



فصلنامه علمی پژوهشی اخلاق پژوهی

سال نهم • شماره اول • بهار ۱۴۰۵

Quarterly Journal of Moral Studies  
Vol. 9, No. 1, Spring 2026



## فرا تحلیل ادبیات پژوهشی اخلاقِ هوش مصنوعی در ایران: روندها، خلأهای پژوهشی و جهت گیری های آینده

یوسف خجیر\* ID یعقوب مرادیان\* ID

doi 10.22034/ethics.2026.52312.1853

### چکیده

پژوهش حاضر با هدف ترسیم سیمای ادبیات پژوهشی «اخلاق هوش مصنوعی» در ایران و شناسایی روندها و خلأهای آن انجام شد. بدین منظور از رویکرد کیفی فرا تحلیل ادبیات استفاده شد. مقالات علمی پژوهشی مرتبط پس از جست و جوی نظام مند در پایگاه های داخلی و اعمال معیارهای ورود و خروج، مطابق منطق رهنمودهای پریسما غربال و در نهایت ۲۵ مقاله برای تحلیل انتخاب شد. این مقالات بر اساس چارچوبی چندبعدی با استفاده از آمار توصیفی و تحلیل محتوای مقایسه ای بررسی شدند. یافته ها نشان داد ادبیات ایرانی اخلاق هوش مصنوعی پدیده ای نوظهور و میان رشته ای است که عمدتاً پس از سال ۱۴۰۰ شکل گرفته و در مرز فلسفه و الهیات، سلامت، ارتباطات و رسانه و حوزه های حرفه ای تولید می شود و مضامینی مانند حریم خصوصی و حفاظت داده، سوگیری و تبعیض الگوریتمی، عدالت، شفافیت، کرامت و خودآیینی انسان برجسته شده اند. در جمع بندی، بر ضرورت گذار از بحث های مفهومی به پژوهش های میدانی بومی شده و تدوین چارچوب ها و استانداردهای اجرایی برای حکمرانی اخلاقی هوش مصنوعی در ایران تأکید دارد.

### کلیدواژه ها

هوش مصنوعی، اخلاق، فرا تحلیل ادبیات پژوهشی، اخلاقِ هوش مصنوعی، پژوهش های اخلاقی.

\* دانشجوی دکتری علوم ارتباطات، دانشگاه بین المللی سوره، تهران، ایران. (نویسنده مسئول) | y.moradian@soore.ac.ir

\*\* دانشیار دانشکده فرهنگ و ارتباطات دانشگاه بین المللی سوره، تهران، ایران. | khojir@soore.ac.ir

تاریخ دریافت: ۱۴۰۳/۹/۱۴ □ تاریخ تأیید: ۱۴۰۵/۰۲/۲۸

■ مرادیان، یعقوب؛ خجیر، یوسف. (۱۴۰۵). فرا تحلیل ادبیات پژوهشی اخلاقِ هوش مصنوعی در ایران: روندها، خلأهای پژوهشی و جهت گیری های آینده. فصلنامه اخلاق پژوهی. ۹(۳۰)، ۱۲۸-۱۰۵. doi: 10.22034/ethics.2026.52312.1853

رشد شتابان فناوری‌های هوش مصنوعی (هومص) در دهه‌های اخیر، عرصه‌های اجتماعی، فرهنگی، اقتصادی و سیاسی را به‌طور گسترده تحت تأثیر قرار داده و شیوه‌های تصمیم‌گیری و کنش انسانی را در سطوح فردی و ساختاری دگرگون کرده است. هومص دیگر صرفاً یک توانمندی فنی یا ابزاری نیست، بلکه به‌مثابه زیرساخت پنهان بسیاری از فرایندهای روزمره - از الگوریتم‌های توصیه‌گر و نظام‌های رتبه‌بندی تا سامانه‌های پیش‌بینی، نظارت و تصمیم‌یار - در حال بازتعریف مرز میان «انسان» و «ماشین» است. این تحولات، پرسش‌های بنیادی فزاینده‌ای را درباره عدالت، شفافیت، پاسخ‌گویی، حریم خصوصی، کرامت انسانی و حدود مسئولیت اخلاقی بازیگران مختلف از طراحان و توسعه‌دهندگان و پلتفرم‌ها تا نهادهای حاکمیتی و کاربران برانگیخته و اخلاقی هوش مصنوعی (آهومص) را به یکی از محورهای کانونی در مطالعات معاصر فناوری و جامعه بدل کرده است.



در سطح جهانی، ادبیات مربوط به آهومص به سرعت رو به گسترش است و هر روز اسناد سیاستی، چارچوب‌ها و دستورالعمل‌های متعددی برای حکمرانی و تنظیم‌گری اخلاقی این فناوری پیشنهاد می‌شود. با این حال، انتقال و به‌کارگیری این چارچوب‌ها در سطح ملی صرفاً امری فنی یا حقوقی نیست و به‌شدت به زمینه‌های فرهنگی، ارزشی و نهادی هر جامعه وابسته است. در ایران نیز طی سال‌های اخیر، مجموعه‌ای از پژوهش‌ها در پیوند با آهومص در حوزه‌های مختلف از ارتباطات و رسانه و علوم اجتماعی گرفته تا حقوق، فلسفه فناوری، علوم کامپیوتر و سیاست‌گذاری علم و فناوری شکل گرفته است. این بدنه پژوهشی به تدریج، رو به گسترش است و می‌کوشد ابعاد گوناگون پیامدهای اخلاقی هومص را در بستر ایرانی صورت‌بندی کند، با این حال بیشتر این مطالعات به‌صورت پراکنده و ناپیوسته و در قالب بررسی‌های نظری منفرد، مطالعات موردی یا تحلیل‌های موضوعی جداگانه منتشر شده است.

چنین پراکندگی مفهومی و روشی خود به مسئله‌ای چالش برانگیز تبدیل شده است، در یک حوزه نوظهور و میان‌رشته‌ای مانند آهومص، تنوع مفهومی و روش‌شناختی میان پژوهش‌ها امری طبیعی است. با این حال، وقتی این تنوع بدون یک چارچوب حداقلی مشترک برای تعریف مفاهیم، طبقه‌بندی حوزه‌های کاربردی و گزارش روش پژوهش پیش می‌رود، امکان مقایسه نظام‌مند نتایج، جمع‌بندی روندهای غالب و شناسایی حوزه‌های کمتر مطالعه‌شده کاهش می‌یابد. از همین رو، این پژوهش به‌جای انتظار همسانی کامل، بر ضرورت ترسیم یک تصویر

کلان و قابل اتکا از وضعیت موجود ادبیات داخلی تأکید می‌کند.

افزون بر این، نبود یک جمع‌بندی نظام‌مند از وضعیت موجود ادبیات آهومص در ایران، امکان اتکای سیاست‌گذاران، نهادهای تنظیم‌گر، انجمن‌های علمی و جامعه دانشگاهی و حوزوی بر یک تصویر مستند و مبتنی بر شواهد از این حوزه را محدود کرده است. هر پژوهش عمدتاً در افق و چارچوب خاص خود به شناسایی و اولویت‌بندی مسائل اخلاقی مرتبط با هومص می‌پردازد و پیوند آن با سایر مطالعات به‌ندرت آشکار می‌شود. در چنین شرایطی، شکاف میان سطح گفتمان علمی و سطح تصمیم‌سازی و سیاست‌گذاری پابرجا می‌ماند؛ در حالی که بحث درباره تنظیم‌گری هومص، تدوین اسناد بالادستی و طراحی چارچوب‌های بومی در حوزه اخلاق فناوری در دستور کار نهادهای مختلف قرار گرفته است (بشیر و راوی، ۱۴۰۴، ص ۱۲). و غیبت یک تصویر کلان، منسجم و فراتحلیلی از ادبیات موجود در عمل به خلأی معرفتی و سیاستی تبدیل می‌شود.

