


# A systematic review of the ethical implications of using artificial intelligence in digital technologies and its relationship with the ethics of flourishing

**Somayeh Bolboli Qadikolaei**

University of Mazandaran, Babolsar, Iran (Corresponding author).

bolboli.s@ut.ac.ir

 0000-0003-4780-8229

**Hamid Parsania**

Associate Professor, Department of Sociology, Faculty of Social Sciences, University of Tehran, Tehran, Iran.

h.parsania@ut.ac.ir

 0000-0000-0000-0000

## **Abstract**

The widespread influence of artificial intelligence in various aspects of human life raises major concerns in terms of the degree to which this technology adheres to ethical standards. Therefore, the aim of this study is to identify the key concepts of ethics for the use of artificial intelligence in digital technologies and its relationship with flourishing ethics. In this research, a systematic review of the literature has been used as a research methodology and the three-stage method of Mustak et al. has been used as a research method. After five stages of screening in the two databases "Scopus" and "Web of Science", 38 articles from 25 journals were analyzed. From their meta-qualitative synthesis, two central concepts under the title of "main patterns of digital technologies" and "ontological layers of digital technologies" were identified. 16 concepts of digital ethics related to the main patterns and ontological layers of digital technologies were also extracted from the intersection of the matrix of central concepts in the articles. Then, a conceptual map was prepared to interpret the findings, classify the identified ethical concepts under ontological patterns and layers, their correspondence with the principles of artificial intelligence ethics and their importance in the articles. Finally, the ratio of identified ethical concepts with flourishing ethics in the existing literature was investigated.

**Keywords:** Ethics, Artificial Intelligence, Digital Technologies, Flourishing Ethics.

# مروری نظام‌مند بر دلالت‌های اخلاقی استفاده از هوش مصنوعی در فناوری‌های دیجیتال و نسبت آن با اخلاق شکوفایی

سمیه بلبلی قادیکلایی

دانشگاه مازندران، بابلسر، ایران (نویسنده مسئول).

bolboli.s@ut.ac.ir

0000-0003-4780-8229

حمید پارسانیا

دانشیار گروه جامعه‌شناسی دانشکده علوم اجتماعی دانشگاه تهران، تهران، ایران.

h.parsania@ut.ac.ir

0000-0000-0000-0000

## چکیده

نغوذ گسترده هوش مصنوعی در ابعاد مختلف حیات انسان نگرانی‌های عمده‌ای به لحاظ میزان پابندی این فناوری به استانداردهای اخلاقی در پی دارد. از این رو هدف پژوهش حاضر شناسایی مفاهیم کلیدی اخلاق برای استفاده از هوش مصنوعی در فناوری‌های دیجیتال و نسبت آن با اخلاق شکوفایی است. در این پژوهش از مرور نظام‌مند ادبیات به‌عنوان روش شناسی پژوهش و از روش سه‌مرحله‌ای مستاک و دیگران، به‌عنوان روش تحقیق استفاده شده است. پس از پنج مرحله غربالگری در دو پایگاه داده «اسکوپوس» و «وب آو ساینس» ۳۸ مقاله از ۲۵ مجله تجزیه و تحلیل شدند. از فراترکیب کیفی آنها دو مفهوم مرکزی تحت عنوان «الگوهای اصلی فناوری‌های دیجیتال» و «لایه‌های هستی‌شناختی فناوری‌های دیجیتال» شناسایی شدند. ۱۶ مفهوم اخلاق دیجیتال مرتبط با الگوهای اصلی و لایه‌های هستی‌شناختی فناوری‌های دیجیتال نیز از تقاطع ماتریسی مفاهیم مرکزی در مقالات استخراج شدند. سپس یک نقشه مفهومی برای تفسیر یافته‌ها، طبقه‌بندی مفاهیم اخلاقی شناسایی و ذیل الگوها و لایه‌های هستی‌شناختی، تناظر آنها با اصول موضوعه اخلاق هوش مصنوعی و میزان اهمیت آنها در مقاله‌ها تنظیم شد. درنهایت نیز نسبت مفاهیم اخلاقی شناسایی شده با اخلاق شکوفایی در ادبیات موجود بررسی شد.

کلیدواژه‌ها: اخلاق، هوش مصنوعی، فناوری‌های دیجیتال، اخلاق شکوفایی.

شاپای الکترونیک: ۶۵۵X-۲۵۸۸ / پژوهشکده تحقیقات راهبردی / فصلنامه علمی راهبرد اجتماعی فرهنگی

CC BY 4.0



doi 10.22034/SCS.2022.160772

## مقدمه و بیان مسئله

شاید تاریخچه پیدایش هوش مصنوعی به افسانه‌ها یا داستان‌های علمی - تخیلی و تمایل انسان به برگزیدن دستیاری برای خود که قادر باشد ناممکن‌ها را برای او ممکن سازد بازگردد؛ اما با گسترش استفاده همگانی از رایانه، افزایش قدرت محاسباتی و قدرت یادگیری زبان در رایانه‌ها، دهه ۱۹۵۰ تا ۱۹۷۰ را به‌طور مشخص می‌توان «عصر طلایی هوش مصنوعی» دانست. پس از آن، از سال ۱۹۸۰ شاهد انقلاب دیجیتال هستیم که تا به امروز ادامه داشته و منجر به هستی‌شناسی مجدد جهان شده است (Floridi, 2011, p. 78). این رخداد عظیم به‌واسطه ادغام واقعیت‌های مجازی و فیزیکی موجب جدایی مفاهیم مرتبط با هم مانند «مالکیت»، «مکان»، «حضور» و «استفاده» شده است (Floridi, 2017, p. 63). سیطره گسترده فناوری‌های دیجیتال بر حوزه‌های متنوعی از زندگی انسان اعم از زنجیره‌های تأمین سلامت، شبکه‌های انرژی، بانکداری برخط، کشاورزی و...، همچنین پیوند آن با هوش مصنوعی تغییرهای بزرگی در حیات اجتماعی را به‌دنبال داشته است (Brynjolfsson & McAfee, 2014, p. 205). هوش مصنوعی این امکان را فراهم می‌آورد تا ابزارها بتوانند وظایف شناختی را که تا پیش از آن بر عهده انسان بوده ضمن به‌حداقل رساندن نقش انسانی و یا حتی حذف انسان، با دقت بالا انجام دهند. دستورالعمل‌های طراحی شده در آزمایشگاه‌ها که اساس کار هوش مصنوعی می‌باشند، با سرعت زیادی از فضای آزمایشگاهی پا را فراتر نهاده و در حال استقرار در تمامی سطوح جامعه هستند. چنین سطحی از گسترش هوش مصنوعی پیامدهای منفی مانند تبعیض‌های نژادی و جنسیتی وسیع‌تر به‌واسطه سوگیری‌های پنهان در داده‌های ورودی، از بین رفتن عاملیت انسان، گسترش جرائم سایبری، سوءاستفاده از کلان‌داده‌ها، نقض حریم خصوصی، بیکارشدن نیروی انسانی، نفوذ سیاسی غیرقابل کنترل شرکت‌های چندملیتی و انحصار منابع ثروت و قدرت در دست گروه‌هایی اندک و... را به‌دنبال خواهد داشت. و رای تمامی سوگیری‌های پنهان و پیامدهای منفی، با توجه به ظرفیت تأثیرگذاری بالای پژوهش‌های حوزه هوش مصنوعی همواره این سؤال مطرح است که آیا متخصصان به هنگام توسعه ابزارها و برنامه‌های خود، به تأمین بیشترین منافع انسان و جامعه می‌اندیشند یا خیر؟ اینجا دقیقاً همان

نقطه‌ای است که پای اخلاق و ارزش‌های اخلاقی به‌عنوان یکی از چالش‌برانگیزترین دغدغه‌های پژوهشگران این حوزه به میان می‌آید. آیا خروجی دستورالعمل‌های زیبا، برای زیبا کردن زندگی انسان و ارتقاء سطح رفاه او مفید و کارآمد است یا اینکه برعکس، به ضررش تمام می‌شود؟ پایبندی طراحان دستورالعمل‌ها و توسعه‌دهندگان هوش مصنوعی به ارزش‌های اخلاقی موجب شکل‌گیری هوش مصنوعی اخلاقی می‌شود که به‌هنگام به‌کارگیری در فناوری‌های دیجیتال، مسیر شکوفایی انسان را فراهم می‌آورد؛ زیرا آنچه که به تعبیر استال<sup>۱</sup> (۲۰۲۱) اساس اخلاق هوش مصنوعی را تشکیل می‌دهد، «اخلاق شکوفایی» است که با هدف «شکوفایی انسان» سعی در ارتقاء قابلیت‌های انسانی اعم از سلامت جسمانی، تخیل، تفکر، حواس پنج‌گانه، احساسات، کنترل بر محیط اطراف، ارتباط با سایر گونه‌های حیات طبیعی و در یک کلام «زندگی» دارد (Nussbaum, 2011, p. 79). اهمیت اخلاق شکوفایی در فناوری‌های - مبتنی بر هوش مصنوعی و همچنین پیامدهای مخاطره‌آمیز سوءاستفاده از آن، توجه به ابعاد اخلاقی استفاده از هوش مصنوعی را در فناوری‌های مذکور پررنگ می‌سازد. فلوریدی<sup>۲</sup> (۲۰۱۸) نیز معتقد است دانش موجود در حوزه هوش مصنوعی فاقد یک جهت‌گیری سیاسی - اجتماعی و نیازمند یک اتفاق نظر جمعی در مورد آینده این حوزه مطالعاتی است تا از سرعت توسعه فناوری جلوگیری باشد. از این‌رو پژوهش‌هایی همه‌جانبه‌نگر از جنس مرور نظام‌مند که بتواند ضمن ایجاد یک اشراف علمی بر پژوهش‌های انجام شده و تولید سطح بالاتری از دانش موجود، چشم‌اندازی روشن برای پژوهش‌های آتی فراهم آورد امری ضروری است. پژوهش‌هایی که می‌توانند برای توسعه‌دهندگان هوش مصنوعی و طراحان دستورالعمل از لحاظ توجه به ارزش‌های اخلاقی به‌هنگام توسعه این فناوری کارآمد باشند. از این‌رو مسئله‌مندی پژوهش حاضر، معطوف به اهمیت جایگاه اخلاق در فناوری‌های دیجیتال و میزان تعقیب اهداف منتهی به شکوفایی انسان از سوی این فناوری‌ها به‌هنگام استفاده از هوش مصنوعی است. دستاوردهای این پژوهش می‌تواند ضمن مشخص نمودن شکاف‌های پژوهشی این حوزه، مروری جامع بر شواهد موجود در خصوص موضوع پژوهش باشد و با استخراج مجموعه مضامین اخلاقی مرتبط با به‌کارگیری هوش مصنوعی در فناوری‌های دیجیتال به خلق یک الگوی مفهومی جدید بپردازد.

