سامانه نشریه:https://cph.samt.ac.ir

شاپا الکترونیکی:۹۴۶۸-۳۰۴۱

فصلنامه علمی فلسفه تطبیقی

# دیدگاه لوچانو فلوریدی در باره چارچوب اخلاقی هوش مصنوعی و اصول اعتمادبخش آن در جوامع خیر و بررسی و نقد آن

حمید شهریاری[1]

۱- دانشیار، گروه حقوق و فقه، پژوهشکده تحقیق و توسعه علوم انسانی (سمت)

**اطلاعات مقاله**

**نوع مقاله:** پژوهشی

**تاریخ ارسال:** ۱۴۰۴/۰۷/۲۹

**تاریخ پذیرش:** ۱۴۰۴/۰۸/۱۴

**کلیدواژه‌ها:** هوش مصنوعی، لوچانو فلوریدی، چارچوب اخلاقی، مسئولیت، توضیح‌پذیری

**چکیده**

مقاله به بررسی چارچوب اخلاقی لوچانو فلوریدی برای هوش مصنوعی (AI) می‌پردازد. وی با تحلیل شش سند بین‌المللی در حوزه اخلاق فناوری، از میان ۴۷ اصل مذکور در این اسناد پنج اصل اخلاقی شامل خیرخواهی، منع ضرر، خودمختاری، عدالت و توضیح‌پذیری را استخراج کرده و آن‌ها را برای ایجاد اعتماد عمومی و سیاست‌گذاری صحیح در جوامع انسانی ضروری می‌داند.

او ضمن تأکید بر تهدیدهای هوش مصنوعی نظیر کاهش مهارت‌های انسانی، حذف مسئولیت انسان، افت کنترل و از بین رفتن توان تعیین سرنوشت، به فرصت‌هایی چون ارتقای کنشگری، خودشکوفایی و انسجام اجتماعی اشاره می‌کند. مقاله نقدهایی از منظر اخلاق اسلامی و فلسفه اگزیستانسیالیسم نیز بر این چارچوب وارد می‌کند؛ از جمله اینکه خودمختاری در نگاه اسلامی محدود به حدود الهی است و نه مطلق.

همچنین مقاله، ایده فروکاست ۴۷ اصل به ۵ اصل را نادرست دانسته و معتقد است این کار، تعارض‌های اخلاقی را پنهان کرده و امکان اجرای مؤثر را کاهش می‌دهد. نویسنده نتیجه می‌گیرد که اخلاق هوش مصنوعی نیازمند چارچوب چندسطحی و زمینه‌محور است که قابل تطبیق با نیازهای عملی، اجتماعی و فرهنگی باشد.

**شیوه استناد به این مقاله:** شهریاری، حمید؛ ***فلسفه تطبیقی***،(۱۴۰۴) شماره ۲، صفحات۱۱تا ۲۹ :DOI: 10.30487/cph.2025.2075404.1053.

ناشر: سازمان مطالعه و تدوین کتب دانشگاهی در علوم انسانی (سمت)

سامانه نشریه:https://cph.samt.ac.ir شاپا الکترونیکی:۹۴۶۸-۳۰۴۱

فصلنامه علمی فلسفه تطبیقی

# Critical interpretation of subject and object duality in Kant's philosophy based on the duality of subjective and universal world In the story of descent from the jurisprudential-interpretive perspective of Allameh Tabatabai

**Hamid Shahriari**[1]

1. Associate Professor, Comparative Philosophy, Department of Law and Jurisprudence, Institute for Research and Development in the Humanities (SAMT), Tehran, Iran.

**Article Info**

**Article type:** Review Articl

**Receive**d: 2025-10-21

**Accepted:** 2025-11-05

**Keywords:** Artificial Intelligence, Luciano Floridi, Ethical Framework, Resposibility, Explainability

**ABSTRACT**

The article examines Luciano Floridi's ethical framework for Artificial Intelligence (AI). By analyzing six international documents on the ethics of technology, he identifies five ethical principles - beneficence, non-maleficence, autonomy, justice, and explainability—out of the forty-seven principles mentioned in those documents, and considers them essential for fostering public trust and sound policymaking in human societies.

While emphasizing the threats posed by AI - such as the erosion of human skills, elimination of human responsibility, decline of control, and loss of the capacity for self- determination e - also points to opportunities, including the enhancement of agency, self-fulfillment, and social cohesion. The article also presents critiques of this framework from the perspectives of Islamic ethics and existentialist philosophy, noting, for instance, that autonomy in Islamic thought is confined within divine boundaries and is not absolute.

Furthermore, the article argues that reducing the forty-seven principles to five is misguided, as it obscures ethical conflicts and diminishes the possibility of effective implementation. The author concludes that AI ethics requires a multi-layered and context-sensitive framework that can be adapted to practical, social, and cultural needs

**Cite this article:** Shahriari, Hamid; (2025). *Comparative Philosophy* (3), 11–29. **DOI:** 10.30487/cph.2025.2075404.1053
**Publisher:** Organization for Researching and Composing University Textbooks in the Humanities (SAMT).

**مقدمه**

در دهه‌های اخیر، پرسش از چارچوب اخلاقی هوش مصنوعی در بستر دگرگونی‌های اجتماعی، فرهنگی و فناورانه اهمیتی فزاینده یافته است. از آنجا که هوش مصنوعی خود موضوعی جدید است چارچوب اخلاقی آن نیز موضوعی نوپدید محسوب می‌شود. پژوهش‌های گسترده در این موضوع غالباً به زبان فارسی نیستند و دستاوردهای موجود، غالباً دارای عمقی محدودند و از مطالعات اخلاقی در موضوع مسائل دیگر از جمله اخلاق پزشکی وام گرفته شده‌اند. ازاین‌رو شکاف دانشی معناداری بین اصول کلی اخلاق هوش مصنوعی و لایه‌های عملی هوش مصنوعی پابرجاست. این مقاله با تکیه بر چارچوب نظری اصول کلی اخلاقی و رویکرد روش‌شناختی تحلیل اسنادی، تفسیر متون، و تحلیل گفتمان می‌کوشد ابتدا به تهدیدهای این حوزه و فرصت‌ها مقابل آن اشاره کند. تهدیدهایی چون بی‌ارزش شدن مهارت‌های انسانی، حذف مسئولیت انسان، کاهش کنترل‌های انسانی، و تحلیل رفتن توان تعیین سرنوشت، در مقابل فرصت‌هایی چون خودشکوفایی، ارتقای کنشگری انسان، افزایش توانمندی‌های اجتماعی و تعمیق انسجام اجتماعی قرار می‌گیرند. سپس فلوریدی شش سند اصلی را بررسی می‌کند و مزایا و ویژگی‌های هر یک را برمی‌شمارد و در نهایت از دل آن‌ها پنج اصل کلی را استخراج می‌کند که چهار اصل اول آن همان اصول مشهور در اخلاق زیستی است: اصل خیرخواهی، اصل منع ضرر، اصل خودمختاری، و اصل عدالت. او سپس افزودن اصل دیگری را برای اصول هوش مصنوعی با عنوان توضیح‌پذیری ضروری می‌داند. این مقاله در اینجا رویکرد او را نقد کرده‌ایم و اشکالاتی که به چارچوب اخلاقی او وارد آمده است را بررسی کرده‌ایم. هدف از این مقاله تبیین اصولی است که در سیاست‌گذاری حوزه هوش مصنوعی کاربرد و سابقه دارد و می‌تواند چراغ راهی برای آینده باشد.

## درباره لوچانو فلوریدی

لوچانو فلوریدی، متولد ۱۶ نوامبر ۱۹۶۴ در رم، پایتخت ایتالیا، فیلسوف برجسته‌ای است که به دلیل کارهای پیشگامانه‌اش در فلسفه اطلاعات و اخلاق دیجیتال شناخته شده است. او مدرک کارشناسی ارشد خود را از دانشگاه ساپینزا رم دریافت کرد و سپس مدرک دکتری خود را از دانشگاه وارویک انگلستان اخذ نمود (“Luciano Floridi”, 2025).

فلوریدی در طول دوره کاری خود در سمت‌های مهمی از جمله در سمت استاد فلسفه و اخلاق اطلاعات در مؤسسه اینترنت آکسفور[1] (OII) فعالیت کرده، جایی که رهبری «آزمایشگاه اخلاق دیجیتال» را بر عهده داشت. او در سال ۲۰۲۳، به دانشگاه ییل پیوست و به‌عنوان مدیر بنیان‌گذار مرکز اخلاق دیجیتال و استاد علوم شناختی منصوب شد( “Luciano Floridi| Digital Ethics Center”, 2025).

حوزه‌های پژوهشی اصلی او شامل فلسفه اطلاعات، اخلاق دیجیتال، اخلاق هوش مصنوعی و فلسفه فناوری است. فلوریدی نقش بزرگی در درک پیامدهای اخلاقی فناوری‌های نوظهور، به‌ویژه هوش مصنوعی، داشته است. او بیش از ۳۰۰ اثر در این زمینه‌ها منتشر کرده است، از جمله کتاب‌های برجسته‌ای مانند «*اخلاق هوش مصنوعی: اصول، چالش‌ها و فرصت‌ها*» (۲۰۲۳) و «*سبز و آبی: ایده‌های ساده برای بهبود سیاست در عصر دیجیتال*» (۲۰۲۳) (Digital Ethics Center at Yale).

در طول زندگی حرفه‌ای‌اش، فلوریدی جوایز متعددی به دلیل دستاوردهایش دریافت کرده است. در سال ۲۰۲۳، او بالاترین نشان افتخار کشور ایتالیا، نشان شوالیه بزرگ صلیب از نشان شایستگی، و جایزه علمی فیوجی را برای کارهایش در زمینه اخلاق هوش مصنوعی دریافت کرد (“About – Luciano Floridi | Philosophy of Information”, 2025).

کارهای فلوریدی تأثیر عمیقی بر حوزه اخلاق دیجیتال و فلسفه و اخلاق هوش مصنوعی گذاشته و به تثبیت این حوزه‌ها

1. Oxford Internet Institute

به‌عنوان زمینه‌های مهم پژوهشی کمک کرده است. رویکرد میان‌رشته‌ای او پلی بین فلسفه، علوم رایانه، و اخلاق ایجاد کرده و بحث‌های مهمی در مورد تأثیر فناوری بر توانایی انسانی و جامعه آغاز کرده است ("Professor Luciano Floridi | University of Oxford", 2025).

علاوه بر فعالیت‌های دانشگاهی، فلوریدی در ابتکارات سیاست‌گذاری مرتبط با ارزش‌ها و پیامدهای اجتماعی-اخلاقی فناوری‌های دیجیتال مشارکت داشته است. او با سازمان‌هایی مانند کمیسیون اروپا، یونسکو، و نهادهای دولتی مختلف همکاری کرده تا چارچوب‌های اخلاقی برای توسعه و بهره‌برداری از هوش مصنوعی و فناوری‌های مرتبط تدوین کند ("Professor Luciano Floridi | University of Oxford", 2025).

