

توسعه فردی و تحول سازمانی

حکمرانی اخلاقی هوش مصنوعی در دانشگاه‌ها: چارچوبی برای عدالت آموزشی، پاسخ‌گویی و استقلال علمی در چارچوب فلسفه تعلیم و تربیت

معصومه بیات^۱

شیوه استناددهی: بیات، معصومه. (۱۴۰۶). حکمرانی اخلاقی

هوش مصنوعی در دانشگاه‌ها: چارچوبی برای عدالت

آموزشی، پاسخ‌گویی و استقلال علمی در چارچوب فلسفه

تعلیم و تربیت. توسعه فردی و تحول سازمانی، ۵(۲)، ۲۶-۱.

۱. استادیار، گروه الهیات، دانشگاه فرهنگیان، تهران، ایران

* ایمیل نویسنده مسئول: ma.bayat@cfu.ac.ir

چکیده

تاریخ چاپ نهایی: ۱ خرداد ۱۴۰۶

تاریخ چاپ اولیه: ۳ مهر ۱۴۰۵

تاریخ پذیرش: ۸ شهریور ۱۴۰۵

تاریخ بازنگری: ۱ شهریور ۱۴۰۵

تاریخ ارسال: ۱۲ اردیبهشت ۱۴۰۵

هدف پژوهش حاضر تبیین چارچوب حکمرانی اخلاقی هوش مصنوعی در دانشگاه‌ها با تأکید بر عدالت آموزشی، پاسخ‌گویی و استقلال علمی در چارچوب فلسفه تعلیم و تربیت بود. این پژوهش با رویکرد کیفی و بر اساس تحلیل مضمون انجام شد. مشارکت‌کنندگان شامل ۱۸ نفر از اعضای هیئت علمی، متخصصان فلسفه تعلیم و تربیت، هوش مصنوعی و فناوری آموزشی، اخلاق فناوری، مدیریت و سیاست‌گذاری آموزش عالی و مدیران دانشگاهی شهر تهران بودند که با روش نمونه‌گیری هدفمند و ملاکی انتخاب شدند. داده‌ها از طریق مصاحبه‌های نیمه‌ساختاریافته عمیق ۴۵ تا ۷۵ دقیقه‌ای گردآوری شد و نمونه‌گیری تا دستیابی به اشباع نظری ادامه یافت. مصاحبه‌ها پس از پیاده‌سازی کلمه‌به‌کلمه با استفاده از تحلیل مضمون و فرایند کدگذاری، مقوله‌بندی و استخراج مضامین تحلیل شدند. مدیریت و سازمان‌دهی داده‌ها با نرم‌افزار MAXQDA انجام شد و اعتبار یافته‌ها از طریق بازبینی مشارکت‌کنندگان، بررسی هم‌تایان، مقایسه مستمر داده‌ها و ثبت مسیر تصمیم‌گیری تحلیلی تقویت شد. تحلیل داده‌ها به استخراج یک مضمون فراگیر با عنوان «حکمرانی انسان‌محور و ارزش‌بنیاد هوش مصنوعی در دانشگاه» و هفت مضمون اصلی منجر شد که شامل عدالت آموزشی، پاسخ‌گویی و شفافیت، استقلال علمی و عاملیت انسانی، حکمرانی داده و حریم خصوصی، سواد هوش مصنوعی و ظرفیت اخلاقی، تنظیم‌گری مشارکتی و نظارت نهادی و صیانت از اصالت فرایند تعلیم و تربیت بود. همچنین، چارچوب نهایی در سه لایه به‌هم‌پیوسته شامل اصول هنجاری و فلسفی، الزامات رویه‌ای و عملیاتی و سازوکارهای نهادی و ظرفیت‌ساز سازمان یافت. یافته‌ها نشان داد که عدالت آموزشی باید در ابعاد توزیعی، رویه‌ای و ترمیمی تحقق یابد، مسئولیت تصمیمات حساس همچنان نزد عامل انسانی و نهاد دانشگاه باقی بماند و استفاده از هوش مصنوعی نباید به تضعیف استقلال حرفه‌ای استاد یا عاملیت شناختی دانشجو منجر شود. حکمرانی اخلاقی هوش مصنوعی در دانشگاه‌ها مستلزم عبور از رویکرد صرفاً فناورانه و حرکت به سوی چارچوبی انسان‌محور، پاسخ‌گو و ارزش‌بنیاد است که در آن عدالت آموزشی، مسئولیت‌پذیری، آزادی علمی، نظارت انسانی و صیانت از غایات تربیتی به‌عنوان معیارهای اصلی مشروعیت استفاده از هوش مصنوعی در نظر گرفته شوند.

کلیدواژگان: هوش مصنوعی، حکمرانی اخلاقی، عدالت آموزشی، پاسخ‌گویی، استقلال علمی، فلسفه تعلیم و تربیت، آموزش عالی

این مقاله متعلق به نویسنده است. انتشار این مقاله به صورت دسترسی آزاد مطابق با گواهی (CC BY-NC 4.0) صورت گرفته است.



© ۱۴۰۶ تمامی حقوق انتشار

Personal Development and Organizational Transformation

Ethical Governance of Artificial Intelligence in Universities: A Framework for Educational Justice, Accountability, and Academic Autonomy within the Philosophy of Education

Masumeh Bayat^{1*}

1. Assistant Professor, Department of Theology, Farhangian University, Tehran, Iran

*Corresponding Author's Email: ma.bayat@cfu.ac.ir

How to cite: Bayat, M. (2027). Ethical Governance of Artificial Intelligence in Universities: A Framework for Educational Justice, Accountability, and Academic Autonomy within the Philosophy of Education. *Personal Development and Organizational Transformation*, 5(2), 1-26.

Submit Date: 02 May 2026

Revise Date: 23 August 2026

Accept Date: 30 August 2026

Initial Publish: 25 September 2026

Final Publish: 22 May 2027

Abstract

This study aimed to develop a framework for the ethical governance of artificial intelligence in universities, emphasizing educational justice, accountability, and academic autonomy within the philosophy of education. This qualitative study employed thematic analysis. Participants consisted of 18 faculty members, experts in philosophy of education, artificial intelligence and educational technology, technology ethics, higher education management and policy, and university administrators in Tehran, selected through purposive criterion-based sampling. Data were collected through in-depth semi-structured interviews lasting approximately 45–75 minutes, and sampling continued until theoretical saturation was achieved. Interviews were transcribed verbatim and analyzed through iterative coding, categorization, and theme development. MAXQDA was used to organize and manage the qualitative data. Trustworthiness was enhanced through member checking, peer review, constant comparison, and systematic documentation of analytical decisions. The analysis identified one overarching theme, “human-centered and value-based governance of artificial intelligence in universities,” and seven major themes: educational justice; accountability and transparency; academic autonomy and human agency; data governance and privacy; AI literacy and ethical capacity; participatory regulation and institutional oversight; and preservation of the authenticity of education. The final framework was organized into three interrelated layers: normative and philosophical principles, procedural and operational requirements, and institutional and capacity-building mechanisms. Educational justice was found to include distributive, procedural, and restorative dimensions. The findings further indicated that responsibility for high-stakes academic decisions should remain with identifiable human and institutional actors, while AI should support rather than replace faculty judgment and students’ cognitive agency. Ethical governance of AI in universities requires moving beyond a purely technological approach toward a human-centered, accountable, and value-based governance model in which educational justice, human responsibility, academic freedom, meaningful oversight, and preservation of educational purposes constitute the principal criteria for legitimate AI use.

Keywords: Artificial Intelligence, Ethical Governance, Educational Justice, Accountability, Academic Autonomy, Philosophy of Education, Higher Education



© 2027 the authors. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

مقدمه

گسترش سریع هوش مصنوعی، به‌ویژه سامانه‌های هوش مصنوعی مولد، در سال‌های اخیر ساختارها، فرایندها و مناسبات اساسی آموزش عالی را در معرض دگرگونی قرار داده است. دانشگاه‌ها اکنون با فناوری‌هایی مواجه‌اند که نه تنها می‌توانند در تولید محتوا، تحلیل داده، پشتیبانی از یادگیری، ارزشیابی، مدیریت آموزشی و پژوهش نقش ایفا کنند، بلکه به تدریج در فرایندهایی وارد می‌شوند که پیش‌تر به‌طور انحصاری در قلمرو داور انسانی، قضاوت حرفه‌ای استادان، تصمیم‌گیری مدیران و فعالیت شناختی دانشجویان قرار داشت. این تحول، ظرفیت‌های قابل توجهی برای شخصی‌سازی آموزش، بهبود دسترسی به منابع، تسریع فرایندهای دانشگاهی، افزایش بهره‌وری و توسعه اشکال جدید تعامل آموزشی ایجاد کرده است، اما هم‌زمان پرسش‌هایی بنیادین درباره عدالت، مسئولیت، آزادی علمی، اصالت یادگیری، مالکیت معرفتی، حریم خصوصی و حدود مشروع واگذاری تصمیم به سامانه‌های هوشمند پدید آورده است. از این رو، مسئله اصلی دیگر صرفاً این نیست که دانشگاه‌ها چگونه از هوش مصنوعی استفاده کنند، بلکه پرسش مهم‌تر آن است که این استفاده بر اساس چه اصولی، تحت چه ساختارهای نهادی و با چه سازوکارهای پاسخ‌گویی باید هدایت شود تا توسعه فناوری با رسالت‌های آموزشی، معرفتی و اخلاقی دانشگاه سازگار باقی بماند (Abedin et al., 2026; Mhlono, 2026; Yan & Chau, 2025).

