

Analysis and Evaluation of Ethical Issues Related to the Application of Artificial Intelligence and Neurotechnology in the Workplace: A Scoping Review

Sajjad Balali Shahvari¹, Alireza Amini²

Abstract

Given the increasing significance of artificial intelligence in contemporary society, there is a fundamental need to address ethical issues related to these technologies. This study adopts a targeted literature review approach to systematically analyze existing scholarly work on AI ethics. By examining 53 peer-reviewed articles sourced from the reputable Scopus and Web of Science databases, the research aims to identify prevailing themes, concerns, and recommendations related to ethical issues in AI deployment. The selection criteria focused on works that explicitly address ethical dimensions, ensuring the findings are directly relevant to the core subject matter. The findings reveal that concepts such as justice, fairness, equality, racial and gender stereotypes, responsibility, accountability, and the promotion of well-being hold the highest levels of importance within the existing literature. These topics are predominantly emphasized in the field of AI ethics. To effectively address the associated ethical challenges, enhanced collaboration among stakeholders, greater alignment of ethical guidelines and regulations, standardization efforts, the utilization of ethical assessment mechanisms, and a focused attention to justice in the application of AI within organizational processes are identified as key points of intervention.

Keywords: *Artificial Intelligence, Neurotechnology, Ethical Considerations, Ethical AI.*

1. Ph.D. Student in Public Administration, Management Department, Shiraz University, Shiraz, Iran
balali.sajad@pgs.usb.ac.ir

2. Corresponding Author: Associate Prof., Management Department, Shiraz University, Shiraz, Iran
alirezaamini@shirazu.ac.ir

بررسی و تحلیل موضوعات اخلاقی در مورد کاربرد هوش مصنوعی و فناوری‌های عصبی در محیط کار: یک مرور هدفمند

سجاد بلالی شهوری*، علیرضا امینی**

تاریخ دریافت: ۱۴۰۴/۰۲/۲۵

تاریخ پذیرش نهایی: ۱۴۰۴/۰۸/۰۲

چکیده

باتوجه به اهمیت فزاینده هوش مصنوعی در جامعه کنونی، نیاز اساسی در رسیدگی به مسائل و موضوعات اخلاقی مرتبط با این فناوری‌ها وجود دارد. بنابراین، در این مطالعه مروری که از نوع هدف‌مند است، پژوهشگران، جستجوی رایانه‌ای را با استفاده از کلمات کلیدی مرتبط در پایگاه‌های اطلاعاتی *Scopus* و *Web of Science* در بازه زمانی ۱۹۷۸ تا ماه مارس ۲۰۲۵ انجام دادند. نتیجه این مرحله، دستیابی به ۵۴۵۶ مطالعه از طریق جستجو در پایگاه‌های اطلاعاتی و ۴۱ مطالعه از طریق جستجوی دیگر منابع شامل جستجو از طریق رفرنس فهرست مقالات یافت شده بود که در نهایت ۵۳ مقاله که به موضوع اخلاق در به کارگیری هوش مصنوعی اشاره نموده بودند، مورد بررسی عمیق قرار گرفتند. مطالعه حاضر در پنج مرحله انجام شده است که عبارتند از: طراحی سؤال تحقیق، جستجو و استخراج مطالعات مرتبط با تحقیق، انتخاب مطالعات منتخب و خلاصه کردن اطلاعات و گزارش نتایج آنها. یافته‌های پژوهش نشان داد که عدالت، انصاف، برابری و کلیشه‌های نژادی و جنسیتی، مسئولیت، پاسخگویی، ترویج رفاه، شفافیت و حفظ حریم خصوصی، دارای بیشترین ضریب اهمیت بین مطالعات صورت گرفته می‌باشند؛ یعنی در حوزه اخلاق هوش مصنوعی این موضوعات بیشتر مورد توجه و تأکید قرار گرفته‌اند که به منظور حل چالش‌های مربوط به موضوعات اخلاقی نیازمند همکاری بیشتر بین ذی‌نفعان، هماهنگی بیشتر بین دستورالعمل‌های اخلاقی و قوانین، ضرورت استانداردسازی، استفاده از سازوکارهای اخلاقی پژوهی و توجه ویژه به عدالت در استفاده از هوش مصنوعی در فرایندهای سازمانی، به عنوان نقاط کلیدی مطرح می‌شوند.

کلیدواژه: هوش مصنوعی؛ فناوری‌های عصبی؛ موضوعات اخلاقی؛ هوش مصنوعی اخلاقی.

مقدمه

در دنیای امروز، شاهد دیجیتالی شدن فزاینده جهان فیزیکی هستیم که ناشی از وابستگی روزافزون فرایندها، نظام‌ها و زنجیره‌های تأمین به فناوری‌های دیجیتال با محوریت هوش مصنوعی است (برینجولفسون و مک‌آفی^۱، ۲۰۱۴). هوش مصنوعی با توانایی انجام وظایف شناختی با حداقل یا بدون دخالت انسان، ظرفیت ایجاد تغییرات اساسی را در جامعه دارد (پاتاک و همکاران^۲، ۲۰۲۳). در واقع، هوش مصنوعی، فناوری ارتباطات را بهبود بخشیده، دسترسی به منابع و اطلاعات را تسهیل کرده و امکان توسعه اقتصاد و ایجاد شغل را فراهم ساخته است (کاستلز^۳، ۲۰۰۲). با این حال، استفاده از هوش مصنوعی در فناوری‌های دیجیتال، ما را با یک چالش اخلاقی روبه‌رو می‌کند؛ تضاد بین پیشرفت‌های فناوری و اجتماعی از یک سو و معضلات اخلاقی ناشی از این فناوری‌ها از سوی دیگر (فلوریدی^۴، ۲۰۱۸). ضمن اینکه توسعه هوش مصنوعی غالباً در انحصار گروهی از متخصصان قرار داشته است که نتوانسته‌اند به اندازه کافی تنوع قومی، فرهنگی، جنسیتی، سنی، جغرافیایی و اقتصادی جامعه انسانی را منعکس کنند. این امر خطر وابستگی انسان‌ها به سامانه‌های خودکار را افزایش می‌دهد. در صورت جهت‌گیری مغرضانه هوش مصنوعی، استقلال انسان‌ها به خطر می‌افتد و انسان‌ها به جای بهره‌مندی از این فناوری، در خدمت نیازهای هوش مصنوعی قرار می‌گیرند (اشوک و همکاران^۵، ۲۰۲۲). مسئله مهم دیگر این است که توسعه فناوری با سرعت بیشتری نسبت به تکامل اخلاق پیش رفته است؛ زیرا تعداد کمی از متخصصان به یک تحلیل رسمی از اخلاق و دیجیتالی‌سازی پرداخته‌اند. به عبارت دیگر، چشم‌انداز اخلاقی پراکنده در مورد دستورالعمل‌های هوش مصنوعی، یک مشکل اساسی برای ایجاد یک حکمرانی مبتنی بر فناوری و کاربردهای آن است (جویین و همکاران^۶، ۲۰۱۹).

در قلب این نگرانی‌ها، سؤال اساسی این است که چگونه می‌توان بین کارایی‌های ناشی از هوش مصنوعی و حمایت از حقوق و منافع کارکنان، تعادل برقرار کرد (ارنست و همکاران^۷، ۲۰۱۹). با تکامل و پیشرفت فناوری‌های مبتنی بر هوش مصنوعی، این فناوری‌ها قادر به انجام وظایفی می‌شوند که زمانی

1. Brynjolfsson & McAfee
2. Pathak
3. Castells
4. Floridi
5. Ashok
6. Jobin et al.
7. Ernst

در انحصار کارکنان بود؛ مشاغلی که شامل وظایف تکراری، پیش‌پاافتاده یا حتی وظایف شناختی پیچیده است. این امر نگرانی‌هایی را در مورد جابه‌جایی مشاغل، نابرابری و آینده کاری ایجاد می‌کند. در حالی که برخی استدلال می‌کنند که هوش مصنوعی فرصت‌های شغلی جدیدی ایجاد کرده و به کارکنان اجازه می‌دهد تا بر وظایف خلاقانه، راهبردی و معنادارتر تمرکز کنند و برخی دیگر از این می‌ترسند که هوش مصنوعی به بیکاری گسترده و تشدید شکاف‌های اجتماعی منجر شود. این دیدگاه‌های متناقض، نیاز فوری به یک چارچوب اخلاقی تعریف‌شده را برجسته می‌کند که تضمین کند هوش مصنوعی به گونه‌ای به کار گرفته شود که مزایای اجتماعی را به حداکثر و آسیب را به حداقل برساند (تئو^۱، ۲۰۲۳).

ظهور هوش مصنوعی همچنین چالش‌های جدی را در ارتباط با کرامت انسانی به وجود آورده است. از آنجا که سامانه‌های هوش مصنوعی در فرایندهایی همچون استخدام، ارتقای شغلی، ارزیابی عملکرد و نظارت کارکنان، نقش فزاینده‌ای ایفا می‌کنند، این نگرانی وجود دارد که انسان‌ها صرفاً به‌عنوان اجزای بی‌روح یک ماشین تلقی شوند و شأن و منزلت انسانی آنها نادیده گرفته شود. همچنین الگوریتم‌های هوش مصنوعی ممکن است سوگیری‌های اجتماعی موجود در داده‌های تاریخی را بازتولید کنند و به نتایج تبعیض‌آمیز در استخدام، ارتقا و ارزیابی عملکرد منجر شوند (کایالی^۲، ۲۰۲۶). تهی کردن کار از روح انسانی، همراه با احتمال سوگیری و تبعیض، اصول عدالت، انصاف و برابری را در محیط کار تهدید می‌کند. بنابراین، ملاحظات و موضوعات اخلاقی باید بر اطمینان از طراحی و پیاده‌سازی سامانه‌های هوش مصنوعی به گونه‌ای متمرکز شود که به حقوق و کرامت کارکنان احترام بگذارد، شفافیت و پاسخگویی را ارتقا دهد و یک بازار کار فراگیر و منصفانه را تقویت کند. در واقع، یکی از مهم‌ترین چالش‌های اخلاقی این است که اطمینان حاصل شود هوش مصنوعی برای تداوم یا تشدید سوگیری‌ها و نابرابری‌های موجود استفاده نمی‌شود (فراهانی و قاسمی، ۲۰۲۴) با وجود اهمیت فزاینده مسائل اخلاقی در کاربرد هوش مصنوعی و توجه گسترده به این موضوع در پژوهش‌های پیشین (مانند: تورنتون^۳، ۲۰۱۷؛

1. Teo
2. Kayyali
3. Thornton

اندرسون و اندرسون^۱، ۲۰۱۷؛ سروانتس و همکاران^۲، ۲۰۲۰؛ بونفون و همکاران^۳، ۲۰۱۶؛ اواد و همکاران^۴، ۲۰۱۸، جوبین، و همکاران، ۲۰۱۹، شواهد نشان می‌دهد که این مطالعات تاکنون در دستیابی به اهداف هنجاری و اخلاقی موفق نبوده‌اند. در واقع، پرداختن به چالش‌ها و موضوعات اخلاقی مرتبط با هوش مصنوعی در محیط کار هنوز به‌طور جامع و نظام‌مند مورد بررسی قرار نگرفته است. از این رو، ضرورت دارد تا موضوعات اخلاقی مرتبط با کاربرد هوش مصنوعی و فناوری‌های عصبی در محیط‌های کاری به‌صورت دقیق‌تر و هدف‌مند تحلیل شوند. بر این اساس، این مقاله با رویکرد مرور هدف‌مند به بررسی دیدگاه‌ها و مسائل اخلاقی در زمینه به‌کارگیری هوش مصنوعی و فناوری‌های عصبی در محیط کار می‌پردازد و پاسخ می‌دهد که مهم‌ترین دغدغه‌ها و موضوعات اخلاقی در این حوزه کدامند؟

کاربرد هوش مصنوعی در محیط کار

تحول دیجیتال و ظهور انقلاب صنعتی، مرزهای سنتی کسب‌وکار را دستخوش تغییر قرار داده و سازمان‌ها را با مجموعه‌ای از چالش‌ها و فرصت‌های بی‌سابقه مواجه ساخته است. در این راستا، هوش مصنوعی به‌عنوان یکی از پیشران‌های اصلی این تحول، با قابلیت‌های منحصر به فرد خود در پردازش داده‌های حجیم، یادگیری ماشینی و تصمیم‌گیری خودکار، توجه فزاینده‌ای را به خود جلب کرده است (پدرامی و نرگسیان، ۱۴۰۴). در واقع، هوش مصنوعی به‌عنوان یک فناوری تحول‌آفرین، به‌سرعت در حال بازتعریف شیوه‌های زندگی، کار و تعامل انسان با جهان پیرامون است. این فناوری نه تنها فرصت‌های بی‌سابقه‌ای برای افزایش بهره‌وری، کارایی و نوآوری فراهم آورده، بلکه همزمان چالش‌های پیچیده‌ای در حوزه‌های اخلاقی، اجتماعی و اقتصادی مطرح کرده است؛ از فرایندهای تولید خودکار گرفته تا تصمیم‌گیری‌های مبتنی بر داده در منابع انسانی و امور مالی. کاربرد هوش مصنوعی به تدریج به تمامی بخش‌های اقتصاد نفوذ کرده است (خان و همکاران^۵، ۲۰۲۴) در محیط‌های کاری، سامانه‌های مجهز به هوش مصنوعی قادرند ریسک‌ها را شناسایی، ارزیابی و مدیریت کنند. این سامانه‌ها همچنین می‌توانند وظایف تکراری را خودکارسازی کرده، قابلیت اطمینان عملیاتی در فرایندهای صنعتی را ارتقا و کارایی تجهیزات حفاظت شخصی هوشمند را افزایش دهند. فناوری‌های حسگر پیشرفته که با هوش مصنوعی

1. Anderson and Anderson
2. Cervantes et al.
3. Bonnefon
4. Awad et al.
5. Khan

تلفیق شده‌اند، امکان ارزیابی و کنترل مستمر مواجهه با خطرات محیطی را فراهم می‌سازند. یکی از حوزه‌های پیشرو در این زمینه، رباتیک صنعتی است. ادغام قابلیت‌های هوش مصنوعی با دستگاه‌های رباتیک، موجب شتاب در توسعه «ربات‌های مشارکتی» شده است. این ربات‌ها با بهره‌گیری از ورودی‌های حسگرها و الگوریتم‌های پیشرفته، می‌توانند مهارت‌های انسانی نظیر انعطاف‌پذیری، چابکی و حل مسئله را با قدرت و دقت مکانیکی ترکیب کرده و در قالب همکاری نزدیک با نیروی انسانی عمل کنند. توانایی درک فرمان‌های صوتی و پاسخگویی از طریق زبان طبیعی، یکی از ویژگی‌های نوین این ربات‌ها به‌شمار می‌رود.

افزون بر کاربرد ربات‌های مشارکتی در تولید، هوش مصنوعی در عملیات انبارداری نیز نقش فزاینده‌ای ایفا می‌کند. ربات‌های هدایت‌شونده خودکار که در مسیرهای ازپیش‌تعیین‌شده حرکت می‌کنند، وظایف حمل و جابه‌جایی مواد را با دقت و کارایی بالاتری انجام داده و موجب ساده‌سازی عملیات و افزایش بهره‌وری می‌شوند. مسیر تحقیقات نیز به سوی طراحی ربات‌هایی است که بتوانند فراتر از مسیرهای محدود عمل کرده و تعامل طبیعی‌تری با محیط داشته باشند. در این راستا، مدل‌هایی تحت عنوان مدل‌های هوش مصنوعی بینایی - زبان - عمل^۱ توسعه یافته‌اند که داده‌های متنی و تصویری مدل‌های زبانی بزرگ را با اطلاعات حسگرهای ربات تلفیق می‌کنند تا به‌عنوان «مغز» ربات‌ها عمل نمایند. این مدل‌ها می‌توانند در آینده، به ربات‌ها برای تصمیم‌گیری و حرکت در دنیای فیزیکی آزادی عملی بیشتر اعطا کنند (هاوارد و شولتی، ۲۰۲۴).

یکی از عوامل کلیدی در رشد چشمگیر فناوری‌های هوش مصنوعی، پذیرش فزاینده مصرف‌کنندگان از محصولات مبتنی بر این فناوری در محیط‌های کاری و زندگی روزمره است. نمونه‌های برجسته این روند، شامل وسایل نقلیه خودران، دستیارهای دیجیتال شخصی، سامانه‌های توصیه‌گر و ابزارهای هوشمند در حوزه‌هایی نظیر خرده‌فروشی، تجارت الکترونیک، خدمات مالی و مراقبت‌های بهداشتی است. وسایل نقلیه خودران با بهره‌گیری از فناوری‌هایی نظیر دید رایانه‌ای و یادگیری عمیق، قادر به تصمیم‌گیری در زمان واقعی هستند. گرچه بسیاری از خودروهای موجود تنها از قابلیت‌هایی مانند ترمز خودکار، شتاب‌گیری و فرمان خودکار بهره می‌برند، اما به مرحله خودران کامل نرسیده‌اند. در صنعت خرده‌فروشی، هوش مصنوعی در سامانه‌های توصیه‌گر نقش کلیدی ایفا می‌کند.