مجموعه آثار گسترده فلوریدی همچنان به بحث‌ها و شکل‌گیری دیدگاه‌ها درباره چالش‌های اخلاقی ناشی از هوش مصنوعی و رابطه در حال تحول بین انسان‌ها و فناوری کمک می‌کند. ما در اینجا به بخشی از مباحث او می‌پردازیم که مربوط به چارچوب اخلاقی هوش مصنوعی و اصول حاکم بر آن است. ۱۲ تن از اساتید و دانشجویان در نگارش این بحث به او کمک کرده‌اند که این مقاله به دلیل اختصار، فقط از فلوریدی نام می‌برد. این بدان جهت است که او در این تحقیق محور و مسئول پروژه تحقیق بوده است.

## چارچوب اخلاقی هوش مصنوعی

فلوریدی تلاش کرده مبتنی بر مطالعات پیشین در حوزه اخلاق هوش مصنوعی، فرصت‌ها و تهدیدهای اصلی این حوزه را برای جوامع انسانی برشمرده و ترکیبی از پنج اصل اخلاقی را به عنوان زیربنایی برای پذیرش و توسعه آن عرضه کند و سپس بیست توصیه و پیشنهاد برای ارزیابی، توسعه، و حمایت از هوش مصنوعی خیر بیان کند که سیاست‌گذاران ملی و بین‌المللی یا ذی‌نفعان دیگر باید بدان‌ها متعهد گردند تا تضمینی برای دستیابی به کاربردهای خوب هوش مصنوعی در جوامع بشری باشد.

در اینجا باید توجه داشت که هوش مصنوعی یک ابزار همچون ابزارهای دیگر نیست که فقط نیاز به مقررات‌گذاری جدید داشته باشد. بلکه بیشتر یک تحول‌ساز جامعه بشری است؛ عاملی قدرتمند است که سبک زندگی ما و ارتباطات ما و محیط زیست ما را دگرگون می‌سازد. به همین جهت لوچانو فلوریدی با مدیریت بر یک گروه ۱۲ نفره سندی رسمی تولید کرد تا راهنمایی باشد برای دولت‌ها که چگونه قوانین مربوط به هوش مصنوعی را تدوین کنند. این کار جمعی در گروهی با عنوان «هوش مصنوعی برای مردم» (AI4people) کار خود را با تدوین نقشه راه برای نشستی در اتحادیه اروپا در مورد تأثیرات اجتماعی هوش مصنوعی آغاز کرد. این گروه کاری به‌وسیله «مؤسسه اروپایی اتومیوم برای علم، رسانه و مردم‌سالاری» تأسیس گردید. هدف آن تنظیم مقررات مربوط به هوش مصنوعی نبود بلکه می‌خواست چراغ راهی را فراهم آورد تا در پرتو آن سیاست‌های مناسب برای توسعه هوش مصنوعیِ خوب و اخلاق‌پسند در جوامع انسانی تدوین گردد. هدف این بود که باهوش مصنوعی کرامت انسان رعایت گردد، بستر شکوفایی انسان فراهم آید و جهانی خوب‌تر برای حیات بشر بدست آید. آنچه فلوریدی راهبری کرد تلاش یک گروه ۱۲ نفره بود که در میان آنان متخصصان رشته‌های مختلفی چون حقوق، فناوری و اخلاق حضور داشتند (Floridi, 2021a, pp. 19–20).

گروه کاری «هوش مصنوعی برای مردم» اهداف راهبردی خود را به این شرح اعلام کرده است:

۱. تأسیس یک مجمع جهانی در زمینه هوش مصنوعی، که از نمایندگان دولت‌ها در سراسر جهان، نهادهای اروپایی، سازمان‌های جامعه مدنی، رسانه‌های مرتبط و کسب‌وکارهای پیشرو استقبال می‌کند. و از طریق این مجمع:

۲. ارزش‌های بنیادینی را شناسایی می‌کند که باید چارچوب اخلاقی حمایت از توسعه بهینه هوش مصنوعی را مشخص کند.

۳. یک چارچوب اخلاقی اروپایی برای "جامعه هوش مصنوعی خوب" بر اساس هدف دوم طراحی می‌کند.
۴. در چارچوب طراحی شده در هدف سوم، توصیه‌هایی را در مورد معیارهای عملی برای توفیق آن درج کند.
۵. درون این مجمع، یک کمیته دائمی برای جامعه هوش مصنوعی خوب ایجاد کند تا اطلاعات لازم را در مورد تکامل مداوم فناوری‌ها و برنامه‌های کاربردی مبتنی بر هوش مصنوعی که بر جامعه تأثیر دارند، به سیاست‌گذاران این حوزه ارائه و از آنان حمایت کند.
۶. آگاهی و مشارکت شهروندان در مورد هوش مصنوعی و پیامدهای اخلاقی آن بر حکمرانی را ارتقا دهد.

هدف اصلی این گروه آنست که یک «جامعه هوش مصنوعی خوب» در اروپا خلق کنند. همچنین این گروه تلاش کرده که با همکاری و مشارکت تمامی ذی‌نفعان این حوزه، اصول و سیاست‌هایی تدوین کنند که برای بقیه جهان نیز کاربرد داشته باشد (Floridi, 2018, p. 5).

## چهار فرصت در برابر چهار تهدید

او چهار فرصت را در مقابله با چهار تهدید به تصویر می‌کشد که در استفاده از هوش مصنوعی بروز می‌یابد.

فرصت اول، مربوط به هویت انسانی است از آن جهت که انسان موجودی خودمختار و خودشکوفا است. تهدید مقابل این فرصت، بی‌ارزش شدن مهارت‌های انسانی است. فرصت دوم، ارتقای کنشگری انسان در مقابل تهدید حذف مسئولیت انسان است. فرصت سوم، افزایش توانمندی‌های اجتماعی در برابر تهدید کاهش کنترل‌های انسانی و فرصت چهارم، تعمیق انسجام اجتماعی در برابر تهدید تحلیل رفتن توان تعیین سرنوشت است. اینک به ترتیب از این فرصت‌ها و تهدیدهای چهارگانه بحث می‌کنیم:

## فرصت اول: خودشکوفایی

**اول:** فرصت خودشکوفایی در برابر تهدید بی‌ارزش شدن مهارت‌های انسانی است. فلوریدی در تبیین نظریه خود، در گام نخست، با نظر به دو اصل کرامت و شکوفایی انسان، تهدیدها و فرصت‌هایی که هوش مصنوعی برای این دو اصل انسانی فراهم می‌آورد برمی‌شمارد. اولین فرصت هوش مصنوعی خودشکوفایی[1] است. این که انسان بتواند تمامی ظرفیت‌ها و استعدادهای خودش را به منصه ظهور برساند. خودشکوفایی به معنای توانانی شکوفایی خصائص، علائق، توانایی‌ها و مهارت‌های بالقوهٔ، آرزوها و برنامه‌هایی است که انسان در زندگی خود در نظر دارد. اختراعاتی مانند ماشین لباسشویی، زنان را از مشقت‌های رختشویی رها کرد آنها فرصت یافتند وقت بیشتری برای تحصیل علم و اشتغال صرف کنند. اره برقی، مردان را از کار سخت روزانه برای هیزم‌شکنی آزاد ساخت. به همین شکل، کاربرد ابزارهای هوشمند می‌تواند وقت بیشتری را برای فعالیت‌های فرهنگی، فکری و اجتماعی در اختیار ما قرار دهد. در عین حال یک تهدید در مقابل این فرصت رخ می‌نماید. این تهدید هم در سطح فردی و هم در سطح اجتماعی وجود دارد. در سطح فردی، مهارت‌های قدیمی افراد، کم‌خاصیت شده و بازار کار این مهارت‌ها کساد می‌شود و شغل‌های قدیمی از بین می‌رود. برخی افراد جایگاه اجتماعی خود را مدیون شغلشان هستند. نجاری که می‌تواند مبلمانی منحصر به فرد بسازد یا تراشکاری که می‌تواند قطعه‌ای کمیاب بتراشد با صرف سالیان دراز شهرتی را در کسب و کار خود به‌دست آورده‌اند. وقتی این نوع کسب و کارها به حاشیه رود آن‌ها اعتبار شخصی خود را از دست می‌دهند و شاید سن و سالشان دیگر اقتضا نکند که هویت خویش را با فضای هوشمند جامعه بازسازی کنند. ممکن است بی‌کار و گوشه‌گیر

1. self realisation

شوند. همچنین در سطح اجتماعی، در برخی از مشاغل حساس مهارت‌هایی وجود دارد که اگر آن‌ها را به ابزارهای هوشمند بسپاریم و احتمال خطای آن ابزارها وجود داشته باشد، آسیب‌ها و خطرات آن برای جامعه انسانی جدی خواهد بود. مشاغل حساسی مانند تشخیص در حوزه سلامت و ناوبری هوایی نیازمند مهارت‌های حساس انسانی هستند. اگر این نوع کارها را به ماشین‌های هوشمند بسپاریم علاوه بر این که نقص آن‌ها ممکن است برای ما مخاطرات جدی به‌وجود آورد، رفته‌رفته فاقد مهارت انسانی در این حوزه‌های حساس خواهیم شد و این بسیار خطرناک است. اگر بخواهیم گسترش هوش مصنوعی را ترویج کنیم تا به توانایی‌ها و مهارت‌های جدید برسیم باید متوجه تأثیر آن بر توانایی‌ها و مهارت‌های قدیمی نیز باشیم و با مطالعه دقیق و طرح ایده‌هایی مثل حمایت از اقشار آسیب‌پذیر در این حوزه بتوانیم حتی المقدور انسجام بین‌نسلی بین افراد غیربرخوردار و برخورداران از این حوزه ایجاد کنیم تا بین آیندگان و پیشینیان انصاف رعایت گردد (Floridi, 2021a: 23).

از منظر اخلاق اسلامی انسان ظرفیتی عظیم درون خود نهفته دارد طوری که می‌تواند با شکوفایی ظرفیت‌های درون خود به کمال برسد و انسان کامل شود. شهید مطهری در کتاب انسان کامل اشاره می‌کند که انسان کامل در قران «امام» نامیده شده و این لقب حضرت ابراهیم علیه السلام است که پس از قبولی در آزمون‌های متعدد از جمله مبارزه با طاغوت زمان و تسلیم بودن در برابر فرمان‌های حق تعالی به این مقام رسیده است. مقام امامت مقام انسان کامل و الگوی مناسبی برای دیگران است. انسان کامل انسانی است که استعدادهای خود را به طور متوازن رشد داده است. عبادت، خدمت به خلق، آزادی از اسارت دیگران، عقل، عشق، محبت، عدالت و ارزش‌های دیگر در انسان کامل به طور متوازن رشد یافته است (مطهری، ۱۳۷۶الف، ج ۲۳: ۱۱۲ـ۱۱۹).