بر این اساس، پرسش اصلی پژوهش حاضر این است که ادبیات پژوهشی «آهومص» در ایران، در مجموع، چه تصویر کلی و چه الگوهای غالبی را نشان می‌دهد و بر پایه این تصویر، مهم‌ترین روندها و خلأهای پژوهشی کدام است؟ برای پاسخ به این پرسش، پژوهش حاضر نخست سیمای کتاب‌شناختی این ادبیات را از حیث روند زمانی انتشار، نوع مدارک و توزیع آنها در رشته‌ها و مجلات علمی توصیف می‌کند؛ دوم تمرکز موضوعی و زمینه‌های کاربردی مورد توجه پژوهش‌ها (از جمله سلامت، آموزش، رسانه و کسب‌وکارهای دیجیتال) را مشخص می‌سازد و حوزه‌های کمتر بررسی‌شده را نشان می‌دهد؛ سوم مضامین و مسائل اخلاقی پرتکرار-مانند حریم خصوصی و حفاظت داده، سوگیری و تبعیض الگوریتمی، عدالت، شفافیت، مسئولیت‌پذیری، اعتماد و کرامت انسانی - را استخراج و الگوی توجه به آنها را تبیین می‌کند؛ و چهارم ویژگی‌های روش‌شناختی مطالعات (نوع مطالعه، شیوه گردآوری و تحلیل داده‌ها و سطح تحلیل) را جمع‌بندی می‌کند. در نهایت، این فراتحلیل می‌کوشد با اتکا به شواهد همین مطالعات، دلالت‌هایی روشن‌تر برای جهت‌گیری پژوهش‌های آینده و نیز برای تصمیم‌سازی و سیاست‌گذاری در حوزه حکمرانی اخلاقی هوش مصنوعی در ایران ارائه کند، و نسبت این دلالت‌ها را با کانون‌های رایج در ادبیات جهانی آهومص به‌صورت تطبیقی روشن سازد.

### پیشینه پژوهش

پژوهش‌های ایرانی درباره «آهومص» در سال‌های اخیر با شتاب قابل توجهی گسترش یافته و از



سطح بحث‌های صرفاً مفهومی به سمت مسائل نهادی و حوزه‌های کاربردی حرکت کرده است. در سطح کلان حکمرانی و سیاست‌گذاری، شماری از مطالعات با رویکرد مرور نظام‌مند یا مرورهای تحلیلی، به بررسی چارچوب‌ها و سازوکارهای حکمرانی اخلاقی هومص پرداخته‌اند و نشان داده‌اند که سامان‌دهی اخلاقی این فناوری در ایران نیازمند پیوند میان مبانی هنجاری، سازوکارهای تنظیم‌گری و الزامات نهادی است (بشیر و راوی، ۱۴۰۴). هم‌زمان، برخی پژوهش‌ها بر «چارچوب‌سازی» و صورت‌بندی الگوهای مطلوب اخلاقی تأکید کرده‌اند و کوشیده‌اند نسبت میان آهومص را از منظرهای فلسفی-هنجاری یا با اتکا بر سنت‌های فکری اسلامی تبیین کنند (مهدوی نور و آرام، ۱۴۰۳؛ میری بالاجورشری و محمودی، ۱۴۰۳؛ هامانی و مختاری، ۱۴۰۴). در امتداد همین رویکردهای مبنایی، مطالعاتی درباره عاملیت اخلاقی ماشین‌ها و معیارهای آن، و نیز امکان و حدود «قضاوت اخلاقی» در سامانه‌های هوشمند طرح شده که نشان می‌دهد بخش مهمی از ادبیات داخلی همچنان دغدغه‌های بنیادین فلسفه اخلاق و فلسفه فناوری را دنبال می‌کند (حسینی سورکی، ۱۴۰۳؛ عاشوری کیسمی و پرویزی، ۱۴۰۳؛ جعفری، ۱۴۰۳).

در سطح حوزه‌های کاربردی، سلامت یکی از پررنگ‌ترین میدان‌های پژوهشی است. مطالعاتی در اخلاق پزشکی، چالش‌های اخلاقی به‌کارگیری هومص در نظام سلامت را از منظرهایی مانند شأن و کرامت بیمار، اتونومی، عدالت، و پیامدهای انسانی تصمیم‌سازی الگوریتمی بررسی کرده‌اند (عباسی و تیموری، ۱۴۰۲؛ بزمی، افشار و موسوی‌نژاد، ۱۴۰۴؛ شفیعیان و منعمی، ۱۴۰۴؛ زاهدپاشا و همکاران، ۱۴۰۴). در کنار سلامت، حوزه‌های رسانه و فرهنگ نیز سهم قابل‌اعتنایی دارند؛ از یک سو مسئله پیامدهای اخلاقی هومص در رسانه‌های نوین و مخاطرات آن صورت‌بندی شده (کلاته آقامحمدی، ۱۴۰۲؛ سیروس کیری و موسوی، ۱۴۰۲). و از سوی دیگر، کاربردهای فرهنگی-هنری و مسائل اخلاقی مرتبط با تولید و گردش محتوای دیجیتال موضوع پژوهش قرار گرفته است (عاشوری کیسمی، ۱۴۰۳).

در حوزه‌های سازمانی و اقتصادی، مطالعاتی به آهومص در حسابداری و حسابرسی، شرایط استفاده از خدمات شرکت‌های مبتنی بر هومص، و چالش‌های اخلاقی اتخاذ هومص در مدیریت منابع انسانی پرداخته‌اند؛ محور مشترک این آثار، دغدغه‌هایی مانند مسئولیت‌پذیری، شفافیت تصمیم‌گیری، عدالت، و مدیریت ریسک‌های مبتنی بر داده است (پاینده و کتعیانی، ۱۴۰۳؛ رحیمی‌اقدم، صالحپور و نامور، ۱۴۰۳). در حوزه آموزش نیز پژوهش‌ها از سطح کاربردهای عمومی به سمت صورت‌بندی «الگوهای اخلاق‌مدارانه» و نیز تحلیل فرااخلاق در محیط‌های یادگیری

مبتنی بر هومص حرکت کرده‌اند (زارعی، ۱۴۰۴؛ مرادی مخلص و همکاران، ۱۴۰۴). در نهایت، بخشی از ادبیات جدید به مسائل تجربه‌محورتر مانند رابطه عاطفی انسان و هومص و پیامدهای اخلاقی آن توجه نشان داده و نشان می‌دهد که کانون‌های پژوهشی در ایران - به تدریج - به سمت مسائل زندگی روزمره و مناسبات انسان-ماشین گسترش یافته است (بابایی، صابری و بحرینی، ۱۴۰۳؛ زرگر، ۱۴۰۴). این بدنه پژوهشی، در مجموع، از یک سو بیانگر رشد میان‌رشته‌ای و گسترش حوزه‌های کاریست است و از سوی دیگر، نشان می‌دهد هنوز نیاز به تقویت پیوند میان چارچوب‌های هنجاری، شواهد تجربی و سازوکارهای اجرایی حکمرانی اخلاقی هومص در ایران وجود دارد.

## روش پژوهش

این پژوهش از نظر هدف، کاربردی و از نظر ماهیت و روش، یک مطالعه مروری نظام‌مند با رویکرد «فراتحلیل محتوای کیفی<sup>۱</sup>» است. واژه فراتحلیل، معادل فارسی واژه مرکب «Meta Analysis» است. واژه «Meta» ریشه یونانی به معنای «فرا» و «ورا» و به مفهوم قرار دادن اجزاء و قسمت‌ها در درون ترکیب‌ها جهت نشان دادن عوامل مورد پژوهش می‌باشد. واژه «Analysis» نیز به معنای «تحلیل» و عبارت است از شکستن کل به متغیرهای مختلف به منظور تعیین ماهیت آن می‌باشد. در روش فراتحلیل این دو عمل با هم انجام می‌گیرد تا نتایج کلی به دست آید. (خنفر، ۱۳۹۷، ص ۴۹۵). اصل اساسی و عملی در فراتحلیل عبارت است از ترکیب نتایج پژوهش‌های مختلف و استخراج نتایج جدید و منسجم و حذف آنچه که موجب سوگیری در نتایج نهایی می‌شود (قاضی طباطبایی، ۱۳۸۹، ص ۱۰۱) غالباً در روش فراتحلیل از یافته‌های پژوهش‌های مرتبط و گذشته به عنوان داده جهت «فراتحلیل» استفاده می‌شود و روش جمع‌آوری آنها در حقیقت احصاء پژوهش‌های مرتبط با موضوع و تمرکز بر یافته‌های آنهاست؛ «روشی که در آن داده‌های مورد استفاده، یافته‌های پژوهش‌های پیشین در یک عرصه خاص است». (خنفر، ۱۳۹۷، ص ۵۰۳).