به‌عنوان مثال، سامانه توصیه‌گر آمازون مسئول تولید حدود ۳۵٪ از درآمد کل شرکت است (فوربس^۱، ۲۰۱۸). نتفلیکس نیز اعلام کرده است که حدود ۸۰٪ از ساعات مشاهده کاربران تحت تأثیر الگوریتم‌های توصیه‌گر مبتنی بر هوش مصنوعی قرار دارد که به‌تنهایی سالانه بیش از ۱ میلیارد دلار برای این شرکت ارزش آفرینی می‌کند (گومز - اوربیه و هانت^۲، ۲۰۱۶). در صنایع خدمات مالی و مراقبت‌های بهداشتی نیز، محصولات و خدمات مبتنی بر هوش مصنوعی به سرعت در حال رشد هستند. از تشخیص زودهنگام بیماری‌ها تا بهینه‌سازی فرایندهای مالی، این فناوری نقش مهمی در بهبود کیفیت خدمات و افزایش بازدهی عملیاتی ایفا می‌کند. هوش مصنوعی در سال‌های اخیر به یکی از مؤلفه‌های کلیدی در تحول خدمات مالی و مراقبت‌های بهداشتی تبدیل شده است. در حوزه مالی، پلتفرم‌های مشاوره مالی خودکار یا مشاوران ربات^۳ از الگوریتم‌های پیچیده برای کمک به مشتریان در طراحی، بهینه‌سازی و مدیریت سبدهای سرمایه‌گذاری استفاده می‌کنند. این ابزارها با تحلیل داده‌های بازار و ویژگی‌های شخصی سرمایه‌گذاران، پیشنهادهای سرمایه‌گذاری خودکار، مقرون‌به‌صرفه و مبتنی بر داده ارائه می‌دهند.

علاوه بر این، شرکت‌های نوظهور فناوری مالی (فین‌تک) مانند Upstart، از هوش مصنوعی برای ارتقای فرایند اعتبارسنجی و وام‌دهی بهره می‌برند. این شرکت‌ها با تحلیل داده‌های غیرسنتی، مانند سوابق تحصیلی، سابقه شغلی و رفتارهای پرداخت گذشته، به مدل‌سازی دقیق‌تر ریسک اعتباری می‌پردازند و فرایند صدور وام را به‌صورت خودکار انجام می‌دهند. در صنعت مراقبت‌های بهداشتی نیز، هوش مصنوعی به‌عنوان یک نیروی تحول‌آفرین ظاهر شده است. جراحی‌های رباتیک مبتنی بر هوش مصنوعی نه تنها دقت در انجام اعمال جراحی را افزایش داده‌اند، بلکه به کاهش قابل توجه عوارض جانبی و کوتاه‌شدن مدت زمان بستری بیماران منجر شده‌اند. یکی دیگر از کاربردهای نوآورانه هوش مصنوعی در این حوزه، توسعه دستیاران پرستار مجازی مانند Care Angel است. این دستیاران با استفاده از فناوری‌های هوش مصنوعی، توانایی نظارت بر علائم حیاتی، ثبت و پیگیری داده‌های سلامت گزارش شده توسط بیمار و شناسایی زودهنگام خطرات احتمالی در زمان واقعی را دارند. همچنین، این سامانه‌ها می‌توانند نقش مؤثری در مدیریت بیماری‌های مزمن، پیشگیری و ارتقای پایبندی به دارو ایفا

1. Forbes
2. Gomez-Urbe & Hunt
3. Robo-Advisors

کنند و بار کاری کادر درمان را کاهش دهند (دی یو و ایکس آی ای^۱، ۲۰۲۰). پژوهشگران نیز در سال‌های اخیر با بهره‌گیری از ظرفیت‌های کلان‌داده و به‌کارگیری الگوریتم‌های پیشرفته هوش مصنوعی، توانسته‌اند ابعاد گوناگون رفتار کاربران، الگوهای مصرف و شاخص‌های مرتبط با رضایت و اعتماد را با دقتی فراتر از روش‌های سنتی مورد تحلیل قرار دهند. این نوع رویکرد تحلیلی زمینه‌ساز طراحی مدل‌های پیش‌بینی‌کننده دقیق‌تری است که می‌توانند در فرایند بهبود تجربه کاربری و ارتقای سطح رضایت ذی‌نفعان نقش بسزایی ایفا کنند (اربابی اصفهانی و همکاران، ۱۴۰۴). با این حال، موفقیت در پیاده‌سازی و بهره‌برداری از چنین ابزارهایی صرفاً به بعد فناورانه محدود نمی‌شود، بلکه به‌شدت وابسته به میزان انطباق آنها با فرهنگ سازمانی، کیفیت و دقت در طراحی ابزارهای هوشمند و همچنین کارآمدی فعالیت‌های مرتبط با مدیریت داده است (مدرسی و همکاران، ۱۴۰۳). در واقع، در سازمان‌هایی که فرهنگ داده‌محور و دیجیتال در آنها نهادینه شده است، تصمیم‌گیری‌ها عمدتاً بر پایه تجزیه و تحلیل داده‌ها و بهره‌گیری از ابزارهای هوش مصنوعی و تحلیل داده صورت می‌گیرد. این رویکرد به سازمان‌ها کمک می‌کند تصمیماتی دقیق‌تر، هوشمندانه‌تر و کارآمدتر اتخاذ کنند (کرمی و همکاران، ۱۴۰۳؛ شیرازی و همکاران، ۱۴۰۰). در کنار این مزایا، چالش‌هایی نیز وجود دارد؛ از جمله نگرانی‌های مربوط به جایگزینی مشاغل، مقاومت و واکنش‌های منفی کارکنان و کمبود زیرساخت‌های لازم که مدیریت دقیق این نگرانی‌ها، تعارضات و مسائل اخلاقی، از جمله الزامات حیاتی برای موفقیت در این حوزه است (احمد شاهی و همکاران، ۱۴۰۴).

هوش مصنوعی اخلاقی^۲

در کنار مفهوم «هوش مصنوعی قابل اعتماد»، رویکرد «هوش مصنوعی اخلاقی» نیز به‌عنوان چارچوبی مهم برای هدایت توسعه و کاربرد این فناوری مطرح شده است. هدف اصلی دستورالعمل‌های هوش مصنوعی اخلاقی، تضمین احترام به حقوق انسانی، حمایت از عدالت در تصمیم‌گیری‌های الگوریتمی، حداکثرسازی منافع اجتماعی هوش مصنوعی و پیشگیری از آسیب‌های احتمالی ناشی از آن است (گرگوری^۳، ۲۰۱۴). همانند دستورالعمل‌های مربوط به هوش مصنوعی قابل اعتماد، در سال‌های اخیر مجموعه‌ای از اصول اخلاقی برای هدایت طراحی و استفاده از سامانه‌های هوش مصنوعی ارائه شده‌اند.

1. Du & Xie
2. ETHICAL AI
3. Greguric

این دستورالعمل‌ها بر اصول مشترکی همچون شفافیت، عدالت، مسئولیت‌پذیری و احترام به کرامت انسانی تأکید دارند و برخی از آنها به‌طور خاص بر کاربردهای سازمانی و محیط کار تمرکز کرده‌اند (هاگندورف^۱، ۲۰۲۰). در فضای سازمانی، سامانه‌های هوش مصنوعی اخلاقی می‌توانند با رعایت اصول زیر، به حمایت از حقوق و ارزش‌های انسانی کارکنان کمک کنند.

۱. خودمختاری: حمایت از حق کارکنان در تصمیم‌گیری مستقل و جلوگیری از سلطه یا کنترل کامل سامانه‌های هوش مصنوعی بر انتخاب‌های فردی آنها؛ ۲. حریم خصوصی: تضمین حق کارکنان برای کنترل داده‌های شخصی خود و آگاهی کامل از نحوه جمع‌آوری، استفاده و پردازش این داده‌ها توسط کارفرما؛ ۳. پرهیز از سوگیری داده‌ها: پیشگیری از تصمیم‌گیری‌های تبعیض‌آمیز و ناعادلانه که ممکن است به آسیب‌های مستقیم یا غیرمستقیم به کارکنان بینجامد؛ ۴. شفافیت و پاسخگویی: فراهم‌سازی امکان درک عملکرد سامانه‌های هوش مصنوعی برای کارکنان و تعریف مسئولیت‌های روشن در قبال نتایج منفی احتمالی حتی در مواردی که تصمیم‌گیری‌ها بسیار پیچیده و مبتنی بر محاسبات ریاضی پیشرفته باشند. این اصول اخلاقی، زیرساختی ضروری برای بهره‌برداری مسئولانه از هوش مصنوعی در سازمان‌ها به‌شمار می‌آیند و رعایت آنها می‌تواند موجب ارتقای اعتماد کارکنان، کاهش ریسک‌های اخلاقی و تقویت سرمایه انسانی شود (هاوارد و شولتی، ۲۰۲۴).

پیشینه تجربی پژوهش

در مطالعات علم و فناوری و ارزیابی فناوری از دیرباز، هوش مصنوعی را موضوعی محوری تلقی کرده (کالینز^۲، ۲۰۱۸). به‌ویژه، ملاحظات اخلاقی پیرامون گسترش وسیع هوش مصنوعی در بخش‌های اجتماعی مرتبط و تأثیرات مختلف آن بر زندگی انسان نمود یافته است. به میزان قابل توجهی، کارهای هنجاری که در دهه‌های گذشته انجام شده است، زمینه اخلاق هوش مصنوعی را شکل داده است. در ادامه به برخی از تحقیقات صورت‌گرفته در خصوص موضوعات اخلاقی هوش مصنوعی پرداخته شده است. مورلی و همکارانش (۲۰۲۱)، در پژوهشی با هدف عملیاتی‌سازی عملی اخلاق هوش مصنوعی، ابزارها و روش‌هایی برای راهنمایی متخصصان هوش مصنوعی در مورد مسائل اصلی اخلاق هوش مصنوعی را شناسایی کرده و این روش‌ها و ابزارها را در یک طبقه‌بندی ترسیم کردند و آنها را با اصول

1. Hagendorff

2. Collins

اخلاقی (مصلحت، عدم آسیب، خودمختاری، عدالت و قابلیت توضیح) و مراحل خط تولید الگوریتم‌ها تطبیق دادند. آنها گزارش دادند که ابزارها و روش‌های متعددی برای کمک به متخصصان در تحقق هوش مصنوعی اخلاقی وجود دارد، اما اکثریت قریب به اتفاق آنها از نظر قابلیت استفاده بسیار محدود هستند. راهنمایی ضعیف در مورد چگونگی تولید هوش مصنوعی اخلاقی ممکن است به هزینه‌های اجتماعی قابل توجهی منجر شود (مورلی، ۲۰۲۱). در پژوهشی دیگر با موضوع چشم‌انداز جهانی دستورالعمل‌های اخلاق هوش مصنوعی جوین (۲۰۱۹) بیان می‌کند که در پنج سال گذشته، شرکت‌های خصوصی، مؤسسات تحقیقاتی و سازمان‌های بخش دولتی، اصول و دستورالعمل‌هایی را برای هوش مصنوعی اخلاقی منتشر کرده‌اند. با این حال، با وجود توافق ظاهری بر اینکه هوش مصنوعی باید «اخلاقی» باشد، بحث‌هایی در مورد اینکه چه چیزی «هوش مصنوعی اخلاقی» را تشکیل می‌دهد و چه الزامات اخلاقی، استانداردهای فنی و بهترین شیوه‌ها برای تحقق آن مورد نیاز است، وجود دارد. نتایج مطالعه نشان‌دهنده یک همگرایی جهانی در اطراف پنج اصل اخلاقی (شفافیت، عدالت و انصاف، عدم سوءاستفاده، مسئولیت‌پذیری و حریم خصوصی) است. با این حال در رابطه با نحوه تفسیر این اصول، دلیل اهمیت آنها، موضوع، حوزه یا بازیگرانی که به آنها مربوط می‌شوند و چگونگی اجرای آنها، اختلاف نظرهای اساسی وجود دارد.

بوستروم و یودکوفسکی^۱ (۲۰۱۴) چندین اصل اخلاقی کلی را شناسایی می‌کنند که فناوری‌های هوش مصنوعی باید از آنها پیروی کنند، مانند شفافیت، قابلیت اطمینان و انصاف. در مطالعه‌ای که آنگراشمید و همکارانش^۲ (۲۰۲۲) با هدف بررسی تأثیرات انصاف و توضیح بر اعتماد در هوش مصنوعی اخلاقی انجام دادند، نشان دادند که انصاف تنها زمانی بر اعتماد کاربر تأثیر می‌گذارد که سطح انصاف پایین باشد و سطح انصاف پایین، اعتماد کاربر را کاهش می‌دهد. با این حال، افزودن توضیحات به کاربران کمک کرد تا اعتماد خود را به تصمیم‌گیری مبتنی بر هوش مصنوعی افزایش دهند. نتایج نشان می‌دهد که استفاده از توضیحات هوش مصنوعی و اظهارات انصاف در برنامه‌های کاربردی هوش مصنوعی پیچیده است: ما باید نه تنها نوع توضیحات، بلکه سطح انصاف ارائه‌شده را نیز در نظر بگیریم. وه^۳ (۲۰۲۴)، در مطالعه‌ای با هدف بررسی مسائل اخلاقی ناشی از تقاطع هوش مصنوعی و فناوری‌های

1. Bostrom & Yudkowsky
2. Angerschmid et al.
3. Weh

عصبی، به‌ویژه در زمینه کاری، می‌پردازد و بر اهمیت درک روابط پیچیده بین این فناوری‌ها، انسان‌ها و جامعه تأکید می‌کند. در این مطالعه چندین اصل اخلاقی را شناسایی می‌کند که فناوری‌های هوش مصنوعی باید از آنها پیروی کنند، مانند حریم خصوصی، عدالت، انصاف و برابری و کلیشه‌های نژادی و جنسیتی، آزادی و خودمختاری اخلاقی. سانتونی د سِو^۱ (۲۰۲۴)، در مطالعه‌ای با موضوع هوش مصنوعی و آینده کار: ترسیم مسائل اخلاقی به بررسی هفت مسئله اخلاقی می‌پردازد که با معرفی هوش مصنوعی در محیط کار مطرح می‌شوند. هفت مسئله عبارتند از: ۱. چگونه باید بر تأثیر هوش مصنوعی بر ازدست‌رفتن مشاغل و سایر مسائل اجتماعی ناشی از تغییر شکل بازار کار نظارت کرد؟ ۲. هوش مصنوعی ممکن است به ایجاد اشکال جدیدی از ستم و نقض حقوق نیروی کار کمک کند؛ ۳. هوش مصنوعی ممکن است بر عاملیت، استقلال یا مسئولیت‌پذیری (اخلاقی) کارکنان تأثیر منفی بگذارد؛ ۴. هوش مصنوعی ممکن است باعث ایجاد "کار پنهان" شود، به‌این‌معناکه وظایف با ارزش اقتصادی توسط عوامل انسانی انجام می‌شود بدون اینکه کار آنها به اندازه کافی شناخته، پاداش داده شود یا محافظت شود. درحالی‌که شرکت‌های (فناوری) سود نا عادلانه‌ای به‌دست می‌آورند و قدرت اقتصادی - اجتماعی خود را بر مردم افزایش می‌دهند؛ ۵. هوش مصنوعی تاجه‌حد می‌تواند بر فرصت افراد برای انجام کار خوب یا معنادار تأثیر بگذارد و چگونه باید کار معنادار را در یک جامعه کثرت‌گرا تعریف کرد؟ ۶. معرفی هوش مصنوعی در محل کار ممکن است تأثیر گسترده‌تری بر ارزش‌ها و هنجارهای اجتماعی داشته باشد؛ ۷. چه کسی مسئول این است که هوش مصنوعی تأثیری مثبت و نه منفی بر ارزش‌های اخلاقی و اجتماعی داشته باشد؟

درخصوص گنجاندن ضوابط رفتاری در صنعت خودروسازی پژوهش‌های متعددی صورت گرفته است. به‌عنوان نمونه بونفون و همکاران (۲۰۱۶) مطالعه‌ای را با هدف بررسی ضوابط رفتاری در خودروهای خودران انجام دادند، نتایج پژوهش نشان داد که وسایل نقلیه خودران باید تصادفات رانندگی را کاهش دهند، اما گاهی اوقات مجبور خواهند بود بین دو گزینه بد، مانند زیرگرفتن عابر پیاده یا فداکردن خود و سرنشینانشان برای نجات عابرین پیاده، یکی را انتخاب کنند. تعریف الگوریتم‌هایی که به وسایل نقلیه خودران در اتخاذ این تصمیمات اخلاقی کمک می‌کنند، یک چالش بزرگ است. شرکت‌کنندگان در این مطالعه، با اعمال مقررات سودمند برای وسایل نقلیه خودران

مخالفتد و تمایل کمتری به خرید چنین وسایل نقلیه خودرانی خواهند داشت. براین اساس، وضع مقررات برای الگوریتم‌های سودمند ممکن است به‌طور متناقضی تلفات را با به تأخیر انداختن پذیرش یک فناوری ایمن‌تر، افزایش دهد. در پژوهشی دیگر با موضوع گنجاندن ملاحظات اخلاقی در کنترل خودکار وسایل نقلیه تورنتون (۲۰۱۷) نیز به بررسی چگونگی پیاده‌سازی اصول اخلاقی در فرایندهای تصمیم‌گیری ماشین‌های خودکار می‌پردازد.