در برخی مکاتب اگزیستانسیالیسم انسان کامل به انسانی اطلاق شده که با اختیار و آزادی می‌تواند تمامی ظرفیت‌های خود را به فعلیت برساند. به نظر ژان پل سارتر، آزادی انسان اساس خودشکوفایی است زیرا انسان با آزادی انتخاب‌های خود، مسئول شکل دادن به زندگی‌اش است. اگرچه واقعیت‌های ناشی از شرایط بیرونی ممکن است افراد را محدود کند، اما نمی‌تواند فرد را مجبور کنند که یکی از گزینه‌های باقی‌مانده را نسبت به دیگری دنبال کند. از این نظر، فرد هنوز تا حدودی آزادی انتخاب دارد. به همین دلیل، یک فرد ممکن است راه غم‌بار و اندوهناکی را انتخاب کند و کاملاً آگاه باشد که انتخاب او چه عواقبی خواهد داشت. از نظر ژان پل سارتر، این ادعا که یکی از گزینه‌های پیش‌روی انسان اجتناب‌ناپذیر است، به این معناست که به جای یک عامل آزاد، نقش یک شیء را در جهان به عهده گرفته است. انسان‌ها از آن جهت مسئول هستند که آزادانه هستی خود را گزینش می‌کنند در حالی که میز و صندلی و اشیای دیگر مسئولیتی در قبال آنچه هستند، ندارند. اگر کسی تصور کند که به لحاظ شرایط و محدودیت‌ها، وجود ـ فی ـ نفسه او صرفاً واقعیتی محتوم است، آزادی خود را انکار کرده و دچار نگرش خود فریبی شده است. او این نوع خودفریبی را ریاکاری[1] می‌داند. به نظر او انسان نباید امور سلبی را بر خود مفروض بگیرد و باید بداند که می‌تواند بدون هیچ محدودیتی به آنچه می‌خواهد برسد. او می‌گوید: « اگر صراحت یا صداقت یک ارزش عام است، بدیهی است که قاعده «آدم باید همان باشد که هست» صرفاً به عنوان یک اصل تنظیم‌گر مفهومی و حکمی عمل نمی‌کند تا با آن آنچه هستم را بیان کنم. این اصل نه تنها آرمان دانایی بلکه آرمان هستی را طرح می‌کند. این اصل برای ما معادل هستی با خودش به عنوان نخستین الگوی هستی است. به این معنا لازم است که خودمان را آن چیزی که هستیم بسازیم.» (Sartre, 1993: 58–59).

این تحلیل از هستی انسانی همانست که دنیای غرب در قرن بیستم بدان روی آورد. در این رویکرد انسان، محور هستی قرار گرفته و خدا و تعالیم انبیا از زندگی انسان یا حذف شده یا به حاشیه رفته است. در حالی که از منظر اخلاق اسلامی انسان آزاد است که هر عملی را به انجام رساند در عین حال مؤظف است که حدود الهی را رعایت کند. به عبارت دیگر آزادی در اخلاق اسلامی محدود به خطوط قرمزی است که در اوامر و نواهی الهی ترسیم شده و در قرآن و سنت بیان شده است. هر کس از این

1. bad faith

حدود الهی تجاوز کند از مسیر عدالت خارج شده و شقاوت را برگزیده است.

از نظر اگزیستانسیالیست‌ها آزادی انسان محدودیت ندارد و اگر کسی این محدودیت را برای خود قائل شود وجود خویش را محدود ساخته است. درحالی که اسلام خودشکوفایی را طوری تفسیر می‌کند که در چارچوب سعادت انسان باشد و او را به شقاوت اخروی نکشاند. به همین جهت امثال سارتر در تلاشند تا ثابت کنند که هم‌جنس‌بازان که وجودشان را در آزادی خویش متبلور ساخته‌اند شایسته تحقق وجود انسانی هستند، در حالی که اگر انتخاب آن را انکار کنند یا آن را به دلیل امور ژنتیکی و مانند آن محصول یک سرنوشت اجتناب‌ناپذیر بدانند دچار ریاکاری شده‌اند. هم‌جنس‌گرایی یک انتخاب است همان طور که دگرجنس‌گرایان انتخاب خود را کرده‌اند (Sartre, 1993: 63–64).

پرسشی که طرح می‌شود این که هر چند از منظر اخلاق اسلامی نیز شقاوت امری اختیاری بوده و محتوم نیست ولی این به معنای خیر بودن این گزینش نیست. بنابر این برخلاف اگزیستانسیالیست‌ها که همه انتخاب‌ها را خیر و خوب می‌دانند اخلاق اسلامی، انتخاب انسان را به خیر و شرّ تقسیم می‌کند. انتخابی خوب است که به سعادت منجر شود و انتخابی شرّ است که به شقاوت بیانجامد. خداوند راه خیر و شرّ را به انسان نشان داده و او مختار است که طبق آن سعادت یا شقاوت خویش را برگزیند (الانسان:۳).

خودشکوفایی انسان اگر به معنای تحقق توانایی‌هایی باشد که به سعادت انسانی منجر می‌شود، مطلوب است ولی اگر صرفاً به معنای تحقق آنچه ممکن الحصول است باشد، خیر مطلق نیست و می‌تواند منجر به شقاوت انسان شود. بنابراین، اینکه چون ما می‌توانیم دستگاهی هوشمند بسازیم، خیر در آن است که به آرزویمان برسیم برای خیر بودن ساخت دستگاه هوشمند کافی نیست، بلکه باید نشان دهیم که این مطلب، از آن نوع خودشکوفایی است که به سعادت انسان می‌انجامد.

در عین حال از منظر اخلاق اسلامی پیشرفت‌های فناوری از جمله هوش مصنوعی فی نفسه امری خیر تلقی می‌شود چون فناوری برخاسته از دانش بشری است و دانش موجب اقتدار جامعه است. کسب دانش و اقتدار در اسلام مزیت است و در قران و سنت بر آن دو تأکید شده است. در قران بر کسب اقتدار برای آمادگی مقابله با دشمن تأکید شده (الانفال:۶۰) و دانشمندان را تنها کسانی برشمرده که نشانه‌ها و داستان‌های حکمت‌آموز الهی را درک می‌کنند (العنکبوت:۴۳؛ الروم:۲۲). علم و قدرت در اخلاق اسلامی حسن اقتضائی دارند. از آنجا که این دو صفت از اوصاف الهی نیز برشمرده شده است، تزین به این دو از این جهت نیز مستحسن است. همچنین شهید مطهری معتقد است که امر به طلب علم در اسلام شامل همه علوم است و تنها علوم دینی را شامل نمی‌شود (مطهری، ۱۳۷۶ب، ج : ۶۹۶-۶۹۷).

## فرصت دوم: ارتقای کنشگری انسان

فرصت دوم ارتقای کنشگری انسان در مقابل تهدید حذف مسئولیت انسان است. فلوریدی در گام دوم چنین بیان می‌کند که هوش مصنوعی منبع عظیم و روبه‌رشدی است که می‌تواند کنشگری انسان را به شدت افزایش داده، ارتقا دهد و سریع‌تر کند. اثرات این ارزش افزوده به حیات انسان را می‌توان با انقلاب صنعتی مقایسه کرد؛ زمانی که با اختراع موتور توانستیم کارهایمان را تسهیل کنیم، سرعت بخشیم و افزایش دهیم. کسانی که در صدر بهره‌مندی از فرصت‌ها و مزایای چنین کنشگری قرار گیرند، جامعه بهتری می‌سازند. اما باید به مسئولیت خودمان در قبال این فناوری نیز توجه کنیم. بسیار اساسی است که ببینیم چه نوع هوشمندی را گسترش می‌دهیم، چگونه از آن استفاده می‌کنیم، و آیا مزایا و منافع آن را با دیگران شریک می‌شویم. بدیهی است که خطر احتمالی در مورد بی‌توجهی به چنین مسئولیت‌هایی ممکن است نه تنها به این دلیل که ما چارچوب سیاسی ـ اجتماعی اشتباهی در جامعه خود داریم، بلکه علاوه بر آن، به این دلیل اتفاق بیفتد که هوش مصنوعی نوعی «جعبه سیاه» است که معلوم

نیست درون آن چه می‌گذرد و بشر قادر نیست درک کند که هوش مصنوعی از چه نظامی برای تصمیم‌سازی استفاده می‌کند. علاوه بر موارد معروفی چون تصادفات مرگبار وسایل نقلیه خودران، مواردی معمولی‌تر چون تصمیم در مورد آزادی مشروط یک متهم یا تعیین اعتبار برای وام به صورت هوشمند، نیز نگرانی‌های برای جامعه ایجاد می‌کند. ما می‌توانیم هنگام استفاده از هوش مصنوعی هم در عمل، مقدار کنشگری انسان را افزایش دهیم و هم با رعایت اصول اخلاقی کیفیت این کنشگری را ارتقاء دهیم. این موجب می‌شود رابطه برد ـ برد بین عمل و اخلاق ایجاد شود. گسترش هوش مصنوعی با مراعات همه جوانب، می‌تواند فرصت‌های جذابی را برای پیشرفت و افزونگی کنشگری انسان پیش‌کش کند. مثلاً می‌توانیم با روش‌های هوشمند نظام وام‌دهی همتا به همتا راه اندازی کنیم (Floridi, 2013, p. 269). کنشگری انسان می‌تواند با طراحی «چارچوب‌های تسهیل‌گر» موجب بهبود نتایج اخلاقی شده و با پالایش عملکرد دستگاه‌های هوشمند آن‌ها را پشتیبانی کند و گسترش دهد. اگر نظام‌های هوشمند به‌طور مؤثر طراحی شوند، می‌توانند نظام‌های اخلاقی مشترک را بسط دهند و آن‌ها را تقویت کنند (Floridi, 2021a: 23–24).

در نقد این کلام فلوریدی می‌توان گفت که هرچند انقلاب صنعتی موجب پیشرفت زندگی بشر شد، اما در عین حال با چالش‌های اجتماعی و اقتصادی زیادی همراه بود، از جمله انقلاب صنعتی موجب پیدایش نابرابری‌های اجتماعی عظیم و آسیب‌های زیست‌محیطی گسترده شده است. بنابراین، این نگرانی وجود دارد که هوش مصنوعی نیز ممکن است به مشکلات مشابهی دامن بزند. دیدگاه فلوریدی بیش از حد خوشبینانه است و نادیده گرفتن چالش‌های واقعی و خطرات ناشی از اتکای بیش از حد به فناوری‌های خودمختار می‌تواند منجر به عواقب ناگواری شود.

این که انسان در قبال چنین پیشرفت اعجاب‌آوری مسئولیت دارد امری روشن است، اما مسئله اساسی تعیین نوع مسئولیتی است که هر یک از ذی‌نفعان در این حوزه دارند. فلوریدی باید به این پرسش پاسخ دهد که چگونه می‌توان این مسئولیت را به طور مؤثر در عمل پیاده‌سازی کرد. به ویژه در جوامع با ساختارهای سیاسی و اجتماعی ناپایدار، به سختی بتوان اطمینان یافت که فناوری به نفع همه افراد جامعه توسعه می‌یابد.

همچنین «جعبه سیاه» بودن سیستم‌های هوش مصنوعی می‌تواند منجر به عدم شفافیت و در نتیجه عدم درک و کنترل انسانی شود. این عدم شفافیت می‌تواند به بی‌اعتمادی عمومی به فناوری‌های هوش مصنوعی منجر شود.

علاوه بر این امور، واگذاری بیش از حد کنشگری به سیستم‌های خودمختار می‌تواند به کاهش توانمندی‌های انسانی منجر شود و در آن نسل‌هایی کم‌کار، سنگین و کم‌تحمل تربیت یابند. نسل‌های پیشین همه بازوانی قوی داشتند تا بتوانند هیزم‌شکنی کنند، نسل کنونی سرانگشتانی ظریف دارند تا بتوانند صفحه کلید را به خوبی لمس کنند، نسل‌های آینده شاید حتی حوصله سخن گفتن و دستور دادن به ربات‌ها را هم نداشته باشند و همه چیز را خودکار و آماده بخواهند. عدم توجه به این آسیب‌ها می‌تواند نسل بشر را کسل کند و شادابی را از او بگیرد.