با توجه به اینکه ادبیات موجود در حوزه اخلاق هوش مصنوعی در ایران عمدتاً شامل مطالعات نظری، تحلیلی و چارچوب‌سازی‌های مفهومی است و کمتر به سمت اندازه‌گیری‌های کمی همگن حرکت کرده است، استفاده از فراتحلیل آماری<sup>۲</sup> امکان‌پذیر نبوده است. بنابراین،

1. Qualitative Meta-Synthesis
2. Quantitative Meta-Analysis



پژوهش حاضر با بهره‌گیری از منطق فراتحلیل کیفی، به دنبال استخراج، ترکیب و بازسازی معنایی یافته‌های مطالعات پیشین برای دستیابی به بینشی جامع‌تر و غنی‌تر از وضعیت کنونی این حوزه است. بدین منظور، پروتکل گزارش‌دهی بر اساس رهنمودهای سیستماتیک مرورها<sup>۱</sup> تنظیم گردید تا شفافیت، تکرارپذیری و قابلیت اعتماد نتایج تضمین شود.

## ۱. جامعه پژوهش و معیارهای ورود و خروج

جامعه آماری این پژوهش را کلیه مقالات علمی-پژوهشی فارسی‌زبان منتشرشده در نشریات معتبر داخلی تشکیل می‌دهد که مستقیماً یا ضمنی به بررسی ابعاد اخلاقی هوش مصنوعی پرداخته‌اند. دوره زمانی جست‌وجو از آغاز نمایه‌سازی پایگاه‌های داده تا پایان سال ۱۴۰۴ در نظر گرفته شد، اما با توجه به نوظهور بودن موضوع، تحلیل نهایی بر مقالات منتشرشده از سال ۱۴۰۲ متمرکز گردید تا تصویر تازه‌تری از آخرین روندها ارائه دهد.

معیارهای ورود و خروج مطالعات به شرح زیر تعریف و اعمال شدند:

معیارهای ورود: (۱) انتشار در مجلات علمی-پژوهشی دارای رتبه وزارت علوم یا جهاد دانشگاهی؛ (۲) دسترسی کامل به متن مقاله؛ (۳) تمرکز صریح بر مفاهیم اخلاقی، هنجاری یا حکمرانی در بستر فناوری هوش مصنوعی؛ (۴) زبان فارسی.

معیارهای خروج: (۱) مقالات غیرپژوهشی (یادداشت، سرمقاله، گزارش همایش، کتاب‌نامه)؛ (۲) مقالاتی که صرفاً به جنبه‌های فنی یا مهندسی هوش مصنوعی پرداخته و بعد اخلاقی آنها مغفول مانده است؛ (۳) متونی که پیوند مشخصی میان اخلاق و هوش مصنوعی برقرار نمی‌کنند.

## ۲. استراتژی جست‌وجو و غربالگری

جستجوی نظام‌مند در پنج پایگاه اطلاعاتی معتبر ایرانی شامل پایگاه استادی علوم جهان اسلام (ISC)، پایگاه اطلاعات علمی جهاد دانشگاهی<sup>۲</sup>، مگیران<sup>۳</sup>، نورمگز<sup>۴</sup> و سیویلیکا انجام گرفت.

1. PRISMA
2. SID
3. Magiran
4. Noormags

همچنین برای اطمینان از عدم مغفول ماندن منابع، جست‌وجو در «گوگل اسکالر» تکمیل شد. واژگان جست‌وجو شامل ترکیبی از کلیدواژه‌های اصلی و مترادف بود: «هوش مصنوعی»، «سامانه‌های هوشمند»، «یادگیری ماشین»، «اخلاق»، «اخلاق فناوری»، «حکمرانی اخلاقی»، «مسئولیت‌پذیری»، «حریم خصوصی». جست‌وجو در فیلدهای عنوان، چکیده و کلمات کلیدی محدود شد.

در این جست‌وجو، ترکیب‌های مختلفی از کلیدواژه‌های «هوش مصنوعی»، «سامانه‌های هوشمند»، «اخلاق»، «اخلاق فناوری» و «اخلاق هوش مصنوعی» در عنوان، چکیده و کلیدواژه‌ها به کار رفت. دوره زمانی جست‌وجو از آغاز نمایه‌سازی هر پایگاه تا ۱۴۰۴ در نظر گرفته شد، اما در مرحله اعمال معیارهای ورود، تنها مقالات منتشر شده از سال ۱۴۰۲ به بعد در نمونه نهایی باقی ماندند.

فرآیند غربالگری طی سه مرحله انجام پذیرفت:

مرحله اول (حذف تکراری‌ها): پس از بازیابی ۳۶۵۲ رکورد، نرم‌افزار EndNote و بررسی دستی منجر به حذف ۳۲۶۰ مورد تکراری شد و تعداد مطالعات یکتا به ۳۹۲ رسید.

مرحله دوم (غربالگری بر اساس عنوان و چکیده): عناوین و چکیده‌ها بر اساس معیارهای ورود بررسی شدند. مقالاتی که ارتباط مستقیم با محور «اخلاق در هوش مصنوعی» نداشتند، حذف گردیدند. در این مرحله ۳۴۵ مقاله حذف و ۴۷ مقاله برای بررسی متن کامل انتخاب شدند.

مرحله سوم (ارزیابی متن کامل و صلاحیت): متن کامل ۴۷ مقاله مطالعه شد. مقالاتی که فاقد عمق تحلیلی کافی بودند، یا تمرکز اصلی آنها خارج از حیطه پژوهش بود، حذف گردیدند. در نهایت، ۲۵ مقاله واجد شرایط وارد فراتحلیل شدند (جدول ۱).

جدول ۱. مراحل غربالگری و معیارهای ورود و خروج مطالعات

| شرح معیار                                             | نوع معیار       |
|-------------------------------------------------------|-----------------|
|                                                       | ■ معیارهای ورود |
| مقاله علمی- پژوهشی یا پژوهشی داور شده در نشریات معتبر | نوع مدرک        |



|                   |                                                                                         |
|-------------------|-----------------------------------------------------------------------------------------|
| زبان و وابستگی    | نگارش به زبان فارسی با وابستگی سازمانی ایرانی                                           |
| موضوع             | تمرکز صریح بر یکی از ابعاد اخلاقی مرتبط با کاربردهای هومص<br>(در عنوان، چکیده یا متن)   |
| دوره زمانی        | انتشار از ابتدای سال ۱۴۰۲ تا ۱۴۰۴                                                       |
| دسترسی            | دسترسی معتبر به متن کامل مقاله برای کدگذاری و تحلیل                                     |
| ■ معیارهای خروج   |                                                                                         |
| عدم ارتباط موضوعی | مقالاتی که فقط به «هومص» یا فقط به «اخلاق» می پرداختند و پیوند مشخص میان این دو نداشتند |
| نوع مدرک          | یادداشت، سرمقاله، گزارش همایش، مقاله ترویجی، مرور غیرداوری شده و متون غیر پژوهشی        |
| تکراری بودن       | رکوردهای تکراری میان پایگاه‌های مختلف                                                   |
| دسترس ناپذیری متن | مقالاتی که به‌رغم ارتباط ظاهری عنوان، متن کامل آنها به‌طور معتبر در دسترس نبود          |

برای استخراج داده‌ها، فرم‌های کدگذاری بر اساس چارچوب مفهومی پژوهش طراحی شد و هر مقاله در پنج محور اصلی (شاخص‌های کتاب‌سنجی و زمینه نهادی، حوزه‌های موضوعی و زمینه‌های کاربردی هومص، دسته‌های مسائل و اصول اخلاقی، ویژگی‌های روش‌شناختی و دلالت‌ها و خلأهای پژوهشی) به‌طور نظام‌مند کدگذاری شد. کدگذاری با بازبینی مکرر و مقایسه درونی مقولات همراه بود تا انسجام و دقت حفظ شود. در نهایت، داده‌های به‌دست آمده در قالب جداول تحلیلی تجمیع و با استفاده از آمار توصیفی و تحلیل محتوای مقایسه‌ای، مبنای تفسیر و بازسازی سیمای کلی ادبیات «آهومص» در ایران قرار گرفت.

در فرایند آماده‌سازی این مقاله، از ابزارهای هوش مصنوعی صرفاً در حد کمک‌های فنی و غیرمحتوایی، مانند ویرایش نگارشی، یکدست‌سازی رسم‌الخط و بهبود روانی متن استفاده شده است. تمامی مراحل علمی پژوهش، از جمله طراحی مطالعه، جست‌وجو و غربال‌گری منابع،

استخراج و کدگذاری داده‌ها، تحلیل و تفسیر یافته‌ها، و نگارش بخش‌های تحلیلی، نتیجه‌گیری و پیشنهادها توسط نویسندگان انجام شده است. نویسندگان مسئولیت علمی و اخلاقی محتوای مقاله، دقت استنادها و صحت نتایج و منابع را بر عهده دارند و هیچ بخش علمی مقاله به‌صورت خودکار و بدون ارزیابی انسانی در متن نهایی درج نشده است.

