فصلنامه علمی- پژوهشی اخلاق پژوهی
سال هشتم • شماره دوم • تابستان ۱۴۰۴
Quarterly Journal of Moral Studies
Vol. 8, No. 2, Summer 2025

کاربرد هوش مصنوعی در مشاورۀ اخلاقی؛ از ظرفیت‌های تحول‌آفرین تا چالش‌های اخلاقی

رضا جعفری* | حسینعلی رحمتی**

doi 10.22034/ethics.2025.51991.1790

چکیده

با گسترش و پیچیده‌تر شدن مسائل اخلاقی در دنیای امروز، نیاز به مشاورۀ اخلاقی دقیق و کارآمد بیش از پیش احساس می‌شود. چت‌بات‌های مبتنی بر هوش مصنوعی با ویژگی‌های منحصربه‌فرد، ظرفیت‌های قابل توجهی برای ارائه مشاوره جهت حلِّ مشکلات و ارتقای سطح اخلاقی جامعه دارند. در این مقاله ضمن بررسی فرصت‌ها و چالش‌های به‌کارگیری هوش مصنوعی برای مشاورۀ اخلاقی به راهکارهایی برای استفاده مسئولانه از این فناوری اشاره شده است. بنا به یافته‌های پژوهش، هوش مصنوعی مزیّت‌های فراوانی برای مشاورۀ اخلاقی دارد، از جمله: افزایش دقت و کارآمدی مشاوره، تسریع در ارائه مشاوره، شخصی‌سازی مشاوره و توسعه ابزارهای نوین برای ارائه مشاوره. با این‌حال، استفادۀ از این فناوری چالش‌هایی چون سوگیری الگوریتمی، فقدان مسئولیت‌پذیری، عدم توجه به ظرافت‌های اخلاقی و سوءاستفاده از داده‌ها را نیز در بر دارد که تدوین چارچوب‌های اخلاقی، افزایش شفافیت و نظارت، آموزش و ارتقای آگاهی و مشارکت‌دادن ذی‌نفعان را ضروری می‌سازد. البته، اعتماد کامل به مشاورۀ اخلاقی ارائه شده از سوی هوش مصنوعی درست نیست و از هوش مصنوعی فعلاً تنها می‌توان به عنوان یک ابزار بهره گرفت و اتخاذ تصمیمات نهایی باید به عهدۀ انسان باشد.

کلیدواژه‌ها

اخلاق نظری هوش مصنوعی، مشاوره اخلاقی، اخلاق هوش مصنوعی، چالش‌های اخلاقی، مسئولیت اخلاقی.

* دانشجوی دکتری اخلاق اسلامی، دانشگاه معارف اسلامی، قم، ایران. (نویسنده مسئول). | rezajafari128@gmail.com
** استادیار پژوهشگاه قرآن و حدیث، قم، ایران. | rahmati.h@riqh.ac.ir

تاریخ دریافت: ۱۴۰۴/۰۲/۱۶ ▣ تاریخ تأیید: ۱۴۰۴/۰۵/۱۲

■ جعفری، رضا؛ رحمتی، حسینعلی. (۱۴۰۴). کاربرد هوش مصنوعی در مشاورۀ اخلاقی؛ از ظرفیت‌های تحول‌آفرین تا چالش‌های اخلاقی. فصلنامه اخلاق‌پژوهی. ۸(۲۷)، ۴۹-۷۲. doi: 10.22034/ethics.2025.51991.1790

مقدمه

امروزه که زندگی بشر با چالش‌های اخلاقی روزافزون در عرصه‌های مختلف، از علم و فناوری تا تجارت و روابط اجتماعی، روبه‌روست، نیاز به اتخاذ تصمیمات آگاهانه اهمیت ویژه‌ای یافته است. درک عمیق مسائل اخلاقی، شناسایی گزینه‌ها و بررسی پیامدهای هر یک جهت اتخاذ تصمیمات اخلاقی در عصر حاضر که با حجم عظیم اطلاعات و پیچیدگی‌های فزاینده مواجهیم، چالشی مضاعف به‌حساب می‌آید.

در این میان، هوش مصنوعی فراتر از یک ابزار ساده، به عاملی قدرت‌مند در ارتقا یا تضعیف اخلاق جوامع تبدیل شده است. این فناوری نوین، ظرفیت‌ها و فرصت‌های فراوانی برای ارائهٔ مشاورهٔ اخلاقی شخصی‌سازی شده و ارتقای سطح اخلاقی در جامعه دارد. هوش مصنوعی با تحلیل داده‌های گسترده و شناسایی الگوهای پنهان، بینش‌های ارزشمندی دربارهٔ مسائل اخلاقی ارائه می‌دهد و امکان ارائهٔ مشاوره‌های دقیق‌تر، کارآمدتر و در دسترس‌تر را فراهم می‌کند.

با این حال، استفاده از هوش مصنوعی در مشاورهٔ اخلاقی بدون چالش هم نیست. چالش‌های متعددی در این زمینه وجود دارد که باید به‌دقت تحلیل و مدیریت شوند. عدم سوگیری، عدم مراعات عدالت، فقدان شفافیت و نقض حریم خصوصی از جمله این چالش‌ها هستند. (Goktas & Grzybowski, 2025, p. 1). با توجه به فراگیرشدن هوش مصنوعی و استفاده روزافزون از چت‌بات‌ها در ایران، همچنین با توجه به نقشی که این فناوری‌ها بر زیست جهان اخلاقی انسان امروز دارد، پژوهش در باب شناخت این فرصت‌ها و چالش‌ها و چاره‌جویی برای آنها امری ضروری و لازم است.

بر این اساس، مقالهٔ حاضر در صدد پاسخ به این سؤالات است که فرصت‌ها و چالش‌های استفاده از هوش مصنوعی در مشاورهٔ اخلاقی چیست؟ برای رفع یا کاهش این چالش‌ها چه باید کرد؟ و در نهایت، آیا می‌توان به چت‌بات‌ها به عنوان مشاوران اخلاقی اعتماد کرد؟

پژوهش حاضر، از نوع کیفی و با رویکردی توصیفی-تحلیلی است که بر پایهٔ روش کتابخانه‌ای به انجام رسیده است. فرایند این تحقیق در سه گام پیوسته سامان یافته است: در گام نخست، از طریق مفهوم‌شناسی «مشاوره اخلاقی» و «هوش مصنوعی»، مبانی نظری و ویژگی‌های هر یک مشخص شد. در گام دوم، با تحلیل تطبیقی ادبیات پژوهشی، «ظرفیت‌های تحول‌آفرین» و «چالش‌های بنیادین» به‌کارگیری هوش مصنوعی در جایگاه مشاور اخلاقی، استخراج و بررسی شد و در گام پایانی نیز بر اساس تحلیل این ظرفیت‌ها و چالش‌ها، راهبردهای

عملی برای استفاده مسئولانه از این فناوری ارائه شد.

یافته‌های این تحقیق می‌تواند به متخصصان اخلاق، مشاوره، سیاست‌گذاران و توسعه‌دهندگان هوش مصنوعی و عموم کاربران این فناوری در درک چالش‌های اخلاقی مرتبط با این فناوری و ارائهٔ راهکارهایی برای استفادهٔ مسئولانه از آن در مشاورهٔ اخلاقی کمک کند.

پیشینه پژوهش

به رغم گسترش پژوهش‌ها در حوزهٔ هوش مصنوعی و اخلاق، بررسی ادبیات موجود نشان می‌دهد که موضوع مشاورهٔ اخلاقی با هوش مصنوعی به طور مستقیم ـ چه در تحقیقات داخلی و چه خارجی ـ مورد توجه قرار نگرفته و خلأ پژوهشی آشکاری در این زمینه وجود دارد. پژوهش‌های مرتبط با این موضوع را می‌توان در دو دسته کلی دسته‌بندی کرد:

نخست، پژوهش‌هایی که به کاربرد هوش مصنوعی در زمینهٔ مشاورهٔ سلامتِ روان توجه کرده‌اند. به عنوان نمونه، پژوهش «مشاورهٔ سلامت روان مبتنی بر هوش مصنوعی: فرصت‌ها، چالش‌ها و پیامدهای اخلاقی» نوشتهٔ دیپتی و رنگاسوامی به پتانسیل این فناوری در بهبود دسترسی و شخصی‌سازی خدمات مشاوره‌ای اشاره دارد (Deepti & Rangaswamy, 2024). همچنین، تسگام و عبداللّهی با بررسی تاریخی، فنی و اخلاقی ابزارهای مشاوره‌گر هوش مصنوعی، در مقالهٔ «هوش مصنوعی در مشاوره سلامت روان: تاریخچه، نوآوری‌ها، اخلاق و چشم‌اندازهای آینده» بر اهمیت رویکردهای تلفیقی انسان-ماشین در افزایش اثربخشی این خدمات تأکید کرده‌اند (Tsagem & Abdullahi, 2025). پژوهش فولمر نیز در مقالهٔ «هوش مصنوعی و مشاوره: چهار سطح از پیاده‌سازی» چهار سطح از ادغام هوش مصنوعی در مشاوره را معرفی کرده و چارچوبی برای درک بهتر این روند فراهم آورده است (Fulmer, 2019). پینگ در مقالهٔ «تجربه در مشاوره روان‌شناختی با پشتیبانی فناوری هوش مصنوعی» نیز به بررسی نقش هوش مصنوعی در مشاوره روان‌شناختی پرداخته و توانایی آن در پیش‌بینی نتایج مشاوره را سنجیده است (Ping, 2024). همچنین، مقالهٔ «چالش‌های اخلاقی و حریم خصوصی در استفاده از هوش مصنوعی در راهنمایی و مشاوره» به مزیّت‌ها و چالش‌های هوش مصنوعی در مشاوره سلامت روان اشاره دارد (Yahya, 2024). در نهایت، گکتاس و گرزیبوفسکی در مقالهٔ «شکل‌دهی آیندهٔ مراقبت‌های بهداشتی: چالش‌های بالینی اخلاقی و مسیرهای دستیابی به هوش مصنوعی

قابل اعتماد» به چالش‌های اخلاقی و بالینی و مسیرهای دستیابی به هوش مصنوعی قابل اعتماد در حوزه سلامت می‌پردازند (Goktas & Grzybowski, 2025).

