



The Concept of Human-Centricity in Sociological Studies of Artificial Intelligence

Somayeh Bolboli Ghadikolaei¹  | Hamid Parsania² 

1. Corresponding Author, Department of Anthropology Faculty of Humanities and Social Sciences, University of Mazandaran, Babolsar, Iran. E-mail: bolboli@umz.ac.ir
2. Department of Sociology, Faculty of Social Sciences, University of Tehran, Tehran, Iran. Email: h.parsania@ut.ac.ir

Article Info

Article type:

Research Article

Article history:

Received: 23 April 2024

Received in revised form: 11 August 2024

Accepted: 12 September 2024

Published online: 21 October 2024

Keywords:

Human-Centricity, Artificial Intelligence, Sociology, Systematic Review, Double Diamond Innovative Design Perspective.

ABSTRACT

Today, we are witnessing a wide range of concerns about the jeopardy of the human position in relation to artificial intelligence-based technologies. The study field of human-centricity artificial intelligence was established in response to these concerns, as it was necessary for human involvement in the development of these systems. The current research is an endeavor to encourage and fortify the convergence of two potential study forces—sociology and artificial intelligence—in order to establish the human-centricity concept.

A systematic review of the existing literature was implemented in this investigation. The projectile sampling strategy was implemented to conduct the sampling. The primer set and articles were selected in accordance with the PRISMA guidelines. To analyze the data, we used Double Diamond innovative design framework.

The research findings were categorized into eight levels, including: Human-centricity artificial intelligence; Problems of human-centricity artificial intelligence in the production of social sciences and sociology; Interdisciplinary research on the human-centricity concept in sociological studies of artificial intelligence; The position of human-centricity artificial intelligence in the methodology of social sciences; Principles and regulations governing the explainability of artificial intelligence; The significance of the human-centricity concept in the comprehension and perception of artificial intelligence by actors; Artificial intelligence and human interaction; Specific disciplines of study that are oriented toward human artificial intelligence.

Two problems related to artificial intelligence and three problems related to humans (in the general sense) were identified through the systematic review of the existing literature. The coexistence of the issues, as well as the adoption of varying perspectives by the articles under investigation in relation to the aforementioned issues, has resulted in the development of 22 distinct approaches to addressing these issues.

Cite this article: Bolboli Ghadikolaei, S. & Parsania, H. (2024). The Concept of Human-Centricity in Sociological Studies of Artificial Intelligence, *Quarterly of Social Studies and Research in Iran*, 13(3): 425- 449.

<https://doi.org/10.22059/jisr.2024.376963.1509>



© The Author(s).

Publisher: University of Tehran Press.

DOI: <https://doi.org/10.22059/jisr.2024.376963.1509>

مفهوم انسان محوری در مطالعات جامعه‌شناختی هوش مصنوعی

سمیه بلبلی قادیکلانی^۱ | حمید پارسانیا^۲

۱. نویسنده مسئول، گروه مردم‌شناسی، دانشکده علوم انسانی و اجتماعی، دانشگاه مازندران، بابلسر، ایران. رایانامه: bolboli@umz.ac.ir
 ۲. گروه جامعه‌شناسی، دانشکده علوم اجتماعی، دانشگاه تهران، تهران، ایران. رایانامه: h.parsania@ut.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: پژوهشی	امروزه شاهد دامنه گسترده نگرانی‌ها از به‌مخاطره‌افتادن موقعیت انسان در نسبت با فناوری‌های مبتنی بر هوش مصنوعی هستیم. در پاسخ به این نگرانی‌ها، حوزه مطالعاتی هوش مصنوعی انسان‌محور به دلیل نیاز انسان به استفاده از چنین سیستم‌هایی شکل گرفت. پژوهش حاضر به تحریک و تقویت دو نیروی بالقوه مطالعاتی - جامعه‌شناسی و هوش مصنوعی - برای هرچه نزدیک‌تر شدن به یکدیگر و شکل‌گیری مفهوم انسان‌محوری می‌پردازد.
تاریخ دریافت: ۱۴۰۳/۲/۴	پژوهش پیش‌رو ذیل جنبش روش‌شناختی کیفی از مرور نظام‌مند مطالعات موجود به‌عنوان روش تحقیق استفاده کرده است. نمونه‌گیری به روش گلوله‌برفی انجام گرفت. فرایند گزینش مقالات و انتخاب مجموعه آغازگر طبق دستورالعمل پریسما انجام شد. تجزیه و تحلیل داده‌ها نیز با چارچوب طراحی نوآورانه الماس دوبر انجام گرفت.
تاریخ بازنگری: ۱۴۰۳/۵/۲۱	یافته‌های پژوهش در هشت سطح طبقه‌بندی شدند که عبارت‌اند از: معضلات هوش مصنوعی انسان‌محور؛ هوش مصنوعی انسان‌محور در تولید علوم اجتماعی و جامعه‌شناسی؛ پژوهش‌های میان‌رشته‌ای درباره مفهوم انسان‌محوری در مطالعات جامعه‌شناختی هوش مصنوعی؛ جایگاه هوش مصنوعی انسان‌محور در روش‌شناسی علوم اجتماعی؛ اصول و مقررات ناظر بر توضیح‌پذیری و تبیین‌پذیری هوش مصنوعی؛ نقش انسان‌محوری در درک و تصور کنشگران از هوش مصنوعی؛ تعامل انسان با هوش مصنوعی؛ هوش مصنوعی انسان‌محور در حوزه‌های مطالعاتی خاص.
تاریخ پذیرش: ۱۴۰۳/۶/۲۲	از مرور نظام‌مند مطالعات موجود، دو معضل ناظر به هوش مصنوعی و سه معضل ناظر به انسان (در معنای عام) شناسایی شدند. هم‌نهشت معضلات، همچنین اتخاذ مواضع متفاوت مقالات درمورد معضلات مذکور، موجب شکل‌گیری رویکردهای مختلف (در مجموع ۲۲ رویکرد) در مواجهه با این معضلات شده است.
تاریخ انتشار: ۱۴۰۳/۷/۳۰	
کلیدواژه‌ها:	
انسان‌محوری، جامعه‌شناسی، چارچوب طراحی نوآورانه الماس دوبر، مرور نظام‌مند، هوش مصنوعی.	

استناد: بلبلی قادیکلانی، سمیه و پارسانیا، حمید (۱۴۰۳). مفهوم انسان‌محوری در مطالعات جامعه‌شناختی هوش مصنوعی. *مطالعات و تحقیقات اجتماعی در ایران*، ۱۳(۳): ۴۲۵-۴۴۹. <https://doi.org/10.22059/jisr.2024.376963.1509>



ناشر: مؤسسه انتشارات دانشگاه تهران. © نویسندگان.

DOI: <https://doi.org/10.22059/jisr.2024.376963.1509>

۱. مقدمه و بیان مسئله

امروزه جامعه‌شناسی به‌عنوان یک علم و رشته دانشگاهی دستخوش دگرگونی‌های اساسی شده است. همچنین اتفاق نظری کلی میان دانشمندان علوم اجتماعی درخصوص کارآمدی علم جامعه‌شناسی برای بهبود وضعیت جامعه به‌دلیل توانایی مطالعه پدیده‌های اجتماعی وجود دارد (کالینز، ۱۹۹۴: ۱۶۳). اکنون این پرسش مطرح می‌شود که آیا می‌توان همین رویکرد را نسبت به هوش مصنوعی به‌مثابه پدیده‌ای که موضوع مطالعه علوم طبیعی و مهندسی طی چند دهه گذشته است، داشت؟ به عبارتی دیگر آیا می‌توان ادعا کرد که علم جامعه‌شناسی می‌تواند به ما در شناخت الگوریتم و فرایند توسعه الگوریتم‌ها و فناوری‌های هوش مصنوعی در جامعه کمک کند؟ و حتی یک گام جلوتر، آیا می‌توان گفت هوش مصنوعی، جامعه‌شناسی را برای همیشه تغییر خواهد داد؟

طرح این پرسش‌ها برای آن دسته از دانشمندان علوم اجتماعی که پیوند فناوری به‌ویژه فناوری‌های نوظهور با علم جامعه‌شناسی را دنبال می‌کنند نه‌تنها عجیب نیست، بلکه پرداخت هرچه بیشتر آنان به موضوع هوش مصنوعی به‌مثابه یکی از مظاهر فناوری‌های نوین به درک هرچه عمیق‌تر مطالعات جامعه‌شناختی هوش مصنوعی کمک خواهد کرد.

منظور از مطالعات جامعه‌شناختی هوش مصنوعی در یک سطح، بررسی رابطه متقابل و اثرگذاری و اثرپذیری دو حوزه «جامعه‌شناسی» و «هوش مصنوعی» است. حال آنکه بررسی در این سطح، مطمح نظر نویسندگان مقاله حاضر نیست. به بیان دیگر مقاله حاضر به هیچ وجه به دنبال تبیین سودمندی جامعه‌شناسی در مطالعه هوش مصنوعی یا تأثیر هوش مصنوعی بر جامعه‌شناسی نیست؛ چرا که پژوهش در این سطح مستلزم مقدمات و مطالعاتی بنیادین است که بتواند دو حوزه مهم دانشی جامعه‌شناسی و هوش مصنوعی را هرچه بیشتر به هم نزدیک کند؛ بنابراین در مقاله حاضر در یک گام عقب‌تر، از زاویه‌ای دیگر به مسئله ورود می‌شود و با تمرکز بر جنبه خاصی از تخیل جامعه‌شناختی به نام «هوش مصنوعی انسان‌محور»^۱ بحث پیش می‌رود. اما احتمالاً این پرسش مطرح می‌شود که چرا نقطه عزیمت پژوهش، ظهور هوش مصنوعی انسان‌محور در مجامع علمی قرار داده شده است.

رسالت اصلی هوش مصنوعی به‌عنوان یکی از رشته‌های حوزه علوم کامپیوتر، پژوهش درباره الگوریتم‌ها و ساخت آن‌ها است؛ الگوریتم‌هایی که خروجی آن‌ها انجام یکسری فعالیت‌ها و نمایش پاره‌ای رفتارها است؛ چنان‌که انتظار می‌رود برای بروز آن فعالیت‌ها و ظهور آن رفتارها هوش انسانی دخیل باشد (ریدل، ۲۰۱۹: ۳۵).

شایان ذکر است امروزه اصول، رویکردها و قواعد ناظر به هوش مصنوعی روی تمامی سیستم‌های رایانه‌ای از دستیار هوشمند اپل سیری^۲ تا دستیار صوتی هوشمند آمازون الکسا^۳ تعبیه شده است. اخیراً نیز زمزمه‌های بسیاری در مجامع علمی دنیا درخصوص جایگزینی کامل نیروی کار انسانی در صنایع مختلف با هوش مصنوعی به گوش می‌رسد (پوپنیک و کر، ۲۰۱۷). همچنین بسیاری از مکاتب فکری مدت‌هاست زنگ خطر امکان دستیابی سیستم‌های مبتنی بر هوش مصنوعی به‌نوعی استقلال و خودسامانی را به صدا درآورده‌اند. از طرف دیگر تصمیم‌گیری‌های سوگیرانه که محصول دریافت اطلاعات نادرست و ناکافی درباره سوژه‌هاست، بر دامنه این نگرانی‌ها افزوده است. همین مسئله زمینه‌ساز ایجاد یک سری آزمایشگاه‌های تحقیقاتی در حوزه هوش مصنوعی انسان‌محور شده است. هدف این آزمایشگاه‌ها بررسی امکان شکل‌گیری یک رویکرد انسان‌محورتر در فرایند توسعه هوش مصنوعی است (زو، ۲۰۱۹: ۴۳). از این‌رو مفهوم «هوش مصنوعی انسان‌محور» به‌دلیل نیاز انسان به چنین سیستم‌هایی شکل گرفت. به عبارت دیگر اساس

1. Human-centered AI

2. Apple Siri

3. Amazon Alexa

شکل‌گیری هوش مصنوعی انسان‌محور که شاخه‌ای از هوش مصنوعی و یادگیری ماشینی است، ناظر بر یک چشم‌انداز کلان‌تر است؛ یعنی سیستم‌های هوشمند باید براساس رویکردی که آن‌ها را جزئی از یک سیستم گسترده‌تر شامل ذی‌نفعان انسانی (کاربران، مشتریان، اپراتورها و...) می‌دانند بنا شوند (ریدل، ۲۰۱۹: ۳۴). این در حالی است که اهمیت پایبندی به اصول هوش مصنوعی و هوش مصنوعی انسان‌محور در حوزه‌های مختلف اعم از آموزش، رباتیک، کشاورزی، انرژی، اقتصاد چرخه‌ای، شهرهای هوشمند، شبکه‌های هوشمند، اتومبیل‌های خودران و خانه‌های هوشمند انکارناپذیر است.