1. Stahl
2. Floridi

## ۱. اهداف و سؤال‌های پژوهش

به تعبیر مورلی<sup>۱</sup> و دیگران (Morley & et al., 2020, p. 2154) اصول هوش مصنوعی باید به «چه» و «چگونگی» اجرا ترجمه شوند تا علاوه بر تنوع فناوری‌های دیجیتال، تأثیر اخلاقی و پیامدهای آن مورد ارزیابی قرار گیرد؛ لذا این پژوهش با هدف شناسایی مفاهیم کلیدی اخلاق برای استفاده از هوش مصنوعی در فناوری‌های دیجیتال و نسبت آن با اخلاق شکوفایی براساس مطالعات انجام گرفته در این حوزه تنظیم شده است. از طرفی اهمیت توجه به گوناگونی فناوری‌های دیجیتال به هنگام مطالعه مفاهیم اخلاقی آن (Adomavicius; Bockstedt; Gupta & Kauffman, 2007) لزوم الگوبایی مناسب برای طبقه‌بندی این فناوری‌ها را نمایان می‌سازد. همچنین عدم وجود هستی‌شناسی مشترک در توسعه فناوری‌های دیجیتال، ارزیابی‌های اخلاقی و توسعه هوش مصنوعی متناظر با آنها را به مخاطره می‌اندازد. از این رو استخراج لایه‌های هستی‌شناختی فناوری‌های دیجیتال در مطالعات مرتبط، در ایجاد درک صحیح از مفاهیم اخلاقی این حوزه راهگشا می‌باشد. بدین ترتیب، پژوهش حاضر به دنبال پاسخگویی به دو پرسش اصلی و همچنین دو پرسش فرعی ذیل پرسش شماره ۱ است:

- ۱- مفاهیم کلیدی اخلاقی برای استفاده از هوش مصنوعی در فناوری‌های دیجیتال چیست؟
- این مفاهیم اخلاقی ذیل کدام الگوهای اصلی از فناوری‌های دیجیتال طبقه‌بندی می‌شوند؟
- کدام لایه‌های هستی‌شناختی از فناوری‌های دیجیتال بر این مفاهیم اخلاقی سایه افکنده‌اند؟
- ۲- مفاهیم اخلاقی شناسایی شده بر اساس پرسش شماره ۱ چه نسبتی با اخلاق شکوفایی دارند؟

## ۲. پیشینه پژوهش

در سال‌های اخیر چندین مرور نظام‌مند بر روی ادبیات هوش مصنوعی در مطالعات رشته‌محور مانند راهبرد کسب‌وکار، توسعه پایدار، بازاریابی دیجیتال و رسانه‌های اجتماعی، اینترنت اشیا، زنجیره بلوکی، و آموزش انجام شده است (Borges; Laurindo; Spinola; Goncalves & Mattos, 2021; Duan; Edwards &

1. Morley

Dwivedi, 2019; Dwivedi & et al., 2021; Kankanhalli; Charalabidis & Mellouli, 2019; Nishant; Kennedy & Corbett, 2020; Sarker; Wu; Cao; (Alam & Li, 2019).

بررسی‌های انجام شده ضمن شناسایی عمده‌ترین چالش‌های پژوهش‌های هوش مصنوعی در حوزه‌های مذکور به ارائه راهکارها و پیشنهادهایی برای فائق آمدن بر این چالش‌ها و شکاف‌های پژوهشی می‌پردازند. استفاده از رویکردهای چندسطحی، رویکردهای پویایی سیستم، تفکر طراحی، ملاحظه‌های جامعه‌شناسانه و روان‌شناسانه، ملاحظه‌های ارزش اقتصادی و استفاده از هوش مصنوعی به روش راهبردها به منظور ایجاد ارزش تجاری از جمله این راهکارهاست.

در میان پژوهش‌های داخلی نیز مطالعات پژوهشگاه فضای مجازی در حوزه‌هایی چون «آموزش فناوری اطلاعات و ارتباطات و هوش دیجیتال» (نائلی، ۱۴۰۱)، «اینترنت اشیاء اجتماعی» (کاظمی و مطلبی کربکندی، ۱۴۰۰) و «هوش مصنوعی و شکاف مسئولیت» (زادیوسفی، ۱۴۰۰) مطالعاتی راهبردی هستند. این مطالعات علاوه بر شرح چهارچوب هوش دیجیتال به‌عنوان اولین استاندارد جهانی برای اندازه‌گیری سواد دیجیتال در حوزه آموزش فناوری اطلاعات و ارتباطات، ارائه تصویری کلی از حوزه‌های رو به رشد اینترنت اشیاء اجتماعی و همچنین ارائه رویکردی حقوقی به مسئله شکاف مسئولیت در هوش مصنوعی، به معرفی پیشنهادهایی راهبردی معطوف به فرصت‌ها و چالش‌های هر حوزه در کشور می‌پردازند. در خصوص آموزش فناوری اطلاعات و ارتباطات و هوش دیجیتال راهبردهایی مانند به‌روزرسانی نظام آموزش و پرورش کشور، ارتقای زیرساخت‌های فناوری اطلاعات مدارس و حمایت و تقویت ظرفیت سازمان‌های مردم‌نهاد در حوزه کودک و نوجوان پیشنهاد شده است. در حوزه اینترنت اشیاء اجتماعی، راهبردهایی چون ورود سریع‌تر کشور به این حوزه، حضور فعال در مجامع بین‌المللی، طراحی حساس به ارزش، آینده‌پژوهی و جدی گرفتن مسائل اخلاقی خاص این حوزه پیشنهاد شده است. در مطالعه هوش مصنوعی و شکاف مسئولیت نیز رفع خلأهای قانونی برای حوادث ناشی از سیستم‌های هوشمند خودسامان توسط قانون‌گذاران کشور و نیز تعریف طرح‌های پژوهشی برای مباحث نظری حوزه مذکور و همکاری فلاسفه و حقوق‌دانان در این عرصه پیشنهاد شده است. بررسی‌ها بیانگر آن است که به‌رغم اهمیت اخلاق در فناوری‌های دیجیتال مبتنی بر هوش مصنوعی به لحاظ پیامدهای اجتماعی و اثرگذاری بر فرایند شکوفایی انسان، مروری نظام‌مند بر ادبیات

موجود در مورد ابعاد اخلاقی استفاده از هوش مصنوعی در فناوری‌های دیجیتال و نسبت آن با اخلاق شکوفایی انجام نشده است. از این رو پژوهش حاضر، ناظر بر این خلأ پژوهشی تنظیم شده است.

### ۳. چهارچوب نظری

مطالعات اخلاق تاریخچه‌ای غنی از زمان فلاسفه باستان دارند. امروزه نیز «اخلاق» مفهومی است که در حوزه‌های مختلف علمی به‌ویژه فناوری‌های نوین، محل مناقشه فراوان می‌باشد. به تعبیر استال (۲۰۲۱) اخلاق منحصرأ بر کنش یا نیت انجام کنش متکی نبوده و سعی دارد با در نظر گرفتن جنبه‌های مشترک طبیعت انسان، بخش مهمی از جهان اجتماعی او را صورتبندی نماید. هایدگر<sup>۱</sup> نیز اخلاق را بخشی از مفهوم «بودن» در جهان می‌داند که ناظر بر حضور آگاهانه کنشگر در جهان است. یعنی توانایی ادراک حالات مختلف از جهان و تصمیم‌گیری میان گزینه‌های غیرقابل تصور و حتی فراتر از آن، توانایی درک معنای چنین تصمیمی هم برای خود و هم برای کلیت جهان از سوی کنشگر (Stahl, 2021). نظریه‌های مرتبط با مفهوم اخلاق تحت عنوان «نظریه‌های اخلاق» به‌طور کلی به‌دنبال پاسخ به این سؤال مهم هستند که از لحاظ اخلاقی چه چیزی موجب بهتر یا بدتر بودن یک کنش در مقایسه با کنشی دیگر می‌شود؟ «نتیجه‌گرایی» با تمرکز بر نتایج کنش و «وظیفه‌گرایی» با تأکید بر انگیزه انجام کنش از برجسته‌ترین نظریه‌های اخلاقی هستند که با طرح قواعدی برای ارزیابی کیفیت اخلاقی یک کنش، در پاسخ به سؤال فوق بسیار راهگشا بوده و با به‌کارگیری در حوزه‌های کاربردی خاص منجر به گفتمان‌های غنی از مفاهیمی مانند اخلاق رایانه‌ای (Van den Hoven, 2010)، اخلاق اطلاعات (Floridi, 2006, p. 31) و اخلاق فناوری (Brey, 2012, p. 11) به‌عنوان بستر نفوذ هوش مصنوعی شده‌اند. «اخلاق فضیلت» نیز با تمرکز بر هدایت افراد به‌منظور کمک به رشد شخصیتی با فضیلت در آنها و به‌عنوان راهنمایی برای چگونه زندگی کردن، از اواخر قرن بیستم مجدداً مورد توجه قرار گرفت و تلاش‌های زیادی برای جانمایی آن در بستر جامعه پیشرفته صورت گرفته است (MacIntyre, 2007, p. 63). در رابطه با اخلاق فضیلت و نسبت آن با فناوری دو نظریه مهم وجود دارد. نخست، نظریه باینوم<sup>۲</sup> (Bynum, 2006, p. 163) است که اصول اخلاق فضیلت