تفاوت این دسته از پژوهش‌ها با تحقیق حاضر در هدف و ماهیت این دو حوزه است. تمرکز آنها عمدتاً بر «بهبود حال درمانی» و «کاهش علائم روان‌شناختی» است، درحالی که این مقاله بر فرایند پیچیده «قضاوت هنجاری» و «تصمیم‌گیری در دوراهی‌های اخلاقی» تمرکز دارد.

با این‌وجود و به رغم این تفاوت ماهوی، این پژوهش‌ها از آن جهت برای ما حائز اهمیت هستند که ظرفیت‌های فنی و عملیاتی هوش مصنوعی را در یک حوزه مشاوره‌ای نشان می‌دهند. بنابراین، آثاری چون دیپتی و رنگاسوامی، تسگام و عبداللّهی و پینگ به عنوان الگوی پایه برای تحلیل امکان‌سنجی و بررسی ظرفیت‌های مشابه در مشاوره اخلاقی، مورد استناد قرار می‌گیرند.

دوم، پژوهش‌هایی که به چالش‌ها و اصول اخلاقی حاکم بر توسعه و به‌کارگیری هوش مصنوعی پرداخته‌اند. پژوهش‌های انجام شده در این زمینه در دهه‌های اخیر بسیار گسترده شده است. این پژوهش‌ها بر مفاهیمی همچون سوگیری الگوریتمی، مسئولیت‌پذیری، شفافیت، پاسخ‌گویی و انصاف تأکید داشته‌اند. آثاری؛ چون ویل و بینز (Veale & Binns, 2017)، مارتین (Martin, 2019)، غلامب (Ghallab, 2019)، بورنستاین و هاوارد (Borenstein & Howard, 2021) و خان و همکاران (Khan et al, 2022)، گاردنر (Gardner, 2022) به تحلیل این چالش‌ها پرداخته‌اند.

تفاوت اصلی این گروه از تحقیقات با مقالهٔ حاضر در سطح تحلیل است. آنها به چالش‌های اخلاقی هوش مصنوعی به صورت عام و فراگیر می‌پردازند، در حالی که پژوهش حاضر این چالش‌ها را به صورت کاربردی و متمرکز در بستر خاص مشاورهٔ اخلاقی بررسی می‌کند. به رغم این تفاوت در سطح تحلیل، این پژوهش‌ها چارچوبی مفهومی و نظری برای این مقاله به‌حساب می‌آیند. مفاهیمی چون سوگیری الگوریتمی، مسئولیت‌پذیری و شفافیت که در آثاری همچون مارتین و گاردنر تحلیل شده‌اند، در این پژوهش جهت تحلیل این مسئله به‌کار گرفته می‌شوند.

در این میان، پژوهش کونستانتینسکو و همکاران با عنوان «آیا تقصیر هوش مصنوعی است؟ دربارهٔ مسئولیت اخلاقی مشاوران اخلاقی مصنوعی»، نزدیک‌ترین پژوهش به موضوع تحقیق حاضر به‌حساب می‌آید (Constantinescu et al., 2022). این مقاله به طور خاص بر روی این پرسش متمرکز است که آیا می‌توان مشاوران اخلاقی مصنوعی را دارای مسئولیت اخلاقی دانست یا خیر. آنها با طراحی یک آزمون مسئولیت اخلاقی مبتنی بر فلسفهٔ ارسطو نتیجه می‌گیرند که هوش مصنوعی فاقد شرایط لازم برای عاملیت اخلاقی است. تفاوت اصلی آن پژوهش با مقالهٔ حاضر

در این است که تحقیق کونستانتینسکو بر چالش فلسفی مسئولیت‌پذیری تمرکز دارد، درحالی کـه پژوهش حاضر به دنبال ارائهٔ یک نگاه جامع به تمامی فرصـت‌ها و چالش‌هـای کـاربردی هـوش مصنوعی در حوزهٔ مشاورهٔ اخلاقی است.

بنابراین، ادبیات پژوهشی موجود یک خلأ آشکار را نمایان می‌سـازد. از یک‌سـو پژوهش‌هـای حوزهٔ سلامت روان به ظرفیت‌های فنی هوش مصنوعی در مشاوره پرداخته‌انـد و از سـوی دیگـر، تحقیقات اخلاقی، چالش‌های کلان این فناوری را برشمرده‌اند. آنچه در این میان مغفـول مانـده، پیوند تحلیلی این دو حوزه برای پاسخ به این پرسش است که آیـا می‌تـوان از هـوش مصـنوعی در حوزهٔ حسّاس و هنجاری «مشاورهٔ اخلاقی»، به عنوان یک مشـاور اخلاقـی کـه قـادر بـه تحلیـل دوراهی‌های پیچیدهٔ اخلاقی و ارائه قضاوت‌های هنجاری باشد بهـره بـرد؟ ایـن مقالـه بـا هـدف بررسی و پاسخ به این پرسش نگاشته شده است.

مفاهیم کلیدی

مشاورهٔ اخلاقی

هر انسانی در زندگی خود با موقعیت‌هایی مواجه است که ابعاد اخلاقی گسـترده‌ای دارنـد و نیازمنـد راهنمایی برای تصمیم‌گیری آگاهانه و اخلاقی است. در این شرایط، مشـاورهٔ اخلاقـی نقـش بسـیار مهمی ایفا می‌کند و می‌تواند به عنوان چراغـی در تـاریکی، مسـیر صـحیح را بـرای او نشـان دهـد. «مشاورهٔ اخلاقی» فرایندی تعاملی و گفت‌وگو محور است که به افـراد و گروه‌هـا کمـک می‌کنـد تـا مسائل اخلاقی را تجزیـه‌وتحلیل کننـد، گزینـه‌های مختلـف را بررسـی کننـد و بهتـرین تصـمیم را بر اساس ارزش‌ها و اصول اخلاقی خود بگیرند. این فرایند کـه توسـط متخصصـانی از حوزه‌هـایی چون فلسفه، علوم تربیتی و روان‌شناسی انجام می‌شود، متکی بر رویکردهای نظری مختلفی اسـت؛ از جمله مدل‌های مبتنی بر «اصل» (Beauchamp & Childress, 2013, p. 6)، «فضیلت» (Slote, 1992, p. 121)، «مراقبت» (نصر اصفهانی، ۱۳۹۹، ص ۸۵) و «روایت» (Winston, 2005, p. 9).

مشاورهٔ اخلاقی شامل مراحلی مشخص است که طی‌کردن آنهـا بـرای دسـتیابی بـه نتیجـه‌ای مطلوب ضروری است. ایـن مراحـل عبارت‌انـد از: ۱. برقـراری رابطـه: ایجـاد اعتمـاد و احتـرام متقابل بین مشاور و مُراجع؛ ۲. شناسایی مسئله: کمک به مُراجع برای درک دقیق و شفاف مسـئله اخلاقی؛ ۳. جمع‌آوری اطلاعات: گردآوری اطلاعات مـرتبط بـا مسـئلهٔ از طریـق گفـت‌وگو بـا

مراجع، بررسی مدارک و انجام تحقیقات؛ ۴. تجزیه‌وتحلیل گزینه‌ها: بررسی گزینه‌های پیشِ روی مُراجع و پیامدهای اخلاقی هر یک؛ ۵. تصمیم‌گیری: حمایت از مُراجع در اتخاذ تصمیمی آگاهانه و متناسب با ارزش‌ها و اصول اخلاقی خود؛ ۶. پیگیری: بررسی وضعیت مراجع پس از تصمیم‌گیری و ارائه حمایت‌های لازم در صورت نیاز (Forester-Miller & Davis, 2016, p. 2).

مشاوران اخلاقی در فرایند یاری‌رساندن به مراجعان خود با چالش‌های متعددی روبه‌رو هستند؛ از جمله پیچیدگی مسائل اخلاقی، تفاوت ارزش‌های مشاور و مُراجع، بار احساسی و فشار زمان. با این وجود، مشاورهٔ اخلاقی نقش بسیار مهمی در کمک به افراد و گروه‌ها برای مواجهه با مسائل دشوار اخلاقی و انتخاب تصمیماتی درست و مسئولانه ایفا می‌کند.

هوش مصنوعی

تعاریف متعددی از «هوش مصنوعی» ارائه شده که حول محور توانایی ماشین‌ها برای انجام وظایف مرتبط با هوش انسانی یا ایجاد برنامه‌های کامپیوتری پیچیده می‌چرخند. جان مک‌کارتی ـ پیش‌گام این حوزه ـ هوش مصنوعی را به عنوان «توانایی یک ماشین برای انجام کارهایی که اگر توسط انسان انجام می‌شد، هوشمندانه تلقی می‌شد» تعریف کرد. به گفتهٔ مک‌کارتی، هدف اصلی هوش مصنوعی توسعه ماشین‌هایی است که بتوانند وظایف را به شیوه‌ای هوشمندانه انجام دهند (McCarthy, 1995, p. 89).