## یافته‌های پژوهش

پژوهش حاضر با هدف ترسیم سیمای ادبیات پژوهشی «آهومص» در ایران و شناسایی روندها و خلا‌های آن انجام شد. یافته‌های حاصل از تحلیل محتوای کیفی ۲۵ مقاله منتخب، در چهار محور اصلی ارائه می‌شود: (۱) ویژگی‌های کتاب‌شناختی ادبیات پژوهشی؛ (۲) تحلیل مضامین محوری و چارچوب‌های نظری؛ (۳) ارزیابی تطبیقی با اسناد بالادستی و قابلیت اجرایی اصول اخلاقی؛ (۴) ماتریس هم‌رخدادی حوزه‌های کاربرد و مسائل اخلاقی.



۱۱۳

### ۱.۱. ویژگی‌های کتاب‌شناختی ادبیات پژوهشی

بررسی روند زمانی انتشار نشان می‌دهد که تولید علم در حوزه اخلاق هوش مصنوعی در ایران پدیده‌ای نوظهور است. بیش از ۷۰ درصد مطالعات در بازه زمانی ۱۴۰۲ تا ۱۴۰۴ منتشر شده‌اند که حاکی از شتاب‌گیری توجه دانشگاهیان به این موضوع پس از گسترش فناوری‌های زبانی بزرگ است. از منظر رشته‌ای، ادبیات حاضر میان‌رشته‌ای است؛ اگرچه فلسفه و الهیات همچنان پیشران مبانی نظری هستند، اما ورود فعال حوزه‌های علوم پزشکی (به‌ویژه انفورماتیک سلامت)، رسانه و ارتباطات، و مدیریت منابع انسانی، نشان‌دهنده گسترش موضوع به عرصه‌های کاربردی است. توزیع نهادی نیز پراکنده است و هنوز کانون تخصصی واحدی شکل نگرفته، هرچند دانشگاه‌های تهران، شهید بهشتی و علوم پزشکی تهران از مراکز تولیدکننده اصلی دانش در این حوزه بوده‌اند.

### ۲.۱. تحلیل مضامین محوری و چارچوب‌های نظری

تحلیل محتوای عمیق نشان داد که مضامین تکرارشونده در ادبیات ایرانی شامل «حریم خصوصی و حفاظت داده»، «عدالت و عدم تبعیض»، «شفافیت و توضیح‌پذیری»، «مسئولیت‌پذیری» و «کرامت انسانی» است. از نظر بنیان نظری، همگرایی جالبی بین سه سنت اخلاقی غربی (فایده‌گرایی،

وظیفه‌گرایی، اخلاق فضیلت) و سنت اخلاق اسلامی (با تأکید بر مقاصد شریعت و عدالت) مشاهده می‌شود. با این حال، غلبه رویکرد «هنجاری-توصیفی» بر رویکرد «تجربی-پیمایشی» مشهود است و اکثر پژوهش‌ها به تعریف مفاهیم پرداخته‌اند تا ارائه راهکارهای اجرایی.

### ۳.۱. ارزیابی تطبیقی با اسناد بالادستی و قابلیت اجرایی اصول اخلاقی

یکی از اهداف کلیدی این پژوهش، پر کردن شکاف میان «گفت‌وگوی نظری» و «سیاست‌گذاری عملیاتی» بود. بدین منظور، مضامین استخراج‌شده از ادبیات پژوهشی با اسناد بالادستی جمهوری اسلامی ایران (شامل سند چشم‌انداز ۱۴۰۴، بیانیه گام دوم انقلاب و قوانین مرتبط با فضای مجازی) تطبیق داده شدند. همچنین، اصول اخلاقی شناسایی‌شده بر اساس معیارهای «قابلیت سنجش»، «ضرورت فوری» و «تناسب با بافت فرهنگی-مذهبی» اولویت‌بندی گردیدند. نتایج این تحلیل در دو جدول زیر ارائه شده است.

جدول ۱. تحلیل تطبیقی مضامین اخلاقی مستخرج از ادبیات پژوهشی با اسناد بالادستی جمهوری اسلامی ایران

| ردیف | محور اخلاقی<br>مستخرج از ادبیات | سند بالادستی مرتبط                                              | همپوشانی<br>مفهومی* | شکاف موجود                                                                |
|------|---------------------------------|-----------------------------------------------------------------|---------------------|---------------------------------------------------------------------------|
| ۱    | عدالت و کاهش<br>شکاف دیجیتال    | سند چشم‌انداز ۱۴۰۴،<br>بیانیه گام دوم                           | بالا                | فقدان شاخص‌های<br>کمی برای پایش عدالت<br>الگوریتمی در خدمات<br>عمومی      |
| ۲    | حفاظت از حریم<br>خصوصی          | قانون حمایت از داده‌های<br>شخصی (مصوب شورای<br>عالی فضای مجازی) | متوسط               | ابهام در تعریف مرزهای<br>داده‌های حساس ملی در<br>مقابل داده‌های پژوهشی    |
| ۳    | کرامت انسانی و<br>خودآیینی      | مبانی انسان‌شناسی<br>اسلامی، قانون مدنی                         | بالا                | عدم ترجمه مفهوم انتزاعی<br>کرامت به پروتکل‌های فنی<br>تعامل انسان و ماشین |

| ردیف | محور اخلاقی<br>مستخرج از ادبیات | سند بالادستی مرتبط                    | همپوشانی<br>مفهومی* | شکاف موجود                                                                 |
|------|---------------------------------|---------------------------------------|---------------------|----------------------------------------------------------------------------|
| ۴    | امنیت ملی و<br>حاکمیت داده      | اسناد امنیت ملی و<br>سایبری           | پایین               | غفلت نسبی از ابعاد<br>ژئوپلیتیک و جنگ شناختی<br>مبتنی بر هوش مصنوعی        |
| ۵    | شفافیت و<br>پاسخ‌گویی           | اصل ۱۱۰ قانون اساسی،<br>قوانین نظارتی | متوسط               | نیاز به تدوین استانداردهای<br>فنی برای حق بر توضیح در<br>سامانه‌های هوشمند |

\* نکته: میزان همپوشانی مفهومی بر اساس تطبیق محتوایی بین مضامین استخراج‌شده از ۲۵ مقاله منتخب و بندهای کلیدی اسناد بالادستی تعیین گردیده است.



۱۱۵

جدول ۲. رتبه‌بندی اصول اخلاقی بر اساس قابلیت اجرایی و شاخص‌های آماری  
(میانگین و انحراف معیار)

| اولویت | اصل اخلاقی                        | میانگین<br>قابلیت<br>اجرایی | انحراف<br>معیار | شاخص‌های<br>عملیاتی پیشنهادی                                                         | ابزار مکانیزمی اجرا                                                   |
|--------|-----------------------------------|-----------------------------|-----------------|--------------------------------------------------------------------------------------|-----------------------------------------------------------------------|
| ۱      | شفافیت و<br>توضیح‌پذیری           | ۴/۸۵                        | ۰/۱۲            | نرخ خطای<br>الگوریتم، وجود<br>سامانه پایش<br>عملکرد، دسترسی<br>به منطق<br>تصمیم‌گیری | الزام به ارائه گزارش<br>اطلاعات الگوریتم<br>برای سامانه‌های<br>پرریسک |
| ۲      | عدم تبعیض و<br>عدالت<br>الگوریتمی | ۴/۶۰                        | ۰/۱۵            | شاخص نابرابری<br>در نتایج خروجی،<br>آزمون‌های<br>تعادل‌سنجی<br>داده‌ها               | استقرار تست‌های<br>اخلاقی پیش از<br>استقرار محصول                     |

| اولویت | اصل اخلاقی                           | میانگین<br>قابلیت<br>اجرایی | انحراف<br>معیار | شاخص‌های<br>عملیاتی پیشنهادی                                              | ابزار مکانیزمی اجرا                                                     |
|--------|--------------------------------------|-----------------------------|-----------------|---------------------------------------------------------------------------|-------------------------------------------------------------------------|
| ۳      | مسئولیت‌پذیری و<br>پاسخ‌گویی         | ۴/۳۰                        | ۰/۱۸            | تعیین مالک نهایی<br>تصمیم، وجود<br>بیمه مسئولیت<br>حرفه‌ای<br>توسعه‌دهنده | ایجاد هیئت‌های<br>نظارتی اخلاق هوش<br>مصنوعی در<br>سازمان‌های بزرگ      |
| ۴      | حریم خصوصی<br>و حفاظت داده           | ۴/۱۰                        | ۰/۲۰            | میزان رمزنگاری<br>داده‌ها، تعداد<br>نقض‌های امنیتی<br>گزارش‌شده           | اجرای استانداردهای<br>حفاظت از داده‌ها                                  |
| ۵      | کرامت انسانی و<br>ارزش‌های<br>اسلامی | ۳/۹۰                        | ۰/۲۵            | عدم تقلیل انسان<br>به داده، حفظ شأن<br>کاربر در تعاملات<br>رباتیک         | تدوین منشور اخلاقی<br>رفتاری برای<br>توسعه‌دهندگان مبتنی<br>بر فقه پویا |

\* نکته: این جدول بر اساس امتیازدهی کیفی به شاخص‌های عملیاتی هر اصل اخلاقی، بر اساس شدت تأثیرگذاری بر حکمرانی فناوری تهیه شده است.