ادغام هوش مصنوعی در آموزش عالی در یک سطح می‌تواند فرصت‌های مهمی برای ارتقای تدریس و یادگیری فراهم کند. کاربرد سامانه‌های هوشمند در ارائه بازخورد فوری، طراحی فعالیت‌های آموزشی متناسب با نیاز دانشجویان، تولید مواد آموزشی، پشتیبانی از حل مسئله، تحلیل الگوهای یادگیری و تسهیل دسترسی به منابع علمی می‌تواند کارآمدی فرایند آموزش را افزایش دهد. در برخی حوزه‌ها، هوش مصنوعی حتی می‌تواند امکان حمایت دقیق‌تر از دانشجویانی را فراهم آورد که به دلایل جغرافیایی، اقتصادی، زبانی یا آموزشی با محدودیت‌هایی در دسترسی به منابع مواجه هستند. با این حال، ظرفیت آموزشی این فناوری تنها زمانی می‌تواند به ارزش تربیتی تبدیل شود که استفاده از آن در نسبت با اهداف اصلی آموزش عالی تعریف شود؛ زیرا ابزارهای هوش مصنوعی، برخلاف ظاهر بی‌طرفانه خود، بر اساس داده‌ها، مدل‌ها، قواعد طراحی و منطق‌های تجاری مشخصی عمل می‌کنند و می‌توانند به‌طور مستقیم یا غیرمستقیم بر محتوای آموزش، فرایند ارزیابی و نحوه شکل‌گیری تجربه یادگیری اثر بگذارند (Abedin et al., 2026; Li et al., 2025; Rao & Suhasini, 2025). بنابراین، رویکردی که صرفاً بر مزایای عملیاتی فناوری تأکید کند، برای تصمیم‌گیری دانشگاهی کافی نیست و لازم است پیامدهای اخلاقی، فلسفی و اجتماعی آن نیز در مرکز سیاست‌گذاری قرار گیرد.

از منظر فلسفه تعلیم و تربیت، مسئله هوش مصنوعی در دانشگاه فراتر از کارآمدی فناورانه است؛ زیرا آموزش دانشگاهی صرفاً انتقال داده، تولید پاسخ یا انجام سریع تکالیف نیست، بلکه فرایندی برای رشد استقلال فکری، قضاوت عقلانی، مسئولیت‌پذیری معرفتی، گفت‌وگوی انتقادی و شکل‌گیری هویت علمی فرد محسوب می‌شود. هنگامی که ابزارهای هوش مصنوعی در نگارش، استدلال، حل مسئله، ارزیابی یا پیشنهاد مسیرهای پژوهشی مداخله می‌کنند، مرز میان «پشتیبانی از یادگیری» و «جایگزینی فعالیت شناختی» می‌تواند مبهم شود. در چنین شرایطی، مسئله اصلی آن است که فناوری تا چه حد باید به‌عنوان تسهیل‌گر یادگیری عمل کند و از چه نقطه‌ای ممکن است به کاهش عاملیت دانشجو یا تضعیف نقش حرفه‌ای استاد منجر شود. برخی مطالعات بر این نکته تأکید کرده‌اند که هوش مصنوعی مولد نباید به‌عنوان یک ابزار همه‌کاره تلقی شود و استفاده آن همواره با مبادله‌هایی میان سرعت، اصالت، استقلال، کیفیت یادگیری و مسئولیت همراه است (Razi et al., 2025).

2025). از همین منظر، چارچوب اخلاقی دانشگاه باید بتواند میان بهره‌گیری از ظرفیت‌های فناورانه و حفظ ماهیت انسانی و تربیتی آموزش تعادل ایجاد کند.

یکی از محوری‌ترین مسائل در این زمینه، عدالت آموزشی است. هوش مصنوعی می‌تواند فرصت‌های جدیدی برای گسترش دسترسی به آموزش ایجاد کند، اما هم‌زمان احتمال بازتولید یا تشدید نابرابری‌های موجود را نیز در خود دارد. تفاوت در دسترسی به نسخه‌های پیشرفته ابزارهای هوش مصنوعی، کیفیت اینترنت، توانایی خرید خدمات اشتراکی، سواد دیجیتال، زبان، رشته تحصیلی و تجربه کار با سامانه‌های هوشمند می‌تواند سبب شود که گروه‌های مختلف دانشجویان از منافع این فناوری به‌طور یکسان بهره‌مند نشوند. افزون بر این، الگوریتم‌هایی که بر داده‌های تاریخی یا مجموعه داده‌های نامتوازن آموزش دیده‌اند ممکن است سوگیری‌های اجتماعی، فرهنگی، زبانی یا جنسیتی موجود را بازتولید کنند. این مسئله در تصمیمات حساس دانشگاهی مانند پذیرش، ارزشیابی، تشخیص عملکرد، تخصیص حمایت، رتبه‌بندی یا تشخیص تخلف اهمیت بیشتری پیدا می‌کند. پژوهش‌های مربوط به اخلاق هوش مصنوعی در آموزش نشان داده‌اند که انصاف، عدم تبعیض، دسترسی‌پذیری و توجه به گروه‌های آسیب‌پذیر باید در زمره اصول بنیادی حکمرانی این فناوری قرار گیرند (Che & Perkins, 2026; Mokadi et al., 2026; Yan & Chau, 2025).

عدالت آموزشی در این زمینه تنها به برابری در دسترسی به ابزار محدود نمی‌شود، بلکه شامل عدالت در طراحی، اجرا، ارزیابی و پیامدهای استفاده از هوش مصنوعی نیز هست. ممکن است همه دانشجویان از نظر صوری به یک سامانه دسترسی داشته باشند، اما توانایی آنان برای استفاده مؤثر، نقد خروجی‌ها و تشخیص خطاهای الگوریتمی یکسان نباشد. از این رو، سیاست دانشگاهی باید میان «دسترسی برابر» و «توان بهره‌مندی برابر» تمایز قائل شود. این تمایز اهمیت سواد هوش مصنوعی را برجسته می‌کند؛ زیرا عدالت در محیط‌های هوشمند بدون توانایی کاربران برای فهم محدودیت‌ها، سوگیری‌ها و منطق عملکرد سامانه‌ها تحقق نمی‌یابد. مسئله آمادگی نهادی نیز در اینجا اهمیت پیدا می‌کند، زیرا دانشگاه‌هایی که فاقد زیرساخت، سیاست روشن، آموزش تخصصی و ظرفیت نظارتی کافی هستند، ممکن است فناوری را پیش از آنکه شرایط اخلاقی استفاده از آن فراهم شده باشد، وارد فرایندهای حساس کنند (Mmari, 2026; Sultana et al., 2026). بنابراین، عدالت در حکمرانی هوش مصنوعی باید هم به سطح فردی و هم به ظرفیت نهادی دانشگاه توجه داشته باشد.

در کنار عدالت، پاسخ‌گویی یکی دیگر از ستون‌های اصلی حکمرانی اخلاقی هوش مصنوعی است. هرچه سامانه‌های هوشمند نقش بیشتری در تصمیمات آموزشی و دانشگاهی ایفا کنند، تعیین مسئولیت در برابر پیامدهای آن‌ها پیچیده‌تر می‌شود. در یک تصمیم الگوریتمی ممکن است طراح سامانه، شرکت ارائه‌دهنده، واحد فناوری دانشگاه، مدیر آموزشی، عضو هیئت علمی و کاربر نهایی هر یک بخشی از فرایند را کنترل کنند. اگر یک سامانه در تشخیص تقلب، ارزیابی عملکرد یا پیشنهاد تصمیم آموزشی دچار خطا شود، مشخص نبودن مسئولیت می‌تواند به نوعی خلأ پاسخ‌گویی منجر شود. از این رو، یکی از الزامات حکمرانی اخلاقی آن است که تصمیمات دانشگاهی حتی در صورت استفاده گسترده از فناوری، همچنان دارای مسئول انسانی و نهادی مشخص باشند. در حوزه‌هایی مانند ارزشیابی مبتنی بر هوش مصنوعی، شفافیت رویه‌ای و پاسخ‌گویی حقوقی اهمیت خاصی دارد، زیرا دانشجو باید بتواند بداند که تصمیم چگونه شکل گرفته و در صورت اعتراض به چه مرجعی مراجعه کند (Fraidan, 2025).

پاسخ‌گویی بدون شفافیت نیز عملاً امکان‌پذیر نیست. دانشگاه باید بتواند برای اعضای خود روشن کند که در چه فرایندهایی از هوش مصنوعی استفاده می‌شود، داده‌های مورد استفاده چیست، خروجی سامانه چه نقشی در تصمیم دارد و چه فرد یا واحدی مسئول تصمیم نهایی است. در

این چارچوب، قابلیت توضیح اهمیت ویژه پیدا می‌کند. البته شفافیت لزوماً به معنای افشای کامل کد یا پیچیدگی‌های فنی الگوریتم نیست، بلکه به این معناست که سطحی از توضیح متناسب با اهمیت تصمیم و نیاز مخاطب فراهم شود. به‌ویژه هنگامی که تصمیمی بر نمره، فرصت تحصیلی، اعتبار علمی یا موقعیت حرفه‌ای فرد اثر دارد، اتکا به خروجی غیرقابل توضیح نمی‌تواند از نظر اخلاقی کافی باشد. پژوهش‌های حوزه سیاست‌گذاری دانشگاهی نیز نشان داده‌اند که دانشگاه‌ها در حال حرکت از سیاست‌های پراکنده و واکنشی به سمت چارچوب‌های منسجم‌تری برای مدیریت هوش مصنوعی مولد هستند، اما تفاوت قابل توجهی در سطح بلوغ، وضوح و جامعیت این سیاست‌ها وجود دارد (Basil et al., 2026; Joshi et al., 2026; Toquero, 2025).