این سیستم‌ها از الگوریتم‌های پیچیده‌ای برای پردازش اطلاعات، حلِّ مسائل و تصمیم‌گیری استفاده می‌کنند. هوش مصنوعی در حال حاضر نقشی حیاتی در طیف وسیعی از حوزه‌ها، از جمله پزشکی، مهندسی، امور مالی، حمل و نقل و دانش‌های نظری ایفا می‌کند.

عملکرد هوش مصنوعی بر پایه اصول زیر استوار است (Russell & Norvig, 2021, p. 20):

۱. **یادگیری:** هوش مصنوعی از طریق الگوریتم‌های یادگیری ماشینی، یادگیری عمیق و شبکه‌های عصبی مصنوعی، قادر به استخراج دانش از داده‌ها و شناسایی الگوهاست. این امر به هوش مصنوعی امکان می‌دهد تا بدون برنامه‌ریزی صریح وظایف مختلف را یاد گرفته و به طور خودکار انجام دهد.

۲. **استدلال:** هوش مصنوعی با تجزیه و تحلیل داده‌ها و استنتاج منطقی، قادر به پیش‌بینی و تصمیم‌گیری است. این توانایی، هوش مصنوعی را به ابزاری قدرت‌مند برای حلِّ مسائل پیچیده

در حوزه‌های مختلف تبدیل می‌کند (Kaplan, 2016, p. 19).

۳. حلِّ مسئله: هوش مصنوعی با شناسایی و درک مسائل، ارائـه راه‌حل‌هـا و انتخـاب بهتـرین گزینه، به انسان در حلِّ مسائل یاری می‌رساند.

هوش مصنوعی را می‌توان به دسته‌های زیر تقسیم کرد (Russell & Norvig, 2021, p. 50-51):

۱. هوش مصنوعی محدود (Narrow AI): این نوع هوش مصنوعی برای انجـام یـک وظیفـه خاص طراحی شده است، مانند تشخیص چهره، ترجمه زبان یا طراحی عکس.

۲. هوش مصنوعی عمومی (Artificial General Intelligence - AGI): ایـن نـوع هـوش مصنوعی که هنوز در مراحل اولیه توسعه خـود قـرار دارد، بـه هـوش مصـنوعی‌ای اطـلاق می‌شود که قادر به انجام طیف گسترده‌ای از وظایف است که انسان می‌تواند انجام دهد.

۳. سوپر هوش مصنوعی یـا هـوش مصـنوعی فوق‌العـاده (Superintelligence): ایـن نـوع هوش مصنوعی از هوش انسان فراتر می‌رود و در حال حاضر یک مفهوم تخیّلی است.

این فناوری به عنوان ابزاری قدرت‌مند، در حال دگرگونی حوزه‌های مختلفی از پزشکی (Rajkomar et al, 2018, p. 2) تا مهندسی (Mumali, 2022, p. 1) اسـت. هـوش مصـنوعی در قلمـرو دانش‌هـای نظری نیز جایگاه یافته و به حلِّ مسائل اخلاقـی پیچیـده، توسـعه چارچوب‌هـای اخلاقـی بـرای استفاده مسئولانه از این فناوری و کاوش در مفاهیم فلسفی و حقوقی مرتبط با آن می‌پردازد.

ظرفیت‌ها و مزیّت‌های هوش مصنوعی در مشاورهٔ اخلاقی

از جمله کاربردهای نـوین هـوش مصـنوعی در حـوزهٔ اخـلاق، اسـتفاده از چت‌بات‌هـا در ارائـهٔ مشاوره‌های اخلاقی است. این فناوری نوین قادر است با تجزیه و تحلیل حجم عظـیم داده‌هـای اخلاقی، شناسایی الگوهای پنهان و ارائه بینش‌هـای ارزشـمند، مشـاورهٔ اخلاقـی را بـه سـطحی جدید ارتقا دهد، همان‌طور که در حوزهٔ مراقبت‌های بهداشتی و مشـاورهٔ سـلامت روان مشـاهده می‌شود (Deepti & Rangaswamy, 2024, p. 1; Yahya, 2024, p. 1; Goktas & Grzybowski, 2025, p. 2).

پینگ، نشان می‌دهد که مدل‌های یادگیری ماشینی در تحلیل وضـعیت‌های روان‌شـناختی بـه دقت بالای ۸۹٪ دست می‌یابند که توانایی هوش مصنوعی در ارائهٔ خـدمات مشـاوره‌ای دقیـق را تأیید می‌کند. این سیستم‌ها بـا شبیه‌سـازی فراینـد مشـاوره انسـانی و ایجـاد پلتفرم‌هـای آنلایـن، محدودیت‌های زمانی و مکانی را کاهش داده و دسترسی افراد بـه پشـتیبانی مشـاوره‌ای را تسـهیل

می‌کنند (Ping, 2024, p. 3871).

مشاوره گرفتن از چت‌بات‌ها در زمینهٔ مسائل اخلاقی، مانند مشاوره در امور دیگر، مزیت‌ها و چالش‌هایی را به دنبال دارد که بایستی به صورت دقیق بررسی شود و نتایج آن در اختیار ذی‌نفعان، به ویژه کسانی قرار گیرد که در صددند از هوش مصنوعی برای مشاوره و تصمیم‌گیری اخلاقی استفاده کنند. ازاین‌رو، به بررسی مزیّت‌های هوش مصنوعی در ارتقای مشاورهٔ اخلاقی می‌پردازیم و به طور خاص بر چهار مورد کلیدی تمرکز می‌کنیم. لازم به ذکر است بیان این مزیت‌ها به معنای نادیده‌گرفتن چالش‌هایی نیست که در ادامهٔ مقاله به برخی از آنها اشاره می‌شود. به هرحال، برخی از ویژگی‌ها و مزیّت‌های چت‌بات‌ها در ارائهٔ مشاورهٔ اخلاقی به انسان‌ها عبارتند از:

افزایش دقت و کارآمدی مشاوره اخلاقی

هوش مصنوعی با قدرت پردازش حجم وسیع داده‌ها، می‌تواند الگوهای پیچیده‌ای را که از دید انسان پنهان می‌مانند، شناسایی کند. این قابلیت، با فراهم آوردن بینش‌های مبتنی بر داده، به مشاوران انسانی کمک می‌کند تا راهکارهای کارآمدتری ارائه دهند و کیفیت مشاوره را ارتقا بخشند (De Cremer & Narayanan, 2023, p. 4). مثلاً در مواجهه سازمان‌ها با تخلّفات اخلاقی، هوش مصنوعی می‌تواند با تحلیل داده‌های مربوطه و شبیه‌سازی پیامدها، راهکارهایی دقیق و کارآمد ارائه دهد. این سیستم‌ها قادرند سناریوهایی دقیق و شبیه‌سازی‌شده طراحی کنند تا نشان دهند تغییرات کوچک در ورودی‌ها چگونه می‌توانند منجر به نتایج تصمیم‌گیری متفاوت و پیامدهای اخلاقی گوناگون شوند و به این ترتیب، تصمیم‌گیرندگان را قادر می‌سازند تا پیامدهای ثانوی تصمیمات خود را پیش‌بینی کنند (De Cremer & Narayanan, 2023, p. 6).

هوش مصنوعی می‌تواند با در نظرگرفتن گسترهٔ وسیعی از عوامل اخلاقی مرتبط با هر موضوع، راهکارهایی مناسب برای هر فرد یا موقعیت خاص را ارائه دهد. به علاوه، هوش مصنوعی می‌تواند احتمال خطاهای انسانی را به شکل چشم‌گیری کاهش دهد. سوگیری و قضاوت نادرست که از جمله خطاهای رایج در مشاورهٔ اخلاقی سنّتی هستند، توسط هوش مصنوعی تا اندازه زیادی کنترل و پیش‌گیری می‌شوند. قابلیت تشخیص الگو در مجموعه‌داده‌های بزرگ که پیش‌تر اثربخشی آن در تشخیص زودهنگام مسائل سلامت روان به نمایش درآمده است (Deepti & Rangaswamy, 2024, p. 1)، در حوزهٔ تحلیل مسائل اخلاقی نیز می‌تواند به شناسایی ظرافت‌هایی کمک کند که از دید مشاور انسانی معمولاً پنهان می‌مانند. همچنین با

پردازش دقیق داده‌ها و استفاده از الگوریتم‌های معتبر، می‌تواند از وقوع این خطاها تا اندازه‌ای جلوگیری کرده و فرایند اتخاذ تصمیمات اخلاقی را هر چه بیشتر بی‌طرفانه و عادلانه‌تر انجام دهد.

ارائهٔ مشاورهٔ اخلاقی در لحظه

هوش مصنوعی با قابلیت ارائه مشاورهٔ اخلاقی در لحظه، تحولی عظیم در این حوزه به وجود آورده است. این قابلیت در مواقع حسّاس و فوری که نیاز به تصمیم‌گیری سریع و اخلاقی وجود دارد، از اهمیت حیاتی برخوردار است (Yahya, 2024, p. 1). ابزارهای مبتنی بر هوش مصنوعی می‌توانند به طور مداوم داده‌های مربوط به فرایند مشاوره را رصد کرده، مسائل اخلاقی را به‌سرعت شناسایی کرده و در لحظه راهکارهای مناسب را ارائه دهند. این قابلیت در مقایسه با روش‌های سنّتی مشاورهٔ اخلاقی که ممکن است زمان‌بر و مستلزم بررسی توسط متخصصان اخلاق باشد، مزیت قابل توجهی دارد (Ping, 2024, p. 3874).