هوش مصنوعی انسان‌محور مبتنی بر طراحی سیستم‌های فناورانه هوش مصنوعی است که حول محور انسان، تعامل او با فناوری و ناظر بر اولویت‌بخشی به ارزش‌ها، نیازها، ملاحظات اخلاقی، رفاه و تجربیات انسانی مفصل‌بندی می‌شود. اساساً تمام قوانین و اهداف مربوط به توسعه هوش مصنوعی انسان‌محور، بر دستیابی به بالاترین سطح کنترل انسانی و همچنین سطوح بالای اتوماسیون برای تولید سیستم‌ها و برنامه‌های کاربردی مطمئن، ایمن و معتمد متمرکز است. درواقع استفاده از استانداردهای فناورانه مناسب برای ارتقای قابلیت اعتماد سیستم، روش‌های مدیریتی برای ترویج و حفظ فرهنگ ایمنی، و وجود مکانیسم‌های نظارتی مستقل برای تضمین قابلیت اطمینان سیستم در اولویت رویکردهای انسان‌محور به مقوله هوش مصنوعی قرار دارد.

بنابراین هدف مقاله حاضر این است که ضمن مرور نظام‌مند مطالعات موجود در حوزه هوش مصنوعی انسان‌محور، به طرح پرسش‌های جدید پردازد و با جمع‌آوری داده‌های علمی بیشتر و فراترکیب کیفی آن‌ها بتواند زمینه تحریک و تقویت دو نیروی بالقوه مطالعاتی جامعه‌شناسی و هوش مصنوعی را برای هرچه نزدیک‌تر شدن به یکدیگر فراهم آورد. از این رو پژوهش حاضر به دنبال پاسخگویی به پرسش‌های اساسی زیر است: ۱. در حوزه هوش مصنوعی انسان‌محور چه مطالعاتی قابل‌شناسایی است؟ ۲. چه طبقه‌بندی‌هایی از مطالعات شناسایی شده می‌توان ارائه کرد؟ ۳. طبقه‌بندی‌های احصاشده را در قالب کدام مدل مفهومی می‌توان به نمایش گذاشت؟ ۴. نظر به تعامل میان سیستم‌های هوشمند و انسان چگونه می‌توانیم سیستم‌های هوشمندی طراحی کنیم که قادر به کمک به آنان در درک تصمیماتشان باشند؟

۲. پیشینه پژوهش

به بیان بوردیو (۱۹۷۵: ۳۸) محصولات تولیدشده در حوزه هوش مصنوعی اعم از محصولات دانشی و روشی، مانند دیگر حوزه‌های علم یک برساخت اجتماعی هستند که به‌وسیله روابط قدرت میان عاملان یا گروه‌هایی از عاملان که در این حوزه حضور دارند و کنشگری می‌کنند شکل می‌گیرند. بررسی مطالعات پیشین در این زمینه بیانگر برداشت‌های متعدد از مفهوم هوش مصنوعی است. این برداشت‌ها را می‌توان ذیل سه رویکرد تحلیلی شامل هوش مصنوعی علمی، هوش مصنوعی فناورانه و هوش مصنوعی فرهنگی که زیربنای مطالعات جامعه‌شناختی در زمینه این فناوری نوظهور است، طبقه‌بندی کرد. به‌منظور ارائه‌ی نمایی روشن‌تر و مقایسه‌ی رویکردهای تحلیلی مذکور و محورهای عمده بحث آن‌ها، جدول ۱ تنظیم شده است.

جدول ۱. تعابیر متفاوت از هوش مصنوعی در مطالعات پیشین

صاحب‌نظران	محور تمرکز	رویکردهای تحلیلی در مطالعات جامعه‌شناسی درباره هوش مصنوعی
بلفلد (۲۰۲۰)؛ لی و لارسن ^۱ (۲۰۱۹)؛ شومان ^۲ (۲۰۰۷)؛	۱. هوش مصنوعی به‌مثابه یک نظام دانشی و تخصص علمی ۲. تمرکز بر عملکرد مشابه انسان و ماشین ۳. پرسش از چگونگی بر ساخت اجتماعی هوش مصنوعی ۴. تأثیرپذیری تحقیقات این حوزه از سنت‌های فکری جامعه‌شناسی علم، جامعه‌شناسی دانش، مطالعات علم و فناوری و تعامل انسان-رایانه ۵. درک منطق اجتماعی تحقیقات هوش مصنوعی از طریق عوامل بیرون‌زا مانند سبک‌های فکری، باورها و اعتقادات ساخت‌یافته اجتماعی ۶. بررسی چگونگی نفوذ و تغییر سیستم‌های هوش مصنوعی در نهادهای اجتماعی به‌عنوان یکی از وظایف جامعه‌شناس هوش مصنوعی ۷. شناخت ماهیت اجتماعی هوش مصنوعی به‌وسیله رد ایده وجود تمایز درونی میان انسان‌ها و ماشین‌ها به‌لحاظ قابلیت‌های اجتماعی‌شان ۸. بهبود عملکرد سیستم‌های تولید دانش توسط توسعه دهندگان هوش مصنوعی به‌وسیله همسان‌سازی پژوهش علوم اجتماعی و علوم انسانی بر روی ماهیت و فرایند تولید دانش	هوش مصنوعی علمی
(بابائیان و همکاران، ۱۴۰۲) (وایسمن ^۳ ، ۲۰۱۷) (مالهوف ^۴ ، ۲۰۲۰) (مارس ^۵ ، ۲۰۲۰)	۱. هوش مصنوعی به‌عنوان ابژه مطالعات علوم اجتماعی ۲. تأثیرپذیری مشاغل از اتوماسیون ۳. توجه به عنصر یادگیری اجتماعی علاوه بر یادگیری ماشینی در فرایند طراحی اتومبیل‌های خودران ۴. بی‌توجهی به پیچیدگی‌های سیاسی-اقتصادی و اجتماعی-فرهنگی اتوماسیون به‌دلیل نقص جبر فناوریانه	هوش مصنوعی فناوریانه
(بلیلی قادیکلای و پارسانیا، ۱۴۰۲) (رجبی و نصرالهی، ۱۴۰۲) (اینون و یانگ ^۶ ، ۲۰۲۱) (ناتال و بالاتور ^۷ ، ۲۰۲۰) (سابانویچ ^۸ ، ۲۰۱۴)	۱. تأکید کمتر بر ابعاد علمی و فناوریانه هوش مصنوعی و تمرکز بیشتر بر پیامدهای گسترده‌تر اجتماعی، اقتصادی، فرهنگی و سیاسی توسعه هوش مصنوعی در بستر انقلاب دیجیتال ۲. استعمار انسان در نتیجه توسعه هوش مصنوعی ۳. غیرانسانی کردن انسان به‌وسیله استفاده از استعاره‌های ملهم از رایانه برای توانایی‌های انسانی (مغز انسان به‌عنوان پردازنده اطلاعات) ۴. نیاز به همکاری جامعه هوش مصنوعی و دانشمند علوم اجتماعی برای پایداری اجتماعی این فناوری ۵. استفاده از دیدگاه‌های جامعه‌شناسی فرهنگی، رسانه، مطالعات فرهنگی و ساخت اجتماعی فناوری برای تجزیه و تحلیل ساخت فرهنگی هوش مصنوعی ۶. گسترش و تعمیق روابط نابرابر قدرت در جامعه با رواج فرهنگ هوش مصنوعی	هوش مصنوعی فرهنگی

1. Lee & Larsen
2. Suchman
3. Wajcman
4. Muhlhuff
5. Marres
6. Eynon & Young
7. Natale & Ballatore
8. Sabanovic

رویکرد تحلیلی، فضایی را برای جامعه‌شناسان در سایر زیرمجموعه‌های جامعه‌شناسی باز می‌کند تا در مورد تأثیر هوش مصنوعی بر موضوعات مورد توجه حوزه خود تحقیق کنند. جامعه‌شناسی هوش مصنوعی در بردارنده جامعه‌شناسی علم و مطالعات علم و فناوری در مورد هوش مصنوعی است، اما به آن محدود نمی‌شود، بلکه مشارکت و توجه حوزه‌های مختلف جامعه‌شناسی برای بحث در مورد روابط هوش مصنوعی و تأثیرات آن بر همه نگرانی‌های اصلی این حوزه دانشی را طلب می‌کند. یکی از این نگرانی‌ها، انسان‌محوری در مطالعات جامعه‌شناختی هوش مصنوعی است که محور بحث پژوهش حاضر است.

دیدگاه‌های تحلیلی مذکور صرفاً گونه‌شناسی‌های مفهومی نیستند، بلکه بیانگر تغییرات اساسی در ماهیت هوش مصنوعی و درک ما از آن در طول زمان است. دیدگاه هوش مصنوعی علمی که بیشتر در پژوهش‌های دهه‌های ۱۹۸۰ و ۱۹۹۰ مشهود است، با رویکرد غالب هوش مصنوعی به‌عنوان یک اختراع علمی جدید در قرن بیستم انطباق دارد. در دهه‌های ۲۰۰۰ و ۲۰۱۰ رویکرد هوش مصنوعی فنی برجسته‌ترین رویکرد در مطالعات این سال‌هاست که بیانگر کاربرد گسترده و تجاری‌سازی فناوری هوش مصنوعی است. در مقابل، دیدگاه «هوش مصنوعی فرهنگی» که بیشتر در بحث‌های متأخر هوش مصنوعی مشاهده می‌شود، منعکس‌کننده رشد اجتماعی و فرهنگی هوش مصنوعی است؛ بنابراین، گونه‌شناسی رویکردهای تحلیلی ناظر بر هوش مصنوعی در مطالعات جامعه‌شناسی نه تنها به پژوهشگران کمک می‌کند تا مرور دقیق‌تری بر مطالعات پیشین داشته باشند، بلکه به‌عنوان مرجعی برای رصد نحوه تکامل اجتماعی هوش مصنوعی مورد استفاده قرار می‌گیرد.

۳. چارچوب مفهومی

مفهوم «هوش مصنوعی انسان‌محور» که نخستین بار به‌صورت رسمی در مراکز تحقیقاتی ایالات متحده آمریکا و اروپا شکل گرفت، اصطلاحی کلی و عام است. حال آنکه در مطالعات تخصصی هوش مصنوعی معانی خاص و متفاوتی پیدا می‌کند. مقاله فورد و همکاران (۲۰۱۵) که به‌عنوان یکی از اولین مقالات در حوزه هوش مصنوعی انسان‌محور در مجله «هوش مصنوعی»^۱ به چاپ رسیده است، از یک تغییر پارادایم در تحقیقات هوش مصنوعی پرده‌برداری می‌کند: تغییر پارادایم از «تقلید هوش مصنوعی» به «اصلاح‌گرها و پروتزه‌های شناختی». آنان در این مقاله به توسعه توانمندی‌های انسانی توسط هوش مصنوعی برای دستیابی مؤثرتر به اهداف فردی می‌پردازند، نه انسان در معنای عام. در همان شماره از مجله مذکور مقاله دیگری توسط راسل و همکاران (۲۰۱۵) درباره هوش مصنوعی انسان‌محور منتشر شد که در آن، نویسندگان از اصطلاح «هوش مصنوعی سودمند»^۲ استفاده کردند. آنان نیز از تغییر پارادایم اما از زاویه‌ای متفاوت سخن می‌گویند: تغییر پارادایم از «هوش مصنوعی کنترل‌نشده» به «توسعه هوش مصنوعی کنترل‌شده». در این مقاله تمرکز نویسندگان بر ابعاد انسانی پیشرفت‌های هوش مصنوعی است، نه جنبه‌های فردی آن. از این‌رو چگونگی ارتقای سودمندی و پیشگیری از پیامدهای ناخواسته و خطرناک استفاده از هوش مصنوعی برای نوع بشر محور بحث آن‌ها است.

رویکرد دیگر مفهوم «انسان‌محوری» در مطالعات زو (۲۰۱۹) است که با تأکید بر جایگاه «کاربر» از «طراحی کاربرمحور» یا «رویکرد کاربرمحور» سخن می‌گوید. به اعتقاد او، ناکامی موج‌های اول و دوم هوش مصنوعی نه‌فقط به‌دلیل فقدان فناوری‌های بالغ در آن زمان، بلکه به‌علت ناتوانی در رفع نیازهای انسان بوده است (همان: ۴۴). زو در پژوهش خود ضمن طرح یک پرسش مهم سعی در تبیین کارآمدی موج سوم هوش مصنوعی به‌لحاظ انسان‌محور بودن دارد؛ اینکه چگونه انجمن تعامل انسان-رایانه^۳ می‌تواند به

1. AI

2. Beneficial AI

3. Human-Computer Interaction (HCI)

توسعه هوش مصنوعی انسان‌محور کمک کند. او در پاسخ، چارچوبی مشتمل بر سه جزء اصلی را پیشنهاد می‌دهد (همان): ۱. طراحی همسو با اخلاق، ۲. انعکاس کامل هوش انسانی در فناوری، و ۳. اطمینان از راه‌حل‌های تبیین‌پذیر، قابل‌درک، مفید و قابل‌استفاده هوش مصنوعی به وسیله طراحی عامل‌های انسانی.