مسئله مهم دیگر، تمامیت علمی و اصالت فعالیت دانشگاهی است. هوش مصنوعی مولد مرزهای سنتی میان نگارش مستقل، کمک فناوریانه، و تولید خودکار محتوا را تغییر داده است. در چنین شرایطی، قواعد کلاسیک مقابله با تقلب علمی به‌تنهایی پاسخ‌گو نیستند و دانشگاه‌ها ناچارند تعریف دقیق‌تری از استفاده مجاز، استفاده نیازمند افشا، و استفاده غیرقابل قبول ارائه کنند. تمرکز صرف بر کشف استفاده از هوش مصنوعی نیز ممکن است به راه‌حلی ناکافی یا حتی ناعادلانه تبدیل شود، زیرا سامانه‌های تشخیص خود نیز محدودیت و احتمال خطا دارند. به همین دلیل، ادبیات جدید بیشتر بر بازطراحی ارزشیابی، شفاف‌سازی انتظارات، آموزش اخلاق پژوهش و تقویت مسئولیت فردی تأکید دارد (Güneş & Kaban, 2025; Jamil et al., 2026; Kangwa et al., 2025). این دیدگاه نشان می‌دهد که مسئله تمامیت علمی را نباید صرفاً به ممنوعیت ابزار تقلیل داد، بلکه لازم است رابطه میان فناوری، یادگیری اصیل و مسئولیت معرفتی از نو تعریف شود.

در این میان، اعضای هیئت علمی نقشی دوگانه دارند. آنان از یک سو کاربران بالقوه ابزارهای هوش مصنوعی در تدریس، طراحی محتوا، ارزشیابی و پژوهش هستند و از سوی دیگر، مسئول هدایت اخلاقی استفاده دانشجویان از این ابزارها محسوب می‌شوند. بنابراین، دانشگاه نمی‌تواند انتظار داشته باشد اعضای هیئت علمی صرفاً با صدور دستورالعمل‌های کلی بتوانند در محیطی پیچیده و سریع‌التغییر تصمیمات مناسبی اتخاذ کنند. آنچه ضروری است، ایجاد چارچوب‌های اخلاقی عملیاتی برای استادان، آموزش درباره محدودیت‌ها و خطرهای فناوری، و فراهم کردن فضای گفت‌وگو درباره موقعیت‌های خاکستری است. چارچوب‌های اخلاقی برای اعضای هیئت علمی باید هم بر حفظ اصالت علمی و هم بر استفاده مسئولانه از ظرفیت‌های هوش مصنوعی تمرکز کنند (McGinty, 2025). چنین رویکردی می‌تواند از تبدیل سیاست دانشگاه به مجموعه‌ای از ممنوعیت‌های غیرواقع‌بینانه جلوگیری کرده و در عوض، تصمیم‌گیری حرفه‌ای و مسئولانه را تقویت کند.

استقلال علمی نیز یکی دیگر از حوزه‌هایی است که تحت تأثیر گسترش هوش مصنوعی قرار می‌گیرد. آزادی استاد در انتخاب روش تدریس، طراحی محتوا، ارزیابی و پژوهش یکی از ارکان سنت دانشگاهی است، اما استفاده گسترده از پلتفرم‌های فناوریانه ممکن است به تدریج بخشی از این اختیار را به قواعد ازپیش‌طراحی‌شده سامانه‌ها منتقل کند. اگر دانشگاه‌ها به پلتفرم‌هایی وابسته شوند که معیارهای تصمیم‌گیری، نحوه پردازش داده و اولویت‌بندی محتوا در آن‌ها شفاف نیست، امکان شکل‌گیری نوعی وابستگی معرفتی و فناوریانه وجود دارد. علاوه بر این، الگوریتم‌های توصیه‌گر می‌توانند بر مسیر جست‌وجوی علمی، انتخاب منابع و حتی شکل‌گیری پرسش‌های پژوهشی اثر بگذارند. از این منظر، استقلال علمی فقط به آزادی فردی استاد محدود نیست، بلکه استقلال نهادی دانشگاه از زیرساخت‌های تجاری و الگوریتم‌هایی را نیز دربر می‌گیرد که ممکن است منطق خود را بر فرایند دانش تحمیل کنند (Razi et al., 2025; Souli & Pierrakeas, 2025).

حکمرانی داده نیز پیوند مستقیمی با استقلال و اخلاق دانشگاهی دارد. سامانه‌های هوش مصنوعی برای عملکرد خود به حجم قابل توجهی از داده متکی هستند و دانشگاه‌ها منابع گسترده‌ای از داده‌های آموزشی، پژوهشی، هویتی و رفتاری را در اختیار دارند. استفاده از این داده‌ها بدون

تعیین دقیق هدف، حداقل گرایی، رضایت، مدت نگهداری و سطح دسترسی می‌تواند حریم خصوصی دانشجویان و استادان را تهدید کند. از سوی دیگر، انتقال داده به ارائه‌دهندگان بیرونی ممکن است کنترل دانشگاه بر داده‌های خود را کاهش دهد. از این رو، سیاست‌گذاری در حوزه هوش مصنوعی باید شامل قواعد صریح درباره مالکیت، پردازش، ذخیره‌سازی، اشتراک‌گذاری و حذف داده باشد. ادغام فناوری‌های جدید در مدیریت دانشگاهی نیز نشان داده است که دیجیتالی‌شدن فرایندهای اداری و آموزشی بدون طراحی ساختارهای حکمرانی مناسب می‌تواند وابستگی و آسیب‌پذیری نهادی ایجاد کند (Souli & Pierrakeas, 2025). بنابراین، حفاظت از داده را باید نه صرفاً مسئله‌ای فنی، بلکه بخشی از استقلال و مسئولیت دانشگاه دانست.

در سطح سیاست‌گذاری، بررسی تجربه دانشگاه‌های مختلف نشان می‌دهد که هنوز الگوی واحدی برای تنظیم استفاده از هوش مصنوعی وجود ندارد. برخی دانشگاه‌ها رویکردهای محدودکننده اتخاذ کرده‌اند، برخی بر افشای استفاده و مسئولیت فردی تأکید دارند و برخی دیگر در حال تدوین چارچوب‌های جامع‌تر هستند که ابعاد آموزشی، پژوهشی، اخلاقی، حقوقی و مدیریتی را هم‌زمان پوشش می‌دهد. تحلیل سیاست‌های دانشگاهی نشان داده است که اثربخشی حکمرانی به میزان انسجام میان سیاست رسمی، آموزش کاربران، حمایت سازمانی و سازوکارهای نظارتی بستگی دارد (Basil et al., 2026; Joshi et al., 2026). همچنین، مطالعات مربوط به تنظیم‌گری هوش مصنوعی در آموزش عالی نشان می‌دهد که سیاست‌ها باید انعطاف‌پذیر و بازنگری‌پذیر باشند؛ زیرا تحول سریع فناوری موجب می‌شود قواعد ثابت در فاصله کوتاهی کارایی خود را از دست بدهند (Toquero, 2025). بنابراین، حکمرانی مطلوب باید سازوکاری برای یادگیری سازمانی و اصلاح مستمر سیاست‌ها داشته باشد.

از سوی دیگر، حکمرانی هوش مصنوعی در دانشگاه را نمی‌توان صرفاً مسئله‌ای درون‌سازمانی دانست. توسعه فناوری در بستری از تعامل دانشگاه، دولت، صنعت و جامعه صورت می‌گیرد و آمادگی کشورها و نهادهای آموزشی برای پذیرش مسئولانه هوش مصنوعی به کیفیت همین تعامل بستگی دارد. مطالعات مربوط به آمادگی هوش مصنوعی نشان داده‌اند که هماهنگی میان سیاست عمومی، زیرساخت، ظرفیت علمی، نیروی انسانی و بخش خصوصی برای استفاده مسئولانه از فناوری ضروری است (Sultana et al., 2026). این موضوع در آموزش عالی اهمیت بیشتری دارد، زیرا دانشگاه هم مصرف‌کننده فناوری و هم یکی از نهادهای اصلی تولید دانش، توسعه نیروی انسانی و نقد اجتماعی فناوری است. در نتیجه، دانشگاه نباید صرفاً سیاست‌های شرکت‌های فناور را بپذیرد، بلکه باید در تعریف قواعد، استانداردها و انتظارات اخلاقی نقش فعال داشته باشد.

مسئله حکمرانی مشارکتی نیز در این زمینه اهمیت فزاینده‌ای یافته است. هنگامی که تصمیمات هوش مصنوعی بر گروه‌های متنوعی از دانشجویان، استادان، مدیران و کارکنان اثر می‌گذارد، طراحی سیاست بدون مشارکت این گروه‌ها می‌تواند به کاهش مشروعیت و اثربخشی آن منجر شود. چارچوب‌های جدید حکمرانی آموزشی بر مفهوم سرپرستی دانشگاهی و مسئولیت نهادی تأکید دارند و پیشنهاد می‌کنند که دانشگاه‌ها از ساختارهای چندرشته‌ای برای ارزیابی و هدایت سامانه‌های هوشمند استفاده کنند (Che & Perkins, 2026). این ساختارها می‌توانند متخصصان آموزش، اخلاق، حقوق، فناوری اطلاعات، داده، اعضای هیئت علمی و نمایندگان دانشجویان را دربر گیرند. چنین ترکیبی امکان می‌دهد که تصمیمات فناورانه از انحصار نگاه فنی خارج شده و پیامدهای تربیتی و اخلاقی نیز در نظر گرفته شود.

در حوزه‌های تخصصی نیز ضرورت سازگار کردن کاربرد هوش مصنوعی با اهداف رشته‌ای مورد تأکید قرار گرفته است. برای مثال، ادغام هوش مصنوعی در آموزش علوم نیازمند توجه هم‌زمان به راهنماهای جهانی، پایداری، مسئولیت علمی و توانایی دانشجویان برای درک و نقد

خروجی‌های ماشینی است (Li et al., 2025). در آموزش پزشکی نیز تحول دیجیتال نیازمند بازطراحی سازوکارهای حکمرانی و مدیریت آموزشی است تا فناوری به‌صورت هماهنگ با اهداف تربیت حرفه‌ای و مسئولیت اجتماعی مورد استفاده قرار گیرد (Zhang & Shi, 2026). این شواهد نشان می‌دهد که حکمرانی اخلاقی نمی‌تواند صرفاً مجموعه‌ای عمومی از اصول باشد، بلکه باید امکان ترجمه این اصول به نیازهای حوزه‌های تخصصی و موقعیت‌های متفاوت آموزشی را نیز فراهم کند.