سیستم‌های هوش مصنوعی می‌توانند با تعامل فعّال با افراد، مسائل اخلاقی را شناسایی کرده و راهکارهای مناسب ارائه دهند. برای مثال، چت‌بات‌ها در موقعیت‌های فوری مانند افکار خودکشی، پشتیبانی فوری فراهم کرده و افراد مرتبط را مطلع می‌سازند. این توانایی در مقایسه با روش‌های سنّتی که زمان‌بر هستند، می‌تواند حتی جان انسان‌ها را نجات دهد و نشان می‌دهد چگونه هوش مصنوعی سریع‌تر از مشاور انسانی می‌تواند راهنمایی لازم را ارائه کند. همچنین، هوش مصنوعی می‌تواند به یک فرد در مورد پیامدهای اخلاقی یک تصمیم خاص، مانند تأثیر آن بر دیگران یا نقض احتمالی قوانین اخلاقی، مشاورهٔ بی‌طرفانه و بدون قضاوت ارائه دهد (Deepti & Rangaswamy, 2024, p. 4). علاوه بر کاهش زمان، هوش مصنوعی موجب کاهش هزینه‌ها نیز می‌شود؛ برنامه‌های درمانی شناختی رفتاری مبتنی بر هوش مصنوعی مانند MoodTools و Sanvello، مراقبت را با کمتر از ۱۰٪ هزینه درمان حضوری ارائه می‌دهند و این مزیت اقتصادی، دسترسی به خدمات مشاوره‌ای را، به‌ویژه برای افراد کم برخوردار، به شدّت افزایش داده و از بروز مشکلات و تصمیم‌های غلط جلوگیری می‌کند (Tsagem & Abdullahi, 2025, p. 4).

شخصی‌سازی مشاورهٔ اخلاقی

سیستم‌های هوش مصنوعی با تحلیل داده‌های گسترده فردی (شامل سابقه و تعاملات قبلی)،

مشاورهٔ اخلاقی را شخصی‌سازی می‌کنند (Yahya, 2024, p. 1; Tsagem & Abdullahi, 2025, p. 5). این مدل‌ها با استفاده از یادگیری ماشینی، رویکرد مشاوره را با نیازهای هر فرد تطبیق می‌دهند؛ رویکردی که موفقیت آن در حوزهٔ سلامت روان برای انطباق برنامه‌های درمانی با ویژگی‌های فردی بیماران به اثبات رسیده است (Tsagem & Abdullahi, 2025, p. 5)، در حوزهٔ اخلاق نیز می‌تواند به ارائه توصیه‌هایی متناسب با نظام ارزشی هر فرد منجر شود.

تجزیه و تحلیل این داده‌ها به سیستم هوش مصنوعی امکان می‌دهد تا الگوهای رفتاری و تمایلات اخلاقی فرد را شناسایی کند و براین‌اساس، مشاورهٔ اخلاقی متناسب با او را عرضه کند. این به هوش مصنوعی امکان می‌دهد تا به دقت تشخیصی بالایی دست یابد و توصیه‌های درمانی را به صورت شخصی‌سازی شده عرضه کند (Deepti & Rangaswamy, 2024, p. 4). ارائه مشاورهٔ اخلاقی شخصی‌سازی شده با هوش مصنوعی، مزیّت‌های متعددی برای افراد و جامعه به ارمغان می‌آورد. افراد با دریافت مشاورهٔ متناسب با شرایط و موقعیت خاص خود، می‌توانند تصمیمات اخلاقی آگاهانه‌تر و مسئولانه‌تری اتخاذ کنند.

علاوه بر این، دریافت بازخورد و راهنمایی شخصی‌سازی شده، حس مسئولیت افراد را نسبت به انتخاب‌های اخلاقی خود افزایش می‌دهد و مقاومت در برابر اجرای آن را کاهش می‌دهد. تعامل با سیستم هوش مصنوعی به افراد کمک می‌کند تا ارزش‌ها، باورها و تعهدات اخلاقی خود را به طور عمیق‌تر درک کنند و به ارتقای خودشناسی اخلاقی آنها کمک می‌کند (Floridi et al, 2018, p. 691).

در مجموع، هوش مصنوعی با ارائهٔ مشاورهٔ اخلاقی شخصی‌سازی شده، ابزاری قدرتمند برای ارتقای سطح اخلاقی فرد و جامعه است. این قابلیت با توانمندسازی افراد برای اتخاذ تصمیمات اخلاقی آگاهانه‌تر و مسئولانه‌تر، به ساختن جهانی عادلانه‌تر و اخلاقی‌تر کمک می‌کند. این تحوّل نشان می‌دهد که هوش مصنوعی چگونه می‌تواند مفاهیم خودشناسی و عاملیت انسانی را تقویت کند و به انسان‌ها در فهم عمیق‌تر و استفاده از پتانسیل‌هایشان یاری رساند (Floridi et al, 2018, p. 692).

توسعهٔ ابزارهای نوین برای مشاوره اخلاقی

هوش مصنوعی با پتانسیل قابل توجهی برای توسعهٔ ابزارهای نوین در زمینه مشاوره اخلاقی، در حال دگرگونی این حوزه است. این ابزارها به افراد در شناسایی و حلّ مسائل اخلاقی در زندگی

روزمرّه، ارتقای مهارت‌های تفکر اخلاقی و ترویج فرهنگ گفت‌وگو و بحث در مورد مسائل اخلاقی کمک می‌کنند. در واقع، هوش مصنوعی می‌تواند به عنوان یک ابزار آموزشی عالی برای کمک به انسان‌ها در شناسایی نقاط کور اخلاقی نقش ایفا کند (De Cremer & Narayanan, 2023, p. 5).

نمونه‌هایی از ابزارهای نوین مشاورهٔ اخلاقی مبتنی بر هوش مصنوعی عبارتند از:

۱. چت‌بات‌های اخلاقی: این چت‌بات‌ها قادرند به افراد در مورد مسائل اخلاقی روزمرّه، مانند نحوه برخورد با دروغ گفتن، تقلب یا فریب، مشاوره ارائه دهند. آنها با پرسش و ارائهٔ اطلاعات، افراد را در اتخاذ تصمیمات اخلاقی آگاهانه یاری می‌رسانند. نخستین کاربردهای هوش مصنوعی در سلامت روان نیز با چت‌بات‌هایی مانند ELIZA در دههٔ ۱۹۶۰ آغاز شد که با شبیه‌سازی روان‌درمانی، توهّم درک را ایجاد می‌کردند؛ هرچند بدون درک واقعی. مدل‌های مدرن‌تر مانند Woebot نیز کاربران را در گفت‌وگوهای درمانی درگیر می‌کنند (Tsagem & Abdullahi, 2025, p. 2).

۲. سیستم‌های تشخیص اخلاقی: این سیستم‌ها قادر به شناسایی اقدامات یا تصمیمات غیراخلاقی در محیط‌های مختلف، مانند تجارت، سیاست یا مراقبت‌های بهداشتی، هستند. این امر به افراد و سازمان‌ها کمک می‌کند تا از رفتارهای غیراخلاقی جلوگیری کرده و به تعهدات اخلاقی خود عمل کنند. به عنوان مثال، در حوزهٔ سلامت روان، ابزارهای تشخیصی مبتنی بر هوش مصنوعی می‌توانند با تحلیل مجموعه‌داده‌های بزرگ، الگوهایی را شناسایی کنند که به تشخیص زودهنگام و دقیق‌تر اختلالات کمک کنند (Deepti & Rangaswamy, 2024, p. 5).

۳. شبیه‌سازی‌های تعاملی: هوش مصنوعی مشاور با شبیه‌سازی سناریوهای اخلاقی مختلف در محیطی امن، به افراد کمک می‌کند تا پیامدهای انتخاب‌هایشان را تجربه کرده و مهارت‌های تفکر اخلاقی خود را ارتقا دهند (Green, 2018, p. 19). فناوری واقعیت مجازی نیز می‌تواند برای شبیه‌سازی اثرات تصمیمات، پیش از انتخاب نهایی، به کار رود (Ping, 2024, p. 3873). این قابلیت می‌تواند تصمیم‌گیرندگان را قادر سازد تا پیش از اتخاذ یک انتخاب خاص، اثرات پایین‌دستی تصمیمات مختلف ممکن را شبیه‌سازی کنند (De Cremer & Narayanan, 2023, p. 6).

این موضوع در موارد تصمیم‌گیری‌های خطیر اخلاقی از قبیل دوراهه‌ها یا تزاحم‌های اخلاقی بسیار کارساز است (برای آگاهی بیشتر، نک: رحمتی، ۱۴۰۱). برای مثال، در شرایطی که فرد میان دو گزینهٔ اخلاقی مانند کمک به دو گروه نیازمند مختلف مردد است، هوش مصنوعی می‌تواند پیامدهای هر یک از وضعیت‌ها را طراحی کند و به فرد ارائه کند تا او تصمیم خود را بگیرد و از

تزاحم نجات پیدا کند.

هوش مصنوعی با ارائهٔ ابزارهای نوین مشاورهٔ اخلاقی، می‌تواند دسترسی به این گونه خدمات را برای افراد در سراسر جهان ممکن می‌سازد (Yahya, 2024, p. 1). این ابزارها به لطف دسترسی آسان از طریق اینترنت یا برنامه‌های تلفن همراه، امکان دریافت مشاورهٔ اخلاقی در هر زمان و مکانی را فراهم می‌آورد (Ping, 2024, p. 3873). این امر به‌ویژه برای افرادی که به دلیل موقعیت جغرافیایی، ناتوانی جسمی، فقر مادی و مانند آن، به مشاورهٔ اخلاقی سنّتی دسترسی ندارند، بسیار کارآمد است. برای مثال، سازمان جهانی بهداشت از کمبود جهانی ۲. ۵ میلیون متخصص سلامت روان خبر می‌دهد و چت‌بات‌های هوش مصنوعی می‌توانند این کمبود را جبران کنند (Tsagem & Abdullahi, 2025, p. 4). از این جهت، استفاده از هوش مصنوعی برای مشاوره به نوعی برابری در جامعه می‌انجامد و شکاف‌های موجود در دسترسی به مشاورهٔ اخلاقی را از بین می‌برد یا کاهش می‌دهد.