این ایده که با طراحی "چارچوب‌های تسهیل‌گر" می‌توان به بهبود نتایج اخلاقی کمک کرد، زمانی ایده مفیدی است که تلاش عملی برای تعیین و تبیین اجزاء مفهومی این چارچوب صورت گیرد. پرسش این است که چه نوع چارچوب‌هایی باید طراحی شوند و چگونه می‌توان اطمینان حاصل کرد که این چارچوب‌ها به درستی عمل می‌کنند.

### فرصت سوم: افزایش توانمندی‌های اجتماعی

فرصت سوم افزایش توانمندی‌های اجتماعی در برابر تهدید کاهش کنترل‌های انسانی است. فلوریدی در گام سوم گوشزد می‌کند که هوش مصنوعی فرصت‌های بی‌شماری را برای بهبود و افزایش قابلیت‌های فردی و اجتماعی در اختیار ما قرار می‌دهد. کاربرد فناوری‌های هوشمند، چه در پیشگیری و درمان بیماری‌ها و چه در بهینه‌سازی حمل و نقل و تدارکات، امکانات

بی‌شماری را برای بازآفرینی جامعه به‌دست می‌دهد و توانایی‌های بشر را به نحو بنیادینی ارتقا می‌دهد. هر چه بیشتر از هوش مصنوعی استفاده کنیم، بهتر می‌توانیم روابط خود را هماهنگ کنیم و در نتیجه اهداف بلندپروازانه‌تری در دسترس ما قرار می‌گیرد. هوش انسان اگر با هوش مصنوعی تقویت گردد راه‌حل‌های جدیدی برای مشکلات قدیمی و جدید پیدا می‌کند، از توزیع عادلانه‌تر یا کارآمدتر منابع گرفته تا رویکردی پایدارتر به مصارف. اما در عین حال چون چنین فناوری‌هایی ظرفیت بسیار قدرتمند و مخربی دارند، به همان اندازه خطراتی را نیز در پی دارند. اگر برای تقویت توانایی‌های خود، با روش‌های نادرست از فناوری‌های هوشمند استفاده کنیم، باید امیدوارم باشیم که وظایف مهم و بالاتر از همه، تصمیم‌گیری‌های خود را طوری به دستگاه‌های خودمختار واگذار کنیم که حداقل تا حدودی تابع نظارت و گزینش انسانی باشند. اگر نظارت‌ها و گزینش‌های انسانی کنار برود دیگر نمی‌توانیم از آسیب‌های احتمالی پیشگیری کنیم یا خطاهای محتمل را اصلاح کنیم. همچنین ممکن است روزبه‌روز کارهای بیشتری را به دستگاه‌های هوشمند بسپاریم و این کار خود موجب شود آسیب‌های احتمالی انباشته و ریشه‌دار شوند. با این توضیحات روشن می‌شود که ضرورت دارد بین دنبال کردن فرصت‌های بلندپروازانه و دستاوردهایی که هوش مصنوعی برای بهبود زندگی انسان در اختیار ما قرار می‌دهد و اطمینان به این که این تحولات عمده و اثرات آنها را می‌توانیم کنترل کنیم، تعادل ایجاد کنیم (Floridi, 2021a: 24).

یکی از نقدهایی که به این بیان فلوریدی وارد است، مربوط به نگرانی‌های اخلاقی و اجتماعی است که با حذف کنترل انسانی بر فرآیندهای تصمیم‌گیری پیش می‌آید. در حالی که فلوریدی به درستی به خطرات واگذاری تصمیمات به دستگاه‌های خودمختار اشاره کرده، ولی به‌طور کافی به ابعاد اجتماعی و اخلاقی این مسأله توجه نکرده است. حذف یا کاهش کنترل انسان‌ها بر تصمیم‌گیری‌های کلیدی می‌تواند منجر به آسیب‌های اجتماعی و بی‌اعتمادی به فناوری‌ها شود. این نگرانی‌ها به‌ویژه در رابطه با الگوریتم‌های مغایر با عدالت اجتماعی و تبعیض‌های نظام‌مند بیشتر است. رویکرد سرمایه‌داری امریکایی و تلاش برای دستیابی به رویای آن، موجب گردیده اختلاف طبقاتی شدیدی در جوامع غربی روی دهد که فقط بر فقیرترین فقرا تأثیر نمی‌گذارد، بلکه بر تعداد بی‌شماری از افراد کم‌درآمد تأثیر می‌گذارد، به نظر ویرجینیا یوبانکس، در عصر دیجیتال خودکارسازی نابرابری‌های اجتماعی و فاصله فقیر و غنی را افزون کرده، زیرا افزایش اتوماسیون نه تنها آن‌هایی را که در حال حاضر فقیر هستند بیشتر به حاشیه می‌برد، بلکه فقر را برای کسانی که در آستانه آن هستند به ارمغان می‌آورد (Eubanks, 2017; 6–13). در تأیید آنچه یوبانکس گفته باید اشاره کرد که طبق اطلاعات موجود ضریب جینی امریکا طی سال ۲۰۲۰ حدود ۴۸,۲ بوده که حدود ۹ واحد بیشتر از کشور ما ایران است («اقتصاد ایالات متحده آمریکا»، ۲۰۲۵).**

نقد دیگری که بر فلوریدی وارد است این که او هر چند به درستی به لزوم حفظ توازن و تعادل بین بهره‌برداری از فرصت‌های هوش مصنوعی و حفظ کنترل انسانی تأکید کرده، اما به‌طور مفصل‌تر به جزئیات و چالش‌های عملی در رسیدن به این تعادل نپرداخته و معلوم نیست که چگونه باید به آن دست یافت. منتقدان بر این باورند که این مفهوم از «تعادل» ممکن است در عمل به‌سادگی قابل پیاده‌سازی نباشد، زیرا در بسیاری از موارد، انسان‌ها عملاً کنترل و نظارت کامل بر فرآیندهای هوش مصنوعی نخواهند داشت و این مسئله به چالش‌های پیچیده‌ای منجر می‌شود که به راهکارهای حقوقی و اجتماعی عمیق، راهبردهای قوی و راهکارهای پیچیده نیاز دارد. نیک بوستروم به خوبی نشان داده که چگونه یک موجود ابرهوشمند می‌تواند حیات بشر را به مخاطره اندازد (Bostrom, 2014: 115–125).

اگر خط سیر هوش مصنوعی به سیستم‌هایی برسد که سطح هوش انسانی دارند، این سیستم‌ها خودشان توانایی توسعه سیستم‌های هوش مصنوعی را خواهند داشت و از سطح هوش انسانی پیشی می‌گیرند و فوق هوشمند می‌شوند. چنین سیستم‌هایی می‌توانند به سرعت خود را بهبود بخشند یا حتی سیستم‌های هوشمندتری را توسعه دهند. این چرخش شدید رویدادها پس از

رسیدن به هوش مصنوعی فوق‌هوشمند، موجب می‌شود توسعه هوش مصنوعی خارج از کنترل انسان باشد و پیش‌بینی آن دشوار گردد (Kurzweil 2005: 487). اگر ماشین فوق هوشمند را ماشینی تعریف کنیم که می‌تواند از تمام فعالیت‌های فکری هر انسانی هرقدر باهوش باشد، پیشی بگیرد، از آنجایی که طراحی ماشین‌ها یکی از این فعالیت‌های فکری است، یک ماشین فوق هوشمند می‌تواند ماشین‌های حتی بهتری طراحی کند. در آن صورت بدون شک یک «سونامی هوشمندسازی» رخ می‌دهد طوری که هوش انسان بسیار عقب خواهد ماند. آنگاه اولین ماشین فوق هوشمند آخرین اختراعی است که بشر باید انجام دهد، مشروط بر اینکه دستگاه آن‌قدر مطیع باشد که به ما بگوید چگونه آن را تحت کنترل داشته باشیم (Good, 2005: 33).

نقد دیگری که به فلوریدی وارد است این که مفهوم «کنترل انسانی» مفهومی مبهم و کش‌دار است و به روشن‌سازی بیشتری نیاز دارد. به‌ویژه، این پرسش مطرح می‌شود که آیا نظارت بر الگوریتم‌ها کافی است یا باید مداخلات انسانی در مراحل مختلف تصمیم‌گیری وجود داشته باشد؟ برخی منتقدان معتقدند که فلوریدی به‌طور دقیق به این پیچیدگی‌ها نپرداخته است. ما می‌توانیم نمونه خوبی از پرداخت به مفهوم کنترل در کتاب بوستروم ببینیم جایی که او به مشکلات ناشی از کنترل پرداخته و ظرفیت‌های روش‌های کنترل را برشمرده است (Bostrom, 2014: 127–143).

اشکال دیگر این است که فلوریدی فرض می‌کند که استفاده از هوش مصنوعی همیشه می‌تواند به نفع اجتماع باشد، در حالی که ممکن است در عمل، این‌طور نباشد. جیمی ساسکایند در سه فصل از کتابش توضیح می‌دهد که در بسیاری از موارد، هوش مصنوعی می‌تواند منجر به نابرابری‌های اجتماعی و اقتصادی شود و به جای کمک به جامعه، تهدیدهایی برای عدالت و دسترسی عادلانه به منابع ایجاد کند (Susskind, 2018, pp. 257–312).

## فرصت چهارم: تعمیق انسجام اجتماعی

فرصت چهارم تعمیق انسجام اجتماعی در برابر تهدید تحلیل رفتن توان تعیین سرنوشت است. در گام چهارم عمق‌بخشی به انسجام و اتحاد اجتماعی فرصتی برای انسجام و وحدت اجتماعی برشمرده شده و در مقابل آن این تهدید طرح شده که اگر مانعی برای پیشرفت هوش مصنوعی ایجاد گردد ممکن است توان تعیین سرنوشت ما تحلیل رفته یا از بین برود.