علاوه بر این، مباحث مربوط به علوم اجتماعی و انسانی نشان می‌دهد که ورود هوش مصنوعی به فضای دانشگاهی صرفاً مسئله‌ای فنی یا آموزشی نیست، بلکه با پرسش‌های معرفت‌شناختی و فلسفی درباره تولید دانش، سوگیری، نمایندگی و قدرت نیز پیوند دارد. سامانه‌های هوش مصنوعی در انتخاب، پردازش و بازنمایی اطلاعات نقش ایفا می‌کنند و ممکن است برخی دیدگاه‌ها، زبان‌ها یا سنت‌های معرفتی را بیش از دیگران بازنمایی کنند. از این‌رو، نقد سوگیری و بررسی پیامدهای معرفتی فناوری برای حفظ تکثر و عدالت معرفتی ضروری است (Mokadi et al., 2026). در چنین شرایطی، استقلال علمی باید شامل حق دانشگاه برای نقد منطق الگوریتمی و جلوگیری از تبدیل ابزارهای تجاری به مرجع ضمنی تعیین اعتبار معرفتی نیز باشد.

مرور ادبیات موجود همچنین نشان می‌دهد که چالش‌های اخلاقی هوش مصنوعی مولد اغلب به‌صورت جداگانه بررسی شده‌اند؛ برای مثال، بخشی از مطالعات بر تمامیت علمی تمرکز دارند (Güneş & Kaban, 2025; Kangwa et al., 2025)، برخی به مسائل استفاده مسئولانه و چالش‌های آموزشی پرداخته‌اند (Abedin et al., 2026; Mhlono, 2026)، بخشی بر سیاست‌ها و مقررات دانشگاهی متمرکز بوده‌اند (Fraidan, 2025; McGinty, 2025) و برخی دیگر موضوع شفافیت، پاسخ‌گویی و مسئولیت را برجسته کرده‌اند (Basil et al., 2026; Toquero, 2025). این پراکندگی نشان می‌دهد که هنوز نیاز به چارچوبی یکپارچه وجود دارد که بتواند ابعاد اخلاقی، تربیتی و نهادی را در یک ساختار منسجم به هم مرتبط سازد. چنین چارچوبی باید علاوه بر پاسخ به پرسش «چه استفاده‌ای از هوش مصنوعی مجاز است»، به پرسش‌های عمیق‌تری مانند «چه کسی تصمیم می‌گیرد»، «چه کسی پاسخ‌گوست»، «چه ارزش‌هایی باید حفظ شوند»، «چه کسانی ممکن است متحمل زیان شوند» و «حدود مشروع و واگذاری قضاوت انسانی به فناوری چیست» نیز پاسخ دهد.

همچنین، بخش قابل‌توجهی از ادبیات موجود در بسترهای متفاوت ملی و نهادی شکل گرفته و نشان می‌دهد که آمادگی و ظرفیت دانشگاه‌ها برای تدوین و اجرای حکمرانی اخلاقی یکسان نیست. تفاوت در زیرساخت، فرهنگ دانشگاهی، قوانین داده، دسترسی به فناوری و ساختار مدیریت می‌تواند نحوه اجرای اصول اخلاقی را تغییر دهد (Mmari, 2026; Sultana et al., 2026). بنابراین، انتقال مستقیم یک الگوی سیاستی از یک نظام آموزش عالی به نظام دیگر لزوماً مناسب نیست و ضروری است چارچوب‌ها با زمینه نهادی و فرهنگی سازگار شوند. این ضرورت به‌ویژه در دانشگاه‌هایی که در مرحله توسعه سیاست‌های هوش مصنوعی قرار دارند اهمیت دارد، زیرا تصمیمات اولیه آن‌ها می‌تواند جهت‌گیری آینده استفاده از فناوری را تعیین کند.

از منظر فلسفه تعلیم و تربیت، می‌توان گفت مسئله بنیادی در مواجهه دانشگاه با هوش مصنوعی، حفظ تقدم غایت بر ابزار است. فناوری زمانی از نظر تربیتی مشروع است که در خدمت اهدافی مانند رشد عقلانی، خودآیینی، عدالت، حقیقت‌جویی، مسئولیت‌پذیری و شکل‌گیری توان داور باشد. اگر منطق بهره‌وری و خودکارسازی جای این اهداف را بگیرد، دانشگاه ممکن است بخشی از رسالت تربیتی خود را به سامانه‌هایی واگذار کند که اساساً برای تحقق اهداف آموزشی یا اخلاقی طراحی نشده‌اند. به همین دلیل، حکمرانی اخلاقی باید نه پس از استقرار فناوری، بلکه پیش از آن و در مرحله انتخاب، طراحی، خرید، پیاده‌سازی و ارزیابی حضور داشته باشد. مطالعات جدید نیز بر اهمیت

رویکرد پیش‌نگر و مسئولانه در ادغام هوش مصنوعی تأکید دارند و نشان می‌دهند که سیاست‌های صرفاً واکنشی نمی‌توانند پیچیدگی این فناوری را مدیریت کنند (Che & Perkins, 2026; Jamil et al., 2026).

در مجموع، ادبیات پژوهش نشان می‌دهد که هوش مصنوعی در آموزش عالی هم‌زمان واجد ظرفیت تحول‌آفرین و مخاطرات اخلاقی قابل توجه است. بهره‌گیری مسئولانه از این فناوری مستلزم آن است که عدالت آموزشی، شفافیت، پاسخ‌گویی، تمامیت علمی، حریم خصوصی، استقلال دانشگاهی و عاملیت انسانی به صورت یکپارچه در ساختارهای سیاست‌گذاری لحاظ شوند. دانشگاه باید بتواند از ظرفیت‌های هوش مصنوعی برای بهبود آموزش و مدیریت استفاده کند، بدون آنکه اختیار حرفه‌ای استاد، استقلال فکری دانشجو، حق اعتراض، امنیت داده یا تنوع معرفتی تضعیف شود. همچنین، استفاده مسئولانه از هوش مصنوعی نیازمند آن است که همه ذی‌نفعان درک روشنی از نقش و محدودیت فناوری داشته باشند و ساختارهای رسمی برای ارزیابی، ممیزی، بازنگری و اصلاح تصمیمات وجود داشته باشد. در چنین رویکردی، هوش مصنوعی نه جایگزین کنشگر انسانی، بلکه ابزاری تحت نظارت و مسئولیت انسانی خواهد بود و دانشگاه نیز از مصرف‌کننده منفعل فناوری به نهادی فعال در تعریف قواعد اخلاقی استفاده از آن تبدیل می‌شود (McGinty, 2025; Rao & Suhasini, 2025; Zhang & Shi, 2026). بر این اساس، هدف پژوهش حاضر تبیین چارچوب حکمرانی اخلاقی هوش مصنوعی در دانشگاه‌ها با تأکید بر عدالت آموزشی، پاسخ‌گویی و استقلال علمی در چارچوب فلسفه تعلیم و تربیت است.

روش‌شناسی

پژوهش حاضر با رویکرد کیفی و با هدف تبیین و صورت‌بندی چارچوبی برای حکمرانی اخلاقی هوش مصنوعی در دانشگاه‌ها با تأکید بر عدالت آموزشی، پاسخ‌گویی و استقلال علمی در چارچوب فلسفه تعلیم و تربیت انجام شد. با توجه به ماهیت مفهومی، تفسیری و هنجاری مسئله پژوهش و ضرورت دستیابی به فهمی عمیق از دیدگاه‌ها و تجارب متخصصان درباره پیامدهای اخلاقی و تربیتی کاربرد هوش مصنوعی در محیط‌های دانشگاهی، از طرح تحلیل مضمون کیفی استفاده شد. جامعه مشارکت‌کنندگان شامل اعضای هیئت علمی، متخصصان فلسفه تعلیم و تربیت، صاحب‌نظران اخلاق فناوری و هوش مصنوعی، متخصصان مدیریت و سیاست‌گذاری آموزش عالی و مدیران دانشگاهی شاغل در دانشگاه‌ها و مراکز آموزش عالی شهر تهران بود. مشارکت‌کنندگان به صورت هدفمند و با استفاده از راهبرد نمونه‌گیری ملاکی انتخاب شدند تا افرادی وارد پژوهش شوند که از دانش نظری یا تجربه اجرایی کافی درباره آموزش عالی، سیاست‌گذاری دانشگاهی، فلسفه تعلیم و تربیت، اخلاق فناوری یا کاربرد هوش مصنوعی در محیط‌های آموزشی برخوردار باشند. معیارهای ورود شامل داشتن حداقل مدرک دکتری در رشته‌های مرتبط برای اعضای دانشگاهی، حداقل پنج سال سابقه علمی، پژوهشی یا مدیریتی مرتبط با آموزش عالی، داشتن سابقه مطالعه، تدریس، پژوهش، سیاست‌گذاری یا فعالیت اجرایی در یکی از حوزه‌های هوش مصنوعی، اخلاق حرفه‌ای، فلسفه تعلیم و تربیت، مدیریت دانشگاهی یا فناوری‌های آموزشی و تمایل آگاهانه به مشارکت در پژوهش بود. برای مدیران و کارشناسان تخصصی نیز برخورداری از حداقل پنج سال سابقه فعالیت مرتبط با سیاست‌گذاری یا مدیریت آموزش عالی در نظر گرفته شد. نمونه‌گیری تا دستیابی به اشباع نظری و عدم ظهور مضامین مفهومی جدید ادامه یافت و در نهایت ۱۸ نفر در پژوهش مشارکت کردند. تلاش شد ترکیب مشارکت‌کنندگان از نظر حوزه تخصصی و موقعیت حرفه‌ای متنوع باشد تا مسئله حکمرانی اخلاقی هوش مصنوعی صرفاً از منظر فناورانه بررسی نشود و ابعاد فلسفی، تربیتی، اخلاقی و مدیریتی آن نیز در فرایند تحلیل منعکس شود. پیش از انجام مصاحبه‌ها، هدف پژوهش، شیوه استفاده از داده‌ها، اصل محرمانگی اطلاعات و اختیار مشارکت‌کنندگان برای انصراف از ادامه همکاری برای آنان توضیح داده شد و مشارکت بر مبنای رضایت آگاهانه صورت گرفت.