چالش‌های پیشِ روی هوش مصنوعی در مشاورهٔ اخلاقی

به رغم مزیّت‌های هوش مصنوعی در مشاورهٔ اخلاقی، این فناوری با چالش‌های اخلاقی متعددی مواجه است که باید پیش از استفاده گسترده از آن، به طور کامل بررسی شوند (Megdad et al, 2024, p. 1; Green, 2018, p. 12). همان‌طور که اندیشمندان نیز هشدار داده‌اند، هوش مصنوعی پتانسیل ایجاد چالش‌های بزرگی را دارد و ازاین‌رو، تدابیر ایمنی شامل افزایش آگاهی و درک عمیق‌تر از خطرات، چالش‌ها و تأثیرات کوتاه‌مدت و بلندمدت توسعه آن ضروری است (Fulmer, 2019, p. 1).

در این بخش، به برخی از چالش‌های کلیدی استفاده از هوش مصنوعی در مشاورهٔ اخلاقی می‌پردازیم. این چالش‌ها شامل سوگیری الگوریتمی، فقدان مسئولیت‌پذیری، عدم توجه به ظرافت‌های اخلاقی و خطرات سوءاستفاده از هوش مصنوعی می‌شود. بررسی دقیق هر یک از این چالش‌ها برای توسعه چارچوب‌های اخلاقی مناسب و تضمین استفاده مسئولانه از هوش مصنوعی در مشاورهٔ اخلاقی ضروری است.

۱. سوگیری الگوریتمی

یکی از بزرگ‌ترین چالش‌های اخلاقی استفاده از هوش مصنوعی در مشاوره اخلاقی، احتمال وجود

سوگیری در الگوریتم‌های هوش مصنوعی است (Yahya, 2024, p. 1; Megdad et al, 2024, p. 1). این سوگیری‌ها می‌توانند تأثیرات منفی قابل توجهی بر عدالت و انصاف در ارائه مشاورهٔ اخلاقی داشته باشند (Goktas & Grzybowski, 2025, p. 7; لیندار، ۱۳۹۷، ص ۸۰). همان‌طور که مارتین تأکید می‌کند، الگوریتم‌ها خنثی نیستند، بلکه دارای ارزش هستند و پیامدهای اخلاقی ایجاد می‌کنند، اصول اخلاقی را تقویت یا تضعیف می‌کنند و حقوق و کرامت ذی‌نفعان را محقّق یا محدود می‌سازند (Martin, 2019, p. 835).

سوگیری در هوش مصنوعی به معنای تمایل الگوریتم‌ها به قضاوت یا تصمیم‌گیری به نفع یا علیه یک گروه خاص از افراد بر اساس ویژگی‌های غیرمرتبط است (لیندار، ۱۳۹۷، ص ۷۶). این سوگیری‌ها می‌توانند ناشی از عوامل مختلفی، مانند سوگیری‌های موجود در داده‌های آموزشی که الگوریتم‌ها بر روی آنها آموزش دیده‌اند (سوگیری داده‌ها)، الگوریتم‌های معیوب یا فرض‌های نادرست در مورد نحوه عملکرد الگوریتم‌ها (سوگیری الگوریتمی)، یا حتی سوگیری‌های ناخواسته توسعه‌دهندگان و ذی‌نفعان انسانی (سوگیری انسانی) باشند (Megdad et al, 2024, p. 1; Veale & Binns, 2017, p. 4). به عبارت دیگر، هوش مصنوعی آینه‌ای است که سوگیری‌ها، الگوهای تبعیض‌آمیز و نقص‌های اخلاقی عمیقاً ریشه‌دار در شناخت انسانی و نهادهای اجتماعی ما را بازتاب می‌دهد؛ بنابراین، شکایت از سوگیری‌های هوش مصنوعی، مانند شکایت از تصویر خودمان در آینه است! (De Cremer & Narayanan, 2023, p. 4). این سوگیری‌ها می‌توانند منجر به مشاوره‌های ناعادلانه، تبعیض و کاهش اعتماد به سیستم‌های مشاورهٔ اخلاقی مبتنی بر هوش مصنوعی شوند.

بونستاین و هاوارد نیز هشدار می‌دهند که اگر یک الگوریتم توسط انسان ساخته می‌شود، انسان‌ها اشتباه می‌کنند، از جمله در هنگام طراحی، برنامه‌نویسی، کالیبراسیون و ارزیابی عملکرد الگوریتم (Borenstein & Howard, 2021, p. 61). آنها مثالی ملموس می‌آورند که یک سیستم هوش مصنوعی مورد استفاده برای توصیه خدمات مراقبت‌های بهداشتی، بیماران سیاه‌پوست را با نرخ کمتری نسبت به همتایان سفیدپوستشان ارجاع داده است (Borenstein & Howard, 2021, p. 62).

برای مثال، چت‌بات مشاوری که الگوریتم‌هایش توسط یک مهندس با تمایل به رویکرد پیامدگرایانه طراحی شده باشد، ممکن است پاسخ‌ها و راهکارهایی که ارائه می‌دهد بیشتر ناظر به پیامدها باشد؛ درحالی که اگر همان مهندس قائل به اخلاق فضیلت یا وظیفه‌گرایی بود، چه‌بسا پاسخ‌های متفاوتی به سؤال‌ها و مشکلات یکسان بدهد. این نشان می‌دهد ارزش‌ها و باورهای

توسعه‌دهندگان انسانی به طور ناخواسته در طراحی الگوریتم‌ها و در نتیجه در خروجی‌های مشاوره‌ای آنها تعبیه می‌شوند (Borenstein & Howard, 2021, p. 63; Martin, 2019, p. 838). سوگیری نهفته، چالشی کلیدی در گنجاندن اخلاق در هوش مصنوعی است (Pant et al, 2024, p. 17).

۲. فقدان مسئولیت‌پذیری

یکی از مهم‌ترین چالش‌ها در استفاده از هوش مصنوعی برای مشاوره اخلاقی، فقدان مسئولیت‌پذیری شفاف است؛ زیرا به وقتِ بروز خطا یا ارائه مشاورهٔ نادرست توسط هوش مصنوعی و وارد شدن ضرر مادی یا معنوی به مراجعه‌کننده، تعیین فرد یا گروه مسئول برای پاسخ‌گویی به عواقب منفی دشوار است (Goktas & Grzybowski, 2025, p. 11). برای مثال، اگر فردی پس از مشاوره با یک چت‌بات در نهایت، دست به خودکشی بزند، چه کسی مسئول است؟ آیا توسعه‌دهندگان سیستم، ارائه‌دهندگان خدمات، کاربران یا افراد دیگری باید مسئول باشند؟ این ابهام که از آن به عنوان «سردرگمی در انتساب مسئولیت» در ادبیات اخلاق هوش مصنوعی یاد شده است (Constantinescu et al, 2022, p. 2; Floridi et al, 2018, p. 692)، در حوزهٔ مشاورهٔ اخلاقی به دلیل پیامدهای مستقیم تصمیمات بر زندگی افراد، به مراتب حادتر می‌شود. همان‌طور که مارتین بیان می‌کند، اگر الگوریتمی طراحی شود که افراد را از پذیرش مسئولیت در یک تصمیم بازدارد، طراح الگوریتم باید در قبال پیامدهای اخلاقی الگوریتم در حال استفاده پاسخ‌گو باشد (Martin, 2019, p. 835). این مسئله به پیچیدگی‌ای به نام «سردرگمی در انتساب مسئولیت» می‌انجامد. تمایل به انسان‌انگاری هوش مصنوعی می‌تواند این تصور غلط را ایجاد کند که خود سیستم مسئول است و در نتیجه، مسئولیت‌پذیری عاملان انسانی در هاله‌ای از ابهام قرار گیرد و کارکنان از پیامدهای اخلاقی کار خود فاصله بگیرند (De Cremer & Narayanan, 2023, pp. 3-4). مثلاً در مشاورهٔ اخلاقی ژنتیک پزشکی، اگر هوش مصنوعی به مطالعه‌ای تأیید دهد که منجر به آسیب یا نقض حریم خصوصی شود، قابلیت ردیابی آسیب و پراکندگی مسئولیت، چالش‌برانگیز است (Goktas & Grzybowski, 2025, p. 11). پراکندگی مسئولیت می‌تواند پیامدهای ناگواری همچون عدم جبران خسارت آسیب‌دیدگان، شناسایی ناقص آسیب و ریشه‌های آن و فرسایش احتمالی اعتماد اجتماعی به چنین فناوری‌هایی را در پی داشته باشد؛ زمانی که به نظر می‌رسد نه توسعه‌دهندگان و نه کاربران نمی‌توانند پاسخ‌گو باشند (Goktas & Grzybowski, 2025, p. 11).

این چالش پیچیده‌ای است که نیازمند بررسی دقیق و تدوین چارچوب‌های قانونی و اخلاقی شفاف برای تعیین مسئولیت در زمینهٔ استفاده از هوش مصنوعی برای ارائهٔ مشاورهٔ اخلاقی است.