تنها در صورتی می‌توان با موفقیت با گرفتاری‌هایی که اینک جهان با آن درگیر است، مقابله کرد که همه ذینفعان از دولت‌ها، سیاست‌گذاران و نمایندگان مناطق آسیب‌پذیر گرفته تا دانشمندان، فناوران و همه گیتی‌نشینان، در طراحی راه‌حل‌های آن مشارکت و حس مالکیت مشترک داشته باشند و برای تحقق آن‌ها همکاری کنند. حل گرفتاری‌های بزرگ جهانی مثل تغییرات اقلیمی، اشاعه هسته‌ای، مقاومت بدن در برابر ضدمیکروبها و سخت‌اندیشی دینی، به انسجام، تشریک مساعی و هماهنگی‌های پیچیده‌ای نیاز دارد. هوش مصنوعی می‌تواند با استفاده از راه‌حل‌های مبتنی بر ابرداده‌ها و راه‌حل‌های ریاضی‌وار (الگوریتمی) به کاهش گرفتاری‌های عالمگیر کمک کند. وقتی راه‌حل چنین مشکلاتی نیازمند مشارکت جهانی است، تعیین و هماهنگی بین اقدامات محلی، ملی و بین‌المللی کار پیچیده‌ای می‌شود و هوش مصنوعی به خوبی می‌تواند راهگشا باشد. با این حال، خطر این است که نظام‌های هوش مصنوعی ممکن است خودتعیین‌گری انسان را از بین ببرند و استقلال رأی را از او بگیرد و حق تعیین سرنوشت او را تضعیف کند. وقتی مردم به ماشین‌های خودکار و نظام‌های هوشمند عادت کنند، ممکن است ناخواسته مبتلا به تغییری ناخوشایند در سبک زندگی خود گردند. قدرت پیش‌بینی هوش مصنوعی که بی‌رحمانه ما را به سوی خود سوق می‌دهد، می‌تواند کرامت انسانی و شکوفایی ما را تضعیف کند. ما نباید استقلال و انسجام اجتماعی خود را دچار چنین تهدیدی کنیم. در نهایت او بیان می‌کند که ایجاد توازن بین فرصت‌ها و تهدیدهای چهارگانه مذکور و استفاده از فرصت‌ها در کنار پیش‌بینی و کاهش تهدیدهای پیش‌رو یا اجتناب از آن‌ها، احتمال توفیق فناوری‌های هوش مصنوعی را برای ارتقای

کرامت و شکوفایی انسان افزایش می‌دهد (Floridi, 2021a: 24–25).

مشخص نیست که فلوریدی چه معنایی را از «تعیین سرنوشت» و «استقلال» اراده می‌کند. عدم شفافیت در تعریف این مفاهیم می‌تواند موجب برداشت‌های متفاوت و حتی اغراق‌آمیز از خطرات احتمالی هوش مصنوعی گردد. همچنین ابعاد این مفاهیم نیز روشن نیست که آیا منظور استقلال اجتماعی، یا سیاسی یا فرهنگی یا فردی یا همه این موارد است و این که در چه حالت یا شرائطی این استقلال از دست رفته است. اگر ما با دست خودمان آنها را مجهز با الگوریتمی کنیم که اختیار را از خودمان سلب کند، مستقل عمل کرده‌ایم و سرنوشت خودمان را خودمان تعیین کرده‌ایم یا سرنوشت خودمان را به‌دست هوش مصنوعی داده‌ایم و دیگر استقلالی نداریم؟ اگر هنگامی که سوار ماشین می‌شویم راننده خودکار را روشن کنیم و در ماشین بخوابیم، سرنوشت خود را تعیین کرده و مستقل عمل کرده‌ایم یا حق تعیین سرنوشت و استقلال را از خود سلب کرده‌ایم؟ اگر به عملمان توجه کنیم مستقل عمل کرده‌ایم ولی اگر نتیجه را ببینیم استقلال را از خود سلب کرده‌ایم. حال پرسش این است که فلوریدی کدام را منظور کرده است. این نکات در کلام فلوریدی به ابهام رها شده است. به همین دلیل برخی از نویسندگان متأخر تلاش کرده‌اند تا معنای استقلال را توصیف کنند و ابعاد آن را تعیین کنند (IEEE, 2017).

همچنین باید دانست که از نگاه اخلاق اسلامی انسان هر چند آزاد و مختار است که آینده خود را بسازد ولی این آزادی و استقلال از نگاه دینی محدود به حدود الهی است و نباید از اصول و چارچوب‌های دینی خارج گردد. هر دستکاری صنعتی در طبیعت که تخریب جامعه و طبیعت را در پی داشته باشد هر چند به دلیل حس استقلال انسانی تحقق یابد مذموم است و بشر باید از آن اجتناب کند. به عبارت دیگر انسان آزاد است تا انتخاب کند ولی این آزادی با مسئولیتی همراه است. عدالت و نیکوکاری و نیز پرهیز از فحشاء (محرمات الهی)، از منکر (زشتی‌هایی که عقلا نمی‌پسندند)، و از ظلم به خود، دیگران و به طبیعت از جمله این مسئولیت‌ها هستند (النحل:۹۰).

اگر روزی بیاید که همه ذینفعان از دولت‌ها و نمایندگان آسیب‌پذیر تا دانشمندان و فناوران و همه مردم جهان با هم مشارکت کنند و به عنوان راه‌حلی برای مقابله با چالش‌های جهانی، دست به دست یکدیگر دهند، دنیا مدینه فاضله‌ای می‌شود و قاعدتاً خلیفه راستین الهی حاکم چنین جهانی خواهد بود، اما قبل از این که آن زمان فرارسد، از منظر عملی و سیاسی، تحقق چنین مشارکت همگانی و هماهنگ در دنیای چندقطبی و پیچیده امروزی با چالش‌های جدی روبه‌روست. این آرمان‌گرایی غیرعملی یکی دیگر از اشکال‌هایی است که به کلی‌گویی فلوریدی وارد است. ما نیازمند تدوین اصول و چارچوب‌هایی هستیم که با آن بتوان نظارت کافی بر رفتار ذی‌نفعان اعمال کرد و آن‌ها را در قبال اعمالشان، مسئول و پاسخگو نمود. همچنین باید راهکارهای مناسبی برای کنترل این فناوری در راستای ارتقای انسجام اجتماعی ارائه داد. تأسیس نهادهای تنظیم‌گر محلی، ملی و بین‌المللی از جمله این راهکارهاست که در اینجا به آن اشاره نشده است.

فلوریدی معتقد است برای رسیدن به نتایج اجتماعی مطلوب، باید هم‌زمان ضمن جلوگیری از سوء استفاده (misuse) از هوش مصنوعی به ترس بی‌دلیل که موجب کم‌استفاده‌کردن (underuse) از فناوری‌های هوش مصنوعی شود دچار نشویم. همچنین هر چند رعایت قوانین وضع شده برای مهار هوش مصنوعی شرطی ضروری است، ولی قوانین و مقررات حداقلِ لازم را تعیین می‌کند و کافی نیست، اما رویکرد اخلاقی مزایایی فراتر از الزامات قانونی ارائه می‌دهد. ممکن است شما قوانین شطرنج را بدانید ولی نتوانید خوب بازی کنید و برنده شوید. ما باید هم قوانین را بدانیم هم در مواجهه با هوش مصنوعی برنده بازی باشیم و این برنده شدن چیزی فراتر از علم به قوانین آنست. باید فرصت‌هایی که از نگاه اجتماعی مطلوب است شناسایی و از آن بهره‌برداری کنیم و از اقدامات پرهزینه و غیرمطلوب اجتماعی پیشگیری کنیم حتی اگر قوانین آن‌ها را تجویز کند. رویکرد اخلاقی باعث می‌شود ترس از اشتباه موجب نشود که فرصت‌های بالقوه هوش مصنوعی از دست برود و در عین حال اخلاق می‌تواند همچون

سامانه هشدار اولیه عمل کند و از بروز خطرات بالقوه‌ای که ممکن است کل کار را تهدید کند، جلوگیری نماید. این مزیت دوگانه (استفاده از فرصت و دوری از تهدید) تنها در صورتی محقق می‌شود که بتوانیم در سازمان‌ها و جوامع مختلف، اعتماد عمومی را جلب کنیم و مسئولیت‌ها را به طور شفاف مشخص کنیم. برای جلب اعتماد عمومی باید منافع واقعی و ملموسی را ترسیم کنیم و از خطرات آن پیشگیری کنیم با راه‌هایی چون بیمه یا پرداخت غرامت خطرات ناشی از آن را کاهش دهیم یا جبران کنیم (Floridi, 2021a: 25–26).

## بررسی اسناد شش‌گانه

فلوریدی و همکارانش با ارزیابی شش سند که در مورد هوش مصنوعی ارائه گردیده بود، تلاش کردند تا ۲۰ توصیه اصلی را از درون این اسناد استخراج کنند. اولین سند اصولی است که در ژانویهٔ ۲۰۱۷ در کنفرانسی در زمینهٔ هوش مصنوعی در مکانی به اسم آسیلومار در کالیفرنیا تدوین گردید. در این کنفرانس با عنوان «کنفرانس آسیلومار دربارهٔ هوش مصنوعی مفید» پیشگامان هوش مصنوعی، دانشمندان، پژوهشگران و اندیشمندان، دربارهٔ آیندهٔ هوش مصنوعی و تأثیرات آن بر جامعه بحث و تبادل نظر کردند. هدف اصلی این اصول اطمینان از آن بود که هوش مصنوعی در خدمت بهبود زندگی بشری باشد و از خطرات احتمالی آن کاسته شود. در نتیجهٔ این گردهمایی مجموعه‌ای از ۲۳ اصل تدوین شد که راهنمای اخلاقی و عملی برای توسعهٔ هوش مصنوعی با عنوان «اصول آسیلومار برای هوش مصنوعی» بود. این اصول به موضوعاتی مانند امنیت هوش مصنوعی، شفافیت، مسئولیت‌پذیری، حقوق و آزادی‌های بشر و تضمین بهره‌مندی همگان از مزایای فناوری تأکید دارد. این اصول به عنوان یک نقطه عطف در مسیر شکل‌گیری چارچوب‌های اخلاقی و ایمنی در تحقیقات هوش مصنوعی شناخته می‌شود و تأثیر قابل توجهی در سیاست‌گذاری و روند پژوهش‌های آینده داشته است. امضای این اصول به قلم بیش از ۵۷۰۰ تن از پژوهشگران و متخصصان برجسته، آن را به یک مرجع اخلاقی در حوزه هوش مصنوعی تبدیل کرده است (Future of Life Institute, 2017).

دومین سند «بیانیه مونترال برای هوش مصنوعی مسئول» است که در نوامبر ۲۰۱۷ تحت نظارت دانشگاه مونترال تولید گردید. بیانیه مونترال، بخشی از تلاش جهانی به هدف ایجاد یک چارچوب اخلاقی برای توسعه و به‌کارگیری فناوری‌های هوش مصنوعی، با تمرکز بر حفظ کرامت انسانی، عدالت اجتماعی، و پایداری محیط زیست است. این سند، برخلاف برخی چارچوب‌های صرفاً دانشگاهی یا صنعتی، به دلیل مشارکت گسترده دانشگاهیان، متخصصان، شهروندان، و نهادهای مدنی و تأکید بر عدالت اجتماعی، از جایگاه ویژه‌ای در مباحثات بین‌المللی برخوردار است و نمونه‌ای از فرآیند مشورتی باز (open consultation) برای تدوین اصول اخلاقی در فناوری‌های نو است. این سند دیدگاهی انسان‌محور به توسعه فناوری دارد. در این سند ده اصل برای هوش مصنوعی مسئول ارائه گردیده که عبارتند از: رفاه انسان‌ها، احترام به خودمختاری، حفظ حریم خصوصی، پایداری، دموکراسی، عدالت و انصاف، شفافیت، امنیت، مسئولیت‌پذیری، و پایداری محیط زیست (Université de Montréal, 2018).