برای حفظ محرمانگی نیز مشخصات هویتی مشارکت‌کنندگان در فرایند پیاده‌سازی، کدگذاری و گزارش یافته‌ها حذف و برای هر مصاحبه یک کد اختصاصی در نظر گرفته شد.

ابزار اصلی گردآوری داده‌ها مصاحبه نیمه‌ساختاریافته عمیق بود که پرسش‌های آن بر اساس هدف پژوهش، مبانی فلسفه تعلیم و تربیت و مسائل اخلاقی ناشی از گسترش کاربرد هوش مصنوعی در دانشگاه‌ها طراحی شد. تدوین راهنمای اولیه مصاحبه پس از مرور مفهومی ادبیات مربوط به حکمرانی هوش مصنوعی، اخلاق فناوری، عدالت آموزشی، مسئولیت‌پذیری سازمانی، آزادی و استقلال دانشگاهی و رابطه میان فناوری و غایات تعلیم و تربیت صورت گرفت. پرسش‌های مصاحبه به صورت باز طراحی شدند تا مشارکت‌کنندگان بتوانند برداشت‌ها، تجربه‌ها، نگرانی‌ها و پیشنهادها را خود را بدون محدود شدن به پاسخ‌های ازپیش‌تعیین‌شده بیان کنند. محورهای اصلی مصاحبه شامل برداشت مشارکت‌کنندگان از مفهوم حکمرانی اخلاقی هوش مصنوعی در دانشگاه، پیامدهای استفاده از الگوریتم‌ها و سامانه‌های مولد هوش مصنوعی برای برابری فرصت‌های آموزشی، امکان ایجاد یا تشدید سوگیری و تبعیض الگوریتمی، شفافیت فرایندهای تصمیم‌گیری مبتنی بر هوش مصنوعی، تعیین مسئولیت در برابر خطاها و پیامدهای تصمیمات الگوریتمی، حفاظت از حریم خصوصی و داده‌های دانشجویان و اعضای هیئت علمی، تأثیر هوش مصنوعی بر استقلال علمی و آزادی آموزشی، حدود مشروع مداخله فناوری در تدریس و ارزشیابی، جایگاه قضاوت حرفه‌ای استادان، چگونگی نظارت دانشگاه بر استفاده آموزشی و پژوهشی از هوش مصنوعی و اصول مورد نیاز برای طراحی چارچوبی اخلاق‌محور در این زمینه بود. علاوه بر پرسش‌های اصلی، در طول مصاحبه از پرسش‌های کاوشگر نظیر «چرا»، «چگونه»، «در چه شرایطی» و «چه سازوکاری برای تحقق این اصل لازم است» استفاده شد تا لایه‌های عمیق‌تر دیدگاه مشارکت‌کنندگان آشکار شود. مصاحبه‌ها به صورت فردی انجام گرفت و مدت هر مصاحبه بین ۴۵ تا ۷۵ دقیقه متغیر بود. با رضایت مشارکت‌کنندگان، تمامی مصاحبه‌ها ضبط و سپس به صورت کلمه‌به‌کلمه پیاده‌سازی شد. هم‌زمان با انجام مصاحبه‌ها، یادداشت‌های میدانی و یادداشت‌های تأملی پژوهشگر نیز درباره نکات برجسته، مفاهیم تکرارشونده، روابط احتمالی میان مفاهیم و موضوعات نیازمند پیگیری ثبت شد. برای بررسی کفایت و وضوح راهنمای مصاحبه، پیش از آغاز مرحله اصلی گردآوری داده‌ها، دو مصاحبه مقدماتی با افراد واجد شرایط انجام شد و بر اساس نتایج آن، ترتیب و صورت‌بندی برخی پرسش‌ها اصلاح شد؛ به گونه‌ای که ساختار نهایی مصاحبه ضمن برخورداری از انسجام موضوعی، انعطاف لازم برای پیگیری تجربه‌ها و استدلال‌های خاص هر مشارکت‌کننده را حفظ کند.

تحلیل داده‌ها به صورت هم‌زمان با فرایند گردآوری اطلاعات و با استفاده از تحلیل مضمون انجام شد. پس از انجام هر مصاحبه، فایل صوتی به طور کامل پیاده‌سازی و متن حاصل چندین بار مطالعه شد تا پژوهشگر با محتوای داده‌ها آشنایی عمیق پیدا کند. در مرحله نخست، واحدهای معنایی مرتبط با مسئله پژوهش شناسایی و برای هر یک کد اولیه متناسب با محتوای آشکار یا ضمنی آن اختصاص داده شد. در مرحله بعد، کدهای مشابه یا دارای قرابت مفهومی با یکدیگر مقایسه و در قالب مقوله‌ها و مضامین اولیه سازمان‌دهی شدند. فرایند مقایسه مستمر در سراسر تحلیل ادامه یافت و کدها و مضامین حاصل از هر مصاحبه با داده‌های مصاحبه‌های پیشین مقایسه شدند تا شباهت‌ها، تفاوت‌ها، موارد استثنا و روابط مفهومی میان آنها مشخص شود. در ادامه، مضامین اولیه از نظر انسجام درونی و تمایز بیرونی مورد بازبینی قرار گرفتند و برخی مضامین ادغام، تفکیک یا بازتعریف شدند. سپس مضامین نهایی بر اساس نسبت آنها با پرسش اصلی پژوهش و مبانی فلسفی تعلیم و تربیت در سطوح مفهومی بالاتر سازمان یافتند. در این مرحله توجه ویژه‌ای به روابط میان عدالت آموزشی، شفافیت و پاسخ‌گویی الگوریتمی، مسئولیت انسانی، حریم خصوصی، استقلال دانشگاهی، آزادی علمی، عاملیت استاد و دانشجو و نظارت نهادی معطوف شد تا چارچوب نهایی صرفاً

مجموعه‌ای از اصول مستقل نباشد، بلکه روابط ساختاری میان ابعاد مختلف حکمرانی اخلاقی هوش مصنوعی را نیز نشان دهد. فرایند کدگذاری و سازمان‌دهی داده‌ها با استفاده از نرم‌افزار MAXQDA انجام شد که امکان مدیریت نظام‌مند متون مصاحبه، دسته‌بندی کدها، بازیابی قطعات مرتبط و بررسی روابط میان مضامین را فراهم کرد. تحلیل داده‌ها تا مرحله‌ای ادامه یافت که مصاحبه‌های جدید کد یا مضمون اساسی تازه‌ای به ساختار تحلیلی اضافه نکردند و اشباع نظری حاصل شد.

برای افزایش استحکام و قابلیت اعتماد یافته‌های کیفی، از معیارهای اعتبار، قابلیت اتکا، تأییدپذیری و انتقال‌پذیری استفاده شد. اعتبار یافته‌ها از طریق درگیری مستمر با داده‌ها، مقایسه مداوم مصاحبه‌ها، بررسی موارد ناسازگار و بازبینی بخشی از برداشت‌های تحلیلی توسط مشارکت‌کنندگان تقویت شد. همچنین بخشی از متن‌های پیاده‌شده و کدگذاری‌های انجام‌شده در اختیار دو پژوهشگر آشنا با روش‌شناسی کیفی و حوزه فلسفه تعلیم و تربیت قرار گرفت و درباره تناسب کدها، نحوه تشکیل مضامین و مرزبندی مفهومی آنها تبادل نظر شد. اختلاف‌های احتمالی از طریق بازگشت به متن اصلی مصاحبه‌ها و بحث تحلیلی برطرف شد. برای افزایش قابلیت اتکا و تأییدپذیری نیز مسیر تصمیم‌گیری تحلیلی، تغییرات ایجادشده در ساختار کدها و مضامین، یادداشت‌های تحلیلی و دلایل ادغام یا تفکیک مقوله‌ها ثبت شد تا فرایند رسیدن از داده خام به مضامین نهایی قابل پیگیری باشد. افزون بر این، با ارائه توصیف نسبتاً غنی از ویژگی‌های مشارکت‌کنندگان، زمینه مطالعه، فرایند گردآوری داده‌ها و منطق تحلیل، امکان قضاوت درباره انتقال‌پذیری یافته‌ها به محیط‌های دانشگاهی مشابه فراهم شد. در مرحله نهایی، مضامین استخراج‌شده با یکدیگر یکپارچه شدند و بر مبنای روابط میان اصول، الزامات و سازوکارهای شناسایی‌شده، چارچوب مفهومی حکمرانی اخلاقی هوش مصنوعی در دانشگاه‌ها با محوریت عدالت آموزشی، پاسخ‌گویی و استقلال علمی تدوین شد.