۳. عدم توجه به ظرافت‌های اخلاقی

استفاده از هوش مصنوعی در مشاورهٔ اخلاقی می‌تواند خطرات بالقوه‌ای برای تضعیف توجه به ظرافت‌ها و پیچیدگی‌های قضاوت اخلاقی داشته باشد. چالش‌ها و مسائل اخلاقی، به دلیل پیچیدگی و متنوع بودن علائم آنها، همچنین به خاطر چندوجهی بودن شخصیت انسان‌ها، ممکن است الگوریتم‌های هوش مصنوعی را در تشخیص دقیق آنها محدود کند. به عبارت دیگر، ممکن است جنبه‌های انسانی و عواطف و احساسات و موقعیت‌های خاصی که فرد در آن بوده است و در تشخیص ابعاد مسئله نقش داشته، توسط این الگوریتم‌ها نادیده گرفته شود که این می‌تواند به تشخیص نادرست یا حتی ایجاد تشخیص‌های مثبت یا منفی کاذب منجر شود (Deepti & Rangaswamy, 2024, p. 2). همچنین فناوری‌هایی که مانع انتقال سیگنال‌های روان‌شناختی و احساسات می‌شوند، ممکن است درک متخصص از وضعیت مراجع را مختل کند (Goktas & Grzybowski, 2025, p. 3).

سیستم مشاورهٔ اخلاقی مبتنی بر هوش مصنوعی با بررسی داده‌های مربوط به پرونده‌های مشابه، اصول اخلاقی پزشکی و اطلاعات موجود، ممکن است پیشنهاد کند که بر اساس وصیت‌نامه پزشکی بیمار، درمان متوقف شود، اما چیزی که سیستم قادر به درک آن نیست، پیچیدگی‌های فرهنگی، عاطفی و روان‌شناختی این موقعیت خاص است که خانواده بیمار با آن مواجه است و این تصمیم را دچار چالش می‌کند. این امر به دلیل ناتوانی سیستم‌های هوش مصنوعی در درک کامل حالات عاطفی پیچیده و ظرافت‌هایی است که ممکن است منجر به تشخیص نادرست یا مداخلات درمانی نامناسب شود (Tsagem & Abdullahi, 2025, p. 5).

علاوه بر این، پیچیدگی مسائل اخلاقی و گاهی همپوشانی مصداقی برخی آنها سبب می‌شود که الگوریتم‌های هوش مصنوعی نتوانند به طور کامل با این تنوع سازگار شوند. این سیستم‌ها ممکن است درک کاملی از این موضوع نداشته باشند و از توانایی‌های لازم برای تفسیر صحیح این تنوع و پیچیدگی‌ها برخوردار نباشند (Walsh et al, 2020, p. 12). برای مثال، خطر تفسیر نادرست کنایه‌ها یا نشانه‌های غیرکلامی که به عنوان یک چالش فنی در چت‌بات‌های سلامت

روان نیز شناسایی شـده اسـت (Tsagem & Abdullahi, 2025, p. 6)، در حـوزۀ مشـاورۀ اخلاقـی بـه چالشی اساسی تبدیل می‌شود.

به‌عنوان مثال، در فضای آموزشی، یک سیستم مشاوره مبتنی بر هوش مصنوعی ممکـن اسـت بر اساس عملکرد تحصیلی یک دانش‌آموز و داده‌های مرتبط، پیشنهاد دهد که وضعیت تحصیلی او تغییر یابد یا تغییر رشته دهد. با این‌حال، این سیستم قادر بـه درک جزئیـات وضـعیت روحـی، مشکلات خانوادگی، فشارهای اجتماعی و انگیزه‌های شخصی دانش‌آموز نیست (Constantinescu et al, 2022, p. 10) که ممکن است تأثیر قابل توجهی بـر عملکـرد و وضـعیت تحصـیلی او داشـته باشند. در چنین شرایطی، یک مشاور انسانی می‌توانـد بـا بررسـی دقیـق‌تر شـرایط زنـدگی فـرد، گفت‌وگو با او و خـانواده او و همچنـین ارزیـابی عوامـل روان‌شـناختی و اجتمـاعی، توصـیه‌های مناسب‌تر و دقیق‌تری ارائه دهد.

به تعبیر دیگر، تقلیل اخلاقیات به یک مجموعه از قوانین و محاسبات مکانیکی توسـط هـوش مصنوعی، می‌تواند به بی‌حسی یـا بی‌اطلاعـی یـا بی‌تفـاوتی نسـبت بـه برخـی عوامـل مهـم در تصمیم‌گیری منجر شود و سبب شود که سیستم‌های هوش مصنوعی، بدون در نظرگرفتن شـرایط خاص و تجربیات فردی، تصـمیم‌گیری کننـد. ایـن امـر می‌توانـد سـبب ارزیابی‌هـای اخلاقـی نادرست یا غیردقیق و در نتیجه تصمیم‌گیری‌های نادرست شود (Green, 2018, p. 19).

۴. سوءاستفاده از هوش مصنوعی جهت اهداف غیر اخلاقی

هوش مصنوعی ابزار قدرت‌مندی است که احتمال سوءاسـتفاده از آن بـرای اهـداف غیراخلاقـی وجود دارد (Green, 2018, p. 13). در زمینۀ مشاوره اخلاقی، این خطر به‌ویژه نگران‌کننده است؛ زیرا هوش مصنوعی می‌تواند برای توجیه یا تسهیل اقدامات غیراخلاقی مورداستفاده قرار گیرد (Russell & Norvig, 2021, p. 23). فناوری‌های هوش مصنوعی آسیب‌پذیری انسـان‌ها را در برابـر دسـتکاری رفتار افزایش می‌دهند (Ghallab, 2019, p. 4).

هوش مصنوعی می‌تواند توصیه‌های مغرضانه ارائـه دهـد یـا در اسـتخدام منجـر بـه تبعـیض ناخودآگاه شود (McGraw, 2024, p. 8). همان طور که ابزارهای اسـتخدام آمـازون و الگوریتم‌هـای گوگل، سوگیری جنسیتی نشـان داده‌انـد (Pant et al, 2024, p. 2). گکتـاس و گرزیبوفسـکی نیـز بـه مثال‌های ملموس‌تری از سوءاستفاده از هوش مصنوعی مولّد در سلامت اشاره می‌کنند، از جمله جعـل سـوابق پزشـکی، ارائـۀ اطلاعـات نادرسـت در توصـیه‌های پزشـکی، تقویـت سـوگیری

الگوریتمی (Goktas & Grzybowski, 2025, p. 3).

همچنین در مشاورهٔ اخلاقی با استفاده از سیستم‌های هوش مصنوعی، یکی از موارد حیاتی توجه به دسترسی و امنیت داده‌هاست. این سیستم‌ها به طور معمول نیازمند دسترسی به اطلاعات حسّاس و شخصی کاربران، از جمله سوابق شناختی، روانی و مجموعه‌ای از داده‌های مربوط به مشکلات و نیازهای افراد هستند (Tsagem & Abdullahi, 2025, p. 3; Yahya, 2024, p. 1; Megdad et al, 2024, p. 2). ازاین‌رو، حفاظت از این داده‌ها برای احترام به حریم خصوصی بیماران و جلوگیری از دسترسی یا سوءاستفاده غیرمجاز بسیار حائز اهمیت است (Deepti & Rangaswamy, 2024, p. 2; Yahya, 2024, p. 1; Fulmer, 2019, p. 9).

راهکارهای استفادهٔ مسئولانه

همان‌طور که پیش‌تر گفته شد، هوش مصنوعی در مشاورهٔ اخلاقی هم مزیّت‌ها و هم مشکلات فراوانی دارد؛ به طوری که هیچ یک را نمی‌توان نادیده گرفت. به علاوه که می‌دانیم با توجه به روند رو به گسترش استفاده از هوش مصنوعی در کشور ما (و دیگر کشورهای جهان) خواه ناخواه مردم از آنها برای مشاوره کمک می‌گیرند. ازاین‌رو، رویکردی واقع‌بینانه حکم می‌کند که برای بهره‌گیری مسئولانه از این فناوری، راهکارهایی جهت رفع یا کاهش چالش‌های آن ارائه شود. در این بخش، به بررسی راهکارهای غلبه بر این چالش‌ها و استفاده صحیح از هوش مصنوعی در مشاورهٔ اخلاقی می‌پردازیم.

۱. استفاده از داده‌های متنوع در آموزش

برای بهبود کیفیت و قابلیت اطمینان مشاوره‌های هوش مصنوعی و کاهش سوگیری‌ها، آموزش و ارزیابی دقیق این سیستم‌ها با استفاده از داده‌های متنوع ضروری است. این مجموعه داده‌ها باید نمایانگر گستره‌ای از جوامع، زبان‌ها و فرهنگ‌های مختلف باشند (Megdad et al, 2024, p. 2). در این راستا، اطمینان از تعادل و کامل بودن مجموعه داده‌ها از اهمیت بالایی برخوردار است و باید از نمونه‌گیری کافی از افرادی که این داده‌ها را ارائه می‌دهند، اطمینان حاصل شود (Tsagem & Abdullahi, 2025, p. 6; Goktas & Grzybowski, 2025, p. 7).

برای اطمینان از شمولیت مجموعه داده و درنظرگرفتن پیچیدگی‌های هویتی افراد، لازم است

تا عوامل مختلفی همچون نژاد، جنسیت و سن در نظر گرفته شـود (Koutsouleris et al, 2022, p. 834). این تنوع در داده‌ها می‌تواند به کاهش سوگیری‌ها کمک کند (Tsagem & Abdullahi, 2025, p. 4).

تحقیقات باید عملکرد الگوریتم‌های هوش مصنوعی را در گروه‌های مختلف جمعیتی بررسـی و تأیید کنند تا تفاوت‌ها و تنوع‌های احتمالی مورد توجـه قـرار گیـرد (Koutsouleris et al, 2022, p. 834). در این راستا، بررسی دقیق و منظم عملکرد الگوریتم‌ها و به‌روزرسانی آنهـا بـا در دسـترس بـودن داده‌های جدیـد از اهمیـت چشـم‌گیری برخـوردار اسـت. همچنـین داده‌هـای اخلاقـی هـوش مصنوعی باید بسیار متعدد و متنـوع باشـد. بـه عنـوان مثـال، همـهٔ نظریـه‌های اخلاقـی از قبیـل وظیفه‌گرایی، پیامدگرایی و فضیلت‌گرایی، همچنین نظریه‌های اخلاقی دینی را پوشش دهد.