به اعتقاد زو (۲۰۱۹: ۴۵) در نتیجه پیشرفت‌های اخیر هوش مصنوعی، شاهد شراکت حرفه‌ای انسان و ماشین به مثابه کنشگران عضو یک تیم هستیم. اما بن‌اشنایدرمن (۲۰۲۱: ۴۹۷) ضمن انتقاد از این رویکرد، رایانه‌ها را صرفاً ابزارهایی در خدمت افراد فعال در یک تیم حرفه‌ای یا اعضای مشغول به کار در اجتماعات مختلف قلمداد می‌کند. وی ضمن استفاده از عبارت «انسان‌ها در گروه؛ رایانه‌ها در حلقه»^۱ رویکرد خود را به فناوری‌های آینده چنین توصیف می‌کند: «فناوری‌های آینده به احتمال زیاد همان چیزی هستند که من به جای آنکه نام شریک، همکار یا هم‌تیمی بر آن‌ها بگذارم، از مفاهیمی مانند ابرابزار^۲ یا ابزارهای کنشگر برای آن‌ها استفاده می‌کنم» (اشنایدرمن، ۲۰۲۱: ۵۹). مطالعات اشنایدرمن درباره تعامل انسان و رایانه موجب شد تا توجه او معطوف به خطاها و محدودیت‌های اجتناب‌ناپذیر ماشین‌ها به هنگام استفاده در فضای زندگی واقعی به‌جای محیط کنترل‌شده آزمایشگاه شود. از این‌رو دریافت او از مفهوم هوش مصنوعی انسان‌محور، کنترل و ارتقای انسان به همراه حفظ پیوندهای اجتماعی به‌جای تقلید از هوش انسان و اجتماعی‌بودن او است.

فرانک دیگنوم و وبرجینیا دیگنوم (۲۰۲۰) از منظری متفاوت که مبتنی بر دو موقعیت است به هوش مصنوعی انسان‌محور می‌پردازند: نخست، خلق استدلال اجتماعی برای هوش مصنوعی به وسیله مدلسازی رفتار و الگوی استدلال انسانی. موقعیت دوم، بر جامعه متمرکز است، نه بر پیامدهای فردی هوش مصنوعی. رویکرد آنان برخلاف دریافت اشنایدرمن از هوش مصنوعی، انسان‌محور است که به‌جای تقلید بر تقویت هوش، و به‌جای توسعه انسانی به نیازها و اهداف تک‌تک کاربران مرتبط با پیوندهای اجتماعی می‌پردازد.

آنچه گفته شد، در حقیقت ارائه‌ی نمایی کلی از معضلات و تعلیق‌های بینشی است که مواجهه با مفهوم «هوش مصنوعی انسان‌محور» را پیش پای جامعه علمی می‌گذارد. اتخاذ هر یک از مواضع و استدلال‌ها به نوبه خود بستگی به این دارد که هر یک از متخصصان هوش مصنوعی به هنگام روبه‌رویی با این معضلات چه واکنشی نشان می‌دهد و در کدام سوی خط می‌ایستد. نظر به ماهیت کیفی تحقیق، با این پیش‌درآمد مفهومی وارد میدان پژوهش می‌شویم تا با شناسایی مواجهات مختلف با معضلات پیش‌رو بتوانیم به پرسش‌های پژوهش پاسخ دهیم.

۴. روش‌شناسی پژوهش

پژوهش حاضر ذیل جنبش روش‌شناختی کیفی تنظیم شده است. برای پاسخ به پرسش‌های پژوهش، از روش مرور نظام‌مند مطالعات موجود که به فهم و تفسیر حجم زیادی از اطلاعات در این حوزه کمک می‌کند استفاده شد. نتیجه بهره‌مندی از این روش، هدایت جامعه علمی به سمت خلاصه‌ای متوازن و فارغ از سوگیری در آن حوزه دانشی است. نمونه‌گیری با استفاده از یک راهبرد پرکاربرد برای جست‌وجوی مطالعات در یک حوزه مطالعاتی خاص با عنوان «نمونه‌گیری گلوله‌برفی» انجام شد.

به این منظور با یک «مجموعه داده آغازگر» از تعداد محدودی مقاله مرتبط، فرایند تحقیق شروع شد. سپس پی‌درپی مقالات مرتبط که در مجموعه داده آغازگر مورد استناد قرار گرفته‌اند به مجموعه اضافه شد که به «فرایندهای گلوله‌برفی روبه‌عقب و روبه‌جلو» شهرت دارند (کلس، ۲۰۱۴: ۵).

گوگل اسکالر به‌عنوان یکی از بزرگ‌ترین موتورهای جست‌وجوی برخط آکادمیک که کاربرد گسترده‌ای در مطالعات مرور ادبیات دارد و نیز پایگاه داده ساینس دایرکت برای جست‌وجوی اولیه مورد استفاده قرار گرفتند. فرایند گزینش مقالات اعم از معیارهای ورود و خروج، و انتخاب مجموعه آغازگر طبق دستورالعمل پریسما^۱ انجام گرفت. همچنین عنوان، چکیده و واژه‌های کلیدی مقالات برای جست‌وجوی کلیدواژه‌های زیر مدنظر بود:

“Human-Centered AI”, “Human-Centered Artificial Intelligence”,
 “Human-Centred AI”, “Human-Centred Artificial Intelligence”,
 “Human-Centered ML”, “Human-Centered Machine Learning”,
 “Human-Centred ML”, “Human-Centred Machine Learning”,

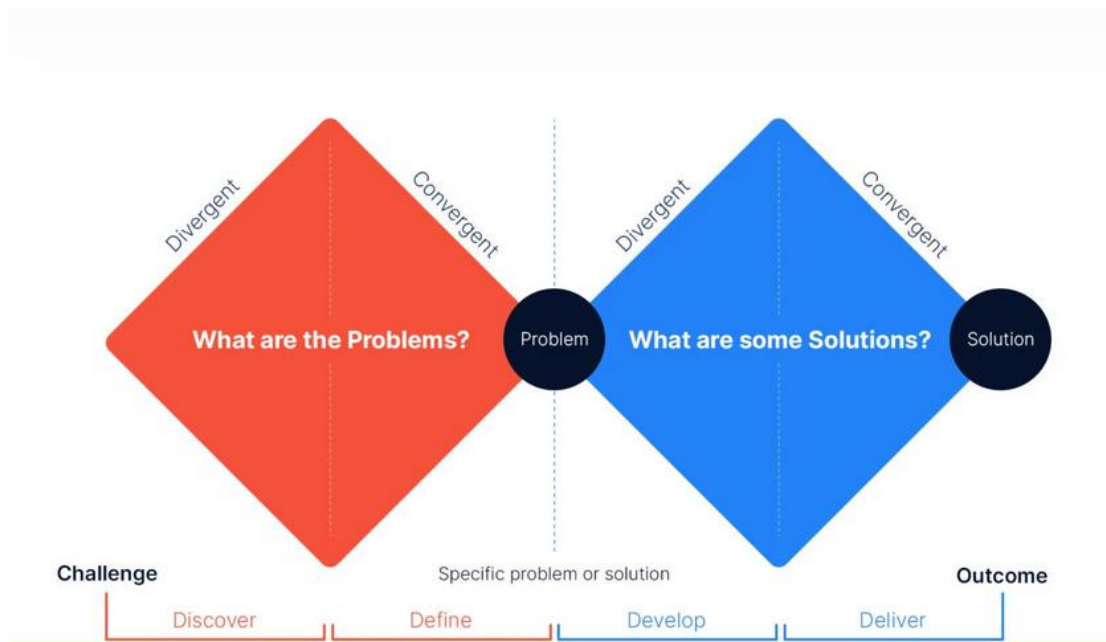
به مقتضای پژوهش، جست‌جوها به مطالعات جامعه‌شناختی بین سال‌های ۲۰۰۴ تا ۲۰۲۳ محدود شد.

در هنگام انتخاب مجموعه داده آغازگر سعی بر انتخاب موضوعات متنوع برای جلوگیری از سوگیری بود. همچنین از ورود یک مجموعه مقالات متوازن که اشاره به تمامی مراحل فرایند الماس دوبر^۲ دارد اطمینان حاصل شد (گوستافسون، ۲۰۱۹؛ شورای طراحی بریتانیا). در پایان، ۱۷ مقاله مرتبط با چهار فاز یا مرحله از الماس دوبر (به‌ترتیب چهار، سه، پنج) و پنج مقاله برای فازهای کشف^۳، تعریف^۴، توسعه^۵ و تحویل^۶ وارد شد؛ بنابراین مجموعه داده آغازگر مشتمل بر ۱۷ مقاله شد. پس از انتخاب مجموعه آغازگر، دو مرتبه دیگر نمونه‌گیری تکرارشونده انجام گرفت. در هر بار تکرار، هم مقالاتی که نمونه تحقیق به آن‌ها استناد کرده (نمونه‌گیری روبه‌عقب مقالات پیشین) و هم مقالاتی که به نمونه تحقیق استناد و اشاره کرده‌اند (نمونه‌گیری روبه‌جلوی مقالات بعدی) بررسی شدند. محققان متن کامل مقالات شناسایی‌شده را برای اطمینان از میزان ارتباط با موضوع، واجد شرایط بودن و کیفیت داده‌ها بررسی کردند. پس از تکرار دومین نمونه‌گیری گلوله‌برفی، نمونه‌گیری متوقف شد؛ زیرا نمونه‌ای بزرگ (شامل ۳۴۷ مورد ثبت‌شده) به‌دست آمد که نماینده مقالات موجود در حوزه پژوهش بودند. همچنین در مرتبه دوم نمونه‌گیری، مشاهده شد که مقالات مرتبه اول نمونه‌گیری در فهرست مقالات مورد علاقه نویسندگان مقاله تکرار می‌شوند. در فرایند تجزیه و تحلیل، پس از بررسی دقیق، ۱۱۸ مقاله به‌دلیل عدم ارتباط نسبی با سؤالات تحقیق حذف شدند. نمونه نهایی حاوی ۲۲۹ مقاله بود. از این تعداد، ۶۶ مقاله در فاز کشف، ۲۴ مقاله در فاز تعریف، ۹۴ مقاله در فاز توسعه و ۴۵ مقاله در فاز تحویل از چارچوب الماس دوبر قرار گرفتند.

به‌منظور تجزیه و تحلیل داده‌ها، پس از آنکه همه مقالات انتخاب و غربال شدند، موضوع اصلی در یکی از چهار مرحله چارچوب طراحی الماس دوبر که دربردارنده فعالیت‌های مختلف در یک فرایند طراحی نوآورانه است، ترسیم شد (شکل ۱). این چارچوب مشتمل بر دو فرایند واگرا و همگرا در طراحی است که در آن، طراحان مشکلات بالقوه یک حوزه را بررسی می‌کنند. سپس به‌سمت موضوعات هدف اصلی همگرا می‌شوند، راه‌حل‌های محدود اولیه را ارائه می‌دهند و درباره‌اثرگذارترین راه‌حل از طریق آزمایش و ارزیابی

1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta Analyses)
2. Double Diamond
3. Discover
4. Define
5. Develop
6. Deliver

تصمیم‌گیری می‌کنند (همان). شایان ذکر است که فرایندهای طراحی مدرن اغلب تکراری هستند؛ بنابراین طراحان می‌توانند بین فازها و مراحل مختلف الماس دوبر به سمت جلو و عقب رفت‌و برگشت داشته باشند. کاربست چارچوب طراحی فوق روی داده‌های پژوهش حاضر با جزئیات در بخش «بحث و نتیجه‌گیری» تشریح خواهد شد.



شکل ۱. چارچوب طراحی الماس دوبر

منبع: گوستافسون، ۲۰۱۹؛ شورای طراحی بریتانیا

۵. یافته‌های تحقیق

در این بخش به دلیل حجم نسبتاً زیاد یافته‌ها، دستاوردهای مرور نظام‌مند مطالعات موجود در هشت سطح تنظیم شدند: ۱. معضلات هوش مصنوعی انسان‌محور؛ ۲. هوش مصنوعی انسان‌محور در تولید علوم اجتماعی و جامعه‌شناسی؛ ۳. پژوهش‌های میان‌رشته‌ای درباره مفهوم انسان‌محوری در مطالعات جامعه‌شناختی هوش مصنوعی؛ ۴. جایگاه هوش مصنوعی انسان‌محور در روش‌شناسی علوم اجتماعی؛ ۵. اصول و مقررات ناظر بر توضیح‌پذیری و تبیین‌پذیری هوش مصنوعی؛ ۶. نقش مفهوم انسان‌محوری در درک و تصور کنشگران از هوش مصنوعی؛ ۷. تعامل انسان با هوش مصنوعی؛ و ۸. هوش مصنوعی انسان‌محور در حوزه‌های مطالعاتی خاص.

۵-۱. معضلات هوش مصنوعی انسان‌محور

عطف به اهداف پژوهش، مشروح در بخش آغازین مقاله، از مرور نظام‌مند مطالعات موجود در مجموع پنج معضل عمده درباره هوش مصنوعی انسان‌محور شناسایی شده است. از این میان، دو معضل (اول و دوم) ناظر به هوش مصنوعی و سه معضل (سوم، چهارم و پنجم) ناظر به انسان (در معنای عام) هستند. نخستین معضل «اتخاذ موضع خوش‌بینانه یا بدبینانه» نسبت به توانایی‌های

هوش مصنوعی است. رویکردهای خوش‌بینانه به توانایی‌های هوش مصنوعی آن را یک کنشگر فعال یا شریک شغلی و همکار می‌دانند؛ حال آنکه رویکردهای بدبینانه، به هوش مصنوعی به‌مثابه یک ابزار یا ابر ابزار می‌نگرند.