پس به‌طور کلی پژوهش‌های انجام‌شده در اخلاق هوش مصنوعی طیف وسیعی از موارد را شامل می‌شود: ۱. بازتاب‌ها و کارهای عملی درمورد چگونگی پیاده‌سازی اصول اخلاقی در فرایندهای تصمیم‌گیری ماشین‌های خودکار (اخلاق ماشین) (تورنتون، ۲۰۱۷؛ اندرسون و اندرسون، ۲۰۱۷؛ سروانتس و همکاران، ۲۰۲۰)؛ ۲. تحلیل تجربی درمورد چگونگی حل مسائل اخلاقی (اخلاق تجربی هوش مصنوعی) (بونفون و همکاران، ۲۰۱۶)؛ اواد و همکاران، ۲۰۱۸) و ۳. اصول و دستورالعمل‌های جامع هوش مصنوعی (اصول و دستورالعمل‌ها) (جویین، و همکاران، ۲۰۱۹). با وجود غنای کارهای انجام‌شده در اخلاق هوش مصنوعی، حجم فزاینده‌ای از ادبیات بیان می‌کند که این تلاش‌ها در تحقق اهداف هنجاری خود ناکام مانده‌اند (سروانتس و همکاران، ۲۰۲۰).

مطالعات اخلاق تجربی هوش مصنوعی اطلاعات غنی درمورد انتظارات و ترجیحات اجتماعی در موقعیت‌های اخلاقی مرتبط با هوش مصنوعی ارائه داده‌اند. باین حال، برخی از این مطالعات، به دلیل پیش‌فرض‌های کلیشه‌ای و غیرواقعی (مسائل اخلاقی) مورد انتقاد قرار گرفته‌اند. تأثیر دستورالعمل‌های اخلاق هوش مصنوعی که برای ارتقای شیوه‌های اخلاقی در هوش مصنوعی تدوین شده‌اند نیز مورد تردید است. بی‌اعتمادی وجود دارد که سازمان‌ها، به‌ویژه سازمان‌هایی که در توسعه فناوری‌های هوش مصنوعی مشارکت دارند، به‌طور داوطلبانه شیوه‌های اخلاقی را اجرا کنند (ویتکر^۲، ۲۰۱۸). مطالعه‌ای که در آن متخصصان هوش مصنوعی درمورد تأثیر ادراک‌شده دستورالعمل‌های اخلاق هوش مصنوعی مورد بررسی قرار گرفتند، گزارش داد که اثربخشی چنین دستورالعمل‌ها یا کدهای اخلاقی تقریباً صفر است و رفتار متخصصان جامعه فناوری را تغییر نمی‌دهد (مک‌نامارا^۳، ۲۰۱۸).

1. Bonnefon
2. Whittaker
3. McNamara

روش پژوهش

این پژوهش از منظر روش‌شناسی، توصیفی است و با استفاده از مرور هدف‌مند، در دسته پژوهش‌های کیفی قرار می‌گیرد. جمع‌آوری داده‌ها از طریق بررسی مطالعات مرتبط در پایگاه‌های معتبر بین‌المللی انجام می‌شود که در مراحل مرور هدف‌مند به آنها اشاره شده است. واحد مطالعه در این مرور هدف‌مند، خود مطالعات است. بنابراین، جامعه آماری این پژوهش شامل مطالعاتی است که از نظر کیفیت و پاسخگویی به سؤالات پژوهشی، شرایط لازم را احراز کرده‌اند. جزئیات این فرایند، در مراحل پنج‌گانه مرور هدف‌مند به تفصیل شرح داده شده است.

مرور هدف‌مند

مرور هدف‌مند، یک روش تحقیق است که تصویری کلی از شواهد پژوهشی موجود در یک زمینه خاص را به روشی نظام‌مند در اختیار کاربران قرار می‌دهد. این نوع مرور، یک بررسی سریع از مفاهیم کلیدی و منابع اصلی در یک موضوع پژوهشی معین است. مرور دامنه‌ای به‌عنوان یک روش تحقیق مبتنی بر اطلاعات ثانویه، با رویکرد اکتشافی و توصیفی، برای ارزیابی وضعیت شواهد موجود، شناسایی شکاف‌های تحقیقاتی و بررسی حجم، کمیت و کیفیت مطالعات انجام‌شده در یک حوزه خاص به کار می‌رود. در این بخش از یک چارچوب دقیق و شفاف استفاده شده که براساس اصول مطرح‌شده توسط طرفداران مطالعات نظام‌مند بنا شده است. هدف پژوهش حاضر این است که تمام مراحل کار را با جزئیات مستند کند تا سایر پژوهشگران بتوانند مطالعه را تکرار کنند. این رویکرد، به افزایش اعتبار نتایج کمک می‌کند و هرگونه انتقادی را که ممکن است در مورد دقت روش‌شناختی مطرح شود، پاسخ می‌دهد. در شناسایی منابع، ما به دنبال دستیابی به یک دید وسیع و جامع هستیم. برخلاف مطالعات نظام‌مند که معمولاً با یک سؤال تحقیقاتی مشخص و محدود آغاز می‌شوند، در اسکوپینگ ریویو، هدف ما شناسایی تمام منابع مرتبط، صرف نظر از نوع مطالعه است.

با پیشرفت مطالعه و آشنایی بیشتر با منابع، ممکن است نیاز باشد که اصطلاحات جستجو را اصلاح و جستجوهای دقیق‌تری انجام دهیم. به همین دلیل، ابتدا محدودیت‌های سختگیرانه‌ای در انتخاب اصطلاحات، شناسایی منابع یا انتخاب مطالعات اعمال نمی‌کنیم. این فرایند، یک مسیر خطی نیست، بلکه یک فرایند تکراری است. ما به‌طور مداوم به هر مرحله بازمی‌گردیم و در صورت نیاز، مراحل را تکرار

می‌کنیم تا اطمینان حاصل شود که تمام منابع مرتبط پوشش داده شده‌اند. در ادامه، مراحل که برای انجام اسکوپینگ ریویو اتخاذ شده است، شرح داده شده است (آرکسی و اومالی^۱، ۲۰۰۵).

- مرحله اول: شناسایی سؤال پژوهش؛

- مرحله دوم: راهبرد جستجو؛

- مرحله سوم: انتخاب مطالعه؛

- مرحله چهارم: ترسیم داده‌ها؛

- مرحله پنجم: گردآوری، خلاصه‌سازی و گزارش نتایج.

مرحله اول: شناسایی سؤال پژوهش

موضوعات اخلاقی درخصوص به‌کارگیری هوش مصنوعی و فناوری‌های عصبی در محیط کار

چیست؟

مرحله دو: راهبرد جستجو

همان‌طور که اشاره شد، هدف اصلی تعیین محدوده موضوع، جامعیت در شناسایی مطالعات اولیه و

مروری‌های مناسب برای پاسخگویی به سؤال اصلی پژوهش است. دو پایگاه داده کتاب‌شناختی زیر

جستجو شدند (جدول ۱).

جدول ۱. پایگاه‌های اطلاعات کتاب‌شناختی مورد جستجو به‌عنوان بخشی از مرور اسکوپینگ

آدرس	پایگاه داده
https://www.scopus.com	Scopus
http://www.webofknowledge.com	Web of Science

برای اطمینان از جستجوی جامع ادبیات مرتبط، از واژه‌های مرتبط نشان‌داده‌شده در جدول ۲ به‌عنوان

اصطلاحات جستجو استفاده شد.

^۱. Arksey and O'Malley

جدول ۲. رشته‌های جستجوی به کاررفته در پایگاه‌های داده

نتیجه	شیوه جست و جو	پایگاه داده
2385 articles	ethic * or moral* (title) and “artificial intelligence” (topic) and year = until 2025 and document type = (peer-reviewed journal) article.	Web of Science Search String 1
2151 articles	ethic * or moral* (title) and “Neurotechnology” (topic) and year = until 2025 and document type = (peer-reviewed journal) article.	Web of Science Search String 2
901 articles	ethic * or moral* (title) and title-abs-key (“artificial intelligence”) and pub year> 1978	Scopus Search String 1
19 articles	ethic * or moral* (title) and title-abs-key (“Neurotechnology”) and pub year> 1978	Scopus Search String 2

مرحله سوم: انتخاب مطالعه

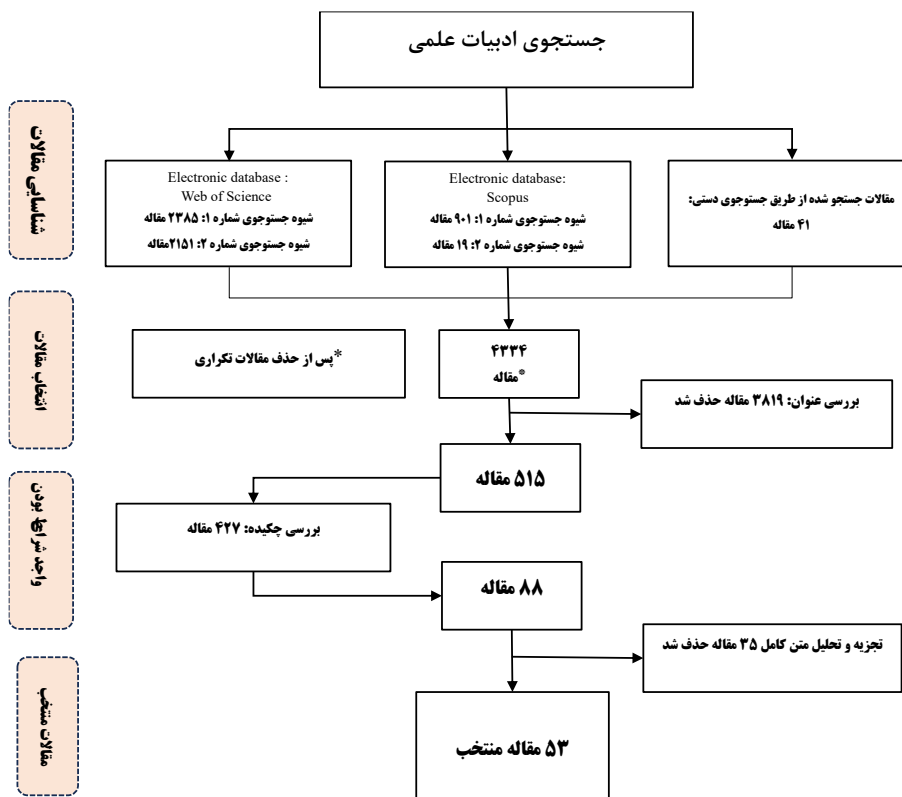
در مرورهای نظام‌مند، معیارهای ورود و خروج معمولاً در ابتدای پروژه و بر پایه پرسش تحقیق تدوین می‌شوند تا تصمیم‌گیری‌ها از انسجام و ثبات برخوردار باشند. مطالعه مروری حاضر نیز از رویکردی مشابه پیروی کرده، با این تفاوت که معیارهای نهایی پس از مرحله غربال‌گری اولیه و با توجه به آشنایی تدریجی با ادبیات موضوع، تنظیم شدند. این معیارها در ادامه بر تمامی منابع بازبایی شده اعمال شدند تا میزان ارتباط آنها با موضوع تحقیق مشخص شود. جدول ۳ معیارهای ورود و خروج این مطالعه را توصیف می‌کند.

جدول ۳. معیارهای انتخاب مرور اسکوپینگ

موضوع	معیارهای ورود
نوع انتشار	مجله معتبر و دانشگاهی - اسناد
زبان	انگلیسی
دوره زمانی	تا سال ۲۰۲۵
ارتباط	مقاله و سند به بررسی و تحلیل دیدگاه‌های اخلاقی در مورد کاربرد هوش مصنوعی و فناوری‌های عصبی پردازد.

نتیجه این مرحله، دستیابی به ۵۴۵۶ مطالعه از طریق جستجو در پایگاه‌های اطلاعاتی و ۴۱ مطالعه از طریق جستجوی دیگر منابع شامل جستجو از طریق رفرنس فهرست مقالات یافت شد و پس از حذف

موارد تکراری در مرحله بعدی، این مقالات (۴۳۳۴) از نظر عنوان بررسی شدند. با شناسایی و حذف مقالاتی که از نظر عنوان مرتبط با تحقیق حاضر نبودند (۳۸۱۹ مقاله)، ۵۱۵ مقاله باقی ماند که از نظر چکیده مورد بررسی قرار گرفتند. در مرحله بعدی (حذف ۴۲۷ مقاله)، ۸۸ مقاله باقیمانده از نظر محتوا ارزیابی شدند و ۳۵ مقاله با توجه به مؤلفه‌های اخلاقی هوش مصنوعی و فناوری‌های عصبی، مورد قبول واقع نشدند. در نهایت، ۵۳ مقاله و سند (شکل ۱) که سطح کیفی خوبی داشتند، باقی ماندند که به‌طور دقیق مورد بررسی قرار گرفتند تا مؤلفه‌های اخلاقی درمورد کاربرد هوش مصنوعی و فناوری‌های عصبی در محیط کار از آنها استخراج شود.



شکل ۱. فرایند اسکوپینگ ریویو در انتخاب مقالات و اسناد

در جدول شماره ۴، فهرست مقالات و اسناد منتخب که سطح کیفی خوبی داشتند، ارائه شده است که به منظور بررسی دقیق تر و استخراج کدهای اخلاقی مورد مطالعه عمیق قرار گرفتند.

جدول ۴. فهرست مقالات و اسناد منتخب

ردیف	نام نویسنده‌گان	سال	عنوان سند
۱	فلوریدی و همکاران ^۱	۲۰۱۸	AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations
۲	کوسکی و هوسو ^۲	۲۰۱۸	Work in the age of artificial intelligence: Four perspectives on the economy, employment, skills and ethics.
۳	ویلاتی و همکاران ^۳	۲۰۱۸	For a meaningful artificial intelligence: Towards a French and European strategy.
۴	موری و همکاران ^۴	۲۰۲۴	A systematic literature review on artificial intelligence in recruiting and selection: A matter of ethics
۵	فنش و همکاران ^۵	۲۰۱۹	Ethical, Social, and Political Challenges of Artificial Intelligence in Health
۶	وارما و همکاران ^۶	۲۰۲۳	artificial intelligence and people management: a critical assessment through the ethical lens
۷	بولاموینی و گبرو ^۷	۲۰۱۸	Gender shades: Intersectional accuracy disparities in commercial gender classification
۸	کاگیل و همکاران ^۸	۲۰۲۰	Biased programmers? Or biased data? A field experiment in operationalizing AI ethics
۹	جویین و همکاران	۲۰۱۹	The global landscape of AI ethics guidelines
۱۰	آویلا و همکاران ^۹	۲۰۱۸	Artificial Intelligence: Open Questions about Gender Inclusion
۱۱	براندج و همکاران ^{۱۰}	۲۰۱۸	The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation
۱۲	فن است و گریتنس ^{۱۱}	۲۰۱۷	Human Rights in the Robot Age: Challenges Arising from the Use of Robotics
۱۳	هافمن و ماسوچی ^{۱۲}	۲۰۱۸	Intel's AI Privacy Policy White Paper: Protecting Individuals' Privacy and Data In The Artificial Intelligence World
۱۴	دمیو ^{۱۳}	۲۰۱۷	How Can Humans Keep the Upper Hand? Report on the Ethical Matters Raised by AI Algorithms

1. Floridi et al.
2. Koski & Husso
3. Villani
4. Mori
5. Fenech
6. Varma
7. Buolamwini & Gebru
8. Cowgill et al.
9. Avila et al.
10. Brundage
11. van Est & Gerritsen
12. Hoffmann, & Masucci
13. Demiaux

ادامه جدول ۴. فهرست مقالات و اسناد منتخب

ردیف	نام نویسنده‌گان	سال	عنوان سند
۱۵	توریلی و فلوریدی ^۱	۲۰۰۹	The ethics of information transparency
۱۶	تورسوناپووا و همکاران ^۲	۲۰۲۲	The ethics of people analytics: risks, opportunities and recommendations”
۱۷	دمیر و گونایدین ^۳	۲۰۲۳	A digital job application reference: how do social media posts affect the recruitment process
۱۸	داموتا ویگا و همکاران ^۴	۲۰۲۳	Seeming ethical makes you attractive: unraveling how ethical perceptions of AI in hiring impacts organizational innovativeness and attractiveness
۱۹	سانتونی د سئو	۲۰۲۴	Artificial Intelligence and the Future of Work: Mapping the Ethical Issues
۲۰	سانتونی د سئو	۲۰۲۰	I’m not a robot: Amazon workers condemn unsafe, grueling conditions at warehouse
۲۱	سانتونی د سئو	۲۰۲۱	The future of work: Freedom, justice and capital in the age of artificial intelligence
۲۲	وردنبرگ ^۵	۲۰۲۲	Freedom at work: Understanding, alienation, and the AI-driven workplace
۲۳	سانتونی د سئو و ون دن هوون ^۶	۲۰۱۸	Meaningful Human Control over autonomous systems: A philosophical account
۲۴	داناها، جان و سون نیهولم ^۷	۲۰۲۱	Automation, work and the achievement gap
۲۵	کریستینس ^۸	۲۰۲۳	Digital working lives: Worker autonomy and the gig economy. Off the fence: Morality, Politics and Society.
۲۶	وروسالیس ^۹	۲۰۱۹	How exploiters dominate
۲۷	کرافورد ^{۱۰}	۲۰۲۱	Atlas of AI: Power, politics, and the planetary costs of artificial intelligence
۲۸	پریگو ^{۱۱}	۲۰۲۳	OpenAI used Kenyan workers on less than \$2 per hour.
۲۹	اسمیلز و همکاران ^{۱۲}	۲۰۱۹	Robots in the workplace: a threat to—or opportunity for—meaningful work?
۳۰	دوسون و همکاران ^{۱۳}	۲۰۱۹	Artificial intelligence: Australia’s ethics framework-a discussion paper