۲. تدوین چارچوب‌های اخلاقی

تدوین و اجـرای چارچوب‌هـای اخلاقـی جـامع و قـوانین مناسـب، گـامی اساسـی در اسـتفاده مسئولانه از هوش مصنوعی در مشاوره اسـت (Floridi et al, 2018, p. 698; Megdad et al, 2024, p. 3). این اصول باید بر پایه ارزش‌های بنیادی مانند عدالت، انصاف، شفافیت، پاسخ‌گویی و احترام بـه حریم خصوصی و نیز حقوق و آزادی‌های افراد بنا شوند (Gardner, 2022, p. 85).

در کنار تدوین اصول و قوانین، تدوین استانداردهای فنی بـرای اطمینـان از کیفیـت، امنیـت و قابلیت اطمینان سیستم‌های هوش مصنوعی مورداستفاده در مشاورهٔ اخلاقی نیز از اهمیت بالایی برخوردار است. این استانداردها باید شامل الزامـات مربـوط بـه جمـع‌آوری و پـردازش داده‌هـا، الگوریتم‌های تصمیم‌گیری و نظارت بر سیستم‌ها باشند. مقرراتـی ماننـد GDPR اتحادیـه اروپـا نمونه‌هایی از تلاش‌ها برای تنظیم حریم خصوصی داده‌ها هستند. چارچوب‌های حکمرانی قـوی برای تقویت پذیرش و پیاده‌سازی موفق هوش مصنوعی در کاربردها ضروری هسـتند و ایـن امـر نیازمند پشـتیبانی‌های اخلاقـی قـوی، مقـررات پیشـگیرانه و همکـاری مـداوم اسـت (Goktas & Grzybowski, 2025, p. 4).

۳. افزایش شفافیت و نظارت

شفافیت و نظارت، دو اصل بنیادین برای انجام مسئولانه فعالیت‌های مرتبط با هوش مصنوعی در زمینـه مشـاورهٔ اخلاقـی هسـتند (Megdad et al, 2024, p. 2-3; Khan et al, 2022, p. 5). شـفافیت در

عملکرد الگوریتم‌های هـوش مصـنوعی بـه معنـای امکـان درک، ارزیـابی و بررسـی فراینـدهای تصمیم‌گیری آنهاست، به‌طوری‌کـه کـاربران و ذی‌نفعـان بتواننـد نحـوه پـردازش داده‌هـا، منطـق خروجی‌ها و معیارهای به‌کاررفته را بفهمند (McGraw, 2024, p. 6). این امر به افـراد و سـازمان‌های مورد نظر اجازه می‌دهد تا این الگوریتم‌ها را به طور دقیق مورد بررسی قرار دهند و به آنهـا اعتمـاد کنند. این اعتماد سبب افزایش اطمینان و اعتمادبه‌نفس در اسـتفاده از هـوش مصـنوعی می‌شـود، به‌ویژه در زمینهٔ مشاوره اخلاقی.

همچنین، به موازات شفافیت، نظارت مستقل بر سیستم‌های هـوش مصـنوعی امـری حیـاتی است. این نظارت باید شامل نظارت مداوم بر عملکرد سیستم‌ها، شناسایی و رفـع سـوگیری‌های بالقوه و اطمینان از انطباق آنها با قوانین و مقررات مربوط به مشاورهٔ اخلاقی باشـد. ایـن نظـارت، به‌منظور حفظ ارزش‌های اخلاقی و تضمین مصلحت عمومی، بسیار اهمیت دارد و باید به طـور مداوم اجرا شود (McGraw, 2024, p. 10).

تدوین استانداردهای بـین المللـی بـرای اسـتفادهٔ مسـئولانه از هـوش مصـنوعی در مشـاوره اخلاقی، گامی ضروری برای تضمین رویکردی هماهنگ و یکپارچـه در سـطح جهـانی اسـت. استانداردهای بین‌المللی باید شامل اصول اخلاقی برای توسعه و استفاده از هـوش مصـنوعی در مشاوره اخلاقـی، الزامـات شـفافیت و نظـارت، دسـتورالعمل‌هایی بـرای کـاهش سـوگیری‌ها و خطرات سوءاستفاده و رویه‌هایی برای تضمین عدالت و انصاف در استفاده از این فناوری باشـند. با تدوین این استانداردها، می‌توان از بروز چالش‌های ناشـی از رویکردهـای متفـاوت در سـطح بین المللی جلوگیری کرد و اطمینان حاصل کرد که هوش مصـنوعی در راسـتای منـافع مشـترک همه کشورهابه‌کار گرفته می‌شود. تشکیل نهادهای نظارتی مستقل برای تضـمین اسـتفاده ایمـن، قابل اعتماد و اخلاقی از هوش مصنوعی در مشاوره ضروری است. این نهادها مسئول نظارت بـر انطباق با قوانین، شناسایی و رفـع خطـرات و ارتقـای اسـتفاده مسـئولانه هسـتند. همچنـین بایـد ساز وکارهایی برای جبران خسارت‌های احتمالی به مراجعین تعیین کنند (Gardner, 2022, p. 87).

۴. ارتقای آگاهی و آموزش

برای استفادهٔ مسئولانه از هوش مصنوعی در مشاوره اخلاقی، ارتقای آگـاهی عمـومی در مـورد ایـن فناوری و چالش‌هـای اخلاقـی مـرتبط بـا آن ضـروری اسـت. ایـن امـر شـامل آمـوزش در مـورد

سوگیری‌های الگوریتمی، سوء استفاده از اطلاعات مراجعه کننده و اهمیت شفافیت و نظارت می‌شود (Borenstein & Howard, 2021, p. 62). با آموزش عموم مردم، می‌توانیم درک و اعتماد آنها را به هوش مصنوعی در زمینهٔ مشاورهٔ اخلاقی افزایش دهیم و از سوءاستفاده احتمالی از این فناوری قدرت‌مند جلوگیری کنیم. همچنین بایستی به کاربران هشدار و آگاهی داد که هیچ‌گاه نظرات هوش مصنوعی را صد در صد درست نپندارند و تا حد امکان آنها را راستی‌آزمایی کنند و دست‌کم، در مقام مشورت جستن به یک چت‌بات اکتفا نکنند و با چند هوش مصنوعی مشورت کنند.

علاوه بر آگاهی‌رسانی عمومی، آموزش متخصصان و کاربران و دیگر ذی‌نفعان نیز از اهمیت بالایی برخوردار است. متخصصان درگیر در توسعه و استفاده از هوش مصنوعی در مشاورهٔ اخلاقی باید آموزش‌های لازم را در مورد اخلاق هوش مصنوعی، قوانین و مقررات مربوطه و بهترین شیوه‌ها دریافت کنند (Fulmer, 2019, p. 9). این امر به آنها کمک می‌کند تا سیستم‌های هوش مصنوعی را به طور مسئولانه و اخلاقی توسعه دهند.

۵. مشارکت‌دادن ذی‌نفعان در فرایند توسعه و طراحی هوش مصنوعی مشاور

مشارکت فعّال ذی‌نفعان مختلف، از جمله متخصصان اخلاق، کاربران سیستم‌ها و توسعه‌دهندگان هوش مصنوعی، در فرایند توسعه و استفاده از هوش مصنوعی در مشاوره اخلاقی، امری ضروری است (Floridi et al, 2018, p. 698; Megdad et al, 2024, p. 3). این امر به تضمین در نظر گرفتن دیدگاه‌ها و تجربیات مختلف در هنگام طراحی و پیاده‌سازی سیستم‌های هوش مصنوعی کمک می‌کند و منجر به راه‌حل‌هایی جامع و همه‌جانبه برای چالش‌های پیش‌گفته خواهد شد.

جمع‌آوری بازخورد از ذی‌نفعان برای شناسایی نقاط ضعف و قوّت هوش مصنوعی و توسعه سیستم‌های کارآمدتر و اخلاقی‌تر ضروری است. سازمان‌هایی؛ چون کمیسیون اروپا، OECD و UNESCO نیز مشارکت عمومی را برای توسعه مسئولانه هوش مصنوعی، حیاتی می‌دانند (Machado, Silva, & Neiva, 2025, p. 2).

توسعهٔ راهکارهای مناسب برای استفاده مسئولانه از هوش مصنوعی در مشاوره اخلاقی، مستلزم همکاری بین‌رشته‌ای بین متخصصان در زمینه‌های مختلف مانند هوش مصنوعی، اخلاق، روان‌شناسی، تعلیم و تربیت، حقوق، علوم اجتماعی و سیاست است. این هم‌افزایی دانش، منجر به راه‌حل‌های نوآورانه و کارآمدتر برای ارتقای سطح اخلاق در جامعه خواهد شد.