براساس یافته‌های پژوهش، معضل و تنگنای دوم به «رویکردهای خوش‌بینانه یا بدبینانه در مورد ماهیت انسان» می‌پردازد. تمرکز رویکردهای خوش‌بینانه بر توسعهٔ انسانی، فرصت‌ها و انتخاب‌های جدید است؛ درحالی‌که رویکردهای بدبینانه بیشتر بر نقاط ضعف و کاستی‌های انسان تأکید دارند.

سومین معضل شناسایی شده براساس یافته‌های پژوهش ناظر بر پیامدهایی است که استفاده از هوش مصنوعی برای «انسان» به معنای ابعاد عینی و زیستی نوع بشر و یا «انسانیت» به معنای ابعاد ذهنی، اخلاقی و فرهنگی نوع بشر دارد. موضع‌گیری‌های متفاوت پژوهش‌ها به‌لحاظ پذیرش جایگاه سوژگی یا ابژگی هوش مصنوعی و انسان، چهارمین معضل احصاء شده از داده‌های پژوهش است.

تمرکز بر روی مفهوم «انسانیت» بجای «انسان» ابعاد دیگری از تنگناهای مفهومی موجود در پژوهش‌ها را به‌لحاظ میزان تعلق و تعهد به مفهوم «انسان‌محوری» در هوش مصنوعی به نمایش می‌گذارد. «طراحی مسئولانهٔ هوش مصنوعی»، «احترام به حریم خصوصی در طراحی»، «چارچوب طراحی و ارزش‌گذاری»، «تعامل انسان با هوش مصنوعی»، «رفاه انسان» و «حکمرانی و نظارت» همگی اجزای پنجمین معضل شناسایی شده با تمرکز بر مفهوم «انسانیت» در مقالات هستند. این رویکردها در جدول ۲ خلاصه شده‌اند.

جدول ۲. معضلات هوش مصنوعی انسان‌محور: معانی «انسان‌محوری»

معضلات ناظر به هوش مصنوعی								
بدبینی به توانایی‌های هوش مصنوعی		خوش‌بینی به توانایی‌های هوش مصنوعی		معضل اول				
ماهیت انسان	خوش‌بینی به ماهیت انسان	بدبینی به ماهیت انسان	خوش‌بینی به ماهیت انسان	معضل دوم				
هدف از ساخت هوش مصنوعی اصلاح اشتباهات انسان است	انسان‌ها در گروه، کامپیوترها در حلقه؛ ارتقای انسان؛ کنترل انسان	جبران ضعف‌ها و کاستی‌های انسان به‌وسیلهٔ هوش مصنوعی	همکاری میان هوش مصنوعی و انسان؛ تقلید از هوش انسانی	هوش مصنوعی در مقام سوژه و انسان در مقام ابژه	معضل چهارم	پیامدهای استفاده از هوش مصنوعی برای انسان	معضل سوم	معضلات ناظر به انسان (در معنای عام)
-	-	-	سیستم‌های هوش مصنوعی که به شناخت انسان از منظر فرهنگی اجتماعی می‌پردازند					
-	-	سیستم‌های هوش مصنوعی که به شناخت هوش مصنوعی یاری می‌رساند	-					

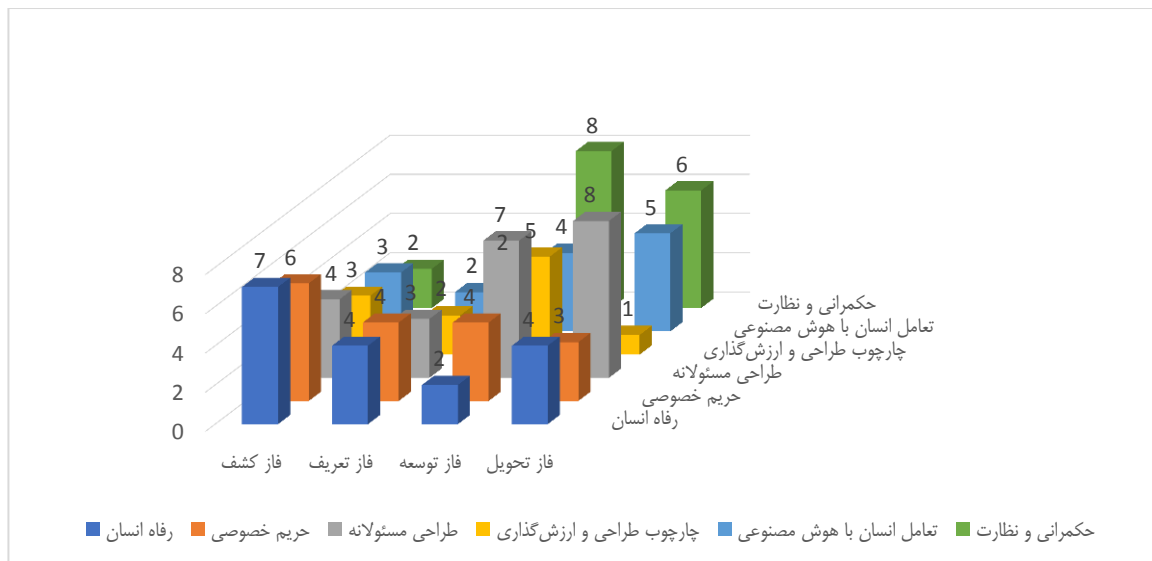
1. Human
2. Humanity

معضلات ناظر به هوش مصنوعی							
بدبینی به توانایی‌های هوش مصنوعی		خوش‌بینی به توانایی‌های هوش مصنوعی		معضل اول			
ماهیت انسان	خوش‌بینی به ماهیت انسان	بدبینی به ماهیت انسان	خوش‌بینی به ماهیت انسان	معضل دوم			
نیاز به تلاش برای جلوگیری از پیامدهای ناخواسته توسعه هوش مصنوعی مانند رشد نابرابری، نقض حریم خصوصی و ...	اتوماسیون کارهای سخت، تکراری و خسته‌کننده؛ کارگران خلاق‌تر و ایجاد اوقات فراغت بیشتر برای انسان؛ تسهیل و حفظ پیوندهای اجتماعی؛	رهایی انسان از بحران توسط هوش مصنوعی؛ مقام داور، حکمران، رهبر، برنامه‌ریز	تحقیق و توسعه دستورالعمل‌های سیاستی انسان‌محور برای هوش مصنوعی	طراحی مسئولانه هوش مصنوعی	معضل پنجم	پیامدهای استفاده از هوش مصنوعی برای انسانیت	
-	اجتناب از دوگانگی کاذب عملکرد پیشگیرانه در طراحی، و در عوض دربرگیری تمامی منافع مشروع؛ ایجاد تنظیمات پیش‌فرض بجای ادغام موارد استثناء در طراحی	تأمین امنیت در طول چرخه حیات فناوری	اتخاذ رویکرد پیشگیرانه بجای رویکرد اصلاحی و واکنشی؛ رؤیت‌پذیری و شفافیت؛				احترام به حریم خصوصی در طراحی
-	توانمندسازی و تقویت عملکرد انسان؛ کنترل انسان با کمک یادگیری ماشینی	پایبندی به اصول هوش مصنوعی اخلاقی در هر مرحله از طراحی؛ امکان شناسایی فرصت‌های جدید و در اولویت جامعه توسط طراحان؛ کاهش احتمال عدم پذیرش فناوری توسعه‌یافته از سوی جامعه	توجه به نیازها، ارزش‌ها، خواسته‌های گروه‌های مختلف کاربران و ذی‌نفعان از فرهنگ‌های متفاوت؛ طراحی محصولات و خدمات هوش مصنوعی با مردم و برای مردم؛ مشاهده کاربران، تعامل با ذی‌نفعان، ارزیابی قابلیت استفاده، اصلاحات مداوم و نظارت مستمر بر عملکرد انسان				چارچوب طراحی و ارزش‌گذاری
-	-	درک و تبیین، برنامه‌ریزی، تصمیم‌گیری‌های متوالی و اقدام، فرایندهای حافظه، توجه و جستجو، مفهوم‌سازی، یادگیری،	برخورداری هوش مصنوعی از عاطفه و خودکنترلی، مدلسازی ذهن انسان و تعامل اجتماعی، فراشناخت و ارزیابی اعتماد، همگی در مواجهه با مردم	تعامل انسان با هوش مصنوعی			

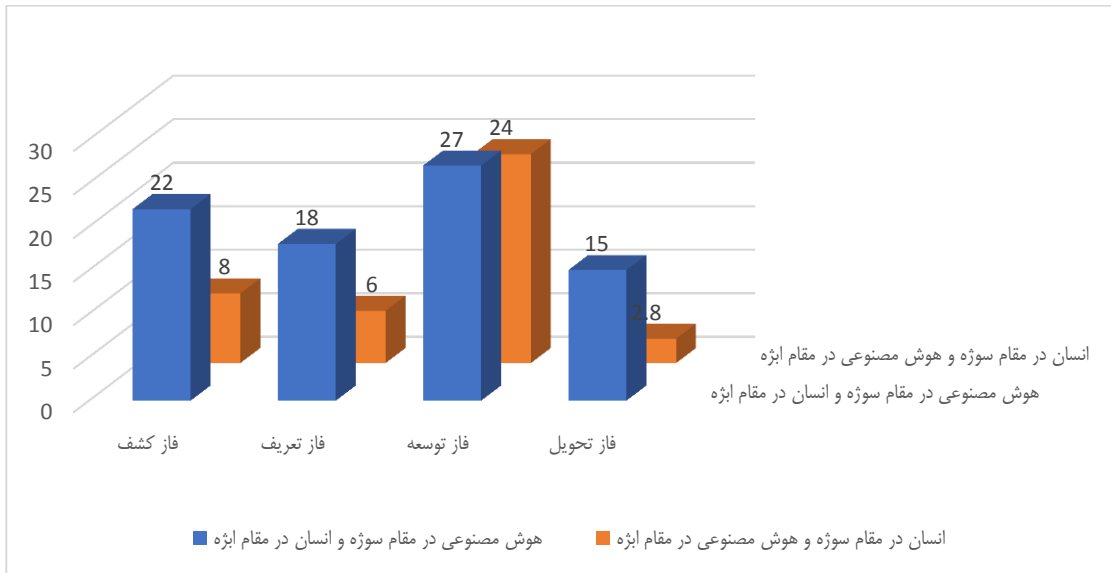
عضلات ناظر به هوش مصنوعی						
بدبینی به توانایی‌های هوش مصنوعی		خوش‌بینی به توانایی‌های هوش مصنوعی		عضل اول		
بدبینی به ماهیت انسان	خوش‌بینی به ماهیت انسان	بدبینی به ماهیت انسان	خوش‌بینی به ماهیت انسان	عضل دوم		
		<p>انتزاعی‌سازی، و استدلال منطقی و آماری، همگی در مواجهه با ایده‌ها و اطلاعات؛</p> <p>تامل حسی - حرکتی، حمل‌ونقل، پردازش بصری و پردازش شنیداری، همگی در مواجهه با اشیای فیزیکی یا مجازی</p>				
-	-	-	<p>هم‌تیمی و همکاری‌بودن هوش مصنوعی با انسان؛ اشکال جدیدی از اجتماعی‌شدن</p> <p>ارتقای اعتماد، پاسخگویی، عاملیت، رفاه کاربر، بهینه‌سازی اولویت‌های انسانی و کاهش سطح آسیب‌پذیری انسان توسط هوش مصنوعی</p>	رفاه انسان		
<p>اجتناب‌ناپذیربودن سوگیری، شناخت آن، شناسایی و به‌کارگیری راهبردهای کاهش برای به‌حداقل‌رساندن هدفمند خطر تصمیم‌گیری‌های غیراخلاقی و پیامدهای ناخواسته آن (عدالت)</p>	-	<p>انطباق و استحکام فناورانه (تاب‌آوری)</p>	<p>ثبات داده‌ها و اعتبار الگوریتمی (انسجام)؛ شفافیت فرایند تصمیم‌گیری الگوریتمی (توضیح‌پذیری)</p>	حکمرانی و نظارت		

شایان ذکر است جدول ۲ از نوع جدول‌های آشیانه‌ای و متشکل از چند جدول ماتریسی تودرتو است. رویکردهای متفاوت به مفهوم انسان‌محوری در هوش مصنوعی محصول مواضع مختلفی است که هریک از پژوهش‌ها به معضلات و تنگنای‌های مذکور اتخاذ می‌کنند. همان‌طور که در جدول ۲ مشاهده می‌شود، برهم‌نهیست مفهومی «معضلات ناظر به هوش مصنوعی» و «معضلات ناظر به انسان»، همچنین اتخاذ مواضع متفاوت مقالات به معضلات مذکور، موجب شکل‌گیری رویکردهای متفاوت (در مجموع ۲۲ رویکرد) در مواجهه با این معضلات شده است.

طبق توضیحات بخش روش‌شناسی مقاله، از کاربست چارچوب طراحی الماس دوبر بر داده‌های پژوهش حاضر، هریک از مواضع و رویکردهای سوی مقالات در مواجهه با معضلات شناسایی‌شده در هوش مصنوعی انسان‌محور در یکی از فازهای چارچوب طراحی الماس دوبر قرار می‌گیرد. شکل‌های ۲ و ۳ فراوانی مقالات را در فازهای چهارگانه چارچوب طراحی الماس دوبر ناظر به رویکردهای مذکور به تفکیک نمایش می‌دهند.