#### ۴.۱. ماتریس هم‌رخدادی مضامین و حوزه‌های کاربرد

برای درک بهتر ساختار دانش و شناسایی زمینه‌های کمتر مطالعه‌شده، رابطه بین مضامین اخلاقی و حوزه‌های کاربردی ترسیم شد. همان‌طور که در جدول ۳ مشاهده می‌شود، برخی مضامین مانند «حریم خصوصی» در تمام حوزه‌ها مشترک هستند، اما برخی دیگر خاص‌ترند (مانند «سوگیری الگوریتمی» که بیشتر در منابع انسانی و عدالت دیده می‌شود).

جدول ۳. ماتریس هم‌رخدادی مضامین اخلاقی و حوزه‌های کاربردی  
در ادبیات پژوهشی ایران

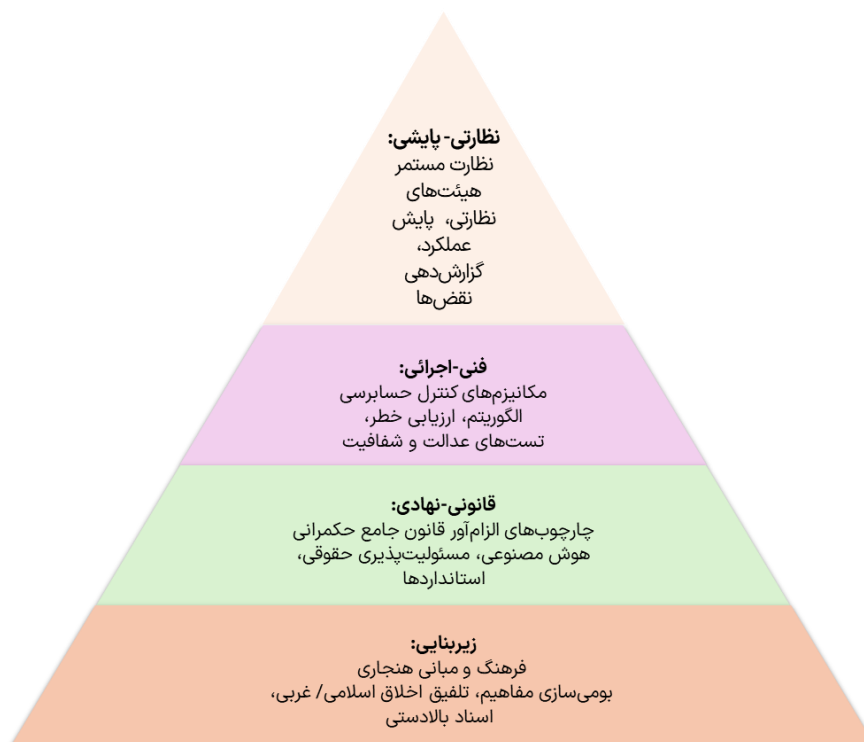
| حوزه کاربردی            | حریم خصوصی | عدالت و تبعیض | شفافیت | مسئولیت‌پذیری | کرامت انسانی | اعتماد |
|-------------------------|------------|---------------|--------|---------------|--------------|--------|
| سلامت و درمان           | زیاد       | متوسط         | متوسط  | زیاد          | زیاد         | متوسط  |
| رسانه و ارتباطات        | متوسط      | زیاد          | زیاد   | متوسط         | کم           | زیاد   |
| منابع انسانی و کسب‌وکار | کم         | زیاد          | متوسط  | زیاد          | متوسط        | متوسط  |
| آموزش                   | متوسط      | متوسط         | کم     | متوسط         | زیاد         | متوسط  |
| حاکمیت و امنیت          | زیاد       | متوسط         | متوسط  | زیاد          | کم           | متوسط  |

(نمادگذاری: زیاد = تمرکز غالب، متوسط = تمرکز متوسط، کم = اشاره گذرا)

### مدل مفهومی حکمرانی اخلاقی هوش مصنوعی در ایران

بر اساس یافته‌های حاصل از فراتحلیل و تحلیل شکاف‌های موجود، مدلی چهارلایه برای حکمرانی اخلاقی هوش مصنوعی در ایران استخراج شد. این مدل بر پایه گذار از «گفت‌وگوی نظری» به «اقدام اجرایی» طراحی شده و شامل لایه‌های زیربنایی تا نظارتی است. شکل ۱ شماتیک این مدل را نشان می‌دهد.





شکل ۱. مدل چهارلایه حکمرانی اخلاقی هوش مصنوعی در ایران

## نسبت‌سنجی تطبیقی ادبیات ایرانی آهومص با کانون‌های رایج در ادبیات جهانی

نتایج این فراتحلیل نشان می‌دهد که ادبیات پژوهشی ایران در بازه زمانی ۱۴۰۲ تا ۱۴۰۴، عمدتاً بر مسائل «نزدیک‌دست»، «ملموس» و «کاربردی» هوش مصنوعی متمرکز بوده است. تمرکز اصلی مطالعات داخلی بر حوزه‌هایی چون نظام سلامت، آموزش عالی، رسانه‌های نوین و کسب‌وکارهای داده‌محور است. در این بستر، محورهای بحث حول اصول پایه‌ای حکمرانی اخلاقی شکل گرفته است؛ از جمله حریم خصوصی و حفاظت داده، سوگیری و تبعیض الگوریتمی، عدالت توزیعی، شفافیت و توضیح‌پذیری،<sup>۱</sup> پاسخ‌گویی نهادی، اعتماد عمومی و صیانت از کرامت و خودآیینی انسان. این هم‌پوشانی مفهومی با دستور کار جهانی نشان‌دهنده پذیرش کلیشه‌های اخلاقی

1. explainability

بین‌المللی در فضای آکادمیک ایران است، چرا که در سطح جهانی نیز همین محورها به عنوان «زیربنای ضروری برای اعتمادپذیری»<sup>۱</sup> سامانه‌های هوشمند شناخته می‌شوند.<sup>۲</sup> با این حال، مقایسه تطبیقی دقیق، دو شکاف راهبردی میان ادبیات ایرانی و کانون‌های پیشرو جهانی را آشکار می‌سازد که ریشه در تفاوت‌های رویکردی و نهادی دارد:

### گذار از «اصول هنجاری» به «مکانیزم‌های اجرایی»

ادبیات جهانی، به ویژه پس از انتشار «قانون هوش مصنوعی اتحادیه اروپا»<sup>۳</sup> و چارچوب‌های سازمان همکاری و توسعه اقتصادی<sup>۴</sup> حرکت سریعی از سطح توصیه‌های اخلاقی انتزاعی به سمت «استانداردهای فنی قابل اندازه‌گیری» و «الزامات قانونی» داشته است (Veale & Zuiderveen Borgesius, 2021). در حالی که ادبیات جهانی بر مکانیزم‌هایی مانند «حسابرسی الگوریتم»<sup>۵</sup>، «ارزیابی تأثیر اخلاقی»<sup>۶</sup> و «مدیریت ریسک مبتنی بر طبقه‌بندی خطر» تمرکز دارد، ادبیات ایرانی همچنان بیشتر در قلمرو «توصیه‌محور» و «چارچوب‌سازی مفهومی» گیر کرده است. این امر باعث می‌شود که خروجی‌های پژوهشی داخلی، اگرچه از نظر فلسفی غنی باشند، اما قابلیت مستقیم تبدیل شدن به سیاست‌گذاری یا استاندارد صنعتی را نداشته باشند.

