011). 13 The inherent dangers of unidirectional emotional bonds between humans and social robots. *Robot ethics: The ethical and social implications of robotics*, 205.
38. Schuller D, Schuller BW (2018) The age of artificial emotional intelligence. *Computer* 51(9):38–46.
39. Stahl, B. C. (2023). Embedding responsibility in intelligent systems: from AI ethics to responsible AI ecosystems. *Scientific Reports*, 13(1), 7586.
40. Schwitzgebel, E., Corklet, B. & Singer, P. (2020) "Do Ethics Classes Influence Student Behavior? Case Study: Teaching the Ethics of Eating Meat", *Cognition*, 203: 104397.
41. Rawas, S. (2024). AI: the future of humanity. *Discover artificial intelligence*, 4(1), 25.
42. Turkle, Sherry. 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books.
43. Vallor, S. (2016). *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford University Press.
44. Verbeek, P. P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. Penn State Press.
45. Whitby, B. (2011). 15 Do You Want a Robot Lover? *The Ethics of Caring Technologies*. *Robot ethics: The ethical and social implications of robotics*, 233.
46. Yashchenko, V. (2024). Artificial intelligence: free will, self-consciousness and ethics. *Artificial Intelligence*, 29(AI. 2024.29 (3)), 32-51.
47. Yearley, Lee H. 1990. *Mencius and Aquinas: Theories of Virtue and Conceptions of Courage*. Albany, NY: SUNY Press.
48. Yearley, Lee H. 2002. "An Existentialist Reading of Book Four of the Analects." In *Confucius and the Analects, New Essays*, edited by Bryan Van Norden, 237–274. New York: Oxford University Press.
49. Zuboff PS (2019) *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. Profile Books, London.