یافته‌ها

در پژوهش حاضر، ۱۸ نفر از صاحب‌نظران و متخصصان دانشگاهی و اجرایی شهر تهران مشارکت داشتند. از این تعداد، ۱۱ نفر مرد و ۷ نفر زن بودند. از نظر حوزه تخصصی، ۵ نفر در حوزه فلسفه تعلیم و تربیت و مبانی نظری آموزش عالی، ۴ نفر در حوزه هوش مصنوعی و فناوری‌های آموزشی، ۳ نفر در حوزه اخلاق فناوری و اخلاق حرفه‌ای، ۳ نفر در حوزه مدیریت و سیاست‌گذاری آموزش عالی و ۳ نفر در سطوح مدیریتی دانشگاه‌ها دارای تجربه اجرایی مرتبط با سیاست‌گذاری آموزشی، پژوهشی یا تحول دیجیتال بودند. سن مشارکت‌کنندگان بین ۳۶ تا ۶۲ سال قرار داشت و میانگین سابقه علمی، پژوهشی یا مدیریتی آنان حدود ۱۵ سال بود. ۱۴ نفر دارای مدرک دکتری تخصصی و ۴ نفر از مدیران و متخصصان اجرایی دارای مدرک کارشناسی ارشد یا دکتری حرفه‌ای مرتبط بودند. از نظر سابقه فعالیت، ۴ نفر بین ۵ تا ۱۰ سال، ۷ نفر بین ۱۱ تا ۱۵ سال و ۷ نفر بیش از ۱۵ سال تجربه حرفه‌ای داشتند. تنوع تخصصی مشارکت‌کنندگان موجب شد مسئله حکمرانی اخلاقی هوش مصنوعی نه صرفاً از منظر فناوری، بلکه در پیوند با فلسفه تعلیم و تربیت، عدالت آموزشی، اخلاق حرفه‌ای، سیاست‌گذاری دانشگاهی، استقلال علمی، حقوق داده و مسئولیت نهادی مورد بررسی قرار گیرد. تحلیل مصاحبه‌ها نشان داد که دیدگاه مشارکت‌کنندگان حول سه حوزه محوری عدالت آموزشی، پاسخ‌گویی و استقلال علمی سازمان یافته است؛ باین‌حال، در سطحی بالاتر، یک مضمون فراگیر با عنوان «حکمرانی انسان‌محور و ارزش‌بنیاد هوش مصنوعی در دانشگاه» شناسایی شد که سه محور یادشده را در قالب یک نظام منسجم اخلاقی و نهادی به یکدیگر پیوند می‌دهد.

جدول ۱. ساختار کلی مضامین حکمرانی اخلاقی هوش مصنوعی در دانشگاه‌ها

توسعه فردی و تحول سازمانی

مضمون فراگیر	مضامین اصلی	زیرمضامین کلیدی	تعداد مشارکت کنندگان اشاره کننده
حکمرانی انسان محور و ارزش بنیاد هوش مصنوعی در دانشگاه	عدالت آموزشی در دسترسی، طراحی و پیامدهای هوش مصنوعی	دسترسی عادلانه، رفع شکاف دیجیتال، پیشگیری از سوگیری الگوریتمی، دسترسی پذیری برای گروه های متنوع، عدالت در ارزیابی، حمایت هدفمند از گروه های آسیب پذیر	۱۸
حکمرانی انسان محور و ارزش بنیاد هوش مصنوعی در دانشگاه	پاسخ گویی، شفافیت و مسئولیت انسانی	شفافیت الگوریتمی، قابلیت توضیح تصمیمات، تعیین مسئولیت، نظارت انسانی، ممیزی الگوریتم ها، سازوکار اعتراض و جبران	۱۸
حکمرانی انسان محور و ارزش بنیاد هوش مصنوعی در دانشگاه	استقلال علمی و صیانت از عاملیت دانشگاهی	آزادی علمی، استقلال حرفه ای استاد، عاملیت دانشجو، عدم جایگزینی قضاوت انسانی، استقلال دانشگاه از پلتفرم ها، تنوع معرفتی	۱۷
حکمرانی انسان محور و ارزش بنیاد هوش مصنوعی در دانشگاه	حکمرانی داده و حریم خصوصی	حداقل گرایی داده، رضایت آگاهانه، امنیت داده، محدودیت استفاده ثانویه، مالکیت و دسترسی، حفاظت از داده های حساس	۱۷
حکمرانی انسان محور و ارزش بنیاد هوش مصنوعی در دانشگاه	سواد هوش مصنوعی و ظرفیت اخلاقی	سواد الگوریتمی، آموزش اعضای هیئت علمی، توانمندسازی دانشجویان، شناخت محدودیت های هوش مصنوعی، سواد انتقادی و مسئولیت حرفه ای	۱۶
حکمرانی انسان محور و ارزش بنیاد هوش مصنوعی در دانشگاه	تنظیم گری مشارکتی و نظارت نهادی	کمیته اخلاق هوش مصنوعی، سیاست های داخلی دانشگاه، مشارکت ذی نفعان، ارزیابی پیش از استقرار، بازنگری مستمر، گزارش دهی عمومی	۱۸
حکمرانی انسان محور و ارزش بنیاد هوش مصنوعی در دانشگاه	صیانت از اصالت فرایند تعلیم و تربیت	تقدم اهداف تربیتی، حفظ رابطه استاد و دانشجو، جلوگیری از وابستگی شناختی، حمایت از تفکر انتقادی، مرزبندی کمک و جایگزینی، اصالت یادگیری	۱۷

همان گونه که در جدول ۱ مشاهده می شود، تحلیل داده های کیفی به شناسایی یک مضمون فراگیر، هفت مضمون اصلی و مجموعه ای از زیرمضامین مرتبط منجر شد. مضمون فراگیر «حکمرانی انسان محور و ارزش بنیاد هوش مصنوعی در دانشگاه» بیانگر این برداشت مشترک در میان مشارکت کنندگان بود که استفاده دانشگاه از هوش مصنوعی نباید صرفاً بر اساس معیارهای فنی مانند سرعت، دقت، کاهش هزینه یا افزایش بهره وری ارزیابی شود، بلکه مشروعیت آن باید بر مبنای میزان انطباق با غایات تعلیم و تربیت، شأن و عاملیت انسانی، عدالت، مسئولیت پذیری و استقلال دانشگاه سنجیده شود. دو مضمون عدالت آموزشی و پاسخ گویی در مصاحبه تمامی ۱۸ مشارکت کننده مطرح شدند و از بالاترین اجماع برخوردار بودند. تنظیم گری مشارکتی و نظارت نهادی نیز مورد تأکید همه مشارکت کنندگان قرار گرفت و نشان داد که از دیدگاه آنان، اصول اخلاقی بدون تبدیل شدن به سازوکارهای رسمی، رویه های اجرایی و ساختارهای پاسخ گو اثربخشی محدودی خواهند داشت. استقلال علمی، حکمرانی داده و صیانت از اصالت تعلیم و تربیت در ۱۷ مصاحبه برجسته شد و سواد هوش مصنوعی و ظرفیت اخلاقی نیز در ۱۶ مصاحبه مورد توجه قرار گرفت. تحلیل روابط میان مضامین نشان داد که این مؤلفه ها مستقل از یکدیگر عمل نمی کنند؛ برای مثال، شفافیت الگوریتمی شرطی برای پاسخ گویی و درعین حال ابزاری برای کشف تبعیض است، استقلال علمی بدون کنترل دانشگاه بر داده ها و سامانه های فناورانه شکننده می شود و عدالت آموزشی بدون توسعه سواد هوش مصنوعی می تواند به دسترسی صوری به فناوری محدود شود. از این رو، یافته ها حکمرانی اخلاقی هوش مصنوعی را به عنوان یک نظام چندسطحی نشان دادند که در آن اصول فلسفی و تربیتی باید هم زمان در طراحی فناوری، تصمیم گیری دانشگاهی، سیاست گذاری، شیوه تدریس و ارزشیابی و رابطه میان استاد، دانشجو و سامانه هوشمند حضور داشته باشند.

جدول ۲. مؤلفه‌های عدالت آموزشی در حکمرانی اخلاقی هوش مصنوعی

زیرمضمون	مؤلفه‌های استخراج شده	دالت برای حکمرانی دانشگاهی	فراوانی مشارکت کنندگان
دسترسی عادلانه به فناوری	دسترسی برابر به ابزارهای هوش مصنوعی، زیرساخت دیجیتال، اینترنت، تجهیزات و خدمات آموزشی هوشمند	جلوگیری از تبدیل توان مالی و فنی دانشجویان به مزیت آموزشی ناعادلانه	۱۷
کاهش شکاف دیجیتال و الگوریتمی	تفاوت در مهارت دیجیتال، زبان، رشته تحصیلی، وضعیت اقتصادی و توان استفاده از ابزارها	ارائه حمایت، آموزش و زیرساخت جبرانی برای گروه‌های کمتر برخوردار	۱۵
پیشگیری از سوگیری الگوریتمی	سوگیری در داده‌های آموزشی، توصیه‌گرها، پذیرش، ارزیابی، تشخیص عملکرد و تحلیل رفتار	آزمون و ممیزی منظم سامانه‌ها پیش و پس از استقرار	۱۸
عدالت در ارزشیابی	جلوگیری از وابستگی کامل نمره‌دهی و تشخیص تخلف به سامانه‌های خودکار	الزام به بررسی انسانی در تصمیمات دارای پیامد تحصیلی مهم	۱۶
دسترس‌پذیری و فراگیری	توجه به دانشجویان دارای معلولیت، تفاوت‌های زبانی و سبک‌های متفاوت یادگیری	طراحی فراگیر و پیش‌بینی گزینه‌های جایگزین غیرالگوریتمی	۱۴
شفافیت تخصیص فرصت	شفاف‌سازی معیارهای توصیه بورسیه، پذیرش، حمایت آموزشی یا فرصت پژوهشی	جلوگیری از تبعیض پنهان در تخصیص منابع و فرصت‌های دانشگاهی	۱۵
عدالت ترمیمی و امکان اصلاح	حق اعتراض، اصلاح داده نادرست، بازبینی تصمیم و جبران پیامدهای ناموجه	ایجاد مسیر رسمی بازنگری تصمیمات متأثر از الگوریتم	۱۶