شکل ۲. فراوانی مقالات در فازهای چهارگانه چارچوب الماس دوبر به تفکیک معضلات هوش مصنوعی انسان‌محور (تمرکز بر مفهوم انسانیت)



شکل ۳. فراوانی مقالات در فازهای چهارگانه چارچوب الماس دوبر به تفکیک معضلات هوش مصنوعی انسان محور (تمرکز بر مفهوم انسان)

۲-۵. هوش مصنوعی انسان محور در تولید علوم اجتماعی و جامعه‌شناسی

از آغاز دهه ۱۹۹۰ جامعه علمی شاهد توجه علوم اجتماعی به فناوری‌های هوش مصنوعی بوده است. این تحول عظیم در مطالعات اجتماعی مرهون پژوهش‌های کالینز (۱۹۹۴: ۶۵) درباره ضرورت جامعه‌شناسی برای مطالعه هوش مصنوعی است. ناظر بر این مسئله، در نتیجه مرور نظام‌مند مطالعات پیشین، پنج نوع تحقیق شناسایی شده است:

۱. مطالعات جامعه‌شناختی فناوری‌های هوش مصنوعی که ذیل عنوان «مطالعات علم و فناوری» طبقه‌بندی می‌شوند: در سال‌های اخیر مطالعات علم و فناوری، هم به‌لحاظ نظری و هم به‌لحاظ روش‌شناختی، انتخاب اول جامعه‌شناسان علاقه‌مند به حوزه هوش مصنوعی است. آنان به‌منظور مطالعه هوش مصنوعی، اغلب تحقیقات خود را در قالب مطالعه الگوریتم‌ها تعریف می‌کنند (زویتز، ۲۰۱۶: ۱۰):

۲. مطالعات جامعه‌شناختی هوش مصنوعی در حوزه ارتباطات و رسانه (گژمن و لوپس، ۲۰۲۰: ۷۵): اخیراً نظر به استفاده گسترده از هوش مصنوعی و الگوریتم‌ها به‌صورت برخط، صاحب‌نظران حوزه ارتباطات و مطالعات رسانه، توجه ویژه‌ای به آن نشان داده‌اند. از این‌رو جامعه‌شناسان و دیگر دانشمندان علوم اجتماعی تا حد زیادی وام‌دار ایده‌های نظری و طرح‌های پژوهشی این حوزه هستند؛

۳. مطالعاتی که بازتاب رویکردهای جامعه‌شناختی کلاسیک روی هوش مصنوعی، هم به‌لحاظ نظری و هم به‌لحاظ روش‌شناختی است: آلن ولف (۱۹۹۳)، رندال کالینز (۱۹۹۴) و النا اسپوزیتو (۲۰۱۷) سهم عمده‌ای در تثویز کردن هوش مصنوعی در جامعه‌شناسی دارند. آن‌ها در پژوهش‌هایی که به کاربرد روش‌شناختی هوش مصنوعی در علوم اجتماعی می‌پردازند، تمرکز خود را بر یادگیری ماشین، مدلسازی عامل محور و کلان‌داده قرار داده‌اند؛

۴. مطالعات جامعه‌شناسی انتقادی هوش مصنوعی: امروزه بسیاری از جامعه‌شناسان تمایل زیادی به نظریات انتقادی دارند. اخیراً مطالعات جامعه‌شناختی هوش مصنوعی به مسائلی مانند نابرابری، بی‌عدالتی و تبعیض می‌پردازند (جویس و همکاران، ۲۰۲۱: ۸)؛
۵. مطالعات متمرکز بر اشکال نوین سازمان اقتصادی-اجتماعی مرتبط با توسعه هوش مصنوعی: سرمایه‌داری نظارتی (زیاف، ۲۰۱۹) و سرمایه‌داری پلتفرمی (شرنیک، ۲۰۱۶) از جمله پژوهش‌های انجام‌شده با این رویکرد هستند.

۳-۵. پژوهش‌های میان‌رشته‌ای درباره مفهوم انسان‌محوری در مطالعات جامعه‌شناختی هوش مصنوعی

یافته‌های مستخرج از مطالعه دقیق مقالات جامعه‌شناختی در حوزه هوش مصنوعی انسان‌محور بیانگر تنوع زیاد این پژوهش‌ها است. برخی از این مقالات اساساً در حوزه علوم کامپیوتر قرار دارند که به مسائل اجتماعی مرتبط با توسعه هوش مصنوعی می‌پردازند. برخی دیگر از مقالات، تمرکز مطالعاتی خود را بر مشکلات عملیاتی و محدودیت توسعه اجتماعی و پیشرفت هوش مصنوعی قرار داده‌اند. گروه دیگری از پژوهش‌ها نیز به مسائل نظری و طرح‌های تحقیقاتی نزدیک‌تر به رویکردهای جامعه‌شناختی می‌پردازند. به این ترتیب براساس یافته‌های پژوهش شاهد نوعی گرایش به تحقیقات میان‌رشته‌ای در مطالعات انجام‌شده به شرح زیر هستیم: ۱. ترکیب علوم اجتماعی با علوم یا مهندسی کامپیوتر؛ ۲. ترکیب علوم اجتماعی و علوم انسانی؛ و ۳. ترکیب علوم اجتماعی و مدیریت.

۴-۵. جایگاه هوش مصنوعی انسان‌محور در روش‌شناسی علوم اجتماعی

براساس یافته‌های پژوهش، عمده نویسندگان مقالات مرتبط با هوش مصنوعی انسان‌محور در علوم اجتماعی از ایالات متحده آمریکا و اروپا هستند. برخی از این مقالات نیز پیشرفت در روش‌شناسی علوم اجتماعی را مورد بحث قرار می‌دهند. شایان ذکر است مفهوم انسان‌محوری در مقالات روش‌شناختی دلالت بر عباراتی مانند «نرم‌افزار کاربرپسند» (هو و همکاران، ۲۰۲۰: ۶۲۷۵)، «روش‌های تبیین‌پذیر» (اختر، ۲۰۲۰: ۳۹۸۸)، «استفاده از تکنیک‌های سودمند برای انسان» (هانگ و همکاران، ۲۰۲۰: ۵۴۰) یا ترکیبی از همه این‌ها دارد.

براساس یافته‌های پژوهش، مدل‌سازی رفتار انسان توسط عواملان هوش مصنوعی، یکی از گرایش‌های مهم در مطالعات روش‌شناختی هوش مصنوعی محسوب می‌شود. این موضوع، خود با چالش مدل‌سازی احساسات انسانی به‌وسیله رایانه پیوند دارد (گرچ و همکاران، ۲۰۰۵: ۲۴۵).

۵-۵. اصول و مقررات ناظر بر توضیح‌پذیری و تبیین‌پذیری هوش مصنوعی

در مرور نظام‌مند مطالعات پیشین، مقالات متعددی وجود دارد که اصول مورد نیاز برای تولید و تنظیم هوش مصنوعی انسان‌محور را به بحث می‌گذارند. مقالات مذکور پرسش‌هایی را در محل تلاقی رویکردهای علوم اجتماعی، فلسفی، حقوقی و علوم رایانه به شرح زیر مطرح می‌کنند:

- آیا باید بر توسعه هوش مصنوعی انسان‌محور براساس اصول حقوقی و اخلاقی موجود تکیه کنیم؟ (کازوناتو، ۲۰۱۹)؛
- چگونه می‌توان قواعد و مقررات مرتبط با هوش مصنوعی را در موقعیت‌های خاصی از حالت انتزاعی به حالت عملیاتی و دارای قابلیت اجرایی تبدیل کرد؟ (هیگمن و پترین، ۲۰۲۱)؛
- آیا اصول جهان‌شمول ناظر بر قواعد و مقررات هوش مصنوعی، امری توضیح‌پذیر است؟ (هیروشی، ۲۰۲۱)؛

- چگونه می‌توان تصمیم‌گیری‌های الگوریتمی را شفاف و توضیح‌پذیر کرد؟ (لوی و همکاران، ۲۰۲۱)؛
- آیا باید در استفاده از هوش مصنوعی محدودیت ایجاد کنیم؟ (اوراک، ۲۰۲۲)؛
- چگونه توسعه‌های اخیر در هوش مصنوعی انسان‌محور و مقررات ناظر بر آن با مشکلات محیط زیستی پیوند پیدا می‌کند؟ (پروسیکا و آنجلکوویچ، ۲۰۲۲).

مقالات مذکور مبتنی بر اصول و مقرراتی هستند که عمدتاً از فلسفه روشنگری و پسا روشنگری وام گرفته شده‌اند. در این میان آنچه اختصاصاً به هوش مصنوعی مرتبط می‌شود، چگونگی توضیح و تبیین کنش‌ها و تصمیمات آن است. براساس یافته‌های پژوهش، بیشتر مقالاتی که اصول و مقررات خاصی را مورد بحث قرار می‌دهند، مبتنی بر قوانین اروپایی هستند. علاوه بر این، تعدادی از مقالات نیز مفهوم هوش مصنوعی انسان‌محور را از منظر محیط زیستی یا از موضع عمل‌گرایانه‌تر مورد انتقاد قرار داده‌اند.

۶-۵. نقش مفهوم انسان‌محوری در درک و تصور کنشگران از هوش مصنوعی

از مرور نظام‌مند مطالعات پیشین، گروه دیگری از مقالات شناسایی شدند. دغدغه اصلی این مقالات آن است که کنشگران مختلف چه درک و تصویری از هوش مصنوعی دارند و مفهوم انسان‌محوری چه نقشی در این تصور ایفا می‌کند. نویسندگان مقالات مذکور برای پاسخ به پرسش‌هایشان، منابع متعددی را تجزیه و تحلیل کرده‌اند که عبارت‌اند از: پیمایش‌ها (ون برکل و همکاران، ۲۰۲۲)، قوانین حوزه هوش مصنوعی (ناپولی، ۲۰۲۰)، برنامه‌های پخش‌شده در رسانه‌های جمعی (هانسن، ۲۰۲۲)، مصاحبه با متخصصان در سازمان‌های فعال در زمینه هوش مصنوعی (لئونارد و تیرز، ۲۰۲۱)، و داده‌های مستخرج از تبلیغات و مصاحبه با کاربران هوش مصنوعی (هوتامو، ۲۰۲۰).

بیشتر مقالاتی که به موضوع درک و تصور از هوش مصنوعی توجه نشان داده‌اند، داده‌های خود را از جوامع اروپایی جمع‌آوری کرده‌اند. براساس یافته‌های پژوهش، در دو مقاله، درباره مفهوم انسان‌محوری ارزیابی منفی ارائه شده است. دیگر مقالات نیز با رویکردی مثبت اما متفاوت، انسان‌محوری را مورد توجه قرار داده‌اند؛ به گونه‌ای که آن را برای جامعه مفید می‌دانند، جایگزین انسان در تعاملات اجتماعی نمی‌کنند، به لحاظ اجتماعی، بافت‌مند تلقی می‌کنند، تقویت‌کننده توانایی‌های انسان به‌شمار می‌روند، دنباله‌رو اصول اخلاقی معرفی می‌کنند.

۷-۵. تعامل انسان با هوش مصنوعی

از مرور نظام‌مند مطالعات موجود، محور تحقیقاتی دیگری نیز در میان مقالات قابل‌احصا است که چالش‌ها و مشکلات هوش مصنوعی انسان‌محور را از دریچه تعامل انسان با هوش مصنوعی پیگیری می‌کند.

برخی از این مقالات، جریان‌های ارتباطی و مبادلات را در تعامل انسان و ماشین، محور تجزیه و تحلیل خود قرار می‌دهند (دمیر و همکاران، ۲۰۱۹). چندین مقاله نیز بر کنش مشترک و همکاری در فرایند تعامل تأکید دارند (کاسین و همکاران، ۲۰۲۲). سایر مقالات نیز موضوعاتی مانند اعتماد^۱ (دیکمن و برنز، ۲۰۲۲)، انصاف^۲ و انسان‌گرایی^۳ (یانگ، ۲۰۲۲) را در تعامل با هوش مصنوعی مورد بحث قرار می‌دهند. به این ترتیب جهت‌گیری کلی بیشتر مقالات این حوزه درباره چگونگی توضیح‌پذیر کردن هوش مصنوعی در مواجهه با افرادی است که با آن تعامل دارند.