1. Heidegger  
2. Bynum

را با ملاحظه‌های اخلاقی فناوری اطلاعات مرتبط می‌داند. وی با بهره‌گیری از مفهومی به نام «اخلاق شکوفایی» اصول قدیمی اخلاق فضیلت را به زبان یک جامعه اشباع‌شده از فناوری‌های پیشرفته برگردان نموده است. دیگری، رویکرد استال (۲۰۲۱) است که در بیان اهداف استفاده از فناوری‌های مبتنی بر هوش مصنوعی علاوه بر کنترل اجتماعی و بهره‌وری، به هدف سومی به نام «هوش مصنوعی برای شکوفایی انسان» اشاره دارد و آن را لازمه یک «زندگی فضیلت‌مند» می‌داند. به این معنا که هوش مصنوعی به‌گونه‌ای توسعه‌یافته و به‌کار گرفته می‌شود که شکوفایی انسان را ارتقا می‌دهد. وی بر ضرورت بیان این اهداف برای فهم اخلاق هوش مصنوعی تأکید داشته و هدف «هوش مصنوعی برای شکوفایی انسان» را جایگزین و مکمل دو هدف دیگر می‌داند. استال همچنین شکوفایی انسان را در پیوند با مفهوم هایدگری «بودن در جهان» می‌داند که ناظر بر درک محدودیت شرایط انسانی و لزوم همدلی با او به دلیل برخورداری از زیستی اساساً اجتماعی است. باینوم (Bynum, 2006, p. 161) شکوفایی را مفهومی یکتا ندانسته و معتقد است همچون تن‌پوشی بر تن هر انسان، بسته به نوع مهارت و میزان استمرار در پیگیری اهداف قابل اندازه‌شدن است. به‌منظور درک چگونگی شکوفایی در عمل، توجه به موقعیت‌های دیگری که هدفشان ارتقای شکوفایی انسان است می‌تواند راهگشا باشد. «نظریه انتقادی فناوری»، «نظریه قابلیت» و «نظریه پژوهش و نوآوری مسئولانه» سه موقعیتی هستند که در نتیجه تحقیق و توسعه فناوری مطرح و از اهمیت به‌سزایی برخوردار شده‌اند. نظریه انتقادی فناوری از جمله نظریه‌های مرتبط با هوش مصنوعی است که به‌دنبال شکوفایی انسان می‌باشد. پژوهش فینبرگ<sup>۱</sup> (Feenberg, 1999, p. 74) احتمالاً شناخته‌شده‌ترین نمونه استفاده از نظریه انتقادی برای مطالعه فناوری‌های پیشرفته است. هسته مرکزی تمامی نظریه‌های انتقادی «رهایی‌بخشی» است (Stahl, 2008, p. 141). رهایی‌بخشی تحقق نیازها و استعداد‌های انسان، خوداندیشی انتقادی و به‌دنبال آن خوددگرگون‌سازی را تسهیل نموده و به شکوفایی انسان کمک می‌کند (Myers & Klein, 2011, p. 26). نظریه قابلیت اما با تأثیرپذیری از دیدگاه‌های آمارتیا سن<sup>۲</sup> (۲۰۰۹) و نوسام<sup>۳</sup> (۲۰۱۱) ارتباط مستقیمی با مفهوم ارسطویی شکوفایی و اخلاق هوش مصنوعی داشته و به‌دنبال یافتن راه‌های

1. Feenberg  
2. Sen, A.  
3. Nussbaum



بهتری برای توصیف توسعه انسانی است (Johnstone, 2007, p. 75). نظریه قابلیت دارای سابقه کاربرد در فناوری اطلاعات بوده و اغلب در چهارچوب مطالعات حوزه فناوری‌های اطلاعات و ارتباطات به بررسی جمعیت‌های حاشیه‌نشین و آسیب‌پذیر و تأثیری که هوش مصنوعی می‌تواند بر چنین جمعیت‌هایی داشته باشد پرداخته است (Oosterlaken & Van den Hoven, 2012). «تحقیق و نوآوری مسئولانه» نیز به معنای فرایند مداوم همسوسازی پژوهش‌ها و نوآوری‌ها با ارزش‌ها، نیازها و انتظارات جامعه بوده و فرایندی شفاف و تعاملی است که از طریق آن کنشگران اجتماعی و نوآوران با در نظر داشتن مقبولیت اخلاقی، پایداری و مطلوبیت اجتماعی فرایندهای نوآورانه و محصولات قابل فروش آنها، خود را نسبت به جامعه پاسخگو و مسئول می‌دانند (Von Schomberg, 2013, p. 63) تا آن را به‌عنوان راهی برای اطمینان از حساسیت اخلاقی در پژوهش‌های آینده هوش مصنوعی، توسعه دهند. بدین ترتیب شکوفایی انسان حوزه معنایی گسترده‌ای است که قابلیت استفاده برای همه انسان‌ها را دارد؛ ما را به شیوه زندگی خاصی متعهد نمی‌کند؛ نیاز به اتخاذ یک موقعیت اخلاقی خاص ندارد؛ همچنین ما را از استفاده از سایر نظریه‌های اخلاقی، از جمله وظیفه‌گرایی و فایده‌گرایی، برای ارزیابی سؤال‌های اخلاقی باز نمی‌دارد و با مواضع نظری مختلف سازگاری دارد (Bynum, 2006, p. 165)؛ لذا اخلاق شکوفایی به‌عنوان یک زبان اخلاقی مشترک، دارای انعطاف‌پذیری بالا و توانایی پوشش‌دهی و بهره‌گیری از نظریه‌های مختلف اخلاقی پلی است بر روی شکاف‌های ناشی از فراگیری هوش مصنوعی در تمام جهان. از این‌رو مقاله حاضر تلاش دارد بنابر اقتضای پژوهش‌های کیفی، در پرتو این نظریه مفاهیم اخلاقی مرتبط با استفاده از هوش مصنوعی در فناوری‌های دیجیتال را شناسایی نموده؛ نسبت مفاهیم احصاء شده در پژوهش‌ها را با مفهوم اخلاق شکوفایی مشخص نماید و در نهایت از تجزیه و تحلیل داده‌ها به یک الگوی مفهومی دست یابد.

#### ۴. روش پژوهش

در پژوهش حاضر از مرور نظام‌مند به‌عنوان روش‌شناسی پژوهش استفاده شده است که به تعبیر بوث<sup>۱</sup> و همکاران (۲۰۱۲) با هدف تولید دانش جدید و الگوهای مفهومی از مطالعات انجام گرفته مورد استفاده قرار می‌گیرد. همچنین از روش سه‌مرحله‌ای

مستاک<sup>۱</sup> و همکاران (Mustak & et al., 2016, p. 265) که عبارت‌اند از: ۱- جستجوی ادبیات؛ ۲- ارزیابی پایگاه‌های داده و ۳- تجزیه و تحلیل و فراترکیب یافته‌ها، در فرایند تحقیق پیروی شده است. شناسایی، غربالگری و تحلیل نمونه‌ها نیز به ترتیب مراحل زیر انجام شده است:

مرحله ۱- شناسایی مقاله‌های مرتبط با اخلاق فناوری‌های دیجیتال و نسبت آن با هوش مصنوعی از طریق جست‌وجو در پایگاه‌های داده اسکوپوس و وب آو ساینس<sup>۲</sup> (شناسایی ۱۷۸ مورد ثبت شده)،

مرحله ۲- ترکیب نتایج جست‌وجو در مرحله اول با کلیدواژه «اخلاق شکوفایی» (شناسایی ۱۱۸ مقاله در قالب ۲۵ مجله)،

مرحله ۳- ارزیابی تناسب مقاله‌های شناسایی شده با موضوع پژوهش (غربالگری و خروج ۸۰ مقاله بر اساس ۵ معیار عدم تناسب واژگان جست‌وجو با متن مقاله‌ها، تمرکز بر روی فناوری‌های دیجیتال بدون بحث در مورد اخلاق، تمرکز بر مباحث فنی هوش مصنوعی و کاربردهای رباتیک، تمرکز بر هوش مصنوعی بدون بحث درباره اخلاق شکوفایی و تمرکز بر توسعه هوش مصنوعی)،

مرحله ۴- فراترکیب کیفی مقالات منتخب (۳۸ مقاله).

پس از غربالگری، مقاله‌های منتخب مورد تجزیه و تحلیل و فراترکیب کیفی قرار گرفتند که منجر به توسعه مضامین پایه، خوشه‌بندی آنها در مضامین سازمان‌دهنده، طبقه‌بندی در قالب مضامین فراگیر و درنهایت تولید دو مفهوم مرکزی تحت عنوان «الگوهای اصلی فناوری‌های دیجیتال» و «لایه‌های هستی‌شناختی فناوری‌های دیجیتال» شد (جدول ۱). سپس یک الگوی پیشینی با استفاده از اصول موضوعه اخلاق هوش مصنوعی تنظیم شد. مفاهیم اخلاق دیجیتال مرتبط با الگوهای اصلی فناوری‌های دیجیتال و لایه‌های هستی‌شناختی آن نیز از تقاطع ماتریسی مفاهیم مرکزی در متون شناسایی شدند. در گام بعدی نقشه‌ای مفهومی به‌منظور خلاصه نمودن نتایج، تفسیر یافته‌ها، طبقه‌بندی مفاهیم اخلاقی شناسایی شده و میزان اهمیت آنها در مقاله‌ها تنظیم شد (جدول ۲). درنهایت نیز نسبت مفاهیم اخلاقی شناسایی شده با اخلاق شکوفایی در ادبیات موجود بررسی و مشخص شد چه مفاهیمی در اخلاق دیجیتال و ناظر بر اصول اخلاقی هوش مصنوعی با بحث شکوفایی انسان پیوند دارند (جدول ۳). اعتباربخشی نتایج نیز به این صورت تأیید

1. Mustak

2. Scopus & Web of Science

شد که کاربرگ‌های استخراج شده از مطالعه مقاله‌های در اختیار دو تن از متخصصان، یک نفر از حوزه مهندسی رایانه با تخصص هوش مصنوعی و دیگری از علوم اجتماعی با تخصص اخلاق فناوری قرار گرفت تا یادداشت‌ها، کدها و مضامین با متن مقاله‌های انطباق داده شده و بازتاب‌دهندگی کاربرگ‌ها نسبت به متون تحت بررسی مشخص شود.