1. Turilli & Floridi
2. Tursunbayeva et al.
3. Demir and Gunaydin
4. da Motta Veiga et al.
5. Vredenburg
6. Santoni de Sio and Van den Hoven
7. Danaher, John, and Sven Nyholm.
8. Christiaens
9. Vrousalis
10. Crawford
11. Perrigo
12. Smids
13. Dawson

ادامه جدول ۴. فهرست مقالات و اسناد منتخب

ردیف	نام نویسنده‌گان	سال	عنوان سند
۳۱	وه	۲۰۲۴	An Integrated Embodiment Concept Combines Neuroethics and AI Ethics – Relational Perspectives on Artificial Intelligence, Emerging Neurotechnologies and the Future of Work
۳۲	مارتینیو و همکاران ^۱	۲۰۲۱	Ethical issues in focus by the autonomous vehicles industry
۳۳	مارتینیو	۲۰۲۱	A healthy debate: Exploring the views of medical doctors on the ethics of artificial intelligence,
۳۴	لودرز و پترسون ^۲	۲۰۲۰	Scrutinising COVID Safe: Frameworks for evaluating digital contact tracing technologies
۳۵	ویدجن و همکاران ^۳	۲۰۲۰	Exploring the ethical implications of business analytics with a business ethics canvas
۳۶	مارتین ^۴	۲۰۱۹	Designing Ethical Algorithms
۳۷	نوبیرت و مونتانیز ^۵	۲۰۲۰	Virtue as a framework for the design and use of artificial intelligence
۳۸	ساتون و همکاران ^۶	۲۰۱۸	How Much Automation Is Too Much? Keeping the Human Relevant in Knowledge Work
۳۹	آشوک و همکاران ^۷	۲۰۲۲	Ethical framework for artificial intelligence and digital technologies
۴۰	ریسمن و همکاران ^۸	۲۰۱۸	Igorithmic Impact Assessments: A Practical Framework for Public Agency Accountability.
۴۱	سیبکر ^۹	۲۰۱۹	Making Corporations More Humane Through Artificial Intelligence.
۴۲	میلاتو ^{۱۰}	۲۰۲۰	Recommender systems and their ethical challenges
۴۳	بور و همکاران ^{۱۱}	۲۰۱۸	An Analysis of the Interaction Between Intelligent Software Agents and Human User
۴۴	دِورس ^{۱۲}	۲۰۱۰	Identity, profiling algorithms and a world of ambient intelligence
۴۵	کوئه و همکاران ^{۱۳}	۲۰۱۵	Ethics of Personalized Information Filtering
۴۶	تادئو و فلوریدی ^{۱۴}	۲۰۱۸	How AI can be a Force for Good

1. Martinho
2. Lodders & Paterson
3. Vidgen
4. Martin
5. Neubert & Montañez
6. Sutton
7. Ashok
8. Reisman
9. Siebecker
10. Milano
11. Burr
12. de Vries
13. Koene
14. Taddeo & Floridi

ادامه جدول ۴. فهرست مقالات و اسناد منتخب

ردیف	نام نویسنده‌گان	سال	عنوان سند
۴۷	بوزداغ ^۱	۲۰۱۳	Bias in algorithmic filtering and personalization
۴۸	بوزداغ و ون دن هوون ^۲	۲۰۱۵	Breaking the filter bubble: democracy and design
۴۹	دو و شی	۲۰۲۱	Paradoxes of artificial intelligence in consumer markets
۵۰	بارلو ^۳	۲۰۲۰	Oncology Research: Clinical Trial Management Systems
۵۱	برایدباخ و ماگلیو ^۴	۲۰۲۰	Accountable algorithms? The ethical implications of data-driven business models
۵۲	دیویدی و همکاران ^۵	۲۰۲۱	Setting the future of digital and social media marketing research: Perspectives and research propositions
۵۳	رابرتز و همکاران ^۶	۲۰۲۱	The Chinese approach to artificial intelligence: an analysis of policy

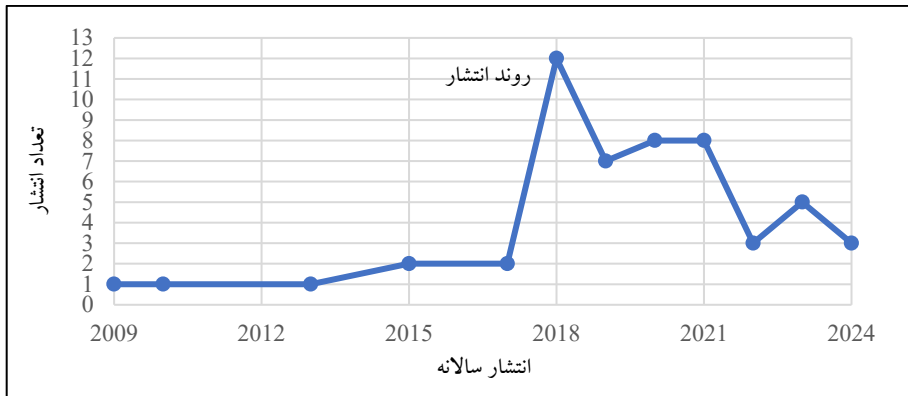
مرحله چهارم: ترسیم داده‌ها

مرحله بعدی کار شامل «ترسیم» موارد کلیدی اطلاعات به دست آمده از گزارش‌های تحقیقات اولیه مورد بررسی بود. «نمودار کردن و ترسیم داده‌ها» فنی برای تجزیه و تحلیل داده‌های کیفی با غربال‌گری، نمودار کردن و مرتب‌سازی مواد با توجه به مسائل و موضوعات کلیدی را توصیف می‌کند، اطلاعات به شرح زیر ارائه می‌گردد:

روند انتشارات؛ سال

در دهه گذشته، تعداد انتشارات سالانه در مورد دیدگاه‌های اخلاقی در مورد کاربرد هوش مصنوعی و فناوری‌های عصبی، تغییرات نسبتاً زیادی داشته است. همان‌طور که در شکل ۱ نشان داده شده است، سال ۲۰۱۸ با انتشار ۱۲ سند بیشترین انتشارات در خصوص موضوعات اخلاقی در به کارگیری هوش مصنوعی و فناوری‌های عصبی در محل کار مطرح شده است پس از آن با شیب نسبتاً کمی این روند کاهش پیدا کرده است؛ به طوری که در سال ۲۰۲۴ تنها ۳ مقاله به این موضوع پرداخته است.

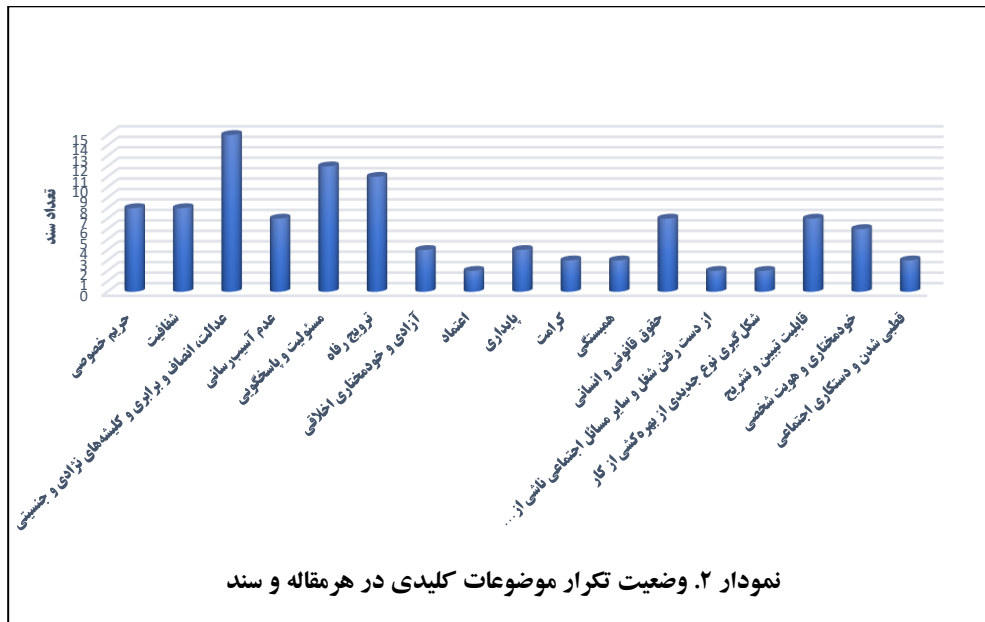
1. Bozdag
2. Bozdag & van den Hoven
3. Barlow
4. Breidbach & Maglio
5. Dwivedi et al.,
6. Roberts



نمودار ۱. روند انتشار مقالات و اسناد

تکرار موضوعات کلیدی در هر سند

با توجه به مرور ادبیات صورت گرفته، مشخص شد که عدالت، انصاف و برابری و کلیشه‌های نژادی و جنسیتی، مسئولیت و پاسخگویی و ترویج رفاه دارای بیشترین ضریب اهمیت می‌باشند؛ یعنی در حوزه اخلاق هوش مصنوعی این موضوعات بیشتر مورد توجه و تأکید قرار گرفته‌اند (نمودار شماره ۲).



نمودار ۲. وضعیت تکرار موضوعات کلیدی در هر مقاله و سند

مرحله پنجم: گردآوری، خلاصه‌سازی و گزارش نتایج

این مرحله از یک مطالعه مروری شامل گردآوری، خلاصه‌سازی و گزارش نتایج است. مقالات براساس منابع هر مقاله از جمله نام نویسنده و سال انتشار طبقه‌بندی شدند. نتایج این فرایند در جدول ۵ نشان داده شده است.

جدول ۵. کدهای استخراجی

ردیف	کدهای اخلاقی	منابع
۱	حریم خصوصی	جابین و همکاران، ۲۰۱۹؛ هافمن و ماسوچی، ۲۰۱۸؛ وه، ۲۰۲۴؛ مارتی‌نهو و همکاران، ۲۰۲۱؛ داوزن و همکاران، ۲۰۱۹؛ میلانو و همکاران، ۲۰۲۰؛ دیویدی و همکاران، ۲۰۲۱؛ رابرتز و همکاران، ۲۰۲۱
۲	شفافیت	فلورییدی، ۲۰۱۸؛ کوسکی، او. و هوسو، ۲۰۱۸؛ جوبین، و همکاران، ۲۰۱۹؛ ویلاتی و همکاران، ۲۰۱۸؛ دمیو، ۲۰۱۷؛ توریلی و فلورییدی، ۲۰۰۹؛ ریسمن و همکاران، ۲۰۱۸؛ داوزن و همکاران، ۲۰۱۹
۳	عدالت، انصاف و برابری و کلیشه‌های نژادی و جنسیتی	لودرز و پترسون، ۲۰۲۰؛ ویدگن و همکاران، ۲۰۲۰؛ فنج و همکاران، ۲۰۱۹؛ موری و همکاران، ۲۰۲۴؛ ورما و همکاران، ۲۰۲۳؛ مارتی‌نهو و همکاران، ۲۰۲۱؛ بولامینی و گنتگو، ۲۰۱۸؛ کوگیل و همکاران، ۲۰۲۰؛ وه، ۲۰۲۴؛ گلدمن و کروپانزانو، ۲۰۱۵؛ فلورییدی، ۲۰۱۸؛ میلانو و همکاران، ۲۰۲۰؛ برایدباخ و مگلیو، ۲۰۲۰؛ دیویدی و همکاران، ۲۰۲۱؛ رابرتز و همکاران، ۲۰۲۱
۴	عدم آسیب‌رسانی	آویلا و همکاران، ۲۰۱۸؛ جوبین و همکاران، ۲۰۱۹؛ برونجد و همکاران، ۲۰۱۸؛ ون است و گرایتسن و همکاران، ۲۰۱۷؛ داوزن و همکاران، ۲۰۱۹؛ دو و شی، ۲۰۲۱؛ دیویدی و همکاران، ۲۰۲۱
۵	مسئولیت و پاسخگویی	جابین و همکاران، ۲۰۱۹؛ مارتی‌نو و همکاران، ۲۰۲۱؛ سانتونی دی سیو و ون دن هوون، ۲۰۱۸؛ سانتونی دی سیو، فیلیپو، و جولینو میکچی، ۲۰۲۱؛ آشوک و همکاران، ۲۰۲۲؛ داناها، جان، و اسون‌نای‌هولم، ۲۰۲۱؛ داوزن و همکاران، ۲۰۱۹؛ دو و شی، ۲۰۲۱؛ بارلو، ۲۰۲۰؛ برایدباخ و مگلیو، ۲۰۲۰؛ دیویدی و همکاران، ۲۰۲۱؛ رابرتز و همکاران، ۲۰۲۱
۶	ترویج رفاه	نویوثر و مونتائز، ۲۰۲۰؛ ویدگن و همکاران، ۲۰۲۰؛ ساتون و همکاران، ۲۰۱۸؛ آشوک و همکاران، ۲۰۲۲؛ جوبین و همکاران، ۲۰۱۹؛ کوسکی و هوسو، ۲۰۱۸؛ سانتونی دی سیو، ۲۰۲۴؛ اسمیدز و همکاران، ۲۰۱۹؛ وان است و گریتسن و همکاران، ۲۰۱۷؛ وریدنبرگ، ۲۰۲۲؛ داوزن و همکاران، ۲۰۱۹
۷	آزادی و خودمختاری اخلاقی	جابین و همکاران، ۲۰۱۹؛ هوفمن و ماسوتچی، ۲۰۱۸؛ وه، ۲۰۲۴؛ مارتی‌نو و همکاران، ۲۰۲۱
۸	اعتماد	جابین و همکاران، ۲۰۱۹؛ دا موتا و گگا و همکاران، ۲۰۲۳

ادامه جدول ۵. کدهای استخراجی

منابع	کدهای اخلاقی	ردیف
جوین، و همکاران، ۲۰۱۹؛ داوسن و همکاران، ۲۰۱۹؛ دویدی، و همکاران، ۲۰۲۱؛ رابرتز و همکاران، ۲۰۲۱	پایداری	۹
جایین و همکاران، ۲۰۱۹؛ دو و شی، ۲۰۲۱؛ دویودی و همکاران، ۲۰۲۱	کرامت	۱۰
اشوک، ۲۰۲۲؛ جوین و همکاران، ۲۰۱۹؛ سبیکر، ۲۰۱۹	همبستگی	۱۱
تورسونبایوا، ۲۰۲۲؛ موری و همکاران، ۲۰۲۴؛ دا موتا ویگا و همکاران، ۲۰۲۳؛ دمیر و گونایدین، ۲۰۲۳؛ سایناتو، ۲۰۲۴؛ سایناتو، ۲۰۲۰؛ داوسن و همکاران، ۲۰۱۹	حقوق قانونی و انسانی	۱۲
سانتونی د سیو، ۲۰۲۴؛ سانتونی د سیو، ۲۰۲۱	ازدست رفتن شغل و سایر مسائل اجتماعی ناشی از بازسازی بازار کار	۱۳
وروساليس، ۲۰۱۹؛ سانتونی ده سیو، ۲۰۲۴	شکل گیری نوع جدیدی از بهره کشی از کار	۱۴
بر اساس مارتینو و همکاران، ۲۰۲۱؛ بارلو، ۲۰۲۰؛ بریدباخ و ماگلیو، ۲۰۲۰؛ داوسن و همکاران، ۲۰۱۹؛ دو و شی، ۲۰۲۰؛ دویدی و همکاران، ۲۰۲۱؛ رابرتز و همکاران، ۲۰۲۱	قابلیت تبیین و تشریح	۱۵
بور و همکاران، ۲۰۱۸؛ د وریس، ۲۰۱۰؛ کوئن و همکاران؛ تدنو و فلوریدی، ۲۰۱۸؛ میلانو و همکاران، ۲۰۲۰؛ دو و شی، ۲۰۲۱؛ بریدباخ و ماگلیو، ۲۰۲۰	خودمختاری و هویت شخصی	۱۶
پوزداغ، ۲۰۱۳؛ میلانو و همکاران، ۲۰۲۰؛ پوزداغ و ون دن هون، ۲۰۱۵	قطبی شدن و دستکاری اجتماعی	۱۷

هفده (۱۷) موضوع اخلاقی از تحلیل محتوای به دست آمده است. این موضوعات، براساس فراوانی تعداد منابعی که در آنها ذکر شده اند، عبارتند از: عدالت، انصاف و برابری و کلیشه های نژادی و جنسیتی، مسئولیت و پاسخگویی، ترویج رفاه، شفافیت، حریم خصوصی، عدم آسیب رساندن، حقوق قانونی و انسانی، قابلیت تبیین و تشریح، خودمختاری و هویت شخصی، پایداری، آزادی و خودمختاری، کرامت، قطبی شدن و دستکاری اجتماعی، همبستگی، اعتماد، شکل گیری نوع جدیدی از بهره کشی از کار، ازدست رفتن شغل و سایر مسائل اجتماعی ناشی از تغییرات بازار کار. قابل ذکر است که هیچ یک از اصول اخلاقی به طور کلی در تمام اسناد و مطالعات مشترک نیستند، هرچند که همگرایی قابل توجهی در حول اصول زیر در حال ظهور است:

عدالت، انصاف و برابری؛ عدالت به عنوان یکی از اصول بنیادین اخلاقی، مستلزم رفتار برابر با همه انسان هاست و فناوری های هوش مصنوعی نباید به تبعیض علیه افراد یا گروه ها منجر شوند (لادرز و

پترسون، ۲۰۲۰؛ ویدجن و همکاران، ۲۰۲۰). در طراحی و توسعه الگوریتم‌ها، تضمین بی‌طرفی و پیشگیری از سوگیری‌ها ضروری است. برخی پژوهشگران مانند فلوریدی (۲۰۱۸) تأکید دارند که هوش مصنوعی می‌تواند ابزار توانمندسازی برای تحقق جامعه‌ای عادلانه‌تر باشد، مشروط بر آنکه در حذف کلیشه‌های نژادی و جنسیتی به درستی به کار رود.