### غفلت از «آینده‌نگری ساختاری» و «ریسک‌های کلان»

جهت‌گیری دوم در ادبیات جهانی، توجه گسترده به پیامدهای بلندمدت فناوری‌های نوظهور، از جمله هوش مصنوعی عمومی<sup>۷</sup>، کنترل‌پذیری سامانه‌های خودمختار، و هم‌ترازی ارزش‌ها<sup>۸</sup>، است (Bostrom, 2014; Amodei et al., 2016). در مقابل، تمرکز غالب پژوهش‌های ایرانی بر حل معضلات فعلی و کوتاه‌مدت است. اگرچه پرداختن به مسائل جاری ضروری است، اما غفلت

1. Trustworthiness
2. European Commission, 2019; UNESCO, 2021
3. EU AI Act
4. OECD
5. algorithmic auditing
6. ethical impact assessment
7. AGI
8. value alignment



نسبی از سناریوهای آینده‌نگر و تحلیل اثرات ساختاری هوش مصنوعی بر امنیت ملی، اقتصاد کلان و تحولات اجتماعی بلندمدت، باعث می‌شود که زیرساخت دانشی لازم برای هدایت فناوری در دهه‌های آینده فراهم نباشد.

## برآیند و دلالت‌ها

این نسبت‌سنجی نشان می‌دهد که ادبیات ایرانی اگرچه در «شناسایی موضوعات پایه» و «ورود به عرصه کاربردی» موفق عمل کرده است، اما برای هم‌گامی با استانداردهای روز جهانی و پشتیبانی مؤثر از تصمیم‌سازی، نیازمند سه گام تکمیلی حیاتی است:

**عملیاتی‌سازی اصول:** تبدیل مفاهیم اخلاقی پرتکرار (مانند عدالت و شفافیت) به شاخص‌های کمی، معیارهای سنجش و الزامات اجرایی در چارچوب‌های استاندارد.

**تولید شواهد میدانی:** افزایش سهم پژوهش‌های تجربی و پیمایشی درباره تجربه زیسته کاربران، بیماران و متخصصان در مواجهه با سامانه‌های هوشمند، به ویژه در حوزه‌های حساس و پرریسک.

**تقویت چشم‌انداز آینده‌نگر:** سرمایه‌گذاری بر مطالعات سناریومحور و آینده‌پژوهی اخلاقی تا ادبیات داخلی بتواند علاوه بر پاسخ به بحران‌های جاری، مسیرهای بلندمدت حکمرانی و توسعه پایدار هوش مصنوعی را ترسیم کند.

جدول ۸. نسبت‌سنجی فشرده ادبیات ایرانی آهومص با کانون‌های رایج جهانی

| دلت و<br>جهت‌گیری<br>پیشنهادی                                                  | کانون‌های رایج<br>در ادبیات جهانی                                   | ادبیات ایرانی<br>(فراتحلیل حاضر)                              | محور مقایسه                     |
|--------------------------------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------------------|---------------------------------|
| حرکت از<br>گفت‌وگوی<br>اصول‌محور به<br>سمت تعریف<br>معیارهای اجرایی و<br>حقوقی | پررنگ: ، اما با تأکید بر<br>«قابلیت سنجش» و<br>«رعایت حقوق بنیادین» | پررنگ: تمرکز بر اصول<br>(حریم خصوصی، عدالت،<br>شفافیت، کرامت) | مسائل پایه<br>حکمرانی<br>اخلاقی |

| محور مقایسه           | ادبیات ایرانی<br>(فرا تحلیل حاضر)             | کانون‌های رایج<br>در ادبیات جهانی                                                            | دلالت و<br>جهت‌گیری<br>پیشنهادی                                                     |
|-----------------------|-----------------------------------------------|----------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| حوزه‌های<br>کار بست   | سلامت، آموزش، رسانه،<br>کسب‌وکارهای داده‌محور | مشابه، اما با تنوع بسیار<br>گسترده‌تر (شامل<br>دفاعی، محیط زیستی،<br>شهری)                   | توسعه<br>استانداردهای<br>تخصصی برای هر<br>بخش                                       |
| سازوکارهای<br>اجرائی  | غالباً مفهومی، توصیه‌محور<br>و هنجاری         | غالباً اجرایی،<br>استاندارد‌پذیر و دارای<br>بازوی نظارتی                                     | تقویت سازوکارهای<br>حسابرسی، ارزیابی<br>ریسک و نظارت<br>نهادی                       |
| پشتوانه<br>روش شناختی | محدودتر؛ غلبه مطالعات<br>نظری-تحلیلی و اسنادی | پررنگ‌تر؛ استفاده<br>گسترده از مطالعات<br>تجربی، داده‌کاوی و<br>آزمایش‌های میدانی            | افزایش سهم<br>پژوهش‌های میدانی<br>و داده‌محور برای<br>بومی‌سازی یافته‌ها            |
| آینده‌نگری و<br>ریسک  | کم‌رنگ‌تر؛ تمرکز بر<br>چالش‌های فعلی و ملموس  | پررنگ‌تر؛ توجه به هوش<br>مصنوعی عمومی،<br>کنترل‌پذیری، هم‌ترازی<br>ارزش‌ها و امنیت<br>سایبری | تقویت مطالعات<br>سناریو‌محور و<br>آینده‌پژوهی برای<br>مدیریت<br>ریسک‌های<br>بلندمدت |

### راهکارهای پیشنهادی برای تقویت حکمرانی اخلاقی هومص در ایران

با توجه به روندها و خلاءهای شناسایی شده در این فرا تحلیل، برخی راهکارهای عملی به صورت اجمالی پیشنهاد می‌شود. نخست، تدوین یک چارچوب ملی حکمرانی اخلاقی هومص با مشارکت نهادهای تنظیم‌گر، انجمن‌های علمی و دانشگاه‌ها ضروری است؛ چارچوبی که اصول حداقلی درباره حفاظت داده، شفافیت و توضیح‌پذیری، پاسخ‌گویی، پیشگیری از سوگیری و تبعیض، و صیانت از کرامت و خودآیینی انسان را به صورت روشن و اجرایی صورت‌بندی کند. دوم، استقرار سازوکارهای ارزیابی و حسابرسی اخلاقی برای سامانه‌های هوشمند در حوزه‌های



حساس مانند سلامت، آموزش، رسانه و کسب و کارهای داده‌محور پیشنهاد می‌شود؛ به‌ویژه با تعریف شاخص‌های عملی برای عدالت الگوریتمی، شفافیت تصمیم‌گیری و مدیریت ریسک. سوم، تقویت پژوهش‌های میدانی و داده‌محور در بافت ایرانی ضروری است؛ برای نمونه، مطالعات پیمایشی و کیفی درباره تجربه کاربران، بیماران، سالمندان، دانشجویان و کارکنان در مواجهه با سامانه‌های هوشمند می‌تواند شکاف میان گفتمان علمی و نیازهای سیاست‌گذاری را کاهش دهد. چهارم، توسعه ظرفیت نهادی در حوزه آهومص باید در دستور کار قرار گیرد؛ از جمله ایجاد کارگروه‌های میان‌رشته‌ای پایدار، حمایت از نشست‌ها و مجلات تخصصی، و طراحی برنامه‌های آموزشی رسمی برای تربیت نیروی انسانی متخصص. پنجم، برای پشتیبانی دانشی از این مسیر، تدوین و ترجمه هدفمند منابع مرجع و کتاب‌های بنیادین و نیز تولید متون دانشگاهی فارسی استاندارد در آهومص پیشنهاد می‌شود تا کمبود منابع اصلی و ضعف آموزشی، که خود زمینه‌ساز کاستی‌های دیگر است، کاهش یابد. در یک نقشه راه حداقلی، گام‌های کوتاه‌مدت می‌تواند بر تدوین چارچوب ملی و تعریف شاخص‌های حسابرسی متمرکز شود، گام‌های میان‌مدت بر اجرای ارزیابی اخلاقی در حوزه‌های حساس و توسعه پژوهش‌های میدانی، و گام‌های بلندمدت بر تربیت نیروی متخصص و تولید منابع مرجع و نهادینه‌سازی سازوکارهای حکمرانی اخلاقی هومص در ایران.