### ۵. یافته‌های پژوهش

در پژوهش حاضر تلاش شد با استفاده از مرور نظام‌مند ادبیات موجود و فراترکیب کیفی آنها به پرسش‌های پژوهش پاسخ داده شود. پرسش اول در رابطه با چیستی مفاهیم کلیدی اخلاقی برای استفاده از هوش مصنوعی در فناوری‌های دیجیتال است. اگرچه این سؤال، نخستین پرسش اصلی پژوهش است؛ اما پاسخی به آن به دلایلی که پیشتر در بخش «اهداف و سؤال‌های پژوهش» بیان شد، نیازمند شناسایی الگوهای اصلی و لایه‌های هستی‌شناختی فناوری‌های دیجیتال در بدو امر می‌باشد؛ لذا منطق پاسخ‌دهی به پرسش‌های پژوهش اولویت یافته‌هاست نه ترتیب سؤال‌ها. دو پرسش فرعی مطرح شده ذیل نخستین سؤال پژوهش، ناظر بر الگوهای اصلی و لایه‌های هستی‌شناختی فناوری‌های دیجیتال می‌باشند. در پاسخ به دو پرسش مذکور، ۵ الگوی اصلی و ۴ لایه هستی‌شناختی در فناوری‌های دیجیتال مورد شناسایی و کدگذاری قرار گرفتند که مضامین فراگیر و سازمان‌دهنده آنها در جدول ۱ نشان داده شده است.

جدول (۱): الگوهای اصلی فناوری‌های دیجیتال، لایه‌های هستی‌شناختی فناوری‌های دیجیتال و

مضامین فراگیر و سازمان‌دهنده آنها

لایه‌های هستی‌شناختی فناوری‌های دیجیتال و مضامین فراگیر و سازمان‌دهنده آن			الگوهای اصلی فناوری‌های دیجیتال و مضامین فراگیر و سازمان‌دهنده آن		
مضامین سازمان‌دهنده	مضامین فراگیر	مفهوم مرکزی	مضامین سازمان‌دهنده	مضامین فراگیر	مفهوم مرکزی
فرستنده، طیف رادیویی، کابل، سیستم عامل، ساختار رایانه‌ای	دستگاه	لایه‌های هستی‌شناختی فناوری-های دیجیتال	پلتفرم مراقبت سالمند از راه دور، پلتفرم مراقبت کودک از راه دور، پلتفرم اطفاء حریق منزل از راه دور، پلتفرم شبکه‌های اجتماعی مجازی، پلتفرم بهداشت و سلامت، جستجوی اینترنتی،	پلتفرم	الگوی اصلی فناوری‌های دیجیتال

لایه‌های هستی‌شناختی فناوری‌های دیجیتال و مضامین فراگیر و سازمان‌دهنده آن			الگوهای اصلی فناوری‌های دیجیتال و مضامین فراگیر و سازمان‌دهنده آن		
مضامین سازمان‌دهنده	مضامین فراگیر	مفهوم مرکزی	مضامین سازمان‌دهنده	مضامین فراگیر	مفهوم مرکزی
تولید محتوا، دستکاری محتوا، ذخیره محتوا، مصرف محتوا	محتوا		محاسبات ابری، رمز ارز، تراکنش بانکی، اشتراک‌گذاری الکترونیکی اسناد	اتصال	
داده‌ها (متن، صدا، تصویر، ویدئو)، فراداده‌ها، مالکیت، حق چاپ، برچسب محتوا، رمزگذاری	داده		حمل‌ونقل هوشمند، انرژی هوشمند، آموزش و پرورش هوشمند، خانه‌های هوشمند، تشخیص پزشکی هوشمند	مصولان مبتنی بر کوشگر	
قوانین رسمی، قوانین غیررسمی، تنظیم رفتار کنشگران، حقوق بشر، حقوق مالکیت معنوی، حکمرانی داده	قانون		سیستم‌های خبره، منطق فازی، الگوریتم، یادگیری عمیق، یادگیری ماشینی، داده، کلان‌داده، استخراج داده، شبکه‌های عصبی	یادگیری تطبیقی	
			تراانسان، بارگذاری ذهن آگاه، آواتارهای تعامل‌گر	هوش مصنوعی	

منبع: (ترسیم محقق)

سپس در پاسخ به سؤال اول پژوهش تحت عنوان «مفاهیم کلیدی اخلاقی برای استفاده از هوش مصنوعی در فناوری‌های دیجیتال چیست؟» از تقاطع ماتریسی الگوها و لایه‌های هستی‌شناختی مذکور، ۱۶ مفهوم اخلاق دیجیتال استخراج شد. هر یک از مفاهیم احصاء‌شده با اصلی از اصول موضوعه اخلاق هوش مصنوعی در تناظر قرار گرفته‌اند. نتایج حاصل در قالب یک الگوی مفهومی در جدول ۲ ارائه شده‌اند.

جدول (۲): الگوی مفهومی اخلاق فناوری‌های دیجیتال

الگوهای اصلی فناوری‌های دیجیتال					اخلاق فناوری‌های دیجیتال	لایه‌های هستی‌شناختی فناوری‌های دیجیتال	اصول موضوعه اخلاق هوش مصنوعی
تعامل تحلیلی (درصد)	تولید بینش تحلیلی (درصد)	محصولات مبتنی بر کنشگر (درصد)	اتصال (درصد)	پلتفرم (درصد)			
۱۸	۱۹	۲۲	۹	۱۲	کرامت انسانی	دستگاه	اصل سودمندی
۳۳	۱۱	۷۵	۹	۱۲	ایمنی	دستگاه	اصل سودمندی
۷	۶	۸	۷	۳	پایداری	دستگاه	اصل سودمندی
۵۷	۶۴	۵۵	۸۱	۵۳	قابل فهم بودن	خدمات	اصل توضیح‌پذیری
۳۳	۴۴	۶۲	۶۶	۴۵	مسئولیت-پذیری	خدمات	اصل توضیح‌پذیری
۳۳	۳۵	۲۸	۲۴	۳۷	خودسامانی	خدمات	اصل توضیح‌پذیری
۲۶	۱۷	۰	۰	۱۲	ارتقاء سعادت و بهروزی	خدمات	اصل عدالت
۱۰	۱۵	۲۸	۲۴	۲۰	انسجام و همبستگی	خدمات	اصل عدالت
۲۳	۲۱	۱۷	۱۵	۱۱	مشارکت مردم‌سالارانه	خدمات	اصل عدالت
۴۹	۴۸	۴۲	۸۱	۴۵	انصاف	خدمات	اصل عدالت
۳۷	۲۵	۳۲	۱۹	۲۸	تنوع‌پذیری	خدمات	اصل عدالت
۸۰	۶۲	۴۸	۸۱	۶۲	حریم خصوصی	محتوا	اصل عدم زیانباری
۳۳	۱۵	۳۵	۲۴	۱۲	امنیت	محتوا	اصل عدم زیانباری
۴۱	۳۵	۵۵	۲۴	۲۸	اثر نظارتی - تنظیمی	نظارت	اصل کنترل
۳۳	۱۳	۲۲	۳۸	۲۵	اثر مالی - اقتصادی	نظارت	اصل کنترل
۶۴	۴۳	۳۴	۷	۲۸	اثر فردی - اجتماعی	نظارت	اصل کنترل

منبع: (ترسیم محقق)

جدول ۲ یک الگوی مفهومی از اخلاق فناوری‌های دیجیتال و طبقه‌بندی آنها ذیل لایه‌های هستی‌شناختی و الگوهای اصلی فناوری‌های دیجیتال ارائه می‌دهد. در این الگوی مفهومی علاوه بر تعیین میزان توجه متون تحت بررسی به هر یک از مفاهیم اخلاق فناوری‌های دیجیتال ذیل الگوهای اصلی و لایه‌های هستی‌شناختی، تناظر هر یک از مفاهیم مذکور با اصول موضوعه اخلاق هوش مصنوعی نیز مشخص شده است. در ادامه به هر یک از این مفاهیم با در نظر گرفتن الگوی طبقه‌بندی آنها پرداخته می‌شود:

### ۵-۱. کرامت انسانی، ایمنی، پایداری

در نتیجه تجزیه و تحلیل نمونه‌های تحت بررسی سه مفهوم اخلاقی «کرامت انسانی»، «ایمنی» و «پایداری» ذیل لایه دستگاه در فناوری‌های دیجیتال طبقه‌بندی شده و ناظر بر اصل سودمندی هوش مصنوعی هستند. «کرامت انسانی» به معنای حفظ شأن و منزلت انسان و ارتقای سطح رفاهی زندگی افراد (Malkiat; Sarunkorn & Stephen, 2019, p. 199) و محصول ایجاد تعادل میان مفاهیم اخلاقی «خودسامانی» و «حریم خصوصی» است. «ایمنی» به معنای محافظت در برابر آسیب‌های فیزیکی ناشی از فناوری‌های مجهز به هوش مصنوعی می‌باشد (Rhim; Lee & Lee, 2020, p. 46). مفهوم مذکور ناظر بر حصول اطمینان از تعامل‌های ایمن و با حداقل آسیب فیزیکی میان انسان و ماشین است. «پایداری» دیگر مفهوم اخلاقی در فناوری‌های دیجیتال است که به معنای ارتقای سلامت کره زمین و رویکرد مثبت به آینده می‌باشد (Floridi & Cowls, 2019, p. 123). هسته مرکزی این مفهوم، تأثیر زیست‌محیطی فناوری‌های دیجیتال، پایداری کره زمین و توجه حداکثری به خیر عمومی است.

### ۵-۲. قابل فهم بودن، مسئولیت‌پذیری، خودسامانی

مفاهیم اخلاقی «قابل فهم بودن»، «مسئولیت‌پذیری» و «خودسامانی» ذیل لایه هستی‌شناختی خدمات در فناوری‌های دیجیتال قرار گرفته و از اصل «توضیح‌پذیری» در هوش مصنوعی حمایت می‌کنند. «قابل فهم بودن» مفهومی است اخلاقی که بیانگر وجود یک احساس معرفت‌شناختی نسبت به نحوه عملکرد و میزان دقت هوش مصنوعی است (Floridi & Cowls, 2019, p. 103) و جزئی از فرایند اعتمادسازی در خروجی‌های هوش مصنوعی محسوب می‌شود. در مرور ادبیات

موجود، به موازات مفهوم « قابل فهم بودن» همواره این پرسش مهم مطرح شده است که در نهایت چه کسی مسئول تصمیم‌های الگوریتمی است؟ مهندسان نرم‌افزار، کاربران یا ماشین‌ها؟ فلورییدی و کالز (Floridi & Cowls, 2019, p. 15) «مسئولیت‌پذیری» را ناظر بر وجود یک احساس اخلاقی مبنی بر اینکه چه کسی مسئول نحوه عملکرد هوش مصنوعی است، می‌دانند. «خودسامانی» دیگر مفهوم اخلاقی است که معتقد است توسعه هوش مصنوعی نباید به قدرت تصمیم‌گیری انسان لطمه بزند (Floridi & Cowls, 2019, p. 16). درحقیقت خودسامانی بر حراست از عاملیت و آزادی انتخاب انسان تأکید داشته و مانع تبعیت محض انسان از طراحی‌های الگوریتمی می‌شود.