سومین سند با عنوان «طراحی هم‌راستا با اخلاق: چشم‌اندازی برای اولویت‌دادن به رفاه انسانی در سامانه‌های خودران و هوشمند» توسط ابتکار جهانی «انجمن بین‌المللی مهندسان برق و الکترونیک» (IEEE) درباره اخلاق در سامانه‌های هوشمند منتشر شد. این سند، که نسخه دوم آن در دسامبر ۲۰۱۷ منتشر گردید، حاصل تلاش جمع‌سپاری‌شده بیش از ۲۵۰ اندیشمند از سراسر جهان است و هدف آن ارائه اصول و توصیه‌هایی برای گسترش اخلاق در سامانه‌های هوشمند و خودران بوده است (IEEE, 2017). در این سند، مجموعه‌ای از اصول کلی درباره هدایت‌های اخلاقی برای توسعه فناوری‌های هوش مصنوعی ارائه شده که به طور خاص بر اولویت‌بخشیدن به رفاه انسانی، شفافیت، مسئولیت‌پذیری و قابلیت پاسخگویی، آگاهی از احتمال سوءاستفاده،

توانمندسازی، کنترل داده‌ها، و بهزیستی اجتماعی و زیست‌محیطی تأکید دارند (IEEE, 2017: 12–15).

چهارمین سند اصول اخلاقی ارائه شده در *بیانیه هوش مصنوعی، رباتیک و سیستم‌های «خود مختار»*، از سوی گروه اروپایی کمیسیون اروپا در زمینه اخلاق در علم و فناوری‌های جدید، در ۹ مارس ۲۰۱۸ در بروکسل منتشر گردید. این بیانیه فهرستی از اسناد پیشین را نام برده و در عین حال بیان می‌دارد که خطرات ناشی از رویکردهای ناهماهنگ و نامتعادل در زمینه مقررات‌گذاری هوش مصنوعی و فناوری‌های «خودمختار» و نیز وجود مقررات پراکنده و جزیره‌ای می‌تواند به پدیده‌ای به نام «گزینش دلبخواهی اخلاقی» منجر شود؛ به این معنا که توسعه و استفاده از هوش مصنوعی به مناطقی منتقل شود که استانداردهای اخلاقی پایین‌تری دارند. اگر اجازه دهیم این بحث‌ها تحت سلطه برخی مناطق، رشته‌های علمی، گروه‌های جمعیتی یا بازیگران صنعتی خاص قرار گیرد، این خطر وجود دارد که منافع و دیدگاه‌های گسترده‌تر اجتماعی نادیده گرفته شوند. همچنین، گفت‌وگوهای کنونی گاهی فاقد چشم‌اندازی جامع نسبت به فناوری‌های «خودمختار» هستند که احتمالاً در دهه آینده مورد مطالعه، توسعه و اجرا قرار خواهند گرفت و همین امر باعث ایجاد نقطه کور در پیش‌بینی‌های مقرراتی می‌شود. این بیانیه مجموعه‌ای از اصول اخلاقی بنیادین را پیشنهاد می‌کند که مبتنی بر ارزش‌های مندرج در معاهدات اتحادیه اروپا و منشور حقوق بنیادین اتحادیه اروپاست و می‌توان آن را راهنمای توسعه این فناوری‌ها دانست. مبتنی بر این نکات، این سند شش اصل را ذکر کرده و هر یک را تعریف می‌کند که عبارتند از: الف) کرامت انسانی، ب) خودمختاری، ج) مسئولیت‌پذیری، د) عدالت، انصاف و همبستگی هـ) مردم‌سالاری، و) حاکمیت قانون و پاسخگویی، ز) امنیت، ایمنی، تمامیت جسمی و روانی، ح) حفاظت از داده‌ها و حریم خصوصی و (ط) پایداری. پس از توصیف هر یک از این مفاهیم در این بیانیه چنین آمده است: «هوش مصنوعی، رباتیک و سامانه‌های «خودمختار» می‌توانند در صورت طراحی و به کارگیری خردمندانه، رفاه به ارمغان آورند، به بهزیستی کمک کنند و در دستیابی به آرمان‌های اخلاقی اروپا و اهداف اجتماعی-اقتصادی آن یاری‌رسان باشند. ملاحظات اخلاقی و ارزش‌های مشترک می‌توانند ابزاری برای شکل دادن به دنیای فردا باشند و باید به‌عنوان محرک و فرصت‌هایی برای نوآوری تلقی شوند، نه به‌عنوان مانع‌ها و سدها. گروه اخلاق اروپا (EGE) از کمیسیون اروپا می‌خواهد بررسی کند که کدام ابزارهای حقوقی موجود می‌توانند به‌طور مؤثر با مسائل مطرح‌شده در این بیانیه رسیدگی کنند و آیا به ابزارهای جدید در حوزه حکمرانی و مقررات‌گذاری نیاز است یا خیر. گروه اخلاق اروپا همچنین خواستار آغاز فرآیندی است که راه را برای دستیابی به یک چارچوب اخلاقی و حقوقی مشترک و دارای اعتبار بین‌المللی برای طراحی، تولید، استفاده و حکمرانی بر هوش مصنوعی، رباتیک و سامانه‌های «خودمختار» هموار سازد ( European Group on Ethics in Science and New Technologies to the European Commission, 2018:13–20).

پنجمین سند گزارشی است که کمیته هوش مصنوعی مجلس اعیان بریتانیا با عنوان «هوش مصنوعی در بریتانیا:آماده، مایل و قادر؟» در آوریل ۲۰۱۸ منتشر کرده است. نخستین بررسی جامع و راهبردی مجلس اعیان درباره هوش مصنوعی بود و نقش مهمی در ترسیم سیاست‌های اخلاقی، حقوقی و اقتصادی این فناوری در بریتانیا ایفا کرد. این سند ۱۸۳ صفحه‌ای حاصل کار کمیته منتخب هوش مصنوعی است و تلاش می‌کند تصویری روشن از فرصت‌ها، چالش‌ها و الزامات پیش روی کشور در مواجهه با تحولات سریع هوش مصنوعی ارائه دهد. در این گزارش بریتانیا ظرفیت‌های علمی، صنعتی و اخلاقی لازم را برای پیشرو بودن در عرصه هوش مصنوعی عرضه کرده و سیاست‌گذاری هوشمندانه، حمایت دولتی پایدار و نظارت مؤثر را شرط آن دانسته است. در متن سند به مسائل کلیدی همچون آموزش و مهارت‌آموزی در دوره هوش مصنوعی، ضرورت تدوین قوانین و مقررات متناسب، شفافیت و مسئولیت‌پذیری شرکت‌های فعال در این حوزه، چالش‌های حریم خصوصی و حفاظت از داده‌ها، و ضرورت هماهنگی بین نهادهای ملی و بین‌المللی برای شکل‌دهی به آینده هوش مصنوعی پرداخته شده است.

یکی از نقاط برجسته گزارش، ارائه پنج اصل اخلاقی برای «اصول راهنمای هوش مصنوعی» است که در بند ۴۱۷ ذکر شده که عبارتند از:

هوش مصنوعی باید در راستای خیر عمومی و منفعت بشریت توسعه یابد.

هوش مصنوعی باید بر اصول شفافیت (قابل فهم بودن) و عدالت پایه‌ریزی شود.

هوش مصنوعی نباید حقوق داده یا حریم خصوصی افراد، خانواده‌ها یا جوامع را کاهش دهد.

همه شهروندان حق آموزش دارند تا بتوانند از نظر روحی، عاطفی و اقتصادی در کنار هوش مصنوعی شکوفا شوند.

قدرت مستقل برای آسیب رساندن، نابود کردن یا فریب انسان‌ها نباید هرگز در اختیار هوش مصنوعی قرار گیرد.

در مجموع، این سند نقشه راهی است برای بهره‌گیری مسئولانه و اخلاق‌مدار از هوش مصنوعی به نفع جامعه (Great Britain, 2018: 125).

آخرین سند متنی است با عنوان «اصول مشارکت در هوش مصنوعی، یک سازمان چندجانبه متشکل از دانشگاهیان، محققان، سازمان‌های جامعه مدنی، شرکت‌های سازنده و استفاده‌کننده فناوری هوش مصنوعی، و سایر گروه‌ها» که در سال ۲۰۱۸ منتشر شده است. «مشارکت در هوش مصنوعی» (Partnership on AI) نام یک سازمان است که در سال ۲۰۱۶ تأسیس شد، نهادی چندجانبه و حقوق‌بنیان برای توسعه مسئولانه هوش مصنوعی. اعضای اصلی آن شامل دانشگاه‌ها، شرکت‌های فناوری (مانند آمازون، سونی، IBM، متا)، سازمان‌های جامعه‌مدنی (مثل PolicyLink، First Draft)، و گروه‌های تحقیقاتی است. اهداف و اصول کلیدی این سند عبارت بود از الف) تمرکز بر منافع عمومی: تضمین اینکه فناوری‌های هوش مصنوعی به نفع همه عموم افراد طراحی و اجرایی شوند و همه جامعه بشری را توانمند ساز، ب) عدم‌ضرر: به معنای محدودسازی خطرات محتمل، جلوگیری از استفاده سوء یا تجاوز به حریم خصوصی افراد، و تضمین امنیت فناوری، ج) خودمختاری انسانی: یعنی حفظ کنترل و تصمیم‌گیری انسان در برابر سیستم‌های هوشمند و عدم انتقال کامل اختیار به ماشین‌ها، د) عدالت و حذف تبعیض: یعنی تلاش برای کاهش سوءگیری و تبعیض در داده‌ها و الگوریتم‌ها و تضمین دسترسی برابر به فرصت‌های ایجادشده در مورد هوش مصنوعی، هـ) شفافیت و پاسخگویی: یعنی اطمینان از اینکه تصمیمات سیستم‌های هوشمند قابل توضیح، بازبینی یا اعتراض هستند، و) امنیت و ایمنی: یعنی اعمال ملاحظات ایمنی در طول چرخه تأمین محصولات و جلوگیری از سوءاستفاده یا خطرات بالقوه، ز) پایداری: یعنی در نظر گرفتن تأثیرات محیطی و اجتماعی بلندمدت و هم‌سو بودن با حقوق بشر و دموکراسی (Partnership on AI, 2018).

## تدوین پنج اصل اخلاقی برای کاربرد هوش مصنوعی در جامعه

در مجموع، این شش سند ۴۷ اصل را مطرح کرده‌اند. بسیاری از این اصول با هم همسانی یا همپوشانی دارد. فلوریدی معتقد است می‌توان این اصول را با چهار اصل محوری که معمولاً در اخلاق زیستی ذکر می‌شود مقایسه کرد: اصل خیرخواهی، اصل عدم ضرر، اصل خودمختاری و اصل عدالت. اخلاق زیستی یکی از حوزه‌هایی است که بیشترین شباهت را به اخلاق دیجیتال در برخورد اکولوژیکی با اشکال جدید عوامل، بیماران و محیط‌ها دارد (فلوریدی ۲۰۱۳). او چهار اصل اخلاقی زیستی را به طرز شگفت انگیزی با چالش‌های اخلاقی جدیدی که هوش مصنوعی ایجاد می‌کند، سازگار می‌داند، اما بیان می‌کند که این اصول جامع نیستند. او بر اساس تحلیل تطبیقی خود استدلال می‌کند که علاوه بر این چهار اصل، به یک اصل جدید دیگر نیز نیاز است که آن را «اصل توضیح‌پذیری» می‌نامد. این اصل شامل معقولیت و پاسخگویی می‌شود.