بر اساس جدول ۲، عدالت آموزشی در دیدگاه مشارکت‌کنندگان مفهومی گسترده‌تر از «دسترسی برابر به یک ابزار فناورانه» داشت. مشارکت‌کنندگان بر این باور بودند که حتی در صورت فراهم بودن یک سامانه واحد برای همه دانشجویان، تفاوت در کیفیت دسترسی، میزان سواد دیجیتال، توان اقتصادی، مهارت زبانی، رشته تحصیلی و آشنایی با شیوه تعامل با سامانه‌های هوش مصنوعی می‌تواند پیامدهای نابرابر ایجاد کند. بنابراین، برابری صوری دسترسی به فناوری برای تحقق عدالت کافی نیست و دانشگاه باید شرایط استفاده منصفانه و مؤثر از فناوری را نیز فراهم کند. بیشترین تأکید در این محور به پیشگیری از سوگیری الگوریتمی اختصاص داشت و همه ۱۸ مشارکت‌کننده به گونه‌ای به این مسئله اشاره کردند. از دیدگاه آنان، ورود سامانه‌های هوشمند به فرایندهایی مانند پذیرش، ارزشیابی، پیش‌بینی افت تحصیلی، پیشنهاد مسیر آموزشی، کشف تقلب یا تخصیص فرصت‌های آموزشی، خطر بازتولید سوگیری‌های موجود در داده‌های تاریخی را در پی دارد. از این رو، ممیزی مستمر داده و الگوریتم، مقایسه پیامدها برای گروه‌های مختلف و ممنوعیت اتکای انحصاری به خروجی الگوریتم در تصمیمات حساس به عنوان الزامات عدالت مطرح شد. دسترسی عادلانه نیز در ۱۷ مصاحبه مورد توجه قرار گرفت و مشارکت‌کنندگان تأکید کردند که دانشگاه باید از شکل‌گیری نظام آموزشی دوگانه‌ای جلوگیری کند که در آن دانشجویان دارای توان مالی بیشتر به نسخه‌های پیشرفته‌تر سامانه‌های هوش مصنوعی دسترسی داشته باشند و مزیت تحصیلی قابل توجهی کسب کنند. در مجموع، داده‌ها نشان داد که عدالت آموزشی در حکمرانی هوش مصنوعی دارای سه سطح توزیعی، رویه‌ای و ترمیمی است؛ عدالت توزیعی به توزیع عادلانه فناوری و فرصت‌ها، عدالت رویه‌ای به منصفانه و شفاف بودن فرایندهای تصمیم‌گیری و عدالت ترمیمی به امکان اعتراض، اصلاح و جبران تصمیمات ناموجه مربوط می‌شود. در این چارچوب، دانشگاه تنها مسئول تأمین فناوری نیست، بلکه باید پیامدهای متفاوت فناوری برای گروه‌های گوناگون را نیز ارزیابی و برای کاهش نابرابری‌های حاصل از آن مداخله کند.

جدول ۳. مؤلفه‌های پاسخ‌گویی، شفافیت و مسئولیت در استفاده دانشگاهی از هوش مصنوعی

زیرمضمون	شاخص‌های اصلی	سازوکار اجرایی پیشنهادی	فراوانی مشارکت کنندگان
----------	---------------	-------------------------	------------------------

توسعه فردی و تحول سازمانی

شفافیت استفاده از هوش مصنوعی	اطلاع‌رسانی درباره وجود، هدف و دامنه استفاده از سامانه	اعلام رسمی کاربرد هوش مصنوعی به دانشجویان و کارکنان	۱۸
قابلیت توضیح تصمیمات	قابل فهم بودن مبنای تصمیم یا توصیه الگوریتمی	ارائه توضیح متناسب با اهمیت تصمیم و مخاطب	۱۶
تعیین مسئولیت انسانی و سازمانی	مشخص بودن مسئول خطا، نظارت و پیامد تصمیم	تعریف زنجیره مسئولیت از طراح تا مدیر و کاربر نهایی	۱۸
نظارت انسانی معنادار	امکان مداخله، توقف، اصلاح یا رد خروجی سامانه	الزام بازبینی انسانی در تصمیمات حساس آموزشی و پژوهشی	۱۸
ممیزی و ارزیابی دوره‌ای اخلاقی	سنجش دقت، سوگیری، امنیت، اثرات آموزشی و انطباق	ممیزی مستقل و بازنگری دوره‌ای سامانه‌ها	۱۵
حق اعتراض و بازبینی	امکان اعتراض به نمره، طبقه‌بندی، تشخیص تخلف یا توصیه الگوریتمی	ایجاد مرجع انسانی مستقل برای رسیدگی به اعتراض‌ها	۱۷
مستندسازی و تصمیمات	ثبت منبع داده، نسخه سامانه، مداخله انسانی و فرایند تصمیم	ایجاد ردپای حسابرسی برای تصمیمات حساس	۱۴
پاسخ‌گویی تأمین‌کنندگان فناوری	مسئولیت شرکت‌های بیرونی در قبال داده، امنیت و عملکرد	قراردادهای شفاف با شروط اخلاقی، امنیتی و قابلیت ممیزی	۱۵

مطابق جدول ۳، پاسخ‌گویی یکی از منسجم‌ترین مضامین استخراج شده از مصاحبه‌ها بود. همه مشارکت‌کنندگان تأکید داشتند که به‌کارگیری هوش مصنوعی نباید به ایجاد «خلأ مسئولیت» منجر شود؛ به‌گونه‌ای که هنگام وقوع خطا، تبعیض، نقض حریم خصوصی یا آسیب آموزشی، دانشگاه مسئولیت را به شرکت سازنده، مدیر به واحد فناوری و کاربر به خروجی الگوریتم منتقل کند. از دیدگاه مشارکت‌کنندگان، اصل بنیادین باید این باشد که مسئولیت نهایی تصمیمات دانشگاهی همچنان انسانی و نهادی باقی بماند، حتی اگر بخشی از تحلیل، پیش‌بینی یا پیشنهاد توسط هوش مصنوعی انجام شده باشد. به همین دلیل، تعیین روشن زنجیره مسئولیت و نظارت انسانی معنادار توسط همه ۱۸ مشارکت‌کننده مورد تأکید قرار گرفت. مفهوم «نظارت انسانی معنادار» در داده‌ها صرفاً به حضور صوری یک فرد برای تأیید خروجی سامانه محدود نبود، بلکه شامل برخورداری ناظر از دانش کافی، دسترسی به اطلاعات لازم، اختیار واقعی برای رد یا اصلاح پیشنهاد الگوریتم و اختصاص زمان کافی برای بررسی تصمیم بود. حق اعتراض و بازبینی نیز در ۱۷ مصاحبه مطرح شد و به‌ویژه در مواردی مانند تشخیص تقلب علمی، نمره‌دهی خودکار، طبقه‌بندی دانشجویان، پیش‌بینی ریسک افت تحصیلی و تخصیص فرصت‌های آموزشی اهمیت بیشتری داشت. مشارکت‌کنندگان معتقد بودند که دانشجو یا عضو هیئت علمی باید بداند چه زمانی تصمیمی تحت تأثیر هوش مصنوعی قرار گرفته، بتواند مبنای آن را در حد معقول درک کند و امکان درخواست بازبینی انسانی داشته باشد. شفافیت نیز به معنای افشای همه جزئیات فنی الگوریتم تلقی نشد، بلکه متناسب بودن سطح توضیح با اهمیت تصمیم، پیامد آن و نیاز مخاطب مورد تأکید قرار گرفت. بدین ترتیب، یافته‌ها نشان داد که پاسخ‌گویی در حکمرانی اخلاقی هوش مصنوعی از ترکیب پنج عنصر «قابلیت شناسایی مسئول»، «شفافیت کاربرد»، «قابلیت توضیح»، «امکان اعتراض» و «قابلیت حسابرسی» تشکیل می‌شود. در غیاب هر یک از این عناصر، دانشگاه ممکن است از نظر ظاهری از یک سامانه پیشرفته و کارآمد استفاده کند، اما از منظر اخلاقی قادر به دفاع از مشروعیت تصمیمات حاصل از آن نباشد.