### ۵-۳. ارتقاء سعادت و بهروزی، انسجام و همبستگی، مشارکت مردم‌سالارانه، انصاف، تنوع‌پذیری

بر اساس یافته‌های پژوهش، مفاهیم اخلاقی «ارتقاء سعادت و بهروزی»، «انسجام و همبستگی»، «مشارکت مردم‌سالارانه» «انصاف» و «تنوع‌پذیری» در الگوهای هستی‌شناختی فناوری‌های دیجیتال ذیل لایه خدمات طبقه‌بندی شده و مضامین مرتبط با آن‌ها، ناظر بر اصل «عدالت» در هوش مصنوعی می‌باشند (Floridi & Cowls, 2019, p. 11). «انصاف» به معنای رفتار یکسان با همه انسان‌ها و عدم استفاده تبعیض‌آمیز از هوش مصنوعی در مواجهه با افراد، جوامع و گروه‌های اجتماعی می‌باشد. مرور ادبیات در پژوهش حاضر بیانگر آن است که به‌رغم وجود سوگیری‌های اجتماعی در داده‌های مورد استفاده در یادگیری ماشینی، همچنان مفهوم «انصاف» در فرایند توسعه هوش مصنوعی و طراحی دستورالعمل‌ها محل بحث و توجه فراوان است؛ تا نسبت به مغرضانه و جانبدارانه نبودن خروجی‌ها اطمینان حاصل شود. «ارتقاء سعادت و بهروزی» به معنای افزایش خیر عمومی، سودمندی اجتماعی و در نظر گرفتن منافع همه گروه‌های ذی‌نفع است (Metcalf; Moss & Boyd, 2019, p. 451). این مفهوم اخلاقی بر اساس یک استدلال فایده‌گرایانه و ناظر بر خیر جمعی مطرح شده است که هدف نهایی آن توسعه نوعی هوش مصنوعی است که بتواند برای بشریت سودمند باشد (Sutton; Arnold & Holt, 2018, p. 18). بر اساس یافته‌های پژوهش، مفهوم اخلاقی «انسجام و همبستگی» ناظر بر حفظ پیوند میان گروه‌های اجتماعی و نسل‌ها، به‌واسطه استفاده از هوش مصنوعی است (Choi & et al., 2019, p. 1988). چوی و دیگران «انسجام و

همبستگی» را به معنای حساسیت اخلاقی، همدلی و احترام به حقوق بشر می‌دانند. آنان معتقدند این مفهوم اخلاقی بر اصل حقوق بشر و درخواست از طراحان و توسعه‌دهندگان هوش مصنوعی برای اطمینان از همدلی و رعایت عدالت اجتماعی برای همه گروه‌های ذی‌نفع، در جهت دستیابی به سعادت و بهروزی تمرکز دارد (Choi & et al., 2019, p. 1988). بر اساس اصل مشارکت مردم‌سالارانه، هوش مصنوعی باید معیارهای قابل فهم‌بودن، توجیه‌پذیری و دسترسی‌پذیری را داشته باشد و باید در معرض بررسی، بحث و کنترل مردم‌سالارانه قرار گیرد (Yoon; Choi; Odlum; Mitchell & Kronish, 2018, p. 255). بر اساس اصل «تنوع‌پذیری» نیز، توسعه و استفاده از هوش مصنوعی باید با حفظ تنوع اجتماعی و فرهنگی سازگار باشد و دامنه انتخاب سبک زندگی یا تجربه‌های شخصی را محدود نکند.

#### ۴-۵. حریم خصوصی، امنیت

مفاهیم اخلاقی «حریم خصوصی» و «امنیت» ذیل لایه هستی‌شناختی محتوا در فناوری‌های دیجیتال طبقه‌بندی شده و مفاهیم مرتبط با آنها ناظر بر اصل «عدم زیانباری» در هوش مصنوعی می‌باشند (Floridi & Cowsls, 2019, p. 13). «حریم خصوصی» مفهومی است حول محور دسترسی، رضایت، حق برخورداری از داده‌های شخصی در عصر کلان‌داده‌ها و نظارت انبوه و به معنای رضایت و حفاظت در برابر نظارت و جمع‌آوری کلان‌داده‌ها است (Schappert & von Hauff, 2020, p. Article 121585). مفهوم «امنیت» نیز به معنای حفاظت از داده‌ها و مدیریت امنیت اطلاعات است (Urquhart; Schnadelbach & Jager, 2019, p. 21) که به کمک ایجاد موانعی، حریم خصوصی و حفاظت از داده‌ها را تضمین می‌کند.

#### ۵-۵. اثر نظارتی - تنظیمی، اثر مالی - اقتصادی، اثر فردی - اجتماعی

بر اساس مرور نظام‌مند ادبیات موجود، مفاهیم «اثر نظارتی - تنظیمی»، «اثر مالی - اقتصادی» و «اثر فردی - اجتماعی» ذیل لایه هستی‌شناختی نظارت در فناوری‌های دیجیتال قرار گرفته و ناظر بر اصل «کنترل» در هوش مصنوعی هستند. اصل کنترل در هوش مصنوعی به معنای «ایجاد و اجرای سیاست‌ها، فرایندها و استانداردها برای توسعه، استفاده و مدیریت مناسب حوزه اطلاعات» است (Floridi, 2018, p. Article 20180081). این حوزه شامل قوانین رسمی و غیررسمی و ارزش‌های اخلاقی اجتماعی است که دلالت بر اثر نظارتی - تنظیمی، مالی - اقتصادی و فردی -



اجتماعی دارد. مفهوم «اثر نظارتی - تنظیمی» به معنای قانون گذاری مناسب و برخوردار از قوانین تدوین شده و اجرا شده از طریق نهادهای اجتماعی یا دولتی است که بتواند رفتار عوامل مرتبط با حوزه اطلاعات را تحت نظارت خود داشته باشد (Floridi, 2018, p. Article 20180081). اثر مالی - اقتصادی نیز به معنای آثار مثبت و منفی ناشی از توسعه هوش مصنوعی بر سازمان ها، جوامع و کشورها از تسلط بر بازار، درآمدهای فراوان و صرفه جویی در هزینه گرفته تا نگرانی در خصوص نحوه درآمدزایی داده ها و اقدام های مالی غیر شفاف و انحصار طلبانه است (Grewal & et, 2020, p. 13). اثر فردی - اجتماعی اما ناظر بر تغییرهای تحول آفرین و دگرگونی در جامعه و عاملیت فرد مانند جابه جایی نیروی کار، بیکاری و مهارت زدایی کارکنان در نتیجه استفاده از هوش مصنوعی است (Akter; McCarthy; Sajib; Michael & Dwivedi, 2021, p. Article 102387).

سطح دوم یافته ها که در پاسخ به دومین پرسش اصلی پژوهش است، به نسبت مفاهیم اخلاقی احصاء شده از مرور نظام مند ادبیات موجود با مفهوم «اخلاق شکوفایی» می پردازد. مفهومی که شکوفایی انسان را یک مطالبه عام می داند. شکوفایی انسان بخشی از رویکرد هنجاری به اخلاق است که بایستی هدف استفاده از فناوری ها به ویژه هوش مصنوعی به لحاظ ایجاد «یک زندگی خوب» باشد (Stahl, 2021). بر اساس یافته های پژوهش، مفاهیم اخلاقی «انصاف»، «کرامت انسانی»، «پایداری»، «مسئولیت پذیری» و «مشارکت مردم سالارانه» مفاهیمی هستند که در نمونه های مورد مطالعه به نسبتی که در جدول ۳ نشان داده شده به مفهوم «اخلاق شکوفایی» پرداخته و «شکوفایی انسان» را هدف از بذل توجه به مفاهیم مذکور در فرایند توسعه هوش مصنوعی می دانند.

جدول (۳): نسبت مفاهیم اخلاق فناوری های دیجیتال با اخلاق شکوفایی در مقاله ها

اخلاق فناوری های دیجیتال	تناظر با اخلاق شکوفایی در مقاله ها (درصد)
انصاف	۸۵
کرامت انسانی	۸۰
پایداری	۷۱
مسئولیت پذیری	۷۶
مشارکت مردم سالارانه	۶۳

## نتیجه‌گیری

نفوذ گسترده برنامه‌های کاربردی مبتنی بر هوش مصنوعی در فناوری‌های دیجیتال و متعاقباً در بسیاری از ابعاد وجودی و زیستی انسان، نیاز فوری جامعه جهانی را به توسعه اخلاقی هوش مصنوعی برجسته می‌سازد. پژوهش حاضر به تشریح مفاهیم کلیدی اخلاق دیجیتال در بستر استفاده از هوش مصنوعی در فناوری‌های دیجیتال پرداخته است. بر اساس یافته‌های پژوهش ۱۶ مفهوم اخلاق دیجیتال شناسایی و ذیل پنج الگوی اصلی فناوری‌های دیجیتال و چهار لایه هستی‌شناختی طبقه‌بندی شده‌اند. در این بخش نتایج حاصل از پژوهش ناظر بر پرسش‌های پژوهشی مطرح شده در ابتدای مقاله، در قالب ۱۳ گزاره استنتاجی به بحث گذاشته می‌شود:

گزاره ۱- اهمیت مفهوم اخلاقی «ایمنی» در لایه «دستگاه» فناوری‌های دیجیتال در مقایسه با مفاهیم «کرامت انسانی» و «پایداری»: در لایه «دستگاه» فناوری‌های دیجیتال سه مفهوم اخلاقی «کرامت انسانی»، «ایمنی» و «پایداری» قرار دارند. در این میان به جز مفهوم «ایمنی» که ذیل الگوی محصولات مبتنی بر کنشگر (کاربرد در ۷۵ درصد مقالات) و تعامل تحلیلی (کاربرد در ۳۳ درصد مقالات) بیشترین توجه را از سوی مطالعات به خود اختصاص داده است، دو مفهوم اخلاقی دیگر در هیچ کدام از الگوهای فناوری‌های دیجیتال محل توجه نبوده‌اند. الگوهای اصلی فناوری‌های دیجیتال به واسطه الزاماتی که ذات معماریشان می‌باشد و همچنین انگیزه‌هایی که در طراحی‌شان وجود دارد ممکن است یک حوزه هستی‌شناختی یا یک مفهوم اخلاقی را بر دیگری اولویت دهند.