در اینجا ابتدا به توضیح هر یک از این اصول از نگاه فلوریدی می‌پردازیم و سپس هر یک را نقد و بررسی می‌کنیم:

اصل اول خیرخواهی است که در ضمن آن ارتقای رفاه، حفظ کرامت و حفظ سیاره نیز گنجانده شده است. فلوریدی بیان می‌کند که از میان چهار اصل اخلاق زیستی، اصل «خیرخواهی» یا «نیکوکاری» شاید روشن‌ترین اصلی باشد که می‌توان آن را در میان شش مجموعه اصولی که در اینجا مورد ترکیب و بازخوانی قرار گرفته‌اند مشاهده کرد. اصل اخلاقی خیرخواهیِ هوش مصنوعی به روش‌های مختلفی بیان می‌شود، و معمولاً در بالای هر فهرست از اصول قرار می‌گیرد. او سپس ارجاع‌هایی را به متون شش‌گانه می‌دهد و با استناد به آن‌ها «اصل خیرخواهی» را از تمامی آن‌ها استخراج می‌کند. ارتقای بهزیستی و رفاه، نفع بشریت و خیر عمومی، توانمندسازی و سودرسانی به بیشترین افراد، او همچنین به مفهوم «کرامت انسانی» و «پایداری» در این اسناد اشاره می‌کند که می‌توان مفاد آن‌ها را در اصل خیرخواهی مشاهده کرد. پایداری به معنای تأکید بر حفظ حیات در سیاره زمین و تداوم شکوفایی بشر برای نسل‌های آینده، از جمله تعبیرهایی است که از اسناد شش‌گانه قابل تطبیق بر اصل خیرخواهی است (Floridi, 2021b: 28).

اصل دوم اصل منع ضرر یا عدم اضرار است که اصل حریم خصوصی، امنیت و احتیاط در گسترش توانمندی‌ها را نیز دربرمی‌گیرد. ممکن است به نظر رسد که وقتی اصل خیرخواهی را ذکر می‌کنیم دیگر نیازی نیست که به «اصل منع ضرر» یا به تعبیر دیگر، «اصل آسیب نرساندن» نیز اشاره کنیم. چون از نظر منطقی این دو بر هم منطبقند و بر ارتقای بهزیستی، بهره‌مندی مشترک و تقویت خیر عمومی تأکید دارند. فلوریدی معتقد است که در عین حال، اصل منع ضرر به پیامدهای بالقوه منفی هوش مصنوعی و سوء استفاده از آن اشاره دارد که هر یک از اصول شش‌گانه به نوعی به آن تصریح کرده‌اند. یکی از ضررهای ناشی از پذیرش هوش مصنوعی، نقض حریم خصوصی اشخاص است که باید از آن اجتناب کرد. گرچه رعایت حریم خصوصی به عنوان یک اصل مستقل در اسناد شش‌گانه از جمله سند IEEE قابل مشاهده است ولی می‌توان آن را در دل «اصل منع ضرر» گنجانید. یک خطر دیگر، تهدیدِ ناشی از رقابت تسلیحاتی و خودبهبوددهیِ اصلاحی در حوزه هوش مصنوعی است. باید نسبت به «حد اعلای قابلیت‌های آینده هوش مصنوعی» «احتیاط» پیشه کرد. اصل منع ضرر این مورد را نیز شامل می‌شود. در عین حال فلوریدی معتقد است که اصل منع ضرر در برخی جهات ابهام دارد. در اینجا معلوم نیست که متعلق منع ضرر کیست، آیا انسان سازنده دستگاه است یا دستگاه هوشمندی که به وسیله انسان ساخته می‌شود. به همین جهت برخی از اسناد مانند اعلامیه مونترال به جای تأکید بر «منع ضرر» بر «مسئولیت‌پذیر ساختن» توسعه‌دهندگان هوش مصنوعی تأکید دارند. ابهام دیگر به نیت عامل بازمی‌گردد. ترویج اصل عدم ضرر می‌تواند هم شامل پیشگیری از آسیب‌های تصادفی و غیرعمدی «حد اعلای هوشمندی ماشین» که پیش‌بینی‌نشده باشد و هم شامل پیشگیری از آسیب‌های عمدی و «سوءاستفاده». اصل عدم ضرر هر دو را به طور یکسان شامل می‌شود (Floridi, 2021b: 28–29).

اصل سوم از اصول کلاسیک در اخلاق زیستی خودمختاری به معنای قدرت تصمیم‌گیری و اختیار است؛ یعنی این ایده که افراد حق دارند درباره رفتاری که انجام می‌دهند یا نمی‌دهند، خود تصمیم بگیرند. در زمینه پزشکی، این اصل خودمختاری اغلب زمانی مختل می‌شود که بیماران فاقد ظرفیت ذهنی برای تصمیم‌گیری در راستای منافع خود باشند. در چنین شرایطی خودمختاری ناخواسته به دیگری واگذار می‌شود. نزد فلوریدی در زمینهٔ هوش مصنوعی، وضعیت پیچیده‌تر است: هنگامی که ما هوش مصنوعی و عاملیت هوشمند آن را به کار می‌گیریم، بخشی از قدرت تصمیم‌گیری خود را آگاهانه به ماشین‌ها واگذار می‌کنیم. بنابراین، تأیید اصل خودمختاری در زمینه هوش مصنوعی به معنای یافتن توازن میان قدرت تصمیم‌گیری‌ای است که برای خود حفظ می‌کنیم و آن بخشی که به عاملان مصنوعی واگذار می‌کنیم. اصل خودمختاری در چهار سند از شش سند به صراحت بیان شده است. طبق این اصل توسعه هوش مصنوعی باید به ارتقای خودمختاری همه انسان‌ها کمک کند و ... خودمختاری نظام‌های رایانه‌ای را کنترل نماید. این انسان‌ها هستند که باید انتخاب کنند چگونگی و اینکه آیا تصمیم‌ها را به

سامانه‌های هوش مصنوعی واگذار کنند؛ آن هم برای تحقق اهدافی که خود برگزیده‌اند. سامانه‌های خودمختار نباید آزادی انسان‌ها را برای تعیین معیارها و هنجارهای خود و زیستن بر اساس آن‌ها تضعیف کنند و قدرت خودمختار برای آسیب رساندن، نابود کردن یا فریب دادن انسان‌ها هرگز نباید به هوش مصنوعی واگذار شود. فلوریدی می‌گوید آنچه در اینجا بیش از همه اهمیت دارد، چیزی است که می‌توان آن را «فراخودمختاری» (meta-autonomy) یا الگوی «تصمیم‌گیری برای واگذاری تصمیم» نامید: انسان‌ها باید همواره قدرت این را داشته باشند که خود تصمیم بگیرند کدام تصمیم‌ها را خود اتخاذ کنند، آزادی عمل برای انتخاب در مواقع ضروری را حفظ نمایند، و تنها در مواردی که دلایل قاطع (مانند کارآمدی) اقتضا کند، اختیار تصمیم‌گیری را به ماشین‌ها بسپارند. البته هر نوع واگذاری باید از اساس قابلیت بازبینی و لغو داشته باشد (یعنی امکان تصمیم‌گرفتن دوباره در مورد تصمیم‌گیری قبلی) (Floridi, 2021b: 29–30).

اصل چهارم رعایت عدالت، ارتقای شکوفایی و حفظ همبستگی است. از آنجا که ظرفیت تصمیم‌گیری (تصمیم‌گرفتن، و دوباره تصمیم‌گرفتن) در جامعه به‌طور برابر توزیع نشده است، نیازمند اصل دیگری هستیم تا این نقیصه را جبران کند. آخرین اصل از چهار اصل کلاسیک اخلاق زیستی، عدالت است؛ اصلی که معمولاً در ارتباط با توزیع منابع از جمله دسترسی به درمان‌های جدید یا محدود طرح می‌شود. به نظر فلوریدی توسعه هوش مصنوعی باید عدالت را ارتقا دهد و در جهت حذف همهٔ صور تبعیض گام بردارد. هوش مصنوعی باید به عدالت جهانی و دسترسی برابر به منافع» فناوری‌های هوش مصنوعی یاری رساند. سوءگیری در داده‌هایی که برای آموزش سامانه‌های هوش مصنوعی به کار می‌روند خطری است که باید از آن اجتناب کرد. نظام‌های حمایت متقابل مانند بیمه‌های اجتماعی و خدمات درمانی نباید تضعیف شوند. هوش مصنوعی نباید مانع شکوفایی شهروندان از نظر ذهنی، عاطفی و اقتصادی گردد. همه باید از هوش مصنوعی برای اصلاح خطاهای گذشته، مانند حذف تبعیض‌های ناعادلانه بهره‌مند گردند. همچنین باید از آسیب‌های جدید مانند تضعیف ساختارهای اجتماعی موجود جلوگیری کرد و همبستگی اجتماعی را حفظ نمود (Floridi, 2021b: 30–31).

ما در اینجا از نقد این چهار اصل صرف‌نظر می‌کنیم و خواننده را به مقالات و کتب مربوط به اصول اخلاق پزشکی ارجاع می‌دهیم و اینجا صرفاً به بحث اصلی خودمان یعنی اخلاق هوش مصنوعی می‌پردازیم

اشکالی که به تطبیق اصول چهارگانه اخلاق پزشکی بر اخلاق هوش مصنوعی وارد آمده از جمله این است که پزشکی بر پایهٔ یک هدف مشترک یعنی ارتقای سلامت بیمار و در چارچوب وظایف امانی و نظام‌نامه حرفه‌ای عمل می‌کند، درحالی‌که توسعه هوش مصنوعی فاقد چنین هدف و همبستگی مشترکی است و یا حداقل تاکنون به چنین هدفی دست نیافته است. در پزشکی، منافع بیماران بر سازمان یا نهاد مقدم است و مقررات سختگیرانه تضمین می‌کند که سود یا پایداری بر رفاه بیماران غلبه نکند، اما در هوش مصنوعی، توسعه‌دهندگان عمدتاً تحت فشار منافع تجاری و سهامداران هستند و چارچوب‌های نظارتی محدود و پراکنده موجود، وظایف امانی روشنی نسبت به کاربران و صاحبان داده ایجاد نمی‌کنند. در نتیجه، اخلاق در هوش مصنوعی بیشتر به فشار افکار عمومی، نگرانی‌های اعتباری شرکت‌ها یا باورهای فردی توسعه‌دهندگان وابسته است که این وضعیت برای حفاظت از منافع بنیادین کاربران (مانند حریم خصوصی و خودمختاری) ناکافی و غیرقابل‌قبول است. علاوه بر آن اخلاق پزشکی دارای یک تاریخ دیرینه است که موجب می‌شود فرهنگ‌سازی لازم در طول قرن‌ها در این رشته تحقق یافته باشد. سوگند بقراطی یک نمونه از این فرهنگ دیرینه است. قوانین و اصول پزشکی طی زمانی طولانی تبدیل به راهنمای عمل شده و ضوابط و چارچوب‌های قانونی و حرفه‌ای تدوین شده، دارد. در حالی که اخلاق هوش مصنوعی هنوز به بلوغ خود نرسیده و ذکر چند اصول کلی نمی‌تواند مانع انحرافات آینده شود (Mittelstadt, 2019: 501–507). کلاوسر و گرت به خوبی بیان کرده‌اند که تمسک به اصول‌گرایی برای حل مشکلات اخلاق پزشکی دارای چه مراحلی است و چه مشکلاتی بر سر راه آن وجود دارد. به

نظر آنان اصول به تنهایی نمی‌توانند راهنمای عمل باشند چون اغلب با هم در تعارضند و چون هر یک از آن‌ها از یک نظریه اخلاقی اخذ شده‌اند و نحوه رفع تعارض آنان مشخص نیست (Clouser & Gert, 1990: 220–236).