1. Trust
2. Fairness
3. Anthropomorphism

همچنین برخی از مقالات، ضمن تجزیه و تحلیل استفاده از هوش مصنوعی در یک زمینه اجتماعی گسترده‌تر به موضوعات زیر توجه نشان می‌دهند:

۱. خلق اطلاعات نادرست به وسیله هوش مصنوعی یا به تعبیری «جعل عمیق»^۱ (علی و همکاران، ۲۰۲۱)؛
 ۲. کاربرد هوش مصنوعی در صنایع خلاق (آناتراسیریچای و بال، ۲۰۲۲) و امور مالی (چمبرز-جونز، ۲۰۲۱)؛
 ۳. تغییر و تحول در قواعد مالکیت معنوی و حق چاپ مرتبط با توسعه هوش مصنوعی (لی و همکاران، ۲۰۲۱)؛
 ۴. استفاده دولت‌ها از داده‌های شهروندان (یورگنسن، ۲۰۲۱)؛
 ۵. نفوذ بین بخش‌های مختلف هوش مصنوعی (اشمیت و همکاران، ۲۰۲۲).
- برخی از این مقالات، فواید مرتبط با توسعه هوش مصنوعی را به بحث می‌گذارند. برخی دیگر به مسائل و آسیب‌های ناشی از آن می‌پردازند. به این ترتیب مفهوم «انسان‌محوری» بار دیگر و به شکلی متفاوت توسط نویسندگان مختلف بررسی می‌شود.

۸-۵. هوش مصنوعی انسان‌محور در حوزه‌های مطالعاتی خاص

در گزارش‌های متعدد، مفهوم هوش مصنوعی انسان‌محور با ارجاع به قواعد و مقررات اروپایی مشخص شده است. علاوه بر این گروهی از پژوهش‌های مرتبط با هوش مصنوعی انسان‌محور در حوزه‌های خاص وجود دارد که تعدادی از آن‌ها به کاربرد هوش مصنوعی در صنایع آموزشی (یانگ و همکاران، ۲۰۲۱)، توسعه شهری (کولسنیچنکو و همکاران، ۲۰۲۱) و حوزه سلامت (بالکومب و دلو، ۲۰۲۲) می‌پردازند. سایر مقالات نیز پروژه‌های تحقیقاتی را در حوزه‌های زیر هدایت می‌کنند: ۱. شهر هوشمند (سونتی و همکاران، ۲۰۱۸)؛ ۲. خانه‌های هوشمند (تیرسن و همکاران، ۲۰۲۱)؛ ۳. هوش مصنوعی برای خلق اجتماع (شن، ۲۰۱۹)؛ ۴. ربات‌چت‌ها برای ارتباطات چندزبانه (باوا و همکاران، ۲۰۲۰)؛ ۵. هوش مصنوعی برای تشخیص پزشکی (گو و همکاران، ۲۰۲۱)؛ ۶. فناوری‌های خودمراقبتی (ایوبی و همکاران، ۲۰۲۱)؛ و ۷. هوش مصنوعی در آموزش (کلوس و همکاران، ۲۰۲۲).

شایان ذکر است مفاهیم «انسان‌محوری» و «طراحی انسان‌محور یا کاربرمحور» در عمده مقالات و پروژه‌های تحقیقاتی فوق محل اعتنا بوده است.

۶. بحث و نتیجه‌گیری

پژوهش حاضر با توجه به تعلیق‌های بینشی و معضلات مطرح‌شده در بخش‌های آغازین مقاله در خصوص محور تمرکز تحقیقات مختلف هوش مصنوعی تنظیم شده است. از موضع‌گیری‌های مختلف در پژوهش‌های این حوزه نسبت به تعلیق‌های بینشی مذکور، برای پاسخگویی به پرسش‌های مطرح‌شده در ابتدای مقاله استفاده شد. در پاسخ به پرسش‌های ۱ و ۲ تحقیق، مقالات علوم اجتماعی را که به مسائل مرتبط با هوش مصنوعی انسان‌محور پرداخته‌اند، می‌توان ذیل شش محور اصلی طبقه‌بندی کرد: ۱. اصول هوش مصنوعی انسان‌محور؛ ۲. تخیل هوش مصنوعی در جامعه؛ ۳. استفاده از هوش مصنوعی در جامعه؛ ۴. انسان‌محوری در تعامل انسان و هوش مصنوعی؛ ۵. توسعه هوش مصنوعی انسان‌محور در حوزه‌های خاص؛ و ۶. توسعه هوش مصنوعی انسان‌محور برای روش‌شناسی علوم اجتماعی.

نبود جهت‌گیری عمومی دربارهٔ چگونگی سازمان‌دهی مطالعات جامعه‌شناختی در عصر هوش مصنوعی موجب ضعف شفافیت در معنای هوش مصنوعی انسان‌محور و روش‌شناسی مناسب برای پژوهش در این حوزه شده است. از دیگر نتایج فقدان جهت‌گیری مذکور، عقب‌ماندن دانشمندان علوم اجتماعی و جامعه‌شناسی از مطالعهٔ هوش مصنوعی است؛ زیرا آنان فقط به اثرات مخرب فناوری‌های مبتنی بر الگوریتم هوش مصنوعی روی انسان و محیط اجتماعی واکنش نشان می‌دهند. حال آنکه صرفاً شناخت تأثیر هوش مصنوعی بر انسان کفایت نمی‌کند، بلکه شناسایی اثرات مخرب احتمالی پیش از طراحی الگوریتم‌ها امری ضروری است. از این‌رو توجه به مسیرهای دستیابی به فناوری، از دیگر الزامات مطالعاتی دانشمندان علوم اجتماعی است.

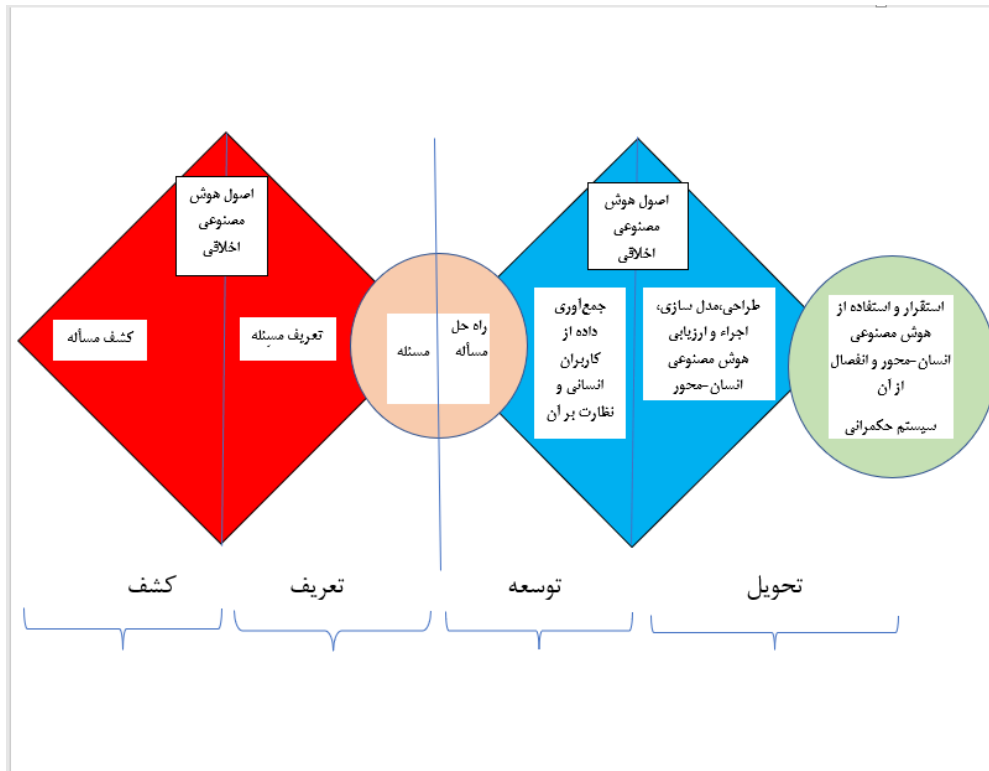
ناظر به سومین پرسش تحقیق دریافتیم که پژوهش‌های طبقه‌بندی‌شده ذیل هریک از محورهای شش‌گانهٔ فوق، در مواجهه با معضلات مربوط به هوش مصنوعی انسان‌محور به دو جنبهٔ عمده توجه دارند: ۱. معضلات ناظر به هوش مصنوعی و ۲. معضلات ناظر به انسان (در معنای عام).

از مرور نظام‌مند مطالعات پیشین، دو معضل (اول و دوم) توضیح‌داده‌شده در بخش یافته‌های سطح ۱ و نشان‌داده‌شده در جدول ۲ ناظر به هوش مصنوعی و سه معضل (سوم، چهارم و پنجم) مشروح در بخش و جدول مذکور ناظر به انسان (در معنای عام) هستند. هم‌نهشتی معضلات مذکور، همچنین اتخاذ مواضع متفاوت از سوی مقالات مورد مطالعه به معضلات مذکور، موجب شکل‌گیری رویکردهای مختلف (در مجموع ۲۲ رویکرد) در مواجهه با این معضلات شده است. این رویکردها و موضع‌گیری‌ها با جزئیات در جدول ۲ بیان شده است.

یافته‌های سطح ۷ پژوهش تحت عنوان تعامل انسان با هوش مصنوعی ناظر به پرسش چهارم تحقیق است. با توجه به ماهیت نوظهور فناوری‌های هوش مصنوعی، فرایند طراحی الگوریتم‌ها چه از منظر پاسخ به نیازهای فنی-مهندسی و چه از منظر سایر علوم مرتبط مشخصاً علوم اجتماعی و جامعه‌شناسی در پژوهش حاضر، مستلزم رویکردهای نوآورانه است. انجمن تعامل انسان-رایانه نیز بر انطباق طراحی الگوریتم‌ها با چالش‌های خاص ایجادشده توسط هوش مصنوعی تأکید دارد (یانگ و همکاران، ۲۰۲۰). با توجه به چالش‌های احصاشده در حوزهٔ تعامل انسان و هوش مصنوعی ذیل یافته‌های سطح ۷ پژوهش، باید چارچوبی برای طراحی هوش مصنوعی انسان‌محور ارائه شود. به اعتقاد فلوریدی و همکاران (۲۰۲۲) سیستم‌های هوش مصنوعی کم‌خطر نیازمند اقدام خاصی به‌لحاظ طراحی‌های انسان‌محورانه نیستند. درحالی‌که سیستم‌های با سطح خطر متوسط و پرخطر به‌شدت محل توجه طراحی الگوریتم‌های هوش مصنوعی انسان‌محور هستند. چنین سیستم‌هایی به‌رغم مخاطرات جدی‌ای که برای انسان دارند، از پتانسیل زیادی به‌لحاظ ارتقای سطح رفاه و بهزیستی او برخوردارند (مک‌گرگور، ۲۰۲۲)؛ زیرا افزایش آگاهی از مخاطرات جامعهٔ نوظهور مبتنی بر داده، به شکل‌گیری جنبش‌هایی با هدف ترغیب دانشمندان به استفاده از رویکردهای انسانی‌تر در طراحی سیستم‌های هوش مصنوعی منجر شده است. در چنین دیدگاهی، کرامت انسان و رفاه جامعه در مرکز فرایند طراحی قرار دارد. برخلاف پارادایم استاندارد هوش مصنوعی که بر داده‌ها و بیشینه‌سازی یک هدف مشخص تمرکز دارد، رویکرد هوش مصنوعی انسان‌محور ضمن تقلیل اهداف صرفاً مهندسی پارادایم استاندارد، اهداف انسانی را مبنای طراحی الگوریتم‌های هوش مصنوعی قرار می‌دهد. به بیان دیگر، هدف طراحی سیستم‌های مهندسی است که ذیل مفهوم کلان «تعامل انسان-رایانه» و با حفظ کرامت انسانی، امنیت، ایمنی و عاملیت انسان، سعی در رفع نیازهای کاربران و ذی‌نفعان مختلف به‌ویژه در اتخاذ تصمیم‌های مناسب و بهنگام دارد. چنین رویکردی به بیان اولسون (۲۰۲۱) مشتمل بر چهار وجه است:

۱. مردم: رویکرد طراحی هوش مصنوعی انسان‌محور بر نیازها، ارزش‌ها و علائق و سلايق گروه‌های مختلف از کاربران، ذی‌نفعان و فرهنگ‌های گوناگون تمرکز دارد. در این رویکرد، محصولات یا خدمات مبتنی بر هوش مصنوعی برای مردم و با مردم طراحی می‌شود؛
 ۲. فرایند: فرایند طراحی هوش مصنوعی انسان‌محور، مراحل طراحی انسان‌محورانه را وارد چرخه حیات توسعه محصول سنتی هوش مصنوعی می‌کند. مهم‌ترین اجزای این مراحل عبارت‌اند از: مشاهده نحوه عملکرد کاربران، تعامل با ذی‌نفعان، سنجش قابلیت استفاده، اصلاح پیاپی و نظارت مستمر بر عملکرد افراد با کمک الگوریتم‌های هوش مصنوعی؛
 ۳. محصول: سیستم‌های هوش مصنوعی انسان‌محور برای توانمندسازی، تقویت و ارتقای عملکرد انسان طراحی می‌شوند. چنین سیستم‌هایی به‌رغم تکیه بر الگوریتم‌های هوش مصنوعی و یادگیری ماشینی و نیز برخورداری از سطوح بالای اتوماسیون، بر قابلیت کنترل‌گری توسط انسان تأکید دارند؛
 ۴. اصول: طراحی هوش مصنوعی انسان‌محور مبتنی بر اصول هوش مصنوعی اخلاقی است. این اصول در هر مرحله از فرایند طراحی دنبال‌شده، به طراحان کمک می‌کنند فرصت‌های جدیدی را که از نظر اجتماعی قابل قبول یا ارجح هستند شناسایی و از آن‌ها استفاده کنند. درعین حال اصول مذکور احتمال این را که فناوری توسعه‌یافته از نظر اجتماعی غیرقابل قبول و در نتیجه رد شود، کاهش می‌دهند.
- شایان ذکر است یکی از چالش‌های کلیدی در طراحی و ارزیابی سیستم‌های هوش مصنوعی انسان‌محور این است که چنین سیستم‌هایی منعکس‌کننده اهداف و معیارهای ارزیابی هستند که اساساً بدون مشارکت فعال کاربران انسانی قابل‌دستیابی نیستند. تحقیقات مرتبط با تعامل انسان-رایانه تاریخچه‌ای غنی از ارزیابی انسانی و تحقیقات انسانی در مراحل اولیه طراحی دارد. تکنیک‌های شناسایی‌شده در مطالعات کلاسیک تعامل انسان-رایانه مانند آشنایدرمن (۲۰۱۶) و نیلسن (۱۹۹۳) مبتنی بر مشارکت فعال کاربران در ارزیابی یک رابط خاص میان انسان و رایانه است.
- حتی زمانی که چنین رویکردی برای استفاده بیشتر از طراحی مبتنی بر اصول و نظریه (استفن، ۲۰۲۱) زمینه‌سازی می‌کند، فناوری‌های نوظهور و تعاملات جدید همچنان نیازمند مشارکت کاربران در ارزیابی هستند. این ارزیابی می‌تواند دربردارنده نوعی بازتنظیم، اعتبارسنجی و تمایز قائل شدن با مدل‌های موجود و نظریات تعامل کاربر باشد.
- از منظر روش‌شناختی، طراحی و روش‌های ارزیابی موجود درخصوص تعامل انسان-رایانه نقشی حیاتی در فعال کردن هوش مصنوعی انسان‌محور دارند. همان‌طور که توضیح داده شد، ارزیابی به‌وسیله مشارکت‌کنندگان انسانی از اهمیت بسیار زیادی برخوردار است. ولی دانشمندان در هنگام طراحی الگوریتم‌های هوش مصنوعی انسان‌محور باید به این مسئله به‌عنوان حداقل ملزومات طراحی توجه کنند. از این‌رو ادغام اقداماتی دیگر در طراحی انسان‌محور با چرخه حیات سنتی هوش مصنوعی به‌ویژه در مراحل اولیه طراحی حائز اهمیت و سودمند خواهد بود. از طرفی ویرث و هیپ^۱ (۲۰۰۰) نیز یک فرایند طراحی سنتی هوش مصنوعی مشتمل بر چهار مرحله ارائه داده‌اند که به‌ترتیب عبارت‌اند از: ۱. تعریف مسئله، ۲. جمع‌آوری داده و مدیریت آن، ۳. مدل‌سازی و اجرا، و ۴. استقرار و استفاده (حکمرانی).

از ادغام رویکرد اولسون (۲۰۲۱) به عنوان رویکردی جدید در طراحی‌های هوش مصنوعی با محوریت منافع انسان و حفظ کرامت و بهزیستی او و رویکرد ویرث و هیپ (۲۰۰۰) به عنوان رویکردی سنتی در طراحی الگوریتم‌های هوش مصنوعی، سپس کاربردی آن بر داده‌های پژوهش حاضر و در نهایت جانمایی در چارچوب طراحی نوآورانه «الماس دوبر» به منظور تجزیه و تحلیل داده‌ها، یک مدل مفهومی به شکل زیر به دست آمد (شکل ۴).



شکل ۴. فرایند طراحی ادغام هوش مصنوعی با تعامل انسان-رایانه

مدل فوق بیانگر آن است که چگونه فرایندهای سنتی طراحی هوش مصنوعی و تعامل انسان-رایانه برای مواجهه با چالش‌های هوش مصنوعی انسان‌محور با یکدیگر ادغام می‌شوند. در این مدل، یک فرایند طراحی سیستم هوش مصنوعی سنتی و یک فرایند طراحی انسان‌محور منطبق بر چارچوب الماس دوبر به یکدیگر پیوند می‌خورند. چارچوب الماس دوبر میان طراحی مفهومی و تحقق در عالم واقع تمایز قائل می‌شود. از این رو برای نشان دادن فرایندهای طراحی انسان‌محور اعم از طراحی تعاملی، طراحی مشارکتی یا طراحی فراگیر کارآمد است. براساس چارچوب الماس دوبر و مشخصاً در مدل مفهومی فوق، جایگزینی فازهای واگرایی کشف و توسعه با فازهای همگرایی تعریف مسئله و تحویل و استقرار هوش مصنوعی، هم در طراحی مفهومی و هم در تحقق در عالم واقع امری ضروری است. به عبارت دیگر، در فاز نخست طراحی هوش مصنوعی انسان‌محور مسائل، چالش‌ها و دغدغه‌های جامعه علمی نسبت به این حوزه اکتشاف می‌شود. در واقع فاز واگرایی اکتشاف مانند پنجره‌ای گشوده رو به مسئله است. در فاز بعدی شاهد همگرایی دیدگاه‌های متنوع دانشمندان در خصوص مسئله و دستیابی به اجماع نظری هستیم. در فاز دوم، جامعه علمی به شناخت و نهایتاً تعریف

همگرا و یکپارچه‌ای از مسئله دست پیدا می‌کند. همان‌طور که در مدل مفهومی مشاهده می‌شود، در هر دو فاز واگرایی کشف و همگرایی تعریف مسئله، لازمه طراحی الگوریتم‌های هوش مصنوعی انسان‌محور پایبندی به اصول هوش مصنوعی اخلاقی است. پس از آنکه طراحان شناخت کافی از مسئله پیدا کردند، باید به دنبال راه‌حل برای آن باشند. در فاز سوم از چارچوب الماس دوبر که به راه‌حل مسئله اختصاص دارد، شاهد رویکردی واگرا و توسعه‌ای به مسئله هستیم. در این مرحله، سیستم‌های فنی هوش مصنوعی در اختیار کاربران انسانی قرار می‌گیرد و نحوه تعامل کاربران با این سیستم‌ها و قوت‌ها و ضعف‌های سیستم‌های مذکور در تأمین منافع کاربران به‌دقت رصد می‌شود. سپس در صورت نیاز، بازاندیشی‌های لازم در طراحی‌های فنی صورت می‌گیرد و بار دیگر تعامل کاربران با سیستم ارزیابی می‌شود. این مرحله (تحویل) آخرین فاز از مراحل چهارگانه طراحی هوش مصنوعی انسان‌محور است. همان‌طور که در مدل مفهومی مشاهده می‌شود، فرایندهای سنتی طراحی هوش مصنوعی و تعامل انسان-رایانه به‌وسیله سه عنصر با یکدیگر پیوند می‌خورند: «اصول هوش مصنوعی اخلاقی»، «کاربران انسانی» و «سیستم حکمرانی هوش مصنوعی». موفقیت مداخلات مرتبط با هوش مصنوعی انسان‌محور نیز تا حد زیادی مرهون کیفیت این عناصر است.

در مدل فوق، کسب اطمینان از رعایت اصول هوش مصنوعی اخلاقی و حفظ کنترل انسان و عاملیت او در طول چرخه حیات سیستم هوش مصنوعی از جمع‌آوری داده و نظارت گرفته تا طراحی، اجرا، ارزیابی، استقرار، استفاده و انفصال بیانگر اهداف اصلی طراحی هوش مصنوعی انسان‌محور و چارچوب ارزیابی آن است. به دلیل پیچیدگی‌های چارچوب مذکور، دانشمندان و طراحان هوش مصنوعی انسان‌محور در آخرین مرحله از چارچوب الماس دوبر به یک همگرایی و اتفاق نظر درخصوص نیاز به وجود یک ساختار حکمرانی دست پیدا می‌کنند که متضمن ایجاد و اجرای چارچوب طراحی هوش مصنوعی انسان‌محور و ارزیابی مخاطرات احتمالی و پیامد اجتماعی طراحی‌ها باشد. از این‌رو می‌توان گفت حکمرانی و سیاست‌های منطبق بر آن، در قبال آسیب‌های ناشی از ادغام رو به گسترش هوش مصنوعی در زندگی روزمره جامعه، مسئولیت و نقشی میانجی‌گرانه دارد (اشنایدرمن، ۲۰۲۰).

در یک جمع‌بندی، مدل مفهومی فوق تلاشی برای کشف مرتبط‌ترین مسائل با حوزه انسان در طراحی هوش مصنوعی است؛ پیشینه‌سازی شناخت به ماهیت انسان و ارائه تعریف درست از مسائل، ارائه کارآمدترین راه‌حل‌ها برای مسائل از طریق قراردادن کاربران انسانی در معرض هوش مصنوعی، مدل‌سازی و اجرا و ارزیابی نحوه تعامل آن‌ها، و درنهایت استقرار و استفاده از هوش مصنوعی ذیل یک سیستم حکمرانی مناسب که توانایی تأمین اهداف انسان‌محورانه طراحی الگوریتم‌ها را داشته باشد و به‌محض به‌مخاطره‌افتادن منافع انسان تصمیم به انفصال از ارائه خدمات بگیرد.

مأخذ مقاله: تألیف مستقل. در این مقاله تعارض منافی وجود ندارد.

منابع

- بابائیان، فاطمه، صفدری رنجبر، مصطفی، و حکیم، امین (۱۴۰۲). واکاوی نقش هوش مصنوعی در چرخه سیاست‌گذاری عمومی: رویکرد فراترکیب. *بهبود مدیریت*، ۱۷(۲)، ۱۱۵-۱۵۰. <https://doi.org/10.22034/jmi.2023.396945.2957>
- بلبلی قادیکلائی، سمیه، و پارسانیا، حمید (۱۴۰۲). مروری نظام‌مند بر دلالت‌های اخلاقی استفاده از هوش مصنوعی در فناوری‌های دیجیتال و نسبت آن با اخلاق شکوفایی. *راهبرد اجتماعی فرهنگی*، ۱۲(۳)، ۷۷۱-۷۹۸. <https://doi.org/10.22034/scs.2022.160772>
- رجبی، محسن، و نصراللهی، محمدصادق (۱۴۰۲). پیام‌شناسی فرهنگی توسعه هوش مصنوعی در رسانه‌های اجتماعی در ایران. *فصلنامه تحقیقات فرهنگی ایران*، ۱۶(۲)، ۹۵-۱۲۵. <https://doi.org/10.22035/jicr.2023.3178.3481>
- Akhtar, J. (2020). An interactive multi-agent reasoning model for sentiment analysis: A case for computational semiotics. *Artificial Intelligence Review*, 53, 3987-4004. <https://link.springer.com/article/10.1007/s10462-019-097856>.
- Ali, S., DiPaola, D., Lee, I., Sindato, V., Kim, G., Blumofe, R., & Breazeal, C. (2021). Children as creators, thinkers and citizens in an AI-driven future. *Computers and Education: Artificial Intelligence*, 2, 100040. <https://doi.org/10.1016/j.caeai.2021.100040>
- Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: A review. *Artificial Intelligence Review*, 55(1), 589-656. <https://link.springer.com/article/10.1007/s10462-021-10039-7>
- Ayobi, A., Stawarz, K., Katz, D., Marshall, P., Yamagata, T., Santos-Rodriguez, R., ... & O'Kane, A. A. (2021). Co-designing personal health? Multidisciplinary benefits and challenges in informing diabetes self-care technologies. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1-26.
- Babaeian, F., Safdari Ranjbar, M., & Hakim, A. (2023). Investigating the role of artificial intelligence in the public policy cycle; Metasynthesis approach. *Journal of Improvement Management*, 17(2), 115-150. <https://doi.org/10.22034/jmi.2023.396945.2957> (In Persian)
- Balcombe, L., & De Leo, D. (2022, February). Human-computer interaction in digital mental health. In *Informatics* (Vol. 9, No. 1, p. 14). MDPI. <https://doi.org/10.3390/informatics9010014>
- Bawa, A., Khadpe, P., Joshi, P., Bali, K., & Choudhury, M. (2020). Do Multilingual Users Prefer Chat-bots that Code-mix? Let's Nudge and Find Out!. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1), 1-23. <http://dx.doi.org/10.1145/3392846>
- Belfield, H. (2020, February). Activism by the AI community: Analysing recent achievements and future prospects. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 15-21). <http://dx.doi.org/10.1145/3375627.3375814>
- Bloomfield, B. P. (1988). Expert systems and human knowledge: A view from the sociology of science. *Ai & Society*, 2, 17-29.
- Bolboli Qadikolaei, S., & Parsania, H. (2023). A systematic review of the ethical implications of using artificial intelligence in digital technologies and its relationship with the ethics of flourishing. *Socio-Cultural Strategy*, 12(3), 771-798. <https://doi.org/10.22034/scs.2022.160772> (In Persian)
- Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the progress of reason. *Social Science Information*, 14(6), 19-47. <http://ssi.sagepub.com>
- British Design Council. [n.d.]. *The Double Diamond - Design Council*. <https://www.designcouncil.org.uk/our-resources/the-double-diamond/>
- Chambers-Jones, C. (2021). AI, big data, quantum computing, and financial exclusion: Tempering enthusiasm and offering a human-centric approach to policy. *FinTech, Artificial Intelligence and the Law*, 193-210.
- Wohlin, C. (2014, May). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering* (pp. 1-10). <https://doi.org/10.1145/2601248.2601268>