گزاره ۲- همسویی مفاهیم اخلاقی «کرامت انسانی» و «پایداری» با شکوفایی انسان: «کرامت انسانی» در ۸۰ درصد و «پایداری» در ۷۱ درصد مقالاتی که به این دو مفهوم پرداخته‌اند ذیل طرح بحث اخلاق شکوفایی، مسیری برای شکوفایی انسان در نظر گرفته شده‌اند. به تعبیر الکساندر<sup>۱</sup> (Alexander, 2019, p. 1031) انسان‌ها ذاتاً موجودهایی اجتماعی هستند و تنها زمانی می‌توانند به شکوفایی دست یابند که اجتماعی بودن خود را درک کرده باشند؛ اما درک و تصور از اجتماعی بودن اتفاق نمی‌افتد مگر هنگامی که فرد برای استقلال شخصیتی دیگران احترام قائل شود. از طرفی کرامت انسانی بازنمایی وجود قابلیت به نام «استقلال» در نوع بشر است؛ لذا همچون استقلال، شایسته احترام می‌باشد.

1. Alexander

گزاره ۳- اهمیت قابل توجه مفاهیم اخلاقی «قابل فهم بودن»، «مسئولیت‌پذیری»، «انصاف»، «خودسامانی» و «تنوع‌پذیری» در تمامی الگوهای فناوری‌های دیجیتال: مفاهیم اخلاقی «قابل فهم بودن»، «مسئولیت‌پذیری»، «انصاف»، «خودسامانی» و «تنوع‌پذیری» تقریباً در تمامی الگوهای فناوری‌های دیجیتال محل توجه فراوان بوده‌اند. در این میان مفهوم اخلاقی «انصاف» در الگوی «اتصال» با کاربرد ۸۱ درصدی در مقاله‌ها بیشترین فراوانی را در نمونه مورد مطالعه داشته است.

گزاره ۴- همسویی قابل توجه مفهوم اخلاقی «انصاف» با شکوفایی انسان: مفهوم «انصاف» در ۸۵ درصد مقاله‌هایی که به اخلاق شکوفایی پرداخته‌اند در تناظر با شکوفایی انسان قرار گرفته است. «انصاف» مفهومی مستقر بر اصل اخلاقی «برابری انسان‌ها» است. لیپتون<sup>۱</sup> و همکاران (Lipton & et al., 2018, p. 8) ذیل مفهوم مرکزی انصاف، به سه مقوله مهم «جانبداری»، «دسترسی» و «مشارکت» اشاره می‌کنند. جانبداری به معنای رفتار تبعیض‌آمیز با افراد یا گروه‌هاست. این مقوله علاوه بر انصاف در رفتار، انصاف در اثرگذاری را نیز شامل می‌شود. عمده نگرانی‌ها در این حوزه، ناظر به جانبداری در مجموعه داده‌ها، بهره‌برداری هدفمند از هوش مصنوعی، سوءاستفاده از افراد و کیفیت خدمات ارائه شده توسط هوش مصنوعی در جوامع است (Lipton; McAuley & Chouldechova, 2018, p. 9). در رابطه با «دسترسی» لازم به ذکر است، اگرچه بسیاری از بحث‌ها در اخلاق هوش مصنوعی، به کاهش سطح آسیب‌های ناشی از به‌کارگیری این فناوری اختصاص دارد، اما پُر واضح است که منافع قابل توجهی نیز از سیستم‌های هوش مصنوعی شامل حال افراد و جوامع می‌شود. از این‌رو دسترسی برابر و متوازن به این فناوری نوظهور حائز اهمیت فراوان بوده و تلاش در طراحی‌های مقرون‌به‌صرفه، دوستدار کاربر و نه به تعبیر لیپتون و همکاران (Lipton & et al., 2018, p. 9) «رویکرد یک‌سایز برای همه» در اولویت قرار دارد. مقوله «مشارکت» ناظر بر ارتباط به زبان بومی و قابل توضیح برای کاربران بوده و تعامل جامعه گسترده‌تر با سیستم‌های هوش مصنوعی را تسهیل می‌کند. این مسئله موجب افزایش امکان یادگیری به‌منظور تعامل با هوش مصنوعی و متعاقباً ارتقای سطح رضایت کاربران و جلب نظر ذی‌نفعان خواهد شد. لیپتون و همکاران (Lipton & et al., 2018, p. 5) معتقدند اخلاق شکوفایی زمانی در جامعه مبتنی

بر هوش مصنوعی مستقر می‌شود که این فناوری به دور از هر گونه جانبداری قومی، نژادی، جنسیتی و با نگاه جهانشمول در دسترس همگان قرار گرفته و منافع تمامی گروه‌های ذی‌نفع را با جلب مشارکت حداکثری تأمین نماید. در چنین شرایطی به‌رغم وجود آسیب‌های احتمالی، می‌توان به شکوفایی انسان امیدوار بود.

گزاره ۵- همسویی مفهوم اخلاقی «مسئولیت‌پذیری» با شکوفایی انسان: مفهوم «مسئولیت‌پذیری» در ۷۶ درصد از مقاله‌هایی که به این مفهوم پرداخته‌اند، یکی از شروط لازم برای شکوفایی انسان در نظر گرفته شده است. قبول مسئولیت در حوزه فناوری‌های نوظهور نیازمند توسعه ساختارهای حکمرانی رسمی و غیررسمی و نظارت بر اجرای آنهاست. در این مسیر وجود تنظیم‌کننده‌ها و عاملانی که بتوانند با زیست‌بوم‌های هوش مصنوعی پیوند برقرار کنند، آن را جهت‌دهی نموده و بر آن نظارت داشته باشند به‌منظور اطمینان از قرارگیری هوش مصنوعی در فرایند شکوفایی انسان ضروری است (Kazim; Denny & Wang, 2019, p. 249). به اعتقاد وانگ (Koshiyama, 2021, p. 8). برای نگهداری انسان درون حلقه مسئولیت، بخشی از فرایند شکوفایی انسان به مثابه یک کنشگر پاسخگو است.

گزاره ۶- اهمیت بسیار بالای مفهوم اخلاقی «حریم خصوصی» در تمامی الگوهای فناوری‌های دیجیتال: مفهوم اخلاقی «حریم خصوصی» ذیل لایه محتوا تقریباً در تمامی الگوهای فناوری‌های دیجیتال بالاترین درصد فراوانی را به خود اختصاص داده و به‌طور گسترده در نمونه ما بحث شده است.

گزاره ۷- اهمیت مفاهیم اخلاقی «اثرنظارتی - تنظیمی»، «اثر مالی - اقتصادی» و «اثر فردی - اجتماعی» تقریباً در تمامی الگوهای فناوری‌های دیجیتال: سه مفهوم اخلاقی «اثر نظارتی - تنظیمی»، «اثر مالی - اقتصادی» و «اثر فردی - اجتماعی» تقریباً در تمامی الگوهای فناوری‌های دیجیتال محل توجه بوده است. «اثر نظارتی - تنظیمی» مفهومی است که به‌واسطه فرهنگ سازمانی، رهبری و ارائه مدل کسب‌وکار موجب تولید ارزش نظارتی برای استفاده از هوش مصنوعی در فناوری‌های دیجیتال می‌شود. «اثر مالی - اقتصادی» ارزشی است که سازمان‌ها را به استفاده از هوش مصنوعی در فناوری‌های دیجیتال تحریک

می‌کند. کومار<sup>۱</sup> و دیگران (Kumar, 2021, p. 19) معتقدند سازمان‌ها برای آنکه بتوانند به ارزش اقتصادی دست پیدا کنند، بایستی مسئولیت اجتماعی را سرلوحه فعالیت‌های سازمانی خود قرار دهند. یعنی نسبت به تأثیر فعالیت‌هایشان بر جامعه از طریق رفتار شفاف و اخلاقی حساس باشند و اثر فردی - اجتماعی توسعه هوش مصنوعی را در تصمیماتشان در نظر بگیرند.

گزاره ۸- همسویی مفهوم اخلاقی «مشارکت مردم‌سالارانه» با شکوفایی انسان: استفاده از هوش مصنوعی بایستی همراه با آگاهی از نیازهای متنوع و متناقض و نیز مشارکت گسترده همه گروه‌های ذی‌نفع باشد. این امر تحت عنوان مشارکت مردم‌سالارانه با هدف شکوفایی انسان، محل توجه ۶۳ درصد از مقاله‌های نمونه مورد مطالعه بوده است. لازم به ذکر است فرایندهای سیاسی موجود در یک دولت مردم‌سالار تا حد زیادی می‌توانند پاسخگوی نیاز به مشارکت گسترده گروه‌های ذی‌نفع در جریان توسعه هوش مصنوعی باشند.

گزاره ۹- اهمیت لایه هستی‌شناختی «محتوا» برای الگوی «پلتفرم»: مضامین مرتبط با الگوهای فناوری‌های دیجیتال مبتنی بر پلتفرم که در حقیقت فناوری‌های زیرساختی هستند به عنوان قطب انتقال داده از یک منبع مشخص و با یک طراحی یک به چند می‌باشند (Berger; Denner & Roeglinger, 2018, p. 5). داده‌های مشخص جزء اساسی این الگو هستند. لذا در نمونه مورد مطالعه ما لایه هستی‌شناختی محتوا به هنگام استفاده از هوش مصنوعی در الگوی مذکور مورد توجه فراوان قرار گرفته است.

گزاره ۱۰- اهمیت قابل توجه لایه‌های هستی‌شناختی «محتوا» و «خدمات» برای الگوی «اتصال»: الگوی اتصال نیز همانند الگوی پلتفرم جزء فناوری‌های زیرساختی محسوب می‌شود و مضامین مرتبط با آن به تعبیر برگر<sup>۲</sup> و همکاران (Berger, 2018, p. 5) از دیگر ابعاد انتقال داده هستند که هدفشان علاوه بر انتقال داده، «توان عملیاتی کارآمد» می‌باشد. از این‌رو در مرور ادبیات موجود نه تنها لایه هستی‌شناختی «محتوا» بلکه لایه «خدمات» نیز در الگوی «اتصال» محل توجه فراوان مقاله‌ها بوده است.