عدالت در زمینه هوش مصنوعی اغلب با مفهوم "انصاف" هم‌معنا تلقی می‌شود و به کاهش سوگیری‌های ناآگاهانه در فرایندهای تصمیم‌گیری اشاره دارد. برخی رویکردها بر دسترسی برابر به فناوری و داده‌ها و برخی دیگر بر شمول‌گرایی و رعایت تنوع فرهنگی و اجتماعی تأکید دارند. در این میان، دولت‌ها نیز به پیامدهای اجتماعی مانند تأثیر بر اشتغال و ضرورت پاسخگویی به دغدغه‌های دموکراتیک توجه دارند.

در حوزه استخدام و منابع انسانی، الگوریتم‌های یادگیری ماشینی تصمیم‌هایی را بر پایه داده‌های تاریخی اتخاذ می‌کنند که ممکن است به تکرار تعصبات گذشته و در نتیجه، ناعادلانه بودن انتخاب‌ها منجر شوند (موری و همکاران، ۲۰۲۴). مطالعه‌ای در این زمینه (بولاموینی و گبرو، ۲۰۱۸) نشان داد که نرخ خطای سامانه‌های شناسایی چهره برای زنان با پوست تیره به‌طور چشمگیری بیشتر از مردان سفیدپوست است. این یافته‌ها بیانگر آن است که تعصب می‌تواند ریشه در داده‌ها، طراحی الگوریتم یا حتی سوگیری‌های ناخواسته توسعه‌دهندگان داشته باشد (کاوگیل و همکاران، ۲۰۲۰؛ وه، ۲۰۲۴). نمونه بارز این چالش، کنار گذاشتن الگوریتم استخدامی شرکت آمازون بود که به دلیل سوگیری جنسیتی، نرخ استخدام زنان را کاهش داده بود (موری و همکاران، ۲۰۲۴).

از منظر نظری، گلدمن و کروپانزانو (۲۰۱۵) تمایز میان عدالت (به‌عنوان رویدادی عینی) و انصاف (به‌عنوان ادراک ذهنی از آن) را مطرح کرده‌اند. این تفکیک می‌تواند راهگشای پژوهش‌های آینده در زمینه طراحی سیاست‌های منابع انسانی باشد. همچنین، مفاهیمی مانند عدالت رویه‌ای، عدالت بین‌فردی و عدالت اطلاعاتی به ابعاد مختلف تجربه کاربران از سامانه‌های تصمیم‌یار هوش مصنوعی اشاره دارند که بر ادراک آنان از انصاف تأثیر می‌گذارد.

در حوزه سلامت نیز نگرانی‌ها درباره پیامدهای تبعیض‌آمیز کاربرد هوش مصنوعی رو به افزایش است. سه نوع جانبداری در این زمینه شناسایی شده است: جانبداری فیزیکی (طراحی نابرابر سخت‌افزار)، جانبداری محاسباتی (داده‌های ناهماهنگ با جمعیت هدف) و جانبداری تفسیری (برداشت نادرست از

خروجی سامانه). این چالش‌ها می‌توانند به افزایش نابرابری در مراقبت‌های بهداشتی منجر شوند (مارتینیو و همکاران، ۲۰۲۱). در پاسخ به این چالش‌ها، مجموعه‌ای از راهکارهای اخلاقی و حاکمیتی پیشنهاد شده است (فنه و همکاران، ۲۰۱۹) که شامل موارد زیر است: ۱. توسعه استانداردهای فنی و تدوین کدهای اخلاقی، ۲. افزایش شفافیت و اطلاع‌رسانی عمومی، ۳. آزمون و ارزیابی مستقل الگوریتم‌ها توسط نهادهای ناظر، ۴. تقویت نظام حقوقی برای جبران و اصلاح تصمیم‌های ناعادلانه، ۵. سیاست‌گذاری‌های جامع با مشارکت نهادهای میان‌رشته‌ای و جامعه مدنی.

در مجموع، عدالت در هوش مصنوعی نیازمند ترکیب رویکردهای فنی، اخلاقی و سازمانی است تا از بازتولید نابرابری‌ها جلوگیری کرده و زمینه‌ساز اعتماد عمومی شود.

مسئولیت و پاسخگویی؛ با وجود تأکید گسترده بر مفهوم «هوش مصنوعی مسئول»، در عمل، ابعاد مشخص مسئولیت و پاسخگویی در تعامل با این فناوری اغلب به روشنی تعریف نشده‌اند. در بسیاری از منابع، بر اهمیت صداقت، شفافیت و تعیین دقیق حدود مسئولیت قانونی تأکید شده است. برخی پژوهشگران، افشای مسئولیت در موارد بروز آسیب را ضروری می‌دانند؛ هدف از این رویکرد، نه تنها افزایش تنوع و ارتقای عدالت، بلکه همچنین ادغام مباحث اخلاقی در آموزش علوم، فناوری، مهندسی و ریاضیات است (جاین و همکاران، ۲۰۱۹). از سوی دیگر، پژوهش‌هایی نیز به کاهش «کنترل معنادار انسانی» به‌عنوان یکی از چالش‌های اصلی در تعامل با سامانه‌های پیشرفته هوش مصنوعی اشاره دارند (مارتینیو و همکاران، ۲۰۲۱؛ سانتونی ده سیو و ون دن هوون، ۲۰۱۸). در بسیاری از موارد، کاربران یا کارکنانی که با این فناوری‌ها در تعامل هستند، تنها نظارت سطحی بر عملکرد سامانه‌ها دارند؛ این امر می‌تواند توانایی آنان را برای جهت‌دهی یا اصلاح تصمیمات سیستم کاهش دهد. عواملی چون پیچیدگی عملکرد، عدم شفافیت، طراحی نامناسب رابط کاربری یا نبود انگیزه و آموزش کافی، ممکن است مانع از شکل‌گیری تعامل معنادار میان انسان و ماشین شوند. نبود کنترل مؤثر انسانی می‌تواند منجر به شکل‌گیری «شکاف‌های مسئولیت» شود؛ به عبارتی، شرایطی ایجاد می‌شود که در آن هیچ فرد یا نهادی به‌طور مستقیم و مشروع پاسخگوی پیامدهای منفی تصمیمات ماشین‌محور نیست. برای مثال، در سامانه‌های رانندگی نیمه‌خودکار، اگر نقص فنی باعث بروز تصادف شود و راننده به‌دلیل طراحی نامناسب یا عدم آگاهی، نتواند به‌موقع مداخله کند، مشخص نیست چه کسی مسئول پیامدهای آن خواهد بود.

نوع دیگری از شکاف، «شکاف پاسخگویی» است که زمانی رخ می‌دهد که کاربران انسانی - مانند پزشکان یا کارمندان دولت - نتوانند توضیح دهند که یک تصمیم فنی چگونه و بر چه اساسی اتخاذ شده است. در این شرایط، درک و ارزیابی تصمیم از سوی افراد آسیب‌دیده نیز دشوار خواهد بود (سانتونی ده سیو، فیلیپو و جولینو مکاچی، ۲۰۲۱؛ آشوک و همکاران، ۲۰۲۲). در نهایت، مسئله‌ای دیگر که در ادبیات مربوط به مسئولیت‌پذیری مطرح می‌شود، «شکاف دستاورد» است. به این معنا که کارکنان، به دلیل مشابه، ممکن است نتوانند برای نتایج مثبت حاصل از استفاده از سامانه‌های هوش مصنوعی، اعتبار حرفه‌ای یا مالکیت فکری کسب کنند (دانهر، جان، و سوین نایهلم، ۲۰۲۱).

ترویج رفاه^۱؛ ترویج رفاه در ادبیات اخلاقی هوش مصنوعی، به معنای افزایش خیر عمومی، ارتقای منافع اجتماعی و در نظر گرفتن منافع همه ذی‌نفعان تعریف می‌شود (نوبرت و مونتاز، ۲۰۲۰؛ ویدجن و همکاران، ۲۰۲۰). این اصل، از دیدگاه‌هایی مانند سودگرایی و خیر جمعی مورد توجه قرار گرفته و توسعه فناوری‌های هوش مصنوعی را با هدف نهایی سودرساندن به بشریت دنبال می‌کند (سانتون و همکاران، ۲۰۱۸؛ آشوک و همکاران، ۲۰۲۲).

اگرچه مفهوم «نیکوکاری» یا «ترویج خوبی» در بسیاری از متون به کار رفته، اما کمتر به طور دقیق تعریف شده است. با این حال، برخی مطالعات برجسته این مفهوم را به مواردی چون تقویت توانایی‌های انسانی، افزایش رفاه و شکوفایی، ترویج صلح و شادی و خلق فرصت‌های اقتصادی و اجتماعی جدید ارتباط داده‌اند. از جمله راهبردهای کلیدی برای تحقق این اهداف، می‌توان به همراستاسازی توسعه هوش مصنوعی با ارزش‌های انسانی و درگیر کردن مستقیم افرادی که تحت تأثیر این فناوری‌ها قرار می‌گیرند اشاره کرد (جوبین و همکاران، ۲۰۱۹؛ کوسکی و هوسو، ۲۰۱۸). در حوزه فناوری و محیط کار، پژوهشگران مانند اسمیدز و همکاران (۲۰۱۹) تأکید می‌کنند که کارفرمایان مسئولیت دارند در فرایند پیاده‌سازی فناوری‌های هوش مصنوعی، فرصت‌هایی برای انجام کار معنادار فراهم کنند. این مسئولیت نه تنها از منظر اخلاق دینی، بلکه بر پایه شواهد تجربی در روان‌شناسی سازمانی نیز قابل دفاع است. تحقیقات نشان می‌دهند که طراحی کار معنادار نه تنها ممکن است، بلکه نقشی کلیدی در موفقیت فناوری و بهبود تجربه کاری کارکنان ایفا می‌کند. مزایای کار معنادار شامل افزایش عملکرد کارکنان، بهبود تعهد سازمانی، تقویت رفتار شهروندی سازمانی و افزایش تمایل به ماندن در سازمان است. نکته مهم این

است که فناوری به تنهایی عامل تغییر ماهیت کار نیست؛ بلکه انتخاب‌های آگاهانه‌ی مدیران و تصمیم‌گیران سازمانی در طراحی شغل، انتخاب ابزارها، نحوه‌ی سازمان‌دهی وظایف، نحوه‌ی پاداش‌دهی و ارائه‌ی اطلاعات، نقش اصلی را در معنا و کیفیت تجربه‌ی کاری ایفا می‌کنند. درنهایت، براساس مطالعات موردی، اسمیدز و همکاران (۲۰۱۹) و سانتونی د سیو (۲۰۲۴) اشاره می‌کنند که درعمل، تمایز بین مفهوم «کار معنادار» و مفاهیمی مانند «کار شایسته»، که شامل حریم خصوصی، امنیت شغلی و سلامت روانی است، دشوار است؛ زیرا این عناصر اغلب در قالب مقررات سختگیرانه‌تر و تعهدات حقوقی جدی‌تری تعریف شده‌اند. به همین دلیل، توجه به هر دو بُعد اخلاقی و قانونی، برای تضمین رفاه کارکنان در عصر هوش مصنوعی، ضروری است.

حریم خصوصی! اگرچه نگرانی‌های مربوط به حریم خصوصی پدیده‌ای جدید نیستند، اما ظهور هوش مصنوعی این نگرانی‌ها را تشدید کرده است (داوسون، ۲۰۱۹). در چارچوب هوش مصنوعی اخلاق‌محور، حریم خصوصی هم یک ارزش اخلاقی و هم یک حق انسانی تلقی می‌شود که باید حفظ و حمایت شود. بااینکه تعریف مشخصی از حریم خصوصی ارائه نشده است، اغلب با حفاظت از داده‌ها، امنیت اطلاعات، آزادی فردی و اعتماد اجتماعی مرتبط دانسته می‌شود. راهکارهای پیشنهادی برای صیانت از حریم خصوصی شامل سه رویکرد اصلی است: ۱. راه‌حل‌های فنی همچون حریم خصوصی تفاضلی، طراحی مبتنی بر حریم خصوصی، حداقل‌سازی داده‌ها و کنترل دسترسی؛ ۲. افزایش آگاهی عمومی و پژوهش‌های بین‌رشته‌ای؛ ۳. اصلاح و تکمیل چارچوب‌های قانونی، از جمله تطبیق مقررات با ویژگی‌های خاص سامانه‌های هوش مصنوعی (جابین و همکاران، ۲۰۱۹؛ هافمن و ماسوتچی، ۲۰۱۸؛ وه، ۲۰۲۴؛ مارتی‌نو و همکاران، ۲۰۲۱).

آزادی و خودمختاری اخلاقی؛ یکی از دغدغه‌های اصلی در توسعه هوش مصنوعی، حفظ آزادی فردی و خودمختاری انسان است. توسعه این فناوری نباید توانایی تصمیم‌گیری آگاهانه انسان‌ها را کاهش دهد. برخی رویکردها از «آزادی مثبت» (امکان پیشرفت، خودتعیینی، انتخاب پلنفرم‌ها) حمایت می‌کنند و برخی دیگر بر «آزادی منفی» (عدم دستکاری، اجبار یا نظارت) تأکید دارند (اشوک و همکاران، ۲۰۲۲، وان است و گرریتسن و همکاران، ۲۰۱۷؛ جوین و همکاران، ۲۰۱۹، وردنبرگ، ۲۰۲۲، سانتونی د سیو، ۲۰۲۴). افزایش شفافیت در تصمیم‌گیری‌های الگوریتمی، ارتقای سواد دیجیتال

و ارائه اطلاعات قابل فهم، ابزارهایی برای حمایت از خودمختاری افراد تلقی می‌شوند. از دیدگاه وردنبرگ (۲۰۲۲)، خودمختاری اخلاقی زمانی محقق می‌شود که افراد بتوانند نسبت به نهادهای اجتماعی و ساختارهای کاری که در آن مشارکت دارند، درک انتقادی داشته باشند و موضع‌گیری آگاهانه اتخاذ کنند.

شفافیت^۱؛ شفافیت یکی از اصول محوری در اخلاق هوش مصنوعی است که نقش مهمی در ارتقای پاسخگویی ایفا می‌کند. با این حال، درک عملکرد درونی بسیاری از الگوریتم‌ها برای کاربران و حتی تصمیم‌گیران دشوار است (به نقل از ریسمان و همکاران، ۲۰۱۸). علاوه بر پیچیدگی فنی، ملاحظاتی مانند اسرار تجاری و امکان سوءاستفاده نیز مانع شفاف‌سازی می‌شود. شفافیت می‌تواند از طریق افشای اطلاعات مربوط به داده‌های مورد استفاده، منطق تصمیم‌گیری، محدودیت‌ها، قوانین و تأثیرات اجتماعی فناوری‌ها محقق شود. پیشنهاد شده است که این اطلاعات با زبانی ساده، قابل حسابرسی و دردسترس برای نهادهای نظارتی و عموم قرار گیرد (فلوریدی، ۲۰۱۸؛ کوسکی، او. و هوسو، ۲۰۱۸؛ جوبین و همکاران، ۲۰۱۹؛ ویلانی و همکاران، ۲۰۱۸؛ دمیو، ۲۰۱۷؛ توریلی و فلوریدی، ۲۰۰۹).

عدم آسیب‌رسانی^۲؛ اصل «ضرر نرساندن» در اغلب دستورالعمل‌های اخلاقی بیش از اصل «نیکوکاری» مورد تأکید قرار گرفته است. این اصل، خواستار آن است که هوش مصنوعی از ایجاد آسیب‌های عمدی یا غیرعمدی، از جمله تبعیض، نقض حریم خصوصی یا آسیب‌های جسمی جلوگیری کند. راهبردهای پیشنهادی شامل ابزارهای فنی برای ارزیابی خطر، سازوکارهای مدیریت ریسک و نهادهای نظارتی مستقل است. (آویلا و همکاران ۲۰۱۸؛ جوبین و همکاران ۲۰۱۹؛ براندریج و همکاران ۲۰۱۸؛ وان است و گریسن و همکاران ۲۰۱۷).

اعتماد؛ اعتماد یکی از الزامات کلیدی برای پذیرش اجتماعی و موفقیت سازمانی هوش مصنوعی است. ایجاد و حفظ اعتماد مستلزم توسعه سامانه‌های قابل اطمینان، پاسخگو و قابل نظارت است. راهکارهای پیشنهادی شامل آموزش، شفافیت، طراحی مبتنی بر اصول اخلاقی، و ارزیابی مستمر عملکرد سامانه‌هاست. در برخی منابع، به جای تأکید صرف بر فهم‌پذیری، بر برآورده کردن انتظارات عمومی و انطباق با هنجارهای اجتماعی تأکید شده است (جوبین و همکاران، ۲۰۱۹؛ دا موتا ویگا و همکاران، ۲۰۲۳).