با توجه به روندها و خلاهای شناسایی شده در فراتحلیل حاضر، گذار از «گفت‌وگوی نظری» به «اقدام اجرایی» نیازمند مداخلات سیاستی هدفمند است. بر این اساس، پنج راهکار استراتژیک برای تقویت حکمرانی اخلاقی هوش مصنوعی در ایران به صورت فهرست‌وار پیشنهاد می‌شود که از سطح کلان تا خرد سازمانی را پوشش می‌دهد:

- تدوین «سند ملی حکمرانی آهومص» با رویکرد الزام‌آور؛
- استقرار سازوکارهای «حسابرسی اخلاقی» در حوزه‌های حساس؛
- توسعه پژوهش‌های میدانی و داده‌محور بومی؛
- تقویت زیرساخت‌های نهادی و تربیت نیروی انسانی متخصص؛
- تولید و بومی‌سازی منابع مرجع؛

## بحث و نتیجه‌گیری

مقاله حاضر با به‌کارگیری رویکرد فراتحلیل ادبیات و تحلیل کیفی ۲۵ مقاله علمی-پژوهشی،

تصویری منسجم از وضعیت «آهومص» در ایران ارائه کرده است. نتایج نشان می‌دهد این حوزه در ایران پدیده‌ای نوظهور و در حال گسترش است و بخش عمده آثار در سال‌های اخیر منتشر شده‌اند؛ با این حال، هنوز نمی‌توان از میدانی کاملاً تثبیت‌شده و نهادمند سخن گفت. ترکیب رشته‌ای و نهادی مطالعات نیز بیانگر آن است که آهومص به تدریج به عرصه‌ای میان‌رشته‌ای در مرزهای فلسفه و الهیات، علوم پزشکی و سلامت، سیاست علم و فناوری، رسانه و ارتباطات و حوزه‌های حرفه‌ای شکل گرفته است.

برای شفافیت بیشتر و متناسب با سبک ارائه مقاله، جمع‌بندی نتایج به صورت شماره‌گذاری شده ارائه می‌شود:

۱. گسترش قلمرو کاربست در مقابل کمبود پشتوانه تجربی: ادبیات ایرانی از سطح تأملات انتزاعی فاصله گرفته و به سمت حوزه‌های حساس و ملموس مانند سلامت، آموزش، رسانه و کسب‌وکارهای داده‌محور حرکت کرده است. با این حال، غلبه قاطع مطالعات نظری-تحلیلی بر پژوهش‌های میدانی، باعث شده است که توصیه‌های اخلاقی ارائه‌شده، اغلب فاقد شواهد عینی از تجربه زیسته کاربران و متخصصان ایرانی باشند. این شکاف روش‌شناختی، اعتبار سیاستی بسیاری از این توصیه‌ها را کاهش می‌دهد.
۲. غنای هنجاری در برابر فقر اجرایی: مطالعات موجود از نظر مبانی نظری غنی هستند و هم‌زمانی سنت‌های اخلاق هنجاری غربی، اخلاق زیستی و اخلاق اسلامی را نشان می‌دهند. مضامینی مانند حریم خصوصی، عدالت الگوریتمی و شفافیت به‌طور مکرر تکرار شده‌اند. با این وجود، تمرکز غالب بر «اصول کلی» مانع از توسعه «مکانیزم‌های اجرایی» شده است. یعنی دانش ما درباره «آنچه باید بکنیم» زیاد است، اما درباره «چگونه انجام دادن آن به صورت استاندارد و قابل سنجش» کمبود جدی وجود دارد.
۳. ضعف در گزارش‌دهی روش‌شناختی و شفافیت: ارزیابی کیفیت مطالعات واردشده نشان داد که اگرچه اهداف پژوهش‌ها روشن است، اما شفافیت روش‌شناختی (به‌ویژه در معیارهای جستجو، غربالگری و تحلیل) در بسیاری از مقالات اسنادی پایین است. این امر امکان تکرارپذیری پژوهش‌ها را دشوار ساخته و نیازمند استانداردسازی بیشتر در گزارش‌دهی مطالعات مروری و تحلیلی در ایران است.
۴. جهت‌گیری کاربردی، اما غیراستاندارد: بخش قابل توجهی از مقالات دارای دلالت‌های عملی هستند، اما این دلالت‌ها عمدتاً در قالب توصیه‌های کلی بیان شده‌اند و کمتر به



سمت طراحی ابزارها، شاخص‌های کمی یا چارچوب‌های استاندارد حرکت کرده‌اند. این پژوهش حاضر با هدف پر کردن این خلأ، تلاش کرد تا با ارائه یک نقشه راه تحلیلی، پلی میان این توصیه‌های کلی و اقدامات عملیاتی ایجاد کند.

۵. بحران زیرساخت دانشی و تربیتی: یکی از ریشه‌ای‌ترین چالش‌ها، کمبود نیروی انسانی متخصص در تقاطع فناوری و اخلاق، و همچنین کمبود منابع مرجع فارسی است. این کمبود، باعث وابستگی نسبی به ترجمه‌های ناقص یا تعمیم بی‌احتیاطانه مدل‌های غربی به بافت فرهنگی ایران شده است. بنابراین، سرمایه‌گذاری بر تولید دانش بومی و تربیت نسل متخصص، پیش‌شرط هرگونه حکمرانی مؤثر است.

درباره اعتبار یافته‌ها و میزان تعمیم‌پذیری. این پژوهش بر پایه ۲۵ مقاله علمی پژوهشی فارسی دارای متن کامل در بازه ۱۴۰۲ تا ۱۴۰۴ انجام شده است که پس از جست و جوی نظام مند و غربال‌گری بر اساس معیارهای ورود و خروج انتخاب شده‌اند. بنابراین، یافته‌ها برای همین مجموعه مقالات معتبر است و می‌تواند تصویری مستند از روندهای غالب و محورهای پرتکرار در این بخش از ادبیات ارائه کند. با این حال، چون کتاب‌ها، پایان‌نامه‌ها، طرح‌های پژوهشی و آثار غیرعلمی پژوهشی در این مرور وارد نشده‌اند و نیز برخی مطالعات به دلیل نبود متن کامل یا عدم انطباق با معیارها کنار گذاشته شده‌اند، نتایج این مطالعه نماینده کل تولیدات پژوهشی ایران در موضوع آهومص نیست. از این رو، این پژوهش به جای ادعای تعمیم گسترده، بیشتر بر جمع‌بندی تحلیلی از مقالات علمی پژوهشی منتخب تمرکز دارد. پژوهش‌های بعدی با گسترش دامنه منابع و افزایش بازه زمانی می‌توانند تصویر کامل‌تری از این حوزه ارائه دهند.

در مجموع، ادبیات پژوهشی اخلاق هوش مصنوعی در ایران از ظرفیت نظری و هنجاری بالایی برخوردار است، اما برای عبور از وضعیت فعلی و تبدیل شدن به یک قطب علمی مؤثر، نیازمند گذار سریع از «گفت‌وگو» به «اقدام» است. همان‌طور که در شکل ۱ (مدل مفهومی حکمرانی) ترسیم شد، این گذار مستلزم پیاده‌سازی همزمان لایه‌های زیربنایی، قانونی، فنی و نظارتی است. این گذار مستلزم سه حرکت همزمان است: ۱) تبدیل اصول اخلاقی به استانداردهای اجرایی و قانونی؛ ۲) افزایش سهم پژوهش‌های میدانی و داده‌محور بومی؛ و ۳) تقویت زیرساخت‌های آموزشی و نهادی. پیوند دادن دانشگاه با صنعت و سیاست‌گذاری، تنها راه خروج از بن‌بست توصیفی و رسیدن به حکمرانی اخلاقی کارآمد در هوش مصنوعی است..