گزاره ۱۱- اهمیت لایه‌های هستی‌شناختی «دستگاه» و «خدمات» برای الگوی «محصولات مبتنی بر کنشگر»: مضامین مرتبط با الگوهای «محصولات مبتنی بر

1. Kumar

2. Berger

کنشگر»، در واقع فناوری‌های کاربردی‌ای هستند که کارکرد اصلی‌شان اجرای داده‌ها از ورودی دیجیتال به خروجی فیزیکی است (Berger; Denner & Roeglinger, 2018, p. 18)؛ بنابراین استفاده از هوش مصنوعی در این الگو هم بر روی لایه دستگاه با توانایی اثرگذاری بر محیط فیزیکی و هم بر روی لایه خدمات به‌عنوان یک موتور شناختی تمرکز بیشتری دارد.

گزاره ۱۲- اهمیت لایه‌های هستی‌شناختی «محتوا» و «خدمات» برای الگوی «تولید بینش تحلیلی»: الگوی «تولید بینش تحلیلی» به‌عنوان یکی از الگوهای پرکاربرد در نمونه مورد بررسی ما بوده است. از این‌رو مباحث تفسیری در حوزه دستورالعمل‌ها و یادگیری ماشین جزء مباحث ضروری و با اهمیت در بستر اخلاق دیجیتال و برنامه‌های کاربردی هوش مصنوعی است. فناوری‌های به‌کاررفته در الگوی «تولید بینش تحلیلی» به اعتقاد برگر و همکاران (Berger, 2018, p. 19) فناوری‌های کاربردی هستند که کارکردشان منطبق و تصمیم‌گیری است. این الگو در نمونه مورد مطالعه ما هم از کلان‌داده‌ها و هم از قابلیت‌های پیشرفته یادگیری ماشین استفاده می‌کند تا با تمرکز بر لایه‌های محتوا و خدمات، کارکردهای شناختی انسان را تکرار نماید.

گزاره ۱۳- اهمیت لایه هستی‌شناختی «محتوا» برای الگوی «تعامل تحلیلی»: «تعامل تحلیلی» به‌عنوان یکی از الگوهای فناوری‌های دیجیتال است که با پیشرفت هوش مصنوعی امکان‌پذیر شده است و مضامین آن، در واقع فناوری‌های کاربردی مرتبط با انتقال ورودی دیجیتال، قابلیت‌های تحلیلی و خروجی فیزیکی هستند (Berger; Denner & Roeglinger, 2018, p. 20). نقشه مفهومی ما بیانگر استفاده از هوش مصنوعی در این الگو با تمرکز بر لایه محتوا و با کارکرد انتقال داده است که اغلب به‌خاطر پرداختن به موضوع جایگزینی مشاغل خدماتی، تأثیرهای فردی و اجتماعی فراوانی به‌دنبال دارد.

مدل مفهومی ارائه شده در مقاله حاضر، به پژوهشگران حوزه‌های مختلف به‌ویژه طراحان دستورالعمل‌ها و توسعه‌دهندگان هوش مصنوعی کمک می‌کند تا ضمن توجه به ملاحظه‌های اخلاقی هر الگو برای ایجاد یک جامعه مبتنی بر هوش مصنوعی اخلاقی به توسعه فعالیت‌هایشان بپردازند. همچنین می‌تواند برای سیاست‌گذاران حوزه‌های فرهنگی و اجتماعی در جهت ارائه راهبردهایی با هدف کاربست اخلاقی هر یک از انواع فناوری‌های دیجیتال در جامعه مفید باشد. در این مسیر پژوهشگران به دلیل منحصربه‌فرد نبودن طبقه‌بندی‌های مفهومی صورت گرفته، احتمالاً با

همپوشانی‌هایی نیز روبه‌رو خواهند بود که از محدودیت‌های پژوهش حاضر است. این محدودیت ناشی از ذات معماری فناوری‌های دیجیتال می‌باشد که موجب تداخل‌هایی در الگویابی می‌شود؛ به‌عنوان مثال طراحی چهرک‌ها (آواتارها) هم نیازمند تولید بینش تحلیلی است و هم تعامل تحلیلی. اگرچه سهم تعامل تحلیلی بیشتر است ولی محقق بایستی به دلالت‌های اخلاقی هر دو الگو توجه نماید که گاه موجب ایجاد تنش‌های اخلاقی می‌شود. با پیشرفت سریع فناوری‌های دیجیتال، همچنین کامل‌تر و تفکیک‌پذیرتر شدن معماری این فناوری‌ها در پژوهش‌های آتی این امکان وجود خواهد داشت تا الگوهای جدیدی با همپوشانی‌های مفهومی کمتر توسط پژوهشگران به مدل مفهومی ارائه شده در پژوهش حاضر اضافه شود.

### فهرست منابع

- زادیوسفی، امیرحسین (۱۴۰۰). *هوش مصنوعی و شکاف مسئولیت*. تهران: پژوهشگاه فضای مجازی.
- کاظمی، علیرضا و مطلبی کربکندی، حسین (۱۴۰۰). *اینترنت اشیاء اجتماعی: چشم‌انداز، فرصت‌ها و چالش‌ها*. تهران: پژوهشگاه فضای مجازی.
- ناثلی، مینا (۱۴۰۱). *مروری بر فعالیتهای آموزشی کشورهای پیشرو در حوزه فناوری اطلاعات و ارتباطات جهت افزایش صلاحیتهای هوش دیجیتال در سنین پایه*. تهران: پژوهشگاه فضای مجازی.
- Adomavicius, G.; Bockstedt, J.; Gupta, A. & Kauffman, R. (2007). Technology roles and paths of influence in an ecosystem model of technology evolution. *Information Technology and Management*, 8(2), 185-202.
- Akter, S.; McCarthy, G.; Sajib, S.; Michael, K. & Dwivedi, Y. (2021). Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management*, 60: Article 102387.
- Alexander, G. (2019). Property, Dignity, and Human flourishing. *Cornell Law Review*, 104(4), 991-1046.
- Berger, S.; Denner, M. & Roeglinger, M. (2018). The nature of digital technologies- development of a multi-layer taxonomy. *26th European Conference on Information Systems (ECIS)* (pp. 1-20). Portsmouth, UK.: Finance&Information Management.
- Booth, A.; Papaioannou, D. & Sutton, A. (2012). *Systematic approaches to a successful literature review*. London: SAGE Publications.

- Borges, A.; Laurindo, F.; Spinola, M.; Goncalves, R. & Mattos, C. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management*, (57), Article 102225.
- Brey, P. (2012). Anticipatory ethics for emerging technologies. *Nanoethics*, (6), 1-13.
- Brynjolfsson, E. & McAfee, A. (2014). *The second Machine age: Work, Progress, and Prosperity in a time of Brilliant technologies*. New York: W.W.Norton and Company.
- Bynum, T. (2006). Flourishing ethics. *Ethics Information Technology*, (8), 157-173.
- Choi, H.; Jung, S.; Baek, S.; Lim, W.; Ahn, S. & Yang, I. (2019). Artificial intelligent model with neural network machine learning for the diagnosis of orthognathic surgery. *J Craniofac Surg.*, 30(7), 1986-1989.
- Duan, Y.; Edwards, J. & Dwivedi, Y. (2019). Artificial intelligence for decision making in the era of Big Data-evolution, challenges and research agenda. *International Journal of Information Management*, (48), 63-71.
- Dwivedi, Y.; Hughes L.; Ismagilova E.; Aarts G.; Coomb, L.; Ismagilova, E.; Aarts, G.; Coombs, C. & Crick, T. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, (57), Article101994.
- Feenberg, A. (1999). *Questioning technology*. London: Routledge.
- Floridi, L. (2006). Information ethics, its nature and scope. *ACM SIGCAS Computers and Society*, (36), 21-36.
- Floridi, L. (2018). *Philosophical Transactions of The royal Society, Series A, Mathematical, Physical, and Engineering Sciences*, 376(2133): Article 20180081.
- Floridi, L. & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1), 2-17.
- Floridi, L. (2011). *The Philosophy of Information*. city: Oxford University Press.
- Floridi, L. (2017). The logic of design as a conceptual logic of information. *Minds Machines*, 1-25.
- Grewal, D.; Kroschke, M.; Mende, M.; Mende, M.; Roggeveen, A. & Scot. (2020). Frontline cyborgs at your service: How human enhancement technologies affect customer experiences in retail, sales, and service settings. *Journal of Interactive Marketing*, (51), 9-25.
- Johnstone, J. (2007). Technology as empowerment: a capability approach to computer ethics. *Ethics and Information Technology*, 9(1), 73-87.
- Kankanhalli, A.; Charalabidis, Y. & Mellouli, S. (2019). IoT and AI for smart government: A research agenda. *Government Information Quarterly*, 36(2), 304-309.



- Kazim, E.; Denny, D. & Koshiyama, A. (2021). AI auditing and impact assessment: according to the UK information commissioner's office. *AI and Ethics*, 1-10.
- Kumar, P.; Dwivedi, Y. & Anand, A. (2021). Responsible Artificial Intelligence (AI) for value formation and market performance in healthcare: The mediating role of patient's cognitive engagement. *Information Systems frontiers. Journal of Research and Innovation*, 1-24.
- Lipton, Z.; McAuley, J. & Chouldechova, A. (2018). Does mitigating ML's impact disparity require treatment disparity? *32nd conference on Neural information Processing Systems* (pp. 1-11). Montreal, Canada: Montreal.
- MacIntyre, A. (2007). *After virtue: a study in moral theory*. Notre Dame: University of Notre Dame Press.
- Malkiat, T.; Sarunkorn, C. & Stephen, J. (2019). Balancing information governance obligations when accessing social care data for collaborative research. *Records Management Journal*, 29(1/2), 194-209.
- Metcalf, J.; Moss, E. & Boyd, D. (2019). Owing ethics: Corporate logics, silicon valley, and the institutionalization of ethics. *Social Research*, 86(2), 449-476.
- Morley, J.; Floridi, L.; Kinsey, L. & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, 26(4), 2141-2168.
- Mustak, M.; Jaakkola, E.; Halinen, A. & Kaartemo, V. (2016). Customer participation management: Developing a comprehensive framework and a research agenda. *Journal of Service Management*, 27(3), 250-275.
- Myers, M. & Klein, H. (2011). A set of principles for conducting critical research in information systems. *MIS Quarterly*, 35(1), 17-36.
- Nishant, R.; Kennedy, M. & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. *International Journal of Information Management*, (53), Article 102104.
- Nussbaum, M. (2011). *Creating capabilities: the human development approach*. Cambridge: Harvard university Press.
- Oosterlaken, I. & van den Hoven, J. (2012). *The capability approach, technology and design*. Dordrecht, Netherlands: Springer.
- Rhim, J.; Lee, G. & Lee, J. (2020). Human moral reasoning types in autonomous vehicle moral dilemma: A cross-cultural comparison of Korea and Canada. *Computers in Human Behavior*, (102), 39-56.
- Sarker, M.; Wu, M.; Cao, Q.; Alam, G. & Li, D. (2019). Leveraging digital technology for better learning and education: A systematic literature review. *International journal of Information and Education Technology*, 9(7), 453-461.