- Collins, R. (1994, June). Why the social sciences won't become high-consensus, rapid-discovery science. In *Sociological forum* (Vol. 9, pp. 155-177). Kluwer Academic Publishers-Plenum Publishers. <https://doi.org/10.1007/BF01476360>
- Demir, M., McNeese, N. J., & Cooke, N. J. (2019). The evolution of human-autonomy teams in remotely piloted aircraft systems operations. *Frontiers in Communication*, 4, 50. <https://doi.org/10.3389/fcomm.2019.00050>
- Dignum, F., & Dignum, V. (2020). How to center AI on humans. In *NeHuAI 2020, 1st International Workshop on New Foundations for Human-Centered AI, Santiago de Compostella, Spain, September 4, 2020* (pp. 59-62).
- Dikmen, M., & Burns, C. (2022). The effects of domain knowledge on trust in explainable AI and task performance: A case of peer-to-peer lending. *International Journal of Human-Computer Studies*, 162, 102792. <http://dx.doi.org/10.1016/j.ijhcs.2022.102792>
- Esposito, E. (2017). Artificial communication? The production of contingency by algorithms. *Zeitschrift für Soziologie*, 46(4), 249-265. <https://doi.org/10.1515/zfsoz-2017-1014>
- Eynon, R., & Young, E. (2021). Methodology, legend, and rhetoric: The constructions of AI by academia, industry, and policy groups for lifelong learning. *Science, Technology, & Human Values*, 46(1), 166-191. <http://dx.doi.org/10.1177/0162243920906475>
- Floridi, L., Holweg, M., Taddeo, M., Amaya, J., Mökander, J., & Wen, Y. (2022). CapAI-A procedure for conducting conformity assessment of AI systems in line with the EU artificial intelligence act. *Available at SSRN 4064091*.
- Floridi, L., Holweg, M., Taddeo, M., Amaya, J., Mökander, J., & Wen, Y. (2022). CapAI-A procedure for conducting conformity assessment of AI systems in line with the EU artificial intelligence act. *Available at SSRN 4064091*. <https://doi.org/10.2139/ssrn.4064091>
- Ford, K. M., Hayes, P. J., Glymour, C., & Allen, J. (2015). Cognitive Orthoses: Toward Human-Centered AI. *AI Magazine*, 36(4), 5-8. <https://doi.org/10.1609/aimag.v36i4.2629>
- Gratch, J., Mao, W., & Marsella, S. (2006). *Modeling social emotions and social attributions* (pp. 219-251). Cambridge: Cambridge University Press.
- Gu, H., Huang, J., Hung, L., & Chen, X. A. (2021). Lessons learned from designing an AI-enabled diagnosis tool for pathologists. *Proceedings of the ACM on Human-computer Interaction*, 5(CSCW1), 1-25. <https://doi.org/10.1145/3449084>
- Gustafsson, D. (2019). Analyzing the double diamond design process through research and implementation. *MA Thesis*. Aalto University.
- Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human-machine communication research agenda. *New Media & Society*, 22(1), 70-86. <https://doi.org/10.1177/1461444819858691>
- Hansen, S. S. (2022). Public AI imaginaries: How the debate on artificial intelligence was covered in Danish newspapers and magazines 1956–2021. *Nordicom Review*, 43(1), 56-78. <https://doi.org/10.2478/nor-2022-0004>
- Hickman, E., & Petrin, M. (2021). Trustworthy AI and corporate governance: the EU's ethics guidelines for trustworthy artificial intelligence from a company law perspective. *European Business Organization Law Review*, 22, 593-625. <https://doi.org/10.1007/s40804-021-00224-0>
- Miyashita, H. (2021). Human-centric data protection laws and policies: A lesson from Japan. *Computer Law & Security Review*, 40, 105487. <http://dx.doi.org/10.1016/j.clsr.2020.105487>
- How, M. L., Chan, Y. J., & Cheah, S. M. (2020). Predictive insights for improving the resilience of global food security using artificial intelligence. *Sustainability*, 12(15), 6272. <https://doi.org/10.3390/su12156272>
- Huang, Y., Fei, T., Kwan, M. P., Kang, Y., Li, J., Li, Y., ... & Bian, M. (2020). GIS-based emotional computing: A review of quantitative approaches to measure the emotion layer of human-environment relationships. *ISPRS International Journal of Geo-Information*, 9(9), 551. <http://dx.doi.org/10.3390/ijgi9090551>
- Huhtamo, E. (2020). The self-driving car: A media machine for posthumans?. *Artnodes*, 26, 1-14. <http://dx.doi.org/10.7238/a.v0i26.3374>
- Jørgensen, R. F. (2023). Data and rights in the digital welfare state: the case of Denmark. *Information, Communication & Society*, 26(1), 123-138. <http://dx.doi.org/10.1080/1369118X.2021.1934069>

- Joyce, K., Smith-Doerr, L., Alegria, S., Bell, S., Cruz, T., Hoffman, S. G., ... & Shestakofsky, B. (2021). Toward a sociology of artificial intelligence: A call for research on inequalities and structural change. *Socius*, 7, 2378023121999581. <http://dx.doi.org/10.1177/2378023121999581>
- Kaasinen, E., Anttila, A. H., Heikkilä, P., Laarni, J., Koskinen, H., & Vääänen, A. (2022). Smooth and resilient human-machine teamwork as an industry 5.0 design challenge. *Sustainability*, 14(5), 2773. <https://doi.org/10.3390/su14052773>
- Kloos, C. D., Dimitriadis, Y., Hernández-Leo, D., Alario-Hoyos, C., Martínez-Monés, A., Santos, P., ... & Safont, L. V. (2022, March). H2o learn-hybrid and human-oriented learning: trustworthy and human-centered learning analytics (tahcla) for hybrid education. In *2022 IEEE Global Engineering Education Conference (EDUCON)* (pp. 94-101). IEEE. <http://dx.doi.org/10.1109/EDUCON52537.2022.9766770>
- Kolesnichenko, O., Mazelis, L., Sotnik, A., Yakovleva, D., Amelkin, S., Grigorevsky, I., & Kolesnichenko, Y. (2021). Sociological modeling of smart city with the implementation of UN sustainable development goals. *Sustainability Science*, 16(2), 581-599. <https://doi.org/10.1007/s11625-020-00889-5>
- Lee, F., & Björklund Larsen, L. (2019). How should we theorize algorithms? Five ideal types in analyzing algorithmic normativities. *Big Data & Society*, 6(2), 2053951719867349. <http://dx.doi.org/10.1177/2053951719867349>
- Lee, J. A., Hilty, R., & Liu, K. C. (Eds.). (2021). *Artificial intelligence and intellectual property*. Oxford University Press.
- Leonard, P., & Tyers, R. (2023). Engineering the revolution? Imagining the role of new digital technologies in infrastructure work futures. *New Technology, Work and Employment*, 38(2), 291-310. <http://dx.doi.org/10.1111/ntwe.12226>
- Loi, M., Ferrario, A., & Viganò, E. (2021). Transparency as design publicity: explaining and justifying inscrutable algorithms. *Ethics and Information Technology*, 23(3), 253-263. <https://doi.org/10.1007/s10676-020-09564-w>
- Marres, N. (2020). Co-existence or displacement: Do street trials of intelligent vehicles test society?. *The British journal of sociology*, 71(3), 537-555. <http://dx.doi.org/10.1111/1468-4446.12730>
- McGregor, S. (2022). *AI incident database*. <https://incidentdatabase.ai>
- Mühlhoff, R. (2020). Human-aided artificial intelligence: Or, how to run large computations in human brains? Toward a media sociology of machine learning. *New Media & Society*, 22(10), 1868-1884. <http://dx.doi.org/10.1177/1461444819885334>
- Napoli, A. (2020). The human-centered AI and the EC policies: risks & chances. In *Technological and Digital Risk: Research Issues*. (pp. 67-92). Peter Lang.
- Natale, S., & Ballatore, A. (2020). Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence: The International Journal of Research into New Media Technologies*, 26(1), 3-18. <https://doi.org/10.1177/1354856517715164>
- Nielsen, J. (1993). *Usability engineering*. Imprint: Morgan Kaufmann. Elsevier Inc.
- Olsson, T., & Väänänen, K. (2021). How does AI challenge design practice?. *Interactions*, 28(4), 62-64. <https://doi.org/10.1145/3467479>
- Oravec, J. A. (2022). The emergence of 'truth machines'? Artificial intelligence approaches to lie detection. *Ethics and Information Technology*, 24, 6. <http://dx.doi.org/10.1007/s10676-022-09621-6>
- Perucica, N., & Andjelkovic, K. (2022). Is the future of AI sustainable? A case study of the European Union. *Transforming Government: People, Process and Policy*, 16(3), 347-358. <https://doi.org/10.1108/TG-06-2021-0106>
- Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1), 22. <https://doi.org/10.1186/s41039-017-0062-8>
- Rajabi, M., & Nasrollahi, M. (2023). The cultural impact of artificial intelligence development on social media in Iran. *Journal of Iranian Cultural Research*, 16(2), 95-125. <https://doi.org/10.22035/jicr.2023.3178.3481> (In Persian)
- Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Hum. Behav. Emerg. Technol*, 1(1), 33-36. <http://dx.doi.org/10.1002/hbe2.117>

- Russell, S., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. Harlow: Pearson Education Limited.
- Sabanovic, S. (2014). Inventing Japan's 'robotics culture': The repeated assembly of science, technology, and culture in social robotics. *Social Studies of Science*, 44(3), 342–367. <http://dx.doi.org/10.1177/0306312713509704>
- Schmid, S., Riebe, T., & Reuter, C. (2022). Dual-use and trustworthy? A mixed methods analysis of AI diffusion between civilian and defense R&D. *Science and engineering ethics*, 28(2), 12. <http://dx.doi.org/10.1007/s11948-022-00364-7>
- Shen, Y. (2019). Create synergies and inspire collaborations around the development of intelligent infrastructure for human-centered communities. *Journal of the Association for Information Science and Technology*, 70(6), 596–606. <https://doi.org/10.1002/asi.24150>
- Shneiderman, B. (2016). The dangers of faulty, biased, or malicious algorithms requires independent oversight. *Proceedings of the National Academy of Sciences*, 113(48), 13538–13540. <https://doi.org/10.1073/pnas.1618211113>
- Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 495–504. <https://doi.org/10.1080/10447318.2020.1741118>
- Shneiderman, B. (2021). Human-centered AI. *Issues in Science and Technology*, 56–61.
- Sonetti, G., Naboni, E., & Brown, M. (2018). Exploring the potentials of ICT tools for human-centric regenerative design. *Sustainability*, 10(4), 1217. <https://doi.org/10.3390/su10041217>
- Srnicek, N. C. (2016). *Platform capitalism*. Cambridge: Polity.
- Steffen, D. (2021). Taking the Next Step towards Convergence of Design and HCI: Theories, Principles, Methods. In *HCI International 2021-Posters: 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part I 23* (pp. 67–74). Springer International Publishing.
- Suchman, L. (1987). *Plans and Situated Actions: The Problem of Human-Machine Communication*.
- Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press.
- Tiersen, F., Batey, P., Harrison, M. J., Naar, L., Serban, A. I., Daniels, S. J., & Calvo, R. A. (2021). Smart home sensing and monitoring in households with dementia: user-centered design approach. *JMIR aging*, 4(3), e27047. <https://doi.org/10.2196/27047>
- van Berkel, N., Tag, B., Goncalves, J., & Hosio, S. (2022). Human-centred artificial intelligence: a contextual morality perspective. *Behaviour & Information Technology*, 41(3), 502–518. <https://doi.org/10.1080/0144929X.2020.1818828>
- Wajcman, J. (2017). Automation: Is it really different this time?. *The British Journal of Sociology*, 68(1), 119–127. <http://dx.doi.org/10.1111/1468-4446.12239>
- Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (Vol. 1, pp. 29–39).
- Wolfe, A. (1993). *The Human Difference: Animals, Computers, and the Necessity of Social Science*. Berkley, CA: University of California Press.
- Woolgar, S. (1985). Why not a sociology of machines? The case of sociology and artificial intelligence. *Sociology*, 19(4), 557–572.
- Xu, W. (2019). Toward human-centered AI. *Interactions*, 26(4), 42–46. <https://doi.org/10.1145/3328485>
- Yang, S. J., Ogata, H., Matsui, T., & Chen, N. S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence*, 2, 100008. <https://doi.org/10.1016/j.caeai.2021.100008>
- Sah, Y. J. (2022). Anthropomorphism in human-centered AI: Determinants and consequences of applying human knowledge to AI agents. In *Human-Centered Artificial Intelligence* (pp. 103–116). Academic Press. <http://dx.doi.org/10.1016/B978-0-323-85648-5.00013-X>
- Ziewitz, M. (2016). Governing algorithms: Myth, mess, and methods. *Science, Technology, & Human Value*, 41(1), 3–16. <https://doi.org/10.1177/0162243915608948>
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: Public Affairs