پنجمین اصلی که فلوریدی اضافه می‌کند توضیح‌پذیری به معنای فعال کردن سایر اصول از طریق شفافیت، قابلیت فهم و پاسخگویی است. انسان هم تولیدکننده هوش مصنوعی است و هم بهره‌بردار از آن است. آنان که هوش مصنوعی را تولید می‌کنند در زندگی تمامی بهره‌برداران هوش مصنوعی تأثیرگذارند. این بدان معناست که نوعی نابرابری ذاتی اجتناب‌ناپذیر است. تولیدکنندگان هوش مصنوعی افراد اندکی هستند که مانند همه مهندسان و پزشکان، زندگی همه آحاد جامعه بشری را دچار تحول می‌کنند. پس نمی‌توان عدالت به معنای برابری را در همه جا سریان داد. مثل همه تولیدهای بشری نباید توقع داشته باشیم که تولید کنندگان سهمی یکسان با مصرف‌کنندگان داشته باشند. فلوریدی معتقد است به همین جهت است که چهار اصل پیشین باید با این اصل پنجم تکمیل گردد. تولیدکنندگان باید نسبت به توصیف عملکرد هوش مصنوعی «شفافیت» را رعایت کنند، «پاسخگوی» مسئولیت تولید خود باشند، و کار آنان «فهم‌پذیر» و «تفسیرپذیر» باشد. توضیح‌پذیری اصلی است که این مفاهیم را می‌رساند. تولیدکنندگان هم باید از منظر معرفتی توضیح دهند که هوش مصنوعی «چگونه کار می‌کند» و هم از منظر اخلاقی توضیح دهند که «مسئول این نحو کار آن کیست» (Floridi, 2021b: 31–32).

خلاصه آن که فلوریدی استدلال می‌کند که این پنج اصل، در معنا، همه ۴۷ اصلی را که در شش سند برجسته و مبتنی بر نظر کارشناسان ذکر شده‌اند، در بر می‌گیرد و یک چارچوب اخلاقی پدید می‌آورد که می‌توانیم توصیه‌های خود را در درون آن ارائه کنیم.

اشکالی که میتل‌اِشتات به این طرح فلوریدی وارد کرده است این که فروکاستِ ۴۷ اصلِ شش سند برجسته به ۵ اصل کلی، موجب کاهش‌گراییِ هنجاری و محوِ اختلاف‌های معنی‌دار می‌گردد. اصولی چون کرامت انسان، پایداری، مسئولیت‌پذیری، و حریم خصوصی دارای تنوع مفهومی هستند و ادغام آن‌ها زیر چتر چند عنوان کلی موجب می‌شود اختلاف‌های معنادار آن‌ها پنهان بماند و تعارض‌ها و اولویت‌ها را از دید ما مخفی سازد. این همان خطری است که در سنجش موج «اصول‌نویسیِ اخلاقِ هوش مصنوعی» گوشزد شده: اجماعِ کاذبِ سطح‌بالا که اختلاف‌های عمیق را می‌پوشاند و قابلیت اجرا را تضعیف می‌کند (Mittelstadt, 2019: 501–507).

## نتیجه‌گیری

با توجه به تحلیل مبسوط مبانی و نقدهای وارد بر چارچوب اخلاقی لوچانو فلوریدی، روشن می‌شود که اصول پنج‌گانهٔ او، یعنی خیرخواهی، منع ضرر، خودمختاری، عدالت و توضیح‌پذیری، در عین آنکه گامی مهم در جهت نظام‌مند کردن گفت‌وگوهای اخلاقی درباره هوش مصنوعی محسوب می‌شوند، اما برای پاسخ‌گویی به پیچیدگی‌های چندبعدی این حوزه کفایت ندارند. اخلاق هوش مصنوعی، برخلاف حوزه‌هایی مانند اخلاق زیستی، با کنشگرانی چندسطحی، فناوری‌هایی متکثر و زمینه‌های فرهنگی و سیاسی گوناگون روبه‌روست. بنابراین، نیازمند چارچوبی فراتر و پویاتر است که بتواند میان اصول کلان، هنجارهای میانی و الزامات خُرد ارتباطی توجیه‌پذیر و سازگار برقرار کند.

در واقع، اخلاق هوش مصنوعی را نمی‌توان تنها در قالب اصول عام و انتزاعی تعریف کرد؛ زیرا ترجمهٔ این اصول به دستورالعمل‌های اجرائی مستلزم در نظر گرفتن عناصر محلی، دینی، فرهنگی و سیاسی هر جامعه است. برای مثال، اصل خودمختاری در فرهنگ‌های سکولار ممکن است بر استقلال فردی مطلق تأکید کند، در حالی که در نظام‌های اخلاقی دینی، این استقلال با حدود الهی و مصالح جمعی تنظیم می‌شود. ازاین‌رو، هرگونه انتقال اصول از سطح جهانی به سطح ملی یا

سازمانی نیازمند توجیه اخلاقی جدید و بومی‌سازی آگاهانه است.

به همین ترتیب، نمی‌توان از اجماع نظری میان اندیشمندان و دولت‌ها بر مجموعه‌ای از اصول کلی، انتظار داشت که به خودی خود به هنجارهای میانی و دستورالعمل‌های اجرایی مؤثر منجر شود. هر مرحله از این زنجیره، از تبیین اصول تا تنظیم هنجارها و طراحی سازوکارهای اجرایی، باید به‌طور مستقل و با ارزیابی انتقادی توجیه گردد. این فرآیند نه تنها از بالا به پایین، بلکه از پایین به بالا نیز شکل می‌گیرد؛ یعنی از سطح عمل و تجربه به سطح نظری و سپس به بازتعریف اصول بازمی‌گردد. در چنین چارچوبی، نظام اخلاق هوش مصنوعی تنها زمانی می‌تواند جهان‌شمول و مورد پذیرش همگانی شود که اولویت‌گذاری روشنی برای حل تعارضات میان هنجارهای رقیب، میان خیرهای عمومی و منافع خاص و میان ارزش‌های جهانی و حساسیت‌های فرهنگی فراهم شود.

از این منظر، باید به سوی «اخلاق تطبیقیِ زمینه‌محور» حرکت کرد که نه صرفاً بر استخراج اصول عام، بلکه بر تبیین نسبت میان فناوری، کاربست، و زمینه‌های انسانی استوار باشد. در این چارچوب، سه سطح از توجیه ضروری است: اول سطح کلان، که اصول بنیادین مانند عدالت و خیرخواهی را تبیین می‌کند؛ دوم سطح میانی، که هنجارهای حرفه‌ای، سیاستی و نهادی را صورت‌بندی می‌نماید؛ و سوم سطح خُرد، که الزام‌های اجرایی و تصمیم‌های عملی را در موقعیت‌های خاص هدایت می‌کند. نبود پیوند توجیهی میان این سطوح، اخلاق هوش مصنوعی را به مجموعه‌ای از شعارهای زیبا اما ناکارآمد بدل می‌سازد.

در نهایت، باید پذیرفت که هدف اخلاق در حوزهٔ هوش مصنوعی، صرفاً تنظیم رفتار در مورد فناوری‌ها نیست، بلکه جهت‌دهی به توسعهٔ اجتماعی و انسانی در عصر دیجیتال است. تحقق این هدف در گرو گفت‌وگوی میان‌رشته‌ای میان فلسفه، فقه، علوم رایانه، سیاست‌گذاری عمومی و فرهنگ‌شناسی است تا از خلال آن بتوان نظامی چندلایه و توجیه‌پذیر از اصول، هنجارها و الزامات را سامان داد؛ نظامی که در آن اخلاق نه صرفاً مجموعه‌ای از قاعده‌ها، بلکه سازوکاری پویا برای حفظ کرامت انسان و پایداری جامعه در برابر تحولات شتابان فناوری باشد.

## منابع


قران کریم

اقتصاد ایالات متحده آمریکا. (۲۰۲۵، ۲۴ فروردین). در *ویکی‌پدیا، دانشنامه آزاد*.

مطهری، شهید مرتضی (۱۳۷۶). *انسان کامل (مجموعه آثار استاد شهید مطهری)* (ج ۲۳). تهران: صدرا.

مطهری، شهید مرتضی (۱۳۷۶). *تعلیم و تربیت در اسلام (مجموعه آثار استاد شهید مطهری)* (ج ۲۲). تهران:، صدرا.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Clouser, K D; & Gert, B. (1990). A critique of principlism. *The Journal of Medicine and Philosophy*, *15*(2), 219–36.

Floridi, Luciano (2018). *A roadmap for Europe's first global forum on the social impacts of Artificial intelligence*. Atomium-European Institute for Science, Media and Democracy (EISMD).

Future of Life Institute (2017). Asilomar AI Principles. Asilomar, California: Future of Life Institute.

IEEE. (2017). The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2. *IEEE*.

Luciano Floridi (2025, January 17). In *Wikipedia*.

Luciano Floridi | Digital Ethics Center. (2025). Retrieved November 3, 1403, from https://dec.yale.edu/people/luciano-floridi

Mittelstadt, Brent (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

Partnership on AI (2018). Tenets of the Partnership on AI. Retrieved from https://www.partnershiponai.org/tenets/

Professor Luciano Floridi | University of Oxford. (2025). Retrieved November 3, 1403, from https://www.ox.ac.uk/news-and-events/find-an-expert/professor-luciano-floridi

Université de Montréal (2018). Montreal Declaration for A Responsible Development of AI. Montreal: Université de Montréal.

About – Luciano Floridi | Philosophy of Information. (2025).

Eubanks, Virginia (2017). *Automating inequality : how high-tech tools profile, police, and punish the poor* (First edition). New York, NY: St. Martin's Press.

European Group on Ethics in Science and New Technologies to the European Commission (2018). *Statement on artificial intelligence, robotics and autonomous systems* (Vols. 1–1 online resource (20 pages)). Brussels: Publications Office of the European Union. https://doi.org/10.2777/531856

Floridi, Luciano (2013). *The ethics of information*. Oxford: Oxford University Press.

Floridi, Luciano (2021a). *Ethics, governance, and policies in artificial intelligence*. Cham: SPRINGER NATURE.

Floridi, Luciano (2021b). *Ethics, governance, and policies in artificial intelligence*. Cham: SPRINGER NATURE.

Good, Irving John (2005). *Speculations Concerning the First Ultraintelligent Machine*. NEW YORK: Virginia Tech. Retrieved from WorldCat.

Great Britain (2018). *AI in the UK : Ready, willing and able? ; report of Session 2017-19*.

Sartre, Jean-Paul (1993). *Being and nothingness : an essay of phenomenological ontology*. (H. E. Barnes, Tran.) (Repr). London: Routledge.

Susskind, Jamie (2018). *Future politics : living together in a world transformed by tech*. Oxford: Oxford University Press.