- Schappert, M., & von Hauff, M. (2020). Sustainable consumption in the smart grid: From key points to eco-routine. *Journal of Cleaner Production*, 267: Article 121585.
- Sen, A. (2009). *The idea of justice*. London: Allen Lane.
- Stahl, B. (2008). The ethical nature of critical research in information systems. *Informations systems Journal*, (18), 137-163.
- Stahl, B. (2021). Artificial Intelligence for a Better Future, An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies. city: Springer.
- Stahl, B. (2021). Artificial Intelligence for a Better Future, An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies. city: Springer.
- Sutton, S.; Arnold, V. & Holt, M. (2018). How much automation is too much? Keeping the human relevant in knowledge work. *Journal of Emerging Technologies in Accounting*, 15(2), 15-25.
- Urquhart, L.; Schnadelbach, H. & Jager, N. (2019). Adaptive Architecture: Regulating human building interaction. *International Review of Law, Computers & Technology*, 33(1), 3-33.
- Van den Hoven, J. (2010). The use of normative theories in computer ethics. In L. Floridi, *The Cambridge handbook of information and computer ethics* (pp. 59-76). UK: Cambridge University Press.
- Von Schomberg, R. (2013). A vision of responsible research and innovation. In R. Owen, M. Heintz, & J. Bessant, *Responsible innovation* (pp. 51-74). London: John Wiley.
- Wang, G. (2019). Humans in the Loop: The Design of interactive AI Systems. *Journal of Artificial Intelligence Research*, 64:243-252.
- Yoon, S.; Choi, T.; Odlum, M.; Mitchell, D. & Kronish, I. (2018). Machine learning to identify behavioral determinants of oral health in inner city older Hispanic adults. *Stud Health Technol Inform*, 251(1), 253-256.

## References

- Adomavicius, G., Bockstedt, J., Gupta, A. & Kauffman, R. (2007). Technology roles and paths of influence in an ecosystem model of technology evolution. *Information Technology and Management*, 8(2), 185-202.
- Akter, S., McCarthy, G., Sajib, S., Michael, K. & Dwivedi, Y. (2021). Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management*, 60: Article 102387.
- Alexander, G. (2019). Property, Dignity, and Human flourishing. *Cornell Law Review*, 104(4), 991-1046.
- Berger, S., Denner, M. & Roeglinger, M. (2018). The nature of digital technologies- development of a multi-layer taxonomy. 26th European Conference on Information Systems (ECIS) (1-20). Portsmouth, UK.: Finance&Information Management.

- Booth, A., Papaioannou, D., & Sutton, A. (2012). Systematic approaches to a successful literature review. London: SAGE Publications.
- Borges, A., Laurindo, F., Spinola, M., Goncalves, R., & Mattos, C. (2021). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *International Journal of Information Management*, 57, Article 102225.
- Brey, P. (2012). Anticipatory ethics for emerging technologies. *Nanoethics*, (6), 1-13.
- Brynjolfsson, E., & McAfee, A. (2014). *The second Machine age: Work, Progress, and Prosperity in a time of Brilliant technologies*. New York: W.W.Norton and Company.
- Bynum, T. (2006). Flourishing ethics. *Ethics Information Technology*, (8), 157-173.
- Choi, H., Jung, S., Baek, S., Lim, W., Ahn, S., & Yang, I. (2019). Artificial intelligent model with neural network machine learning for the diagnosis of orthognathic surgery. *J Craniofac Surg.*, 30(7), 1986-1989.
- Duan, Y., Edwards, J., & Dwivedi, Y. (2019). Artificial intelligence for decision making in the era of Big Data-evolution, challenges and research agenda. *International Journal of Information Management*, (48), 63-71.
- Dwivedi, Y., Hughes L, Ismagilova E, Aarts G, Coomb, L., Ismagilova, E., Aarts, G., Coombs, C., & Crick, T. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, (57), Article101994.
- Feenberg, A. (1999). *Questioning technology*. London: Routledge.
- Floridi, L. (2006). Information ethics, its nature and scope. *ACM SIGCAS Computers and Society*, (36), 21-36.
- Floridi, L. (2018). *Philosophical Transactions of The royal Society, Series A, Mathematical, Physical, and Engineering Sciences*, 376(2133): Article 20180081.
- Floridi, L., & Cowl, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1), 2-17.
- Floridi, L. (2011). *The Philosophy of Information*. city: Oxford University Press.
- Floridi, L. (2017). The logic of design as a conceptual logic of information. *Minds Machines*, 1-25.
- Grewal, D., Kroschke, M., Mende, M., Mende, M., Roggeveen, A., & Scot. (2020). Frontline cyborgs at your service: How human enhancement technologies affect customer experiences in retail, sales, and service settings. *Journal of Interactive Marketing*, (51), 9-25.
- Johnstone, J. (2007). Technology as empowerment: a capability approach to computer ethics. *Ethics and Information Technology*, 9(1), 73-87.

- Kankanhalli, A., Charalabidis, Y., & Mellouli, S. (2019). IoT and AI for smart government: A research agenda. *Government Information Quarterly*, 36(2), 304-309.
- Kazemi, A. R., & Motallebi Korbekandi, H. (2021). Social internet of things: Perspectives, Opportunities, and challenges. Tehran: CyberSpace Research Institute. (In Persian)
- Kazim, E., Denny, D., & Koshiyama, A. (2021). AI auditing and impact assessment: according to the UK information commissioner's office. *AI and Ethics*, 1-10.
- Kumar, P., Dwivedi, Y., & Anand, A. (2021). Responsible Artificial Intelligence (AI) for value formation and market performance in healthcare: The mediating role of patient's cognitive engagement. *Information Systems frontiers. Journal of Research and Innovation*, 1-24.
- Lipton, Z., McAuley, J., & Chouldechova, A. (2018). Does mitigating ML's impact disparity require treatment disparity? 32nd conference on Neural information Processing Systems (pp. 1-11). Montreal, Canada: Montreal.
- MacIntyre, A. (2007). *After virtue: a study in moral theory*. Notre Dame: University of Notre Dame Press.
- Malkiat, T., Sarunkorn, C., & Stephen, J. (2019). Balancing information governance obligations when accessing social care data for collaborative research. *Records Management Journal*, 29(1/2), 194-209.
- Metcalfe, J., Moss, E., & Boyd, D. (2019). Owing ethics: Corporate logics, silicon valley, and the institutionalization of ethics. *Social Research*, 86(2), 449-476.
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, 26(4), 2141-2168.
- Mustak, M., Jaakkola, E., Halinen, A., & Kaartemo, V. (2016). Customer participation management: Developing a comprehensive framework and a research agenda. *Journal of Service Management*, 27(3), 250-275.
- Myers, M., & Klein, H. (2011). A set of principles for conducting critical research in information systems. *MIS Quarterly*, 35(1), 17-36.
- Naeli, M. (2022). A review of educational activities of leading countries in the field of information and communication technology to increase the competencies of digital quotient at basic ages. Tehran: CyberSpace Research Institute. (In Persian)
- Nishant, R., Kennedy, M., & Corbett, J. (2020). Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda. *International Journal of Information Management*, (53), Article 102104.
- Nussbaum, M. (2011). *Creating capabilities: the human development approach*. Cambridge: Harvard university Press.

- Oosterlaken, I., & van den Hoven, J. (2012). *The capability approach, technology and design*. Dordrecht, Netherlands: Springer.
- Rhim, J., Lee, G., & Lee, J. (2020). Human moral reasoning types in autonomous vehicle moral dilemma: A cross-cultural comparison of Korea and Canada. *Computers in Human Behavior*, (102), 39-56.
- Sarker, M., Wu, M., Cao, Q., Alam, G., & Li, D. (2019). Leveraging digital technology for better learning and education: A systematic literature review. *International journal of Information and Education Technology*, 9(7), 453-461.
- Schappert, M., & von Hauff, M. (2020). Sustainable consumption in the smart grid: From key points to eco-routine. *Journal of Cleaner Production*, (267), Article 121585.
- Sen, A. (2009). *The idea of justice*. London: Allen Lane.
- Stahl, B. (2008). The ethical nature of critical research in information systems. *Informations systems Journal*, (18), 137-163.
- Stahl, B. (2021). *Artificial Intelligence for a Better Future, An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies*. city: Springer.
- Sutton, S., Arnold, V., & Holt, M. (2018). How much automation is too much? Keeping the human relevant in knowledge work. *Journal of Emerging Technologies in Accounting*, 15(2), 15-25.
- Urquhart, L., Schnadelbach, H., & Jager, N. (2019). Adaptive Architecture: Regulating human building interaction. *International Review of Law, Computers & Technology*, 33(1), 3-33.
- van den Hoven, J. (2010). The use of normative theories in computer ethics. In L. Floridi, *The Cambridge handbook of information and computer ethics* (pp. 59-76). UK: Cambridge University Press.
- Von Schomberg, R. (2013). A vision of responsible research and innovation. In R. Owen, M. Heintz, & J. Bessant, *Responsible innovation* (pp. 51-74). London: John Wiley.
- Wang, G. (2019). Humans in the Loop: The Design of interactive AI Systems. *Journal of Artificial Intelligence Research*, (64), 243-252.
- Yoon, S., Choi, T., Odum, M., Mitchell, D., & Kronish, I. (2018). Machine learning to identify behavioral determinants of oral health in inner city older Hispanic adults. *Stud Health Technol Inform*, 251(1), 253-256.
- Zadyusefi, A. H. (2021). *Artificial intelligence and responsibility gap*. Tehran: CyberSpace Research Institute. (In Persian)