1. Transparency
2. Do no harm

پایداری!؛ هوش مصنوعی باید در راستای توسعه پایدار طراحی و به کارگیری شود، به گونه‌ای که نه تنها از محیط‌زیست حفاظت کند، بلکه به برابری اجتماعی، صلح و عدالت نیز کمک کند. این امر مستلزم کاهش مصرف انرژی، بهینه‌سازی فرایندها و طراحی سیاست‌های سازمانی در مواجهه با چالش‌هایی همچون حذف مشاغل است (جویین و همکاران، ۲۰۱۹).

کرامت؛ اگرچه اغلب به صورت ضمنی و بدون تعریف دقیق مطرح می‌شود، اما به شدت با حقوق بشر، آزادی فردی و مقابله با تبعیض در ارتباط است. از توسعه‌دهندگان و سیاست‌گذاران انتظار می‌رود که تدابیر فنی، حقوقی و نهادی لازم برای صیانت از کرامت انسانی را در نظر بگیرند (جویین و همکاران، ۲۰۱۹).

همبستگی^۲؛ همبستگی به عنوان ارزش اخلاقی در مواجهه با پیامدهای نابرابر فناوری‌های نوظهور، بر توزیع عادلانه منافع، حمایت از اقشار آسیب‌پذیر و تقویت سرمایه اجتماعی تأکید دارد. این ارزش باید از طریق سیاست‌های بازتوزیعی، آموزش اخلاقی و طراحی فراگیر سامانه‌های هوش مصنوعی تقویت شود (اشوک، ۲۰۲۲؛ سیکر، ۲۰۱۹؛ جویین و همکاران، ۲۰۱۹).

حقوق قانونی و انسانی؛ با ورود هوش مصنوعی به فرایندهای جذب و استخدام، مسئله "بررسی سایبری" (استفاده از اطلاعات برخط غیررسمی برای تصمیم‌گیری‌های استخدامی) اهمیت دوچندانی پیدا می‌کند. این عمل که اغلب بدون اطلاع یا رضایت متقاضیان انجام می‌شود، با انتقادهای جدی در مورد نقض حریم خصوصی و بی‌عدالتی همراه است که می‌تواند به کاهش تمایل به پذیرش شغل و طرح دعوی‌های حقوقی منجر شود. به دلیل فقدان قوانین مشخص برای حمایت از منافع کارکنان در این زمینه، محققان به دنبال ایجاد مقرراتی برای حفاظت از داده‌ها و حقوق کارکنان، از جمله استانداردهای بین‌المللی حقوق بشر هستند. رعایت حقوق قانونی و انسانی در استفاده از هوش مصنوعی در استخدام، زیربنای سیاست‌های داده‌های منابع انسانی است (تورسونیاوا، ۲۰۲۲). تحقیقات نشان می‌دهد که الگوریتم‌ها می‌توانند نه تنها به حقوق اساسی افراد آسیب برسانند، بلکه باعث تبعیض و نقض حقوق اخلاقی نیز شوند. این موضوع در بررسی سایبری، سازمان‌ها را با چالش‌های متعددی روبه‌رو می‌کند. آنها باید در مورد تصمیمات مربوط به بررسی سایبری شفاف عمل کنند و تعادلی بین شفافیت، حریم

خصوصی و محرمانگی برقرار سازند. این امر، پیچیدگی‌های استفاده از ابزارهای دیجیتال را نشان می‌دهد (سیناتو، ۲۰۲۴).

از دست رفتن شغل و سایر مسائل اجتماعی ناشی از بازسازی بازار کار؛ یکی از

دغدغه‌های اصلی در مباحث فلسفی، سیاسی و اجتماعی پیرامون آینده کار و گسترش هوش مصنوعی، افزایش بیکاری ناشی از اتوماسیون و بازسازی ساختار بازار کار است. در این زمینه، یک رویکرد رایج بر آن است که برای مقابله با پیامدهای منفی فناوری‌های نوین، به‌ویژه نابرابری درآمد و بیکاری، باید سیاست‌های بازتوزیعی قوی‌تری اتخاذ شود. یکی از پیشنهادهای مطرح، وضع مالیات بر ربات‌هاست؛ یعنی دریافت مالیات از شرکت‌هایی که از فناوری‌های خودکار استفاده می‌کنند تا از طریق آن، از کارکنانی که به واسطه این فناوری‌ها شغل خود را از دست می‌دهند، حمایت شود. در کنار آن، پیشنهاد گسترده‌تری به نام “درآمد پایه همگانی” نیز مطرح شده است؛ بدین معنا که دولت مبلغی ثابت را بدون توجه به وضعیت شغلی یا درآمدی به همه شهروندان پرداخت کند. اگرچه این ایده می‌تواند تاحدی از نظر مالی افراد بیکار را حمایت کند، اما نمی‌تواند به‌صورت پیشگیرانه از بروز بیکاری گسترده جلوگیری کند (سانتونی د سیو، ۲۰۲۴). علاوه بر این، این رویکردها بیکاری ناشی از فناوری را امری اجتناب‌ناپذیر و خارج از کنترل تلقی می‌کنند و توسعه فناوری را به گونه‌ای می‌نگرند که گویی امری مستقل و غیرقابل هدایت اجتماعی و سیاسی است. این نگاه، ریشه در نوعی “تعیین‌گرایی فناورانه” دارد که مانع از مداخله فعالانه برای جهت‌دهی اخلاقی و اجتماعی به توسعه فناوری می‌شود.

شکل‌گیری نوع جدیدی از بهره‌کشی از کار؛ گسترش پلتفرم‌های دیجیتال، نوع جدیدی از

بهره‌کشی از کار شکل گرفته است. این پلتفرم‌ها بخش زیادی از موفقیت خود را مدیون داده‌هایی هستند که توسط کاربران تولید می‌شود؛ داده‌هایی که حاصل فعالیت‌های فکری، عاطفی و اجتماعی کاربران است. این فعالیت‌ها هرچند داوطلبانه و رایگان به نظر می‌رسند، اما در عمل تحت سلطه شرکت‌های مالک پلتفرم قرار دارند؛ چرا که کاربران در بسیاری از ابعاد زندگی روزمره خود - از کار گرفته تا روابط اجتماعی - به این پلتفرم‌ها وابسته شده‌اند. این وابستگی باعث شکل‌گیری نوعی «سلطه نرم» شده است؛ یعنی وضعیتی که در آن افراد اگرچه ظاهراً آزادند، اما در واقع در موقعیت‌هایی قرار دارند که انتخاب‌های واقعی آنها محدود شده است. این مسئله، نشانگر نوعی بهره‌کشی ساختاری است که در آن روابط قدرت ناعادلانه و سلطه‌گرایانه پنهان شده‌اند (ورسالیس، ۲۰۱۹). در این چارچوب، شرکت‌های بزرگ فناوری

از دو مسیر بهره‌برداری می‌کنند: نخست، از کار نهادهای عمومی مانند دانشگاه‌ها و مؤسسات پژوهشی که دانش و فناوری اولیه را تولید می‌کنند، دوم؛ از کاربران عادی که با تولید داده، محتوا و ترافیک، ارزش اقتصادی این پلتفرم‌ها را افزایش می‌دهند. این نوع بهره‌برداری دو گانه در ادبیات «سرمایه‌داری شناختی» و تحلیل‌های انتقادی از اقتصاد دیجیتال، به‌ویژه در رویکردهای مارکسیستی، به‌خوبی تبیین شده است (کریستینز، ۲۰۲۳).

در کنار مسائل ذکر شده، تولید و توسعه فناوری‌های هوش مصنوعی به منابع گسترده‌ای از داده، مواد خام، انرژی و نیروی کار نیاز دارد. برخلاف تصور رایج از اطلاعات و دانش به‌عنوان منابعی «غیررقابتی» و در دسترس همگان، تولید آنها با محدودیت‌های مادی و ساختاری روبه‌روست. تولید سامانه‌های هوشمند نیازمند استخراج معادن، مصرف بالای انرژی و حجم زیادی از داده‌هاست که اغلب از کشورهای در حال توسعه تأمین می‌شوند (کرافورد، ۲۰۲۱). به‌همین ترتیب، کار انسانی نیز بخش مهمی از فرایند آموزش هوش مصنوعی است. در این راستا، شرکت‌های فناوری از طریق برون‌سپاری به شرکت‌های آفریقایی، نیروی کار ارزان و فاقد حقوق کافی را برای انجام کارهایی همچون دسته‌بندی داده‌ها یا تعدیل محتوا استخدام می‌کنند. این واقعیت، نوعی بازتولید ساختار نئو کلونیالیستی در مقیاس جهانی است که در آن منافع فناوری عمدتاً به کشورهای توسعه یافته می‌رسد، اما هزینه‌های زیست‌محیطی و انسانی آن بر کشورهای در حال توسعه تحمیل می‌شود (پریگو، ۲۰۲۳؛ سانتونی د سیو، ۲۰۲۴).

قابلیت تبیین و تشریح؛ یکی از مسائل مهم در اخلاق هوش مصنوعی، به‌ویژه در حوزه سلامت، قابلیت توضیح و تبیین است. یک مدل هوش مصنوعی زمانی قابل توضیح در نظر گرفته می‌شود که اطلاعاتی ارائه دهد که هم قابل تفسیر (یعنی برای کاربران انسانی قابل درک باشد) و هم کامل (یعنی عملکرد واقعی سامانه را به‌درستی بازنمایی کند) باشد.

اما رسیدن به این دو ویژگی به‌طور همزمان چالش‌برانگیز است؛ چراکه هرچه توضیح دقیق‌تر و جامع‌تر باشد، درک آن برای کاربران غیرمتخصص دشوارتر می‌شود و در مقابل، هرچه توضیح ساده‌تر باشد، ممکن است بخشی از دقت یا عملکرد واقعی مدل را نادیده بگیرد. به همین دلیل است که بسیاری از فناوری‌های پزشکی مبتنی بر هوش مصنوعی، به‌دلیل استفاده از الگوریتم‌های پیچیده و دشوارفهم، به‌عنوان «جعبه سیاه» شناخته می‌شوند.

مطالعه‌ای از مارینهو (۲۰۲۱) نشان می‌دهد که در زمینه توسعه و به‌کارگیری سامانه‌های هوش مصنوعی قابل توضیح در سلامت، اختلاف نظرهای جدی وجود دارد. در این میان، قابلیت توضیح نه تنها برای افزایش اعتماد و پذیرش پزشکان اهمیت دارد، بلکه یک اصل اخلاقی بنیادی نیز به‌شمار می‌آید. برای درک بهتر این مسئله، فناوری MRI اغلب به‌عنوان نمونه مقایسه‌ای مطرح می‌شود. اگرچه MRI نیز یک فناوری پیچیده است و پزشکان لازم نیست جزئیات فیزیکی یا ریاضی آن را بدانند، اما الگوریتم‌ها و سازوکار آن از سوی سازندگان و متخصصان قابل توضیح و درک‌پذیر است. در مقابل، بسیاری از سامانه‌های هوش مصنوعی به‌گونه‌ای طراحی شده‌اند که حتی توسعه‌دهندگان آنها نیز دقیقاً نمی‌دانند الگوریتم چگونه به یک تصمیم خاص رسیده است. این وضعیت به‌ویژه در مراقبت‌های بهداشتی می‌تواند مشکل‌ساز باشد؛ زیرا تصمیمات درمانی باید برپایه اطلاعات شفاف و قابل اطمینان اتخاذ شوند. همان‌طور که یکی از شرکت‌کنندگان در پژوهش مارینهو و همکاران (۲۰۲۱) گفته است: «یک متخصص خوب سلامت، هرگز تنها به یک شاخص یا عدد اعتماد نمی‌کند، مگر آنکه سابقه کامل بیمار را نیز در نظر گرفته باشد.» ازای رو، تلاش برای طراحی مدل‌هایی که ضمن حفظ دقت بالا، بتوانند به شیوه‌ای شفاف، تصمیمات خود را توضیح دهند، هم از منظر اخلاقی و هم از نظر کارایی بالینی، امری ضروری است.

خود مختاری و هویت شخصی؛ سامانه‌های توصیه‌گر^۱ با ارائه پیشنهادهای شخصی‌سازی‌شده، می‌توانند تأثیرات قابل توجهی بر خودمختاری افراد بگذارند. این تأثیر می‌تواند از کمک‌های جزئی مانند حذف اطلاعات اضافی و تسهیل فرایند تصمیم‌گیری، تا مداخلات سنگین‌تر مانند فریب‌دادن یا سوق‌دادن افراد به انتخاب‌های خاص گسترش یابد. این سامانه‌ها با تحلیل رفتارهای گذشته کاربران، پروفایل‌هایی الگوریتمی از آنها ایجاد می‌کنند. این پروفایل‌ها گاهی موجب تقویت گروه‌بندی‌های اجتماعی موجود و حتی ایجاد سوگیری در پیشنهادهای می‌شوند. تأثیر این پروفایل‌سازی بر هویت شخصی را می‌توان در دو سطح بررسی کرد: ۱. تأثیر بر هویت در حال‌تغییر: پروفایل‌ها تنها بازتابی از گذشته کاربر نیستند، بلکه با پیشنهادات خود، می‌توانند نقش فعالی در شکل‌گیری هویت آینده افراد ایفا کنند. به بیان دیگر، کاربران تنها دریافت‌کننده پیشنهاد نیستند، بلکه رفتار آنها در تعامل با این پیشنهادات، بر ادراک آنها از خود نیز اثرگذار است. ۲. ناهمخوانی با هویت واقعی: گاهی گروه‌هایی که توسط

الگوریتم‌ها به کاربر نسبت داده می‌شوند، با هویت‌هایی که افراد به‌طور شخصی برای خود قائلند، تطابق ندارند. این ناهماهنگی می‌تواند باعث شود کاربران نسبت به محتوای پیشنهادی احساس بیگانگی یا بی‌ربط بودن داشته باشند.

در نهایت، اگرچه شخصی‌سازی محتوا می‌تواند تجربه کاربران را بهبود ببخشد، اما ممکن است افراد را از شبکه‌های اجتماعی و فرهنگی گسترده‌تری که بخشی از هویت آنها را شکل می‌دهند، جدا کند. به‌عبارت‌دیگر، هزینه‌های اجتماعی و هویتی شخصی‌سازی نباید نادیده گرفته شود (بر و همکاران، ۲۰۱۸؛ ده فریس، ۲۰۱۰؛ کوئن و همکاران؛ تدئو و فلوریدی، ۲۰۱۸).

قطبی‌شدن و دستکاری اجتماعی؛ یکی دیگر از پیامدهای مهم سامانه‌های توصیه‌گر، قطبی‌سازی دیدگاه‌ها و افزایش آسیب‌پذیری در برابر دستکاری اطلاعاتی است. این سامانه‌ها با نمایش محتوا براساس علائق گذشته کاربران، ممکن است آنها را درون «حباب‌های فیلتر» نگه دارند؛ فضایی که در آن کاربران تنها با دیدگاه‌هایی مواجه می‌شوند که با باورهای آنها همسوست. چنین شرایطی می‌تواند به شکل‌گیری تعصبات فکری، کاهش تنوع اطلاعاتی، تضعیف گفت‌وگوی عمومی و درنهایت، آسیب به بنیان‌های دموکراتیک منجر شود.

علاوه‌براین، این سامانه‌ها در برابر دستکاری هدف‌مند نیز آسیب‌پذیرند. گروه‌های کوچک اما سازمان‌یافته می‌توانند با استفاده از تعاملات مصنوعی یا داده‌های ساختگی، الگوریتم‌ها را به سمت نمایش گسترده‌تر محتوای خاص سوق دهند. این ویژگی، آنها را به ابزاری مؤثر برای تبلیغات سیاسی هدف‌مند و تأثیرگذاری بر افکار عمومی تبدیل کرده است. مطالعات پیشین، راهکارهایی برای کاهش این اثرات پیشنهاد کرده‌اند. از جمله آنها می‌توان به افزایش تنوع در محتواهای توصیه‌شده اشاره کرد، به‌گونه‌ای که درعین حفظ ارتباط با علائق کاربران، دیدگاه‌های متنوع نیز در دسترس قرار گیرند. پژوهشگران بر ضرورت ایجاد تعادل میان شخصی‌سازی و تنوع اطلاعاتی تأکید دارند و معتقدند که طراحی سامانه‌های توصیه‌گر باید با ارزش‌های دموکراتیک و پاسخگویی اجتماعی هم‌راستا باشد (به نقل از بوزداغ، ۲۰۱۳؛ میلان و همکاران، ۲۰۲۰؛ بوزداغ و ون دن هوون، ۲۰۱۵).