نتیجه‌گیری

انسان امروز با مشکلات و مسائل روزافزون اخلاقی روبه‌رو است که برای حلِّ آنها نیازمند مشاوره است و هوش مصنوعی به مثابه ابزاری نوظهور و قدرت‌مند، افق‌های تازه‌ای را برای ارتقای سطح مشاورهٔ اخلاقی جامعه می‌گشاید. این فناوری نوین با ارائهٔ راه‌حل‌های متنوع و خلاقانه، ظرفیت قابل توجهی برای دگرگونی حوزه مشاورهٔ اخلاقی دارد و نویدبخش تحولات عمیقی در این زمینه است. مزیّت‌های هوش مصنوعی در مشاورهٔ اخلاقی، گستره‌ای وسیع را در بر می‌گیرد. از جملهٔ این مزیّت‌ها می‌توان به ارائه مشاوره دقیق‌تر و سریع‌تر با تحلیل حجم عظیمی از داده‌های اخلاقی، ارائه مشاوره در لحظه در شرایط حسّاس و فوری، شخصی‌سازی مشاوره با در نظر گرفتن شرایط، ارزش‌ها و باورهای هر فرد، توسعه ابزارهای نوین مانند چت‌بات‌های اخلاقی و سیستم‌های تشخیص اخلاقی و ارتقای شفافیت و عدالت در فرایند مشاوره اشاره کرد. این مزایا که برخی از آنها مشابه دستاوردهای هوش مصنوعی در حوزه مشاورهٔ سلامت روان است، پتانسیل تحوّل در دسترسی به راهنمایی‌های اخلاقی را دارند.

با وجود این مزیّت‌های چشمگیر، چالش‌هایی نیز در مسیر استفاده از هوش مصنوعی در مشاورهٔ اخلاقی وجود دارد که باید به طور جدی مورد توجه قرار گیرند. سوگیری الگوریتمی، فقدان مسئولیت‌پذیری در قبال خطاها یا ارائه مشاوره نادرست، عدم توجه به ظرافت‌های اخلاقی در اتکای بیش از حد به این فناوری و خطرات سوءاستفاده از هوش مصنوعی برای اهداف غیراخلاقی، از جملهٔ این چالش‌ها هستند. این چالش‌ها به دلیل ماهیت هنجاری و پیامدهای تصمیم‌گیری در حوزهٔ اخلاق، ابعاد پیچیده‌تری نسبت به دیگر حوزه‌های کاربردی هوش مصنوعی پیدا می‌کنند.

غلبه بر این چالش‌ها نیازمند یک رویکرد چندوجهی و برنامه‌ریزی دقیق است که شامل تدوین چارچوب‌های اخلاقی و فنی، افزایش شفافیت و نظارت بر فرایند توسعه و عملکرد الگوریتم‌های هوش مصنوعی، ارتقای آگاهی عمومی و تضمین مشارکت فعّال همه ذی‌نفعان در این حوزه می‌شود.

نکته نهایی اینکه گرچه راهکارهای استفادهٔ مسئولانه از هوش مصنوعی در مشاورهٔ اخلاقی بیان شد، اما با توجه به چالش‌هایی که گفته شد به نظر می‌رسد که نباید مسئولیت ارائه مشاورهٔ اخلاقی را به‌تنهایی بر عهدهٔ هوش مصنوعی بگذاریم و بلکه با توسعهٔ آن، می‌توان آن را به عنوان یک ابزار کارآمد در دست مشاوران اخلاقی جهت تقویت مشاوره‌ها معرفی کرد.

در نهایت، بهره‌برداری از فرصت‌های هوش مصنوعی در مشاورهٔ اخلاقی، نیازمند آگاهی از چالش‌ها و همکاری مستمر میان جامعهٔ علمی، توسعه‌دهندگان و سیاست‌گذاران است. تنها از این طریق می‌توان اطمینان حاصل کرد که این ابزار قدرت‌مند در جهت منافع عمومی و به شیوه‌ای مسئولانه به‌کار گرفته می‌شود.

فهرست منابع

رحمتی، حسینعلی. (۱۴۰۱). تزاحم اخلاقی در شبکه‌های اجتماعی؛ از چالش تا تصمیم. قم: پژوهشگاه قرآن و حدیث.

ریچلز، جیمز. (۱۳۹۶). عناصر فلسفهٔ اخلاق. (ترجمه: محمود فتحعلی و علیرضا آل‌بویه). قم: پژوهشگاه علوم و فرهنگ اسلامی.

لیندا الدر؛ ریچارد پل. (۱۳۹۷). آشنایی با مبانی استدلال اخلاقی بر اساس مفاهیم و اصول تفکر انتقادی. (ترجمه: پیام یزدانی). تهران: نشر اختران.

نصر اصفهانی، مریم. (۱۳۹۹). پروای دیگران؛ درآمدی بر فلسفهٔ اخلاق مراقبت. تهران: پژوهشگاه علوم انسانی و مطالعات فرهنگی.

Beauchamp, T. L., & Childress, J. F. (2013). *Principles of biomedical ethics* (7th ed.). New York: Oxford University Press.

Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics*, *1*(1), 61-65. https://doi.org/10. 1007/s43681-020-00010-0

Constantinescu, M., et al. (2022). Blame it on the AI? On the moral responsibility of artificial moral advisors. *Philosophy & Technology*, *35*(2), Article 35. https://doi.org/10. 1007/s13347-022-00539-7

De Cremer, D., & Narayanan, D. (2023). How AI tools can—and cannot—help organizations become more ethical. *Frontiers in Artificial Intelligence*, *6*, 1-11. https://doi.org/10. 3389/frai. 2023. 1093712

Deepti, S., & Rangaswamy, R. (2024). AI-driven mental health counseling: Opportunities, challenges, and ethical implications. *Revista Electronica De Veterinaria*, *25*(1S), 550-558.

Dirican, C. (2015). The impacts of robotics, artificial intelligence on business and economics. *Procedia - Social and Behavioral Sciences*, *195*, 564-573.

https://doi.org/10. 1016/j. sbspro. 2015. 06. 134

Floridi, L., et al. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10. 1007/s11023-018-9482-5

Forester-Miller, H., & Davis, T. E. (2016). *Practitioner's guide to ethical decision making*. Alexandria, VA: American Counseling Association.

Fulmer, R. (2019). Artificial intelligence and counseling: Four levels of implementation. *Theory & Psychology*, *29*(6), 807–819. https://doi.org/10. 1177/0959354319853045

Gardner, A. (2022). Responsibility, recourse, and redress: A focus on the three R's of AI ethics. *IEEE Technology and Society Magazine*, *41*(2), 84–89. https://doi.org/10. 1109/MTS. 2022. 3173342

Ghallab, M. (2019). Responsible AI: requirements and challenges. *AI Perspectives*, *1*(1), 1-7.

Goktas, P., & Grzybowski, A. (2025). Shaping the future of healthcare: Ethical clinical challenges and pathways to trustworthy AI. *Journal of Clinical Medicine*, *14*(5), 1605. https://doi.org/10. 3390/jcm14051605

Green, B. P. (2018). Ethical reflections on artificial intelligence. *Studia Humana*, *7*(2), 9-31. https://doi.org/10. 2478/sh-2018-0008

Kaplan, J. (2016). *Artificial intelligence: What everyone needs to know*. New York: Oxford University Press.

Khan, A. A., et al. (2022). Ethics of AI: A systematic literature review of principles and challenges. In *Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering (EASE '22)* (pp. 383-392). New York: Association for Computing Machinery. https://doi.org/10. 1145/3530019.3531329

Koutsouleris, N., et al. (2022). From promise to practice: towards the realisation of AI-informed mental health care. *The Lancet Digital Health*, *4*(11), e829-e840. https://doi.org/10. 1016/S2589-7500(22)00152-2

Machado, H., et al. (2025). Publics' views on ethical challenges of artificial intelligence: A scoping review. *AI and Ethics*, *5*(1), 139-167. https://doi.org/10. 1007/s43681-023-00366-6

Martin, K. (2019). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, *160*(4), 835–850. https://doi.org/10. 1007/s10551-018-3921-6

McCarthy, J. (1995). Making robots conscious of their mental states. In D. Michie & S. Muggleton (Eds.), *Machine intelligence 15* (pp. 89-96). Oxford: Oxford University Press.

McGraw, D. K. (2024). Ethical responsibility in the design of artificial intelligence (AI) systems. *International Journal on Responsibility*, *7*(1), Article 4. https://doi.org/10. 62365/2576-0955. 1114

Megdad, M. M., et al. (2024). Ethics in AI: Balancing innovation and responsibility. *International Journal of Academic Pedagogical Research (IJAPR)*, *8*(9), 20-25.

Mumali, F. (2022). Artificial neural network-based decision support systems in manufacturing processes: A systematic literature review. *Computers & Industrial Engineering*, *165*, 107964. https://doi.org/10. 1016/j. cie. 2022. 107964

Pant, A., et al. (2024). Ethics in the age of AI: An analysis of AI practitioners' awareness and challenges. *ACM Transactions on Software Engineering and Methodology*, *33*(3), 1-35. https://doi.org/10. 1145/3635715

Ping, Y. (2024). Experience in psychological counseling supported by artificial intelligence technology. *Technology and Health Care*, *32*(S1), 3871-3888. https://doi.org/10. 3233/THC-230809

Rajkomar, A., et al. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, *1*(18), 1-10. https://doi.org/10. 1038/s41746-018-0029-1

Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). London: Pearson.

Slote, M. (1992). *From morality to virtue*. New York: Oxford University Press.

Tsagem, S. Y., & Abdullahi, R. (2025, June). *Artificial intelligence in mental health counseling: History, innovations, ethics, and future prospects*. Paper presented at the 9th Annual National Conference of the Faculty of Education and Extension Services, Usmanu Danfodiyo University, Sokoto.

Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, *4*(2), 1-18. https://doi.org/10. 1177/2053951717743530

Walsh, C. G., et al. (2020). Stigma, biomarkers, and algorithmic bias: recommendations for precision behavioral health with artificial intelligence. *JAMIA Open*, *3*(1), 9–15. https://doi.org/10. 1093/jamiaopen/ooz054

Winston, J. (2005). *Drama, narrative and moral education*. London: Routledge.

Yahya, F. (2024). Ethical and privacy challenges in using AI in guidance & counselling. *BICC Proceedings*, *2*, 238–244.