## فهرست منابع

- بابایی، سعیده؛ صابری، نرجس؛ بحرینی، منیره. (۱۴۰۳). تحلیلی اخلاقی از رابطه احساسی انسان و هوش مصنوعی. سیاست علم و فناوری، ۱۷(۴)، ص ۱۹-۳۰. doi: 10.22034/jstp.2025.11773.1816
- بزمی شبنم، افشار لیلیا؛ موسوی نژاد، شهریار. (۱۴۰۴). هوش مصنوعی و چالش‌های اخلاقی احترام به شأن انسانی بیمار با تأکید بر اتونومی. مجله علمی پژوهشی سازمان نظام پزشکی ایران، ۴۳(۱): ۱۸-۲۹. <https://doi.org/10.22034/43.1.5>
- بشیر، حسن؛ راوی، محمد جواد. (۱۴۰۴). مرور نظام‌مند چارچوب‌ها و سازوکارهای حکمرانی راهبردی هوش مصنوعی در مطالعات سال‌های ۲۰۱۷-۲۰۲۵. مطالعات فضای مجازی و رسانه‌های اجتماعی، ۲(۳)، ۵۴-۹. doi: 10.22083/cssms.2025.538964.1067
- بلبلی قادیکلایی، سمیه؛ پارسانیا، حمید. (۱۴۰۲). مروری نظام‌مند بر دلالت‌های اخلاقی استفاده از هوش مصنوعی در فناوری‌های دیجیتال و نسبت آن با اخلاق شکوفایی. راهبرد اجتماعی فرهنگی، ۱۲(۳)، ۷۷۱-۷۹۸. doi: 10.22034/scs.2022.160772
- پاینده، رضا؛ کنعانی، هستی. (۱۴۰۳). تحلیل اخلاقی شرایط و مقررات استفاده از محصولات و خدمات شرکت‌های مبتنی بر هوش مصنوعی در ایران. سیاست علم و فناوری، ۱۷(۳)، ۲۷-۳۹. doi: 10.22034/jstp.2025.11769.1813
- جعفری، رضا. (۱۴۰۳). تأملی در باب هوش مصنوعی و ارتقای قضاوت اخلاقی؛ واکاوی چالش‌ها و فرصت‌ها. مطالعات دینی رسانه، ۶(۲۴)، ۷۸-۵۵. doi: 10.22034/jmrs.2024.210084
- جعفری، رضا؛ رحمتی، حسینعلی. (۱۴۰۴). کاربرد هوش مصنوعی در مشاوره اخلاقی؛ از ظرفیت‌های تحول‌آفرین تا چالش‌های اخلاقی. فصلنامه اخلاق پژوهی، ۸(۲)، ۷۲-۴۹. doi: 10.22034/ethics.2025.51991.1790
- حسینی سورکی، سید محمد. (۱۴۰۳). معنا و معیار عاملیت اخلاقی ماشین‌های هوشمند. پژوهش‌های فلسفی-کلامی، ۲۶(۴)، ۱۰۸-۸۳. doi: 10.22091/jptr.2025.12466.3256
- خنifer، حسین؛ مسلمی، ناهید (۱۳۹۷). اصول و مبانی روش‌های کیفی. تهران: انتشارات نگاه دانش.
- رحیمی اقدم، صمد؛ صالحپور، پدرام؛ نامور، رقیه. (۱۴۰۳). چالش‌های اخلاقی اتخاذ هوش مصنوعی در مدیریت منابع انسانی. اخلاق در علوم و فناوری، ۱۹(۴): ۱۴۲-۱۴۹.
- زارعی محمدحسین. (۱۴۰۴). طراحی و اعتبارسنجی الگوی کاربرد اخلاق‌مدارانه هوش مصنوعی در آموزش عالی. اخلاق در علوم و فناوری، ۲۰(۳): ۸۱-۸۹. <http://ethicsjournal.ir/article-1-3396-fa.html>



- زاهدپاشا، یداله؛ هدایت پور، بردیا؛ بالغی، مهرانگیر؛ ؛ بالغی دماوندی، صدیقه. (۱۴۰۴). اهمیت اخلاق در استفاده از هوش مصنوعی و چالش های آن در مقالات حوزه سلامت. مجله علمی دانشگاه علوم پزشکی بابل. ۲۷(۱). ۱۸. <http://jbums.org/article-1-12544-fa.html>
- زرگر، زهرا. (۱۴۰۴). رابطه عواطف و ظرفیت اخلاقی در فناوری های هوش مصنوعی. مجله پژوهش های فلسفی، ۵۰(۵)، ۴۰۱۹. doi: 10.22034/jpiut.2024.63626.3875
- سیروس کبیری، سپیده؛ موسوی، سید محمدرضا. (۱۴۰۲). تحولات صنعت رسانه مبتنی بر فناوری هوش مصنوعی: کاربردها، فرصت ها و مخاطرات. بررسی های مدیریت رسانه، ۲(۳)، ۳۶۶-۳۹۰. doi: 10.22059/mmr.2024.368379.1069
- شفیعان سارا؛ منعمی عرفان. (۱۴۰۴). دوروی سکه هوش مصنوعی در علوم پزشکی: نگاهی به جنبه های انسانی و اخلاقی. انفورماتیک سلامت و زیست پزشکی. ۱۲(۱): ۹۶-۱۰۴. <http://jhbmj.ir/article-1-928-fa.html>
- عاشوری کیسمی، محمد علی. (۱۴۰۳). طرح و بررسی برخی از مسائل اخلاقی هوش مصنوعی در هنر. متافیزیک، ۱۶(۳۷)، ۹۳-۱۱۰. doi: 10.22108/mp.2024.138105.1488
- عاشوری کیسمی، محمد علی؛ پرویزی، مریم. (۱۴۰۳). بررسی عاملیت اخلاقی هوش مصنوعی عام. پژوهش های علم و دین، ۱۵(۱)، ۱۲۵-۱۵۱. doi: 10.30465/srs.2024.49282.2162
- عباسی، محمود؛ تیموری، مهرداد. (۱۴۰۲). مروری بر چالش های اخلاقی و حقوقی کاربرد هوش مصنوعی در نظام سلامت. مجله اخلاق پزشکی، دوره ۱۷(۴۸)، ۱-۱۱. <https://doi.org/10.22037/mej.v17i48.44053>
- قاضی طباطبایی، محمود؛ ودادهیر، ابوعلی (۱۳۸۹). روش تحقیق فراتحلیل. تهران: انتشارات جامعه شناسان.
- قوامی پور سرشکه، محدثه؛ محمودی، امیررضا. (۱۴۰۳). واکاوی چالش های پیاده سازی هوش اخلاقی در هوش مصنوعی. فصلنامه اخلاق پژوهی، ۶(۴)، ۲۳-۴۸. doi: 10.22034/ethics.2024.51227.1650
- کلاته آقامحمدی، آمنه. (۱۴۰۲). چالش های اخلاقی هوش مصنوعی در رسانه های نوین و راهکارهای مواجهه با آن. مطالعات دینی رسانه، ۵(۲۰)، ۵۲-۸۷. doi: 10.22034/jmrs.2024.205457
- مرادی مخلص، حسین؛ حاجیلونی محب، محمدمهدی؛ عمونی رازانی، امیرحسین. (۱۴۰۴). کشف ابعاد فرااخلاق در محیط های یادگیری مبتنی بر هوش مصنوعی: مطالعه ای فراترکیب. فرهنگ در دانشگاه اسلامی، ۱۵(۵۵)، ۶۱-۸۴. doi: 10.22034/ciu.2025.55.61
- مرتضوی شاهرودی، سید محمد علی؛ زارعی، عیسی. (۱۴۰۴). اصول و چالش های اخلاقی استفاده از هوش مصنوعی در تحقیقات علمی. فصلنامه اخلاق پژوهی، ۷(۴)، ۵-۲۶. doi: 10.22034/ethics.2025.51545.1701

مهدوی نور، سید حاتم؛ آرام، محمدرضا. (۱۴۰۳). اخلاق هوش مصنوعی: چهارچوبی مبتنی بر اصول ارتباطی انسان با خدا، خود، دیگران و محیط. پژوهش دینی. ۲۴ (۴۹): ۲۰۵-۲۲۴. URL: <http://pdmag.ir/article-1-2176-fa.html>

میری بالاجورشری، سیده مهشیدغ محمودی، امیررضا. (۱۴۰۳). واکاوی مسائل اخلاقی در زمینه هوش مصنوعی با نگاهی به اخلاق اسلامی. مطالعات اخلاق کاربردی، ۲۰ (۱)، ۹۷-۱۲۳. doi: 10.22081/jare.2024.68803.1901

هامانی، رضوان؛ مختاری، مصطفی. (۱۴۰۴). صورت‌بندی الگوی مطلوب در نسبت اخلاق با هوش مصنوعی با تأکید بر مبانی اخلاقی علامه مصباح یزدی (ره). فصلنامه اخلاق پژوهی، ۱۸ (۱)، ۸۹-۱۱۲. doi: 10.22034/ethics.2025.51930.1777

هرندی، عطاءاله؛ ابراهیمی، علیرضا. (۱۴۰۳). شناسایی عوامل مؤثر بر هوش مصنوعی در بازاریابی صنعت بانکداری با رویکرد فراترکیب. بررسی‌های مدیریت رسانه، ۳ (۳)، ۳۱۱-۳۴۱. doi: 10.22059/mmr.2024.384200.1124



۱۲۷

- Amodei, D., Olah, C., Steinhardt, J., Christiano, p. , Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. *Proceedings of the 2016 AAAI Spring Symposium Series*, 1–8.
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- European Commission. (2019). *Ethics guidelines for trustworthy AI*. High-Level Expert Group on Artificial Intelligence. Publications Office of the European Union. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org/ark:/48223/pf0000379707>
- Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the draft EU Artificial Intelligence Act – Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4), 97–115. <https://doi.org/10.29296/clri.2021.0010>