بحث و نتیجه‌گیری

در این پژوهش تلاش شد که دیدگاه‌ها و موضوعات اخلاقی در به‌کارگیری هوش مصنوعی و فناوری‌های عصبی در محیط کار تبیین شود. بدین ترتیب با استفاده از روش مرور اسکوپینگ ۵۳ مقاله حوزه اخلاق هوش مصنوعی مورد بررسی و با کمک تحلیل محتوا، کدهای مربوطه استخراج شد؛ با مرور ادبیات (ژوبین و همکاران، ۲۰۱۹؛ موری و همکاران، ۲۰۲۴؛ تورسونبایوا، ۲۰۲۲) عنوان شد که رشد سریع و تنوع اسناد راهنمای اخلاق هوش مصنوعی، نشان از توجه روزافزون جامعه جهانی و سازمان‌های مختلف به این موضوع دارد. تقریباً برابری تعداد اسناد منتشرشده از سوی بخش عمومی (سازمان‌های دولتی و بین‌المللی) و بخش خصوصی (شرکت‌ها و اتحادیه‌های خصوصی) حاکی از آن است که نگرانی‌های اخلاقی هوش مصنوعی، هم نهادهای عمومی و هم شرکت‌های خصوصی را دربرمی‌گیرد. با این حال، راه‌حل‌های پیشنهادی برای چالش‌های اخلاقی هوش مصنوعی تفاوت‌های قابل توجهی با یکدیگر دارند. علاوه بر این، نمایندگی ناکافی مناطق جغرافیایی مانند آفریقا، آمریکای لاتین و آسیای مرکزی، نشان‌دهنده مشارکت نابرابر مناطق مختلف جهان در بحث اخلاق هوش مصنوعی و عدم تعادل قدرت در گفت‌وگو بین‌المللی است. کشورهای توسعه‌یافته‌تر نقش غالب‌تری در این گفت‌وگو دارند که نگرانی‌هایی در مورد نادیده گرفتن دانش بومی، تنوع فرهنگی و مطالبات عدالت جهانی ایجاد می‌کند. تحلیل موضوعی ما وجود ۱۷ موضوع اخلاقی را آشکار می‌کند و همگرایی فزاینده ذی‌نفعان را در ترویج اصول اخلاقی عدالت، انصاف و برابری نژادی و جنسیتی، شفافیت، حریم خصوصی، عدم آسیب، خودمختاری و هویت شخصی، پایداری، کرامت، همبستگی و اعتماد را نشان می‌دهد که در مطالعات به آنها اشاره شده است. با توجه به تحلیل مطالعات صورت گرفته از اصول اخلاقی حاکم بر هوش مصنوعی اگرچه برخی اصول مثل عدم ضرررسانی و حریم خصوصی اهمیت زیادی دارند، اما بین اصول اخلاقی مختلف اختلافات و تناقضاتی وجود دارد. این تناقضات در نحوه تفسیر، اهمیت، موضوعات مرتبط و روش اجرای اصول اخلاقی دیده می‌شود. مطالعه نشان می‌دهد که اصولی مثل پایداری و همبستگی به اندازه کافی مورد توجه قرار نگرفته‌اند، در حالی که اعتماد به هوش مصنوعی نیز موضوعی بحث‌برانگیز است. برای حل این چالش‌ها، نیاز به همکاری بیشتر بین ذی‌نفعان، هماهنگی بیشتر بین دستورالعمل‌های اخلاقی و قوانین و ایجاد سازوکارهای حل اختلاف بین مناطق مختلف جهان اشاره می‌شود. همچنین ضرورت استانداردسازی، استفاده از سازوکارهای اخلاق پژوهی و توجه ویژه به عدالت

در استفاده از هوش مصنوعی در فرایندهای استخدام، به عنوان نقاط کلیدی مطرح شده‌اند. در نهایت، مطالعات بر ضرورت تعامل بیشتر برای رسیدن به یک چشم‌انداز اخلاقی فراگیرتر برای هوش مصنوعی تأکید دارند. به طور خلاصه، یافته‌های این مقاله به این شرح است که تاکنون هیچ تحقیقی موضوعات اخلاقی را با این جامعیت ارائه نداده است. از آنجا که تحقیقات در مورد دیدگاه‌ها و اصول اخلاقی به کارگیری هوش مصنوعی، هریک بخشی از مؤلفه‌های اخلاقی به کارگیری هوش مصنوعی را به طور جداگانه بررسی کرده‌اند، هریک از متخصصان این موضوع را با دیدگاه خاص خود و شرایط زمینه‌ای متفاوت بررسی کرده و ابعاد و مؤلفه‌های اخلاقی متفاوتی ارائه داده‌اند. بنابراین، تصمیم بر ارائه یک مدل پژوهشی با دیدگاه جامع و فراگیر از دیدگاه و موضوعات اخلاقی هوش مصنوعی در محیط کار و مطالعه گسترده گرفته شد. بنابراین، استفاده از این روش‌شناسی را می‌توان به نوآوری‌های این تحقیق اضافه کرد. دوم، اندیشمندان و محققان می‌توانند خلاصه‌ای از یافته‌های کلیدی محققان پیشین را در چارچوب چنین مرور نظام‌مند ادبیات مرتبط درک کنند.

پیشنهادات تحقیق: با توجه به یافته‌های این پژوهش، روشن است که به کارگیری هوش مصنوعی در سازمان‌ها صرفاً یک تحول فناورانه نیست، بلکه پدیده‌ای پیچیده با پیامدهای گسترده اخلاقی، حقوقی و اجتماعی است. این نتایج نشان می‌دهد که غفلت از ملاحظات عدالت، شفافیت، مسئولیت‌پذیری و کرامت انسانی در فرایند طراحی و پیاده‌سازی سامانه‌های هوش مصنوعی می‌تواند به بروز تبعیض، نقض حریم خصوصی، تضعیف اعتماد سازمانی و نابرابری‌ها منجر شود. بر این اساس، پیشنهاد می‌شود سازمان‌ها در سطوح مختلف تصمیم‌گیری و اجرا، اقدامات زیر را در دستور کار خود قرار دهند:

۱. توسعه فناوری، به‌ویژه هوش مصنوعی، باید با مداخله فعال سیاست‌گذاران و جامعه مدنی همراه باشد تا ارزش‌های اخلاقی مانند عدالت، شفافیت و مسئولیت‌پذیری در آن لحاظ شود و سازمان‌ها باید هنگام استفاده از فناوری‌های هوش مصنوعی، چارچوب‌های اخلاقی مشخصی برای انتخاب، خرید، طراحی و پیاده‌سازی این فناوری‌ها تدوین کرده و کمیته‌های اخلاق فناوری درون‌سازمانی ایجاد کنند.
۲. نهادهای ناظر باید چارچوب‌هایی برای مسئولیت اجتماعی شرکت‌های فعال در حوزه فناوری تدوین کنند تا فعالیت‌های آنها با اصول عدالت اجتماعی و توسعه پایدار هماهنگ باشد. شرکت‌ها باید بخشی از درآمد حاصل از فناوری‌های مبتنی بر داده و هوش مصنوعی را به سرمایه‌گذاری در آموزش، توسعه جوامع محلی و کاهش شکاف دیجیتال اختصاص دهند.

۳. ایجاد نهادهای ناظر مستقل برای بررسی و ارزیابی عملکرد هوش مصنوعی در حوزه‌های حساس، با تمرکز بر عدالت، مسئولیت‌پذیری و شفافیت. این نهادها می‌توانند به کاهش شکاف‌های اخلاقی و حقوقی کمک کنند.

۴. توسعه سامانه‌های هوشمند با قابلیت توضیح‌پذیری، تعیین مسئولیت‌های مشخص حقوقی و انسانی در فرایند تصمیم‌سازی و همچنین و شرکت‌های توسعه‌دهنده ملزم به ارائه مستندات فنی و توضیح‌پذیری تصمیمات الگوریتمی باشند، به‌ویژه در مواردی که کاربران انسانی (مانند پزشکان، قضات یا کارفرمایان) باید به دیگران پاسخ دهند.

۵. تطبیق سیاست‌های سازمانی با اصول نوین حفاظت داده از جمله رضایت آگاهانه، طراحی مبتنی بر حریم خصوصی و کنترل دسترسی در راستای افزایش اعتماد عمومی.

۶. گنجاندن دوره‌های آموزش اخلاق حرفه‌ای و حقوق فناوری در برنامه‌های درسی علوم مهندسی و هوش مصنوعی، برای تقویت آگاهی توسعه‌دهندگان نسبت به پیامدهای اجتماعی و انسانی فناوری.

۷. سازمان‌ها می‌باید در فرایند پیاده‌سازی هوش مصنوعی، به گونه‌ای برنامه‌ریزی کنند که فرصت انجام کار معنادار برای کارکنان حفظ یا تقویت شود. این امر می‌تواند با ایجاد فضای مشارکت در تصمیم‌گیری، واگذاری وظایف خلاقانه و تقویت احساس هدف‌مندی در کار تحقق یابد.

۸. برنامه‌ریزی برای آموزش مستمر در زمینه مهارت‌های دیجیتال، تحلیل داده و همکاری با فناوری‌های هوش مصنوعی، با هدف جلوگیری از شکاف مهارتی و تقویت ظرفیت تطابق کارکنان با تحولات فناورانه.

۹. گنجاندن شاخص‌های پایداری اجتماعی و زیست‌محیطی در ارزیابی فناوری‌های مرتبط با هوش

مصنوعی

۱۰. شرکت‌هایی که از پلتفرم‌های دیجیتال استفاده می‌کنند، باید سیاست‌هایی برای محافظت از حریم خصوصی کاربران، شفافیت الگوریتم‌ها و ایجاد روش‌های بازخورد کاربران تدوین کنند.

کاربردهای مدیریتی نتایج پژوهش

یافته‌های این پژوهش مجموعه‌ای از پیامدها و الزامات مدیریتی را در زمینه توسعه و استفاده اخلاقی و مسئولانه از هوش مصنوعی در سازمان‌ها روشن می‌سازد. براین اساس، مدیران می‌توانند از نتایج پژوهش در جهت طراحی سیاست‌ها و راهبردهایی بهره‌گیرند که بهره‌وری فناورانه را با ارزش‌های

انسانی و اصول عدالت اجتماعی همراستا سازد. کاربردهای مدیریتی نتایج عبارتند از:

۱. طراحی و استقرار چارچوب اخلاقی سازمانی برای هوش مصنوعی: پیشنهاد می‌گردد مدیران نسبت به تدوین دستورالعمل‌ها و کدهای رفتاری در حوزه هوش مصنوعی، چارچوبی اخلاق محور برای انتخاب، طراحی و به کارگیری این فناوری اقدام نمایند. مشارکت کارشناسان اخلاق، حقوق، منابع انسانی و فناوری اطلاعات در این فرایند توصیه می‌شود.

۲. پیاده‌سازی نظام شفافیت و پاسخگویی در تصمیم‌گیری‌های الگوریتمی: پیشنهاد می‌گردد در فرایندهایی که از هوش مصنوعی برای تصمیم‌گیری استفاده می‌شود (مانند استخدام، ارزیابی عملکرد یا تخصیص منابع)، نقش انسان مشخص، سازوکارهای پاسخگویی تعریف و توضیح‌پذیری الگوریتم‌ها تضمین شود.

۳. توسعه سیاست‌های آموزش و بازآموزی فناورانه برای کارکنان: پیشنهاد می‌گردد مدیران منابع انسانی نسبت به تدوین برنامه‌های آموزش مستمر برای تقویت مهارت‌های دیجیتال و تطبیق کارکنان با تحولات فناورانه اقدام نمایند. این اقدام مانع از حذف نیروی انسانی و منجر به ارتقای آنها در نقش‌های فناور محور خواهد شد.

۴. بازنگری در سیاست‌های حریم خصوصی و حفاظت از داده‌ها: کاربردهای هوش مصنوعی مستلزم مدیریت دقیق داده‌های شخصی است. مدیران می‌باید سیاست‌های حفاظت از داده‌ها را به‌روز کرده، آموزش‌های مرتبط را به کارکنان ارائه دهند و اصولی چون رضایت آگاهانه و کنترل دسترسی را اجرا کنند.

۵. ایجاد ساختارهای نظارتی و ارزیابی مستقل در سازمان: پیشنهاد می‌گردد مدیران کمیته‌هایی بین‌رشته‌ای برای نظارت بر عملکرد سامانه‌های هوش مصنوعی تشکیل دهند تا از منظر عدالت، عدم تبعیض، شفافیت و مسئولیت‌پذیری عملکرد این سامانه‌ها را ارزیابی کنند.

۶. طراحی فرایندهای کاربر محور و انسانی محور در تعامل با فناوری: در توسعه سامانه‌های هوشمند، مدیران می‌باید به نیازها، سطح سواد فناورانه و تجربیات کاربران انسانی توجه داشته باشند تا تعامل مؤثر، حس معناداری و کنترل انسانی حفظ شود.

۷. پایش منظم شاخص‌های عدالت، رفاه و سلامت سازمان: استفاده از داده‌های سازمانی برای پایش شاخص‌هایی مانند رضایت شغلی، عدالت ادراک شده، فرصت برابر و سلامت روان می‌تواند به بهبود

- تصمیم‌گیری مدیریتی و شناسایی زود هنگام پیامدهای منفی فناوری کمک کند.
۸. مدیریت تبعات اجتماعی فناوری و توزیع منصفانه منافع آن: پیشنهاد می‌گردد مدیران نسبت به تدوین سیاست‌هایی مانند باز توزیع منافع ناشی از اتوماسیون، پشتیبانی از کارکنانی که شغلشان تحت تأثیر قرار گرفته و ایجاد فرصت‌های جدید شغلی، از شکاف‌های اجتماعی جلوگیری کنند.
۹. کنترل و پیشگیری از تبعیض و سوگیری الگوریتمی: مدیران می‌باید از ابزارهای ارزیابی سوگیری الگوریتم‌ها پیش از بهره‌برداری استفاده کنند و تیم‌های توسعه‌دهنده را با اصول طراحی اخلاقی، داده‌محور و منصفانه آشنا سازند.
۱۰. حفظ کرامت انسانی و اعتماد در محیط کار فناورمحور: در تعاملات سازمانی که توسط هوش مصنوعی پشتیبانی می‌شوند، می‌باید از فرایندهای غیرشخصی و ماشینی شدن روابط انسانی، به‌ویژه در حوزه‌های حساسی چون منابع انسانی یا خدمات مشتریان، پرهیز شود.
- محدودیت‌های تحقیق:** در حالی که این مطالعه به درک موضوعات اخلاقی در به کارگیری هوش مصنوعی در محیط کار کمک می‌کند، دارای محدودیت‌هایی نیز هست. محدودیت‌های دسترسی به پایگاه‌های داده و مجلات خاص، جامعیت مرور ادبیات را محدود کرده است که ممکن است به حذف مطالعات مرتبط و تأثیرگذاری بر تعمیم‌پذیری یافته‌ها منجر شود. تحقیقات آینده باید با استفاده از راهبردهای گسترده‌تر جمع‌آوری داده‌ها، از جمله روش‌های متنوع و دسترسی به طیف وسیع‌تری از منابع علمی، این محدودیت‌ها را برطرف کنند و در نتیجه استحکام و اعتبار بینش‌ها در موضوعات اخلاقی در به کارگیری هوش مصنوعی در محیط کار را افزایش دهند.

منابع

- احمدشاهی، امیرحسین؛ سلیمی، قاسم و یزدان‌دوست، پریسا (۱۴۰۴)، واکاوی نقش فناوری هوش مصنوعی در ارتقای فرایندهای آموزش، بازآموزی و بهسازی حرفه‌ای کارکنان: مرور گستره دانشی، *مطالعات منابع انسانی*، دوره پانزدهم، شماره سوم: ۶۲ - ۹۲.
- اربابی اصفهانی، محمد؛ کوشکی جهرمی، علی‌رضا، زاهدی، شمس‌السادات و خان‌محمدی، هادی (۱۴۰۴)، طراحی مدل ارزیابی خط‌مشی‌های فرهنگ تحول دیجیتال، *مطالعات رفتار سازمانی*، دوره چهاردهم، شماره یکم: ۱۸۱ - ۲۳۲.
- پدرامی، محمد و نرگسیان، عباس (۱۴۰۴)، شناسایی و تحلیل فراترکیب چالش‌های اخلاقی هوش مصنوعی در مدیریت منابع انسانی: ارائه چارچوب مفهومی، *مطالعات منابع انسانی*، دوره پنجم، شماره یکم: ۱۴۳ - ۱۶۶.
- شیرازی، مهرزاد؛ یزدانی، حمیدرضا و زارعی متین، حسن (۱۴۰۰)، ارائه نقشه راه جاری‌سازی فرهنگ سازمانی موردنیاز برای تحول دیجیتال با رویکرد فراترکیب، *مطالعات رفتار سازمانی*، دوره دهم شماره سوم: ۱ - ۲۲.
- کریمی، پیام؛ بسطامی، مسعود و ایراندوست، منصور (۱۴۰۳)، فراترکیب الگوی بسط مفهومی فرهنگ سازمانی با رویکرد تحول دیجیتال، *مطالعات رفتار سازمانی*، دوره سیزدهم، شماره چهارم: ۹ - ۴۰.
- مدرسی، یاسمن؛ سیدنقوی، میرعلی؛ رودساز، حبیب‌الله و رئیسی وانانی، ایمان (۱۴۰۳)، شناسایی مؤلفه‌های تجربه دیجیتال کارکنان و ارائه الگوی حاصل از آنها در شهرداری شهرستان سمنان، *مطالعات رفتار سازمانی*، دوره سیزدهم، شماره یکم: ۱۵۹ - ۱۹۶.

Reference

- Al-Amoudi, I. (2023). The politics of post-human technologies: Human enhancements, artificial intelligence and virtual reality. *Organization*, 30(6), 1238–1245.
- Ali, M. A., Hussin, N., Haddad, H., Al-Araj, R., & Abed, I. A. (2021). A multidimensional view of intellectual capital: The impact on innovation performance. *Journal of Open Innovation: Technology, Market, and Complexity*, 7(4), 216.
- Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology*, 78(4), 772.
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology: Theory and Practice*, 8(1), 19–32.

- Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for artificial intelligence and digital technologies. **International Journal of Information Management**, 62, 102433. <https://doi.org/10.1016/j.ijinfomgt.2022.102433>
- Avila, R., Brandusescu, A., Freuler, J. O., & Thakur, D. (2018). **Artificial intelligence: Open questions about gender inclusion**. World Wide Web Foundation.
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The moral machine experiment. **Nature**, 563, 59.
- Bære, K., & Gundersen, T. (2019). Social impact under severe uncertainty: The role of neuroethicists at the intersection of neuroscience, AI, ethics, and policymaking. **AJOB Neuroscience**, 10(3), 117–119.
- Bankins, S., & Formosa, P. (2023). The ethical implications of artificial intelligence (AI) for meaningful work. **Journal of Business Ethics**, 185(4), 725–740.
- Barlow, C. (2020). Oncology research: Clinical trial management systems, electronic medical record, and artificial intelligence. **Seminars in Oncology Nursing**, 36(2).
- Barocas, S., & Nissenbaum, H. (2014). **Big Data's end run around anonymity and consent**.
- Beltramino, N. S., Garcia-Perez-de-Lema, D., & Valdez-Juarez, L. E. (2020). The structural capital, the innovation, and the performance of the industrial SMES. **Journal of Intellectual Capital**, 21(6), 913–945. <https://doi.org/10.1108/jic-01-2019-0020>
- Bilgin, R. (2024). A machine learning analysis of the value-added intellectual coefficient's effect on firm performance. **Journal of Modelling in Management**.
- Bizoi, A. C., & Bizoi, C. G. (2024). Neuroethical implications of AI-driven productivity tools on intellectual capital: A theoretical and econometric analysis. **Journal of Intellectual Capital**, 26(7), 1–23.
- Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. **Science**, 352, 1573.
- Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In K. Frankish & W. M. Ramsey (Eds.), **The Cambridge Handbook of Artificial Intelligence**, 316–334. Cambridge University Press.
- Bozdag, E. (2013). Bias in algorithmic filtering and personalization. **Ethics and Information Technology**, 15, 209–227.
- Bozdag, E., & van den Hoven, J. (2015). Breaking the filter bubble: Democracy and design. **Ethics and Information Technology**, 17(4), 249–265.
- Breidbach, C. F., & Maglio, P. (2020). Accountable algorithms? The ethical implications of data-driven business models. **Journal of Service Management**.
- Brundage, M., et al. (2018). **The malicious use of artificial intelligence: Forecasting, prevention, and mitigation**. Oxford: Future of Humanity Institute, University of Oxford, Centre for the Study of Existential Risk, University of Cambridge, Center for a New American Security, Electronic Frontier Foundation, OpenAI.
- Brynjolfsson, E., & McAfee, A. (2014). **The second machine age: Work, progress, and prosperity in a time of brilliant technologies**. New York: W. W. Norton & Company.

- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In **Conference on fairness, accountability and transparency**, 77–91.
- Burr, C., Cristianini, N., & Ladyman, J. (2018). An analysis of the interaction between intelligent software agents and human users. **Minds and Machines**, 28(4), 735–774.
- Castells, M. (2002). **The Internet galaxy: Reflections on the Internet, business, and society**. Oxford University Press on Demand.
- Cervantes, J.-A., López, S., Rodríguez, L.-F., Cervantes, S., Cervantes, F., & Ramos, F. (2020). Artificial moral agents: A survey of the current status. **Science and Engineering Ethics**, 26, 501.
- Chondough, S. M., & Chondough, J. T. (2022). The moderating role of intellectual capital on the relationship between artificial intelligence and employee performance of the commercial banks in Nigeria. **Cross-Cultural Management Journal**, 24(2), 115–119.
- Christiaens, T. (2023). **Digital working lives: Worker autonomy and the gig economy**. Lanham: Rowman & Littlefield.
- Collins, H. M. (2018). **Artificial intelligence: Against humanity's surrender to computers**. Chichester: John Wiley & Sons.
- Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N., & Chaintreau, A. (2020). Biased programmers? Or biased data? A field experiment in operationalizing AI ethics. In **Proceedings of the 21st ACM Conference on Economics and Computation**.
- Crawford, K. (2021). **Atlas of AI: Power, politics, and the planetary costs of artificial intelligence**. New Haven: Yale University Press.
- da Motta Veiga, S. P., Figueroa-Armijos, M., & Clark, B. B. (2023). Seeming ethical makes you attractive: Unraveling how ethical perceptions of AI in hiring impacts organizational innovativeness and attractiveness. **Journal of Business Ethics**, 186(1), 199–216.
- Danaher, J., & Nyholm, S. (2021). Automation, work and the achievement gap. **AI and Ethics**, 1(3), 227–237.
- Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. **Harvard Business Review**, 96(1), 108–116.
- Dawson, D., Schleiger, E., Horton, J., McLaughlin, J., Robinson, C., Quezada, G., ... Hajkovicz, S. (2019). **Artificial intelligence: Australia's ethics framework—a discussion paper**.
- de Vries, K. (2010). Identity, profiling algorithms and a world of ambient intelligence. **Ethics and Information Technology**, 12(1), 71–85.
- Demiaux, V. (2017). **How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence**.
- Demir, M., & Gunaydin, Y. (2023). A digital job application reference: How do social media posts affect the recruitment process? **Employee Relations: The International Journal**, 45(2), 457–477.
- Du, S., & Xie, C. (2021). Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. **Journal of Business Research**, 129, 961–974.

- Dwivedi, Y. K., et al. (2021). Setting the future of digital and social media marketing research: Perspectives and research propositions. **International Journal of Information Management**, 59.
- Ernst, E., Merola, R., & Samaan, D. (2019). Economics of artificial intelligence: Implications for the future of work. **IZA Journal of Labor Policy**, 8(1).
- Farahani, M., & Ghasemi, G. (2024). Artificial intelligence and inequality: Challenges and opportunities. **Qeios**.
- Farisco, M., Evers, K., & Salles, A. (2022). On the contribution of neuroethics to the ethics and regulation of Artificial intelligence. **Neuroethics**, 15(1), 4.
- Fenech, M., Strukelj, N., & Buston, O. (2019). **Ethical, social, and political challenges of artificial intelligence in health**.
- Floridi, L., et al. (2018). AI4People—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. **Minds and Machines**, 28, 689–707.
- Forbes. (2018). **How Amazon has reorganized around artificial intelligence and machine learning**.
- Friedman, B., & Hendry, D. G. (2019). **Value sensitive design: Shaping technology with moral imagination**. The MIT Press.
- Gaggioli, A., Riva, G., Peters, D., & Calvo, R. A. (2017). Positive technology, computing, and design: Shaping a future in which technology promotes psychological well-being. **In Emotions and affect in human factors and human-computer interaction** (pp. 477-502). Academic Press.
- Goldman, B., & Cropanzano, R. (2015). ‘Justice’ and ‘fairness’ are not the same thing. **Journal of Organizational Behavior**, 36(2), 313–318.
- Gomez-Uribe, C. A., & Hunt, N. (2015). The Netflix recommender system: Algorithms, business value, and innovation. **ACM Transactions on Management Information Systems (TMIS)**, 6(4), 1-19.
- Greguric, I. (2014). Ethical issues of human enhancement technologies: Cyborg technology as the extension of human biology. **Journal of Information, Communication and Ethics in Society**, 12(2), 133-148.
- Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. **California Management Review**, 61(4), 5–14.
- Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. **Minds and Machines**, 30, 99–120.
- Hoffmann, D., & Masucci, R. (2018). **Intel’s AI privacy policy white paper: Protecting individuals’ privacy and data in the artificial intelligence world**.
- Howard, J., & Schulte, P. (2024). Managing workplace AI risks and the future of work. **American Journal of Industrial Medicine**, 67(11), 980-993.
- Illes, J., & Bird, S. J. (2006). Neuroethics: A modern context for ethics in neuroscience. **Trends in Neurosciences**, 29(9), 511–517.
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. **Nature Machine Intelligence**, 1(9), 389–399.

- Kayyali, M. (2026). Ethical AI and automation in the workplace. In Z. Achour (Ed.), **Leading inclusive workplaces through digital transformation and organizational change**, 103-138. IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3373-5127-8.ch004>
- Kellmeyer, P., Müller, O., & Voigt, J. (2021). Embodiment, movement and agency in neuroethics. **Neuroethics**, 14, 1–3.
- Khan, M., & Ullah, I. (2024). Impact of artificial intelligence on the global economy and technology advancements. In **Advances in Intelligent Systems and Computing**, 1883, 68-79.
- Koene, A., et al. (2015). Ethics of personalized information filtering. In **Proceedings of the 2nd International Conference on Data Science and Advanced Analytics (DSAA)**. Cham: Springer. https://doi.org/10.1007/978-3-319-18609-2_10
- Koski, O., & Husso, K. (2018). **Work in the age of artificial intelligence: Four perspectives on the economy, employment, skills and ethics.**
- Kroes, P., & Verbeek, P. P. (2013). Introduction: The moral status of technical artefacts. In P. Kroes & P. P. Verbeek (Eds.), **The moral status of technical artefacts** (pp. 1-9). Springer.
- Lecher, C. (2019). How Amazon automatically tracks and fires warehouse workers for 'productivity.' **The Verge**. <https://www.theverge.com/2019/4/25/18516004/amazon-warehouse-fulfillment-centers-productivity-firing-terminations>
- Lodders, A., & Paterson, J. M. (2020). Scrutinising COVIDSafe: Frameworks for evaluating digital contact tracing technologies. **Alternative Law Journal**.
- Magnoni, I. (2022). **8 Principles of Sustainable Productivity for Long-Term Well-Being.**
- Martin, K. (2019). Designing ethical algorithms. **MIS Quarterly Executive**, 18(2), 129–142.
- Martinho, A., Herber, N., Kroesen, M., & Chorus, C. (2021). Ethical issues in focus by the autonomous vehicles industry. **Transport Reviews**, 41, 556.
- Martinho, A., Kroesen, M., & Chorus, C. (2021). A healthy debate: Exploring the views of medical doctors on the ethics of artificial intelligence. **Artificial Intelligence in Medicine**, 121, 102190.
- Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. **AI & Society**, 35(4), 957–967.
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. **Nature Machine Intelligence**, 1, 501.
- Mori, M., Sassetti, S., Cavaliere, V., & Bonti, M. (2024). A systematic literature review on artificial intelligence in recruiting and selection: A matter of ethics. **Personnel Review**.
- Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a service: A pragmatic operationalisation of AI ethics. **Minds and Machines**, 31, 239.
- 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, 2141.

- Morrow, D. R. (2014). When technologies makes good people do bad things: Another argument against the value-neutrality of technologies. **Science and Engineering Ethics**, 20, 329–343. <https://doi.org/10.1007/s11948-013-9464-1>
- Müller, V. C. (2007). Is there a future for AI without representation? **Minds and Machines**, 17(1), 101–115.
- Munn, L. (2024). Digital labor, platforms, and AI. In H. Werther, C. Ghezzi, J. Kramer, et al. (Eds.), **Introduction to Digital Humanism. A Textbook**, 557-570. Springer Nature. <https://doi.org/10.1007/978-3-031-45304-5>
- Neubert, M. J., & Montañez, G. D. (2020). Virtue as a framework for the design and use of artificial intelligence. **Business Horizons**, 63(2), 195–204.
- Pathak, Y. V., Saikia, S., & Pathak, S. (Eds.). (2023). **Ethical issues in AI for bioinformatics and cheminformatics**. CRC Press.
- Perrigo, B. (2023). OpenAI used Kenyan workers on less than \$2 per hour. **Time**. <https://time.com/6247678/openai-chatgpt-kenya-workers>
- Popescu, C. R. G., & Sebestova, J. D. (2024). The impact of artificial intelligence on intellectual capital development: Shifting requirements for professions and processes in the non-profit sector. **Journal of Infrastructure, Policy, and Development**, 8(10), 3899. <https://doi.org/10.24294/jipd.v8i10.3899>
- Reisman, D., Schultz, J., Crawford, K., et al. (2018). **Algorithmic impact assessments: A practical framework for public agency accountability**.
- Riva, G., Baños, R. M., Botella, C., Wiederhold, B. K., & Gaggioli, A. (2012). Positive technology: Using interactive technologies to promote positive functioning. **Cyberpsychology, Behavior, and Social Networking**, 15(2), 69–77.
- Roberts, H., Cowsls, J., Morley, J., Taddeo, M., Wang, V. C., & Floridi, L. (2021). The Chinese approach to artificial intelligence: An analysis of policy, ethics, and regulation. **AI & Society**, 36(1), 59–77.
- Sainato, M. (2020, February 5). I'm not a robot: Amazon workers condemn unsafe, grueling conditions at warehouse. **The Guardian**. <https://www.theguardian.com/technology/2020/feb/05/amazon-workers-protest-unsafe-grueling-conditions-warehouse>
- Santoni de Sio, F. (2024). Artificial intelligence and the future of work: Mapping the ethical issues. **The Journal of Ethics**, 28(3), 407–427.
- Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. **Frontiers in Robotics and AI**, 5. <https://doi.org/10.3389/frobt.2018.00015>
- Santoni de Sio, F., Almeida, T., & van den Hoven, J. (2021). The future of work: Freedom, justice and capital in the age of artificial intelligence. **Critical Review of International Social and Political Philosophy**, 1–25.
- Savage, N. (2019). How AI and neuroscience drive each other forwards. **Nature**, 571(7766), S15–S15.
- Siebecker, M. R. (2019). Making corporations more humane through artificial intelligence. **Journal of Corporation Law**, 45(1), 95–149.
- Smids, J., Nyholm, S., & Berkers, H. (2019). Robots in the workplace: A threat to—or opportunity for—meaningful work? **Philosophy & Technology**, 33(3), 503–522.

- Steels, L. (2007). Fifty years of AI: From symbols to embodiment—and back. In M. Lungarella, F. Iida, J. Bongard, & R. Pfeifer (Eds.), **50 years of artificial intelligence**. 18–28. Berlin, Heidelberg: Springer.
- Sutton, S. G., 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.
- Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. **Science**, 361(6404), 751-752.
- Taleb, M., & Kadhum, H. J. (2024). The role of artificial intelligence in promoting the environmental, social and governance (ESG) practices. In **International Conference on Explainable Artificial Intelligence in the Digital Sustainability**, 256–279. Springer Nature Switzerland.
- Teo, S. A. (2023). Human dignity and AI: Mapping the contours and utility of human dignity in addressing challenges presented by AI. **Law, Innovation and Technology**, 15, 1-39. <https://doi.org/10.1080/17579961.2023.2184132>
- Thornton, S. M., Pan, S., Erlien, S. M., & Gerdes, J. C. (2017). Incorporating ethical considerations into automated vehicle control. **IEEE Transactions on Intelligent Transportation Systems**, 18, 1429.
- Tiwari, R. (2024). **The impact of artificial intelligence in the workplace and its effect on the digital well-being of employees**.
- Turilli, M., & Floridi, L. (2009). The ethics of information transparency. **Ethics and Information Technology**, 11, 105–112.
- Tursunbayeva, A., Pagliari, C., Di Lauro, S., & Antonelli, G. (2022). The ethics of people analytics: Risks, opportunities and recommendations. **Personnel Review**, 51(3), 900–921. <https://doi.org/10.1108/pr-12-2019-0680>
- van de Poel, I., & Kroes, P. (2014). Can technology embody values? In P. Kroes & P. P. Verbeek (Eds.), **The moral status of technical artefacts**, 111–125. Springer.
- van Est, R., & Gerritsen, J. (2017). **Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality**. The Hague: Rathenau Institute.
- Varma, A., Dawkins, C., & Chaudhuri, K. (2023). Artificial intelligence and people management: A critical assessment through the ethical lens. **Human Resource Management Review**, 33(1), 100923.
- Verbeek, P. P. (2011). **Moralizing technology: Understanding and designing the morality of things**. University of Chicago Press.
- Verbeek, P. P. (2014). Some misunderstandings about the moral significance of technology. In P. Kroes & P. P. Verbeek (Eds.), **The moral status of technical artefacts** (pp. 71-84). Springer.
- Vidgen, R., Hindle, G., & Randolph, I. (2020). Exploring the ethical implications of business analytics with a business ethics canvas. **European Journal of Operational Research**, 281(3), 491-501.
- Villani, C., Bonnet, Y., & Rondepierre, B. (2018). **For a meaningful artificial intelligence: Towards a French and European strategy**.

- Vredenburg, K. (2022). Freedom at work: Understanding, alienation, and the AI-driven workplace. *Canadian Journal of Philosophy*, 52(1), 78-92.
- Vrousalis, N. (2019). How exploiters dominate. *Review of Social Economy*, 79(1), 103-130.
- Wallach, W., & Allen, C. (2008). **Moral machines: Teaching robots right from wrong**. Oxford University Press.
- Weh, L. (2024). An integrated embodiment concept combines neuroethics and AI ethics-relational perspectives on artificial intelligence, emerging neurotechnologies, and the future of work. *NanoEthics*, 18(2), 1-16. <https://doi.org/10.1007/s11569-024-00457-6>
- Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S. M., Richardson, R., Schultz, J., & Schwartz, O. (2018). **AI Now report 2018**. New York: **AI Now Institute at New York University**.
- Wolkenstein, A., Jox, R. J., & Friedrich, O. (2018). Brain-computer interfaces lessons to be learned from the ethics of algorithms. *Cambridge Quarterly of Healthcare Ethics*, 4, 635-646.
- Yuste, R., Goering, S., Bi, G., Carmena, J. M., et al. (2017). Four ethical priorities for neurotechnologies and AI. *Nature News*, 551(77679), 159.