جدول ۴. مؤلفه‌های استقلال علمی، عاملیت انسانی و الزامات نهادی حکمرانی اخلاقی

حوزه	زیرمضمون	مؤلفه‌های اصلی	فراوانی مشارکت‌کنندگان
استقلال علمی	استقلال حرفه‌ای استاد	حفظ اختیار استاد در طراحی تدریس، انتخاب منابع، ارزشیابی و پذیرش یا رد توصیه هوش مصنوعی	۱۷

Personal Development and Organizational Transformation

استقلال علمی	آزادی پژوهش و اندیشه	جلوگیری از جهت‌دهی الگوریتمی به موضوعات، منابع و مسیرهای پژوهش	۱۵
عاملیت انسانی	عاملیت دانشجو	استفاده از هوش مصنوعی به‌عنوان ابزار تقویت یادگیری و نه جایگزین تفکر، نوشتن و استدلال	۱۷
عاملیت انسانی	تقدم قضاوت انسانی	حفظ نقش قضاوت حرفه‌ای در تصمیمات آموزشی، اخلاقی و پژوهشی پیچیده	۱۸
استقلال نهادی	کاهش وابستگی به پلتفرم‌های تجاری	ارزیابی قراردادهای، قابلیت انتقال داده، جلوگیری از قفل‌شدگی فناوریانه و کنترل دانشگاه بر سیاست استفاده	۱۴
حکمرانی داده	حریم خصوصی و حداقل‌گرایی داده	گردآوری فقط داده ضروری، محدودیت نگهداری و جلوگیری از استفاده ثانویه بدون مجوز	۱۷
ظرفیت اخلاقی	سواد انتقادی هوش مصنوعی	آموزش محدودیت‌ها، سوگیری‌ها، خطاهای مولد، مسئولیت استفاده و ارزیابی خروجی	۱۶
ساختار نهادی	کمیته حکمرانی و اخلاق هوش مصنوعی	حضور متخصصان آموزش، فلسفه، حقوق، فناوری، اعضای هیئت علمی و نمایندگان دانشجویان	۱۸
ساختار نهادی	ارزیابی اثر پیش از استقرار	بررسی اثرات آموزشی، اخلاقی، حقوقی و داده‌ای پیش از استفاده گسترده	۱۶
ساختار نهادی	بازنگری مستمر سیاست‌ها	اصلاح سیاست‌ها متناسب با تغییر فناوری، تجربه کاربران و پیامدهای مشاهده‌شده	۱۵
فلسفه تعلیم و تربیت	اصالت رابطه آموزشی	جلوگیری از جایگزینی ارتباط استاد-دانشجو با تعامل صرفاً ماشینی	۱۶
فلسفه تعلیم و تربیت	صیانت از تفکر انتقادی	استفاده از هوش مصنوعی برای توسعه پرسشگری و نه تولید پاسخ آماده و وابستگی شناختی	۱۷

یافته‌های جدول ۴ نشان داد که استقلال علمی از دیدگاه مشارکت‌کنندگان یکی از نقاط تمایز اساسی میان حکمرانی اخلاقی هوش مصنوعی در دانشگاه و مدیریت عمومی فناوری در سایر سازمان‌هاست. مشارکت‌کنندگان تأکید داشتند که دانشگاه صرفاً یک سازمان ارائه‌دهنده خدمات نیست، بلکه نهادی معرفتی و تربیتی است که مشروعیت آن به امکان پرسشگری آزاد، تنوع دیدگاه‌ها، استقلال حرفه‌ای استادان و رشد توان داوری دانشجویان وابسته است. بر این اساس، هوش مصنوعی هنگامی با فلسفه تعلیم و تربیت دانشگاهی سازگار تلقی می‌شود که ظرفیت انسانی را تقویت کند، نه اینکه به تدریج اختیار معرفتی و تصمیم‌گیری را از استاد و دانشجو سلب کند. همه ۱۸ مشارکت‌کننده بر تقدم قضاوت انسانی در تصمیمات پیچیده و ضرورت ایجاد ساختار رسمی حکمرانی و اخلاق هوش مصنوعی تأکید کردند. از نظر آنان، خروجی سامانه‌های هوشمند می‌تواند نقش مشورتی، تحلیلی یا پشتیبان داشته باشد، اما تصمیم نهایی در حوزه‌هایی مانند ارزیابی علمی، تخلف پژوهشی، پذیرش یا رد دانشجو، تشخیص صلاحیت، طراحی آموزشی و ارزیابی کیفیت تدریس باید تحت مسئولیت انسانی باقی بماند. استقلال حرفه‌ای استاد و عاملیت دانشجو نیز هر یک در ۱۷ مصاحبه برجسته شد. مشارکت‌کنندگان هشدار دادند که استفاده بدون مرز از سامانه‌های مولد می‌تواند از یک سو استاد را به مجری پیشنهادهای الگوریتمی و از سوی دیگر دانشجو را به مصرف‌کننده پاسخ‌های آماده تبدیل کند؛ وضعیتی که با هدف تربیتی دانشگاه در پرورش استقلال فکری، توان استدلال، خلاقیت و مسئولیت‌پذیری معرفتی ناسازگار است.

در سطح نهادی نیز یافته‌ها نشان داد که استقلال دانشگاه تنها در رابطه استاد و دانشجو تعریف نمی‌شود، بلکه وابستگی گسترده دانشگاه به پلتفرم‌های تجاری خارجی می‌تواند شکل جدیدی از وابستگی نهادی ایجاد کند. مشارکت‌کنندگان به مخاطراتی مانند قفل‌شدگی فناوریانه، دشواری انتقال داده‌ها، فقدان شفافیت درباره نحوه آموزش مدل‌ها، استفاده ثانویه از داده‌های دانشگاهی و تغییر یک‌جانبه شرایط خدمات توسط شرکت‌های ارائه‌دهنده اشاره کردند. از این‌رو، کنترل دانشگاه بر داده‌ها، شروط قراردادی، امکان ممیزی و حق توقف استفاده از سامانه در صورت تعارض با ارزش‌های دانشگاهی به‌عنوان بخشی از استقلال علمی و نهادی تلقی شد. همچنین، ۱۷ مشارکت‌کننده بر حریم خصوصی و حداقل‌گرایی داده تأکید داشتند و معتقد بودند که استفاده از هوش مصنوعی نباید دانشگاه را به محیطی برای نظارت دائمی بر رفتار، حضور،

فعالیت شناختی و تعاملات آموزشی دانشجویان تبدیل کند. بر اساس این دیدگاه، هر نوع گردآوری داده باید ضرورت مشخص، هدف محدود و دوره نگهداری تعریف شده داشته باشد و استفاده ثانویه از داده‌های آموزشی بدون اطلاع و مبنای مشروع صورت نگیرد.

تحلیل نهایی روابط میان مضامین نشان داد که چارچوب حکمرانی اخلاقی هوش مصنوعی در دانشگاه‌ها از سه لایه به هم پیوسته تشکیل می‌شود. لایه نخست، «اصول هنجاری و فلسفی» است که عدالت آموزشی، کرامت انسانی، آزادی علمی، استقلال دانشگاهی، مسئولیت‌پذیری و اصالت رابطه تربیتی را دربر می‌گیرد. این لایه تعیین می‌کند که دانشگاه چرا و با چه غایتی باید از هوش مصنوعی استفاده کند و چه ارزش‌هایی نباید در مسیر بهره‌وری فناوری قربانی شوند. لایه دوم، «الزامات رویه‌ای و عملیاتی» است که شامل شفافیت، قابلیت توضیح، نظارت انسانی، حق اعتراض، ممیزی الگوریتمی، حفاظت از داده‌ها، ارزیابی پیامدها و تعیین مسئولیت است. این لایه اصول انتزاعی اخلاقی را به قواعد قابل اجرا تبدیل می‌کند. لایه سوم، «سازوکارهای نهادی و ظرفیت‌ساز» را شامل می‌شود که دربرگیرنده کمیته اخلاق و حکمرانی هوش مصنوعی، آموزش مستمر اعضای هیئت علمی و دانشجویان، سیاست‌های رسمی استفاده، ارزیابی پیش از استقرار سامانه‌ها، نظارت بر قراردادهای فناوری و بازنگری دوره‌ای سیاست‌هاست. بر اساس یافته‌ها، ضعف در هر یک از این سه لایه می‌تواند کارکرد دو لایه دیگر را محدود کند؛ برای نمونه، وجود اصول اخلاقی بدون ساختار نظارتی قابلیت اجرا ندارد، وجود مقررات بدون سواد هوش مصنوعی به اجرای صوری منجر می‌شود و وجود سامانه‌های نظارتی بدون مبنای فلسفی روشن ممکن است به کنترل اداری و محدود کردن استقلال علمی بینجامد.

در مجموع، تحلیل داده‌ها نشان داد که «حکمرانی اخلاقی هوش مصنوعی در دانشگاه» از دیدگاه مشارکت‌کنندگان نه به معنای منع یا محدودسازی کلی فناوری و نه به معنای پذیرش نامشروط نوآوری است، بلکه بر استفاده مشروط، قابل توضیح، قابل اعتراض، تحت نظارت انسانی و همسو با اهداف تربیتی دانشگاه دلالت دارد. در این الگو، عدالت آموزشی مشخص می‌کند که منافع و مخاطرات فناوری چگونه میان افراد و گروه‌ها توزیع شوند؛ پاسخ‌گویی تعیین می‌کند چه کسی در برابر تصمیمات و پیامدهای سامانه مسئول است؛ و استقلال علمی مرزهای مداخله فناوری در قضاوت، آموزش، پژوهش و تولید معرفت را مشخص می‌سازد. حکمرانی داده، سواد هوش مصنوعی، نظارت نهادی و صیانت از رابطه آموزشی نیز به عنوان سازوکارهای پشتیبان این سه محور عمل می‌کنند. بنابراین، یافته‌های پژوهش نشان می‌دهد که چارچوب مطلوب حکمرانی اخلاقی باید از منطق «هوش مصنوعی به جای انسان» فاصله گرفته و بر منطق «هوش مصنوعی تحت هدایت، مسئولیت و داوری انسان» استوار باشد. در چنین چارچوبی، معیار موفقیت فناوری صرفاً میزان خودکارسازی یا صرفه‌جویی در زمان نیست، بلکه میزان توان آن در تقویت فرصت‌های برابر، حفظ آزادی و استقلال معرفتی، افزایش کیفیت تصمیم‌گیری و حمایت از رشد فکری و اخلاقی اعضای جامعه دانشگاهی است.

بحث و نتیجه‌گیری

پژوهش حاضر با هدف تبیین چارچوب حکمرانی اخلاقی هوش مصنوعی در دانشگاه‌ها با تأکید بر عدالت آموزشی، پاسخ‌گویی و استقلال علمی و در چارچوب فلسفه تعلیم و تربیت انجام شد. یافته‌ها نشان داد که حکمرانی اخلاقی هوش مصنوعی در دانشگاه را نمی‌توان به تدوین چند دستورالعمل فنی یا محدود کردن استفاده از ابزارهای مولد تقلیل داد، بلکه این مفهوم واجد ماهیتی چندبعدی، نهادی و هنجاری است که باید هم‌زمان به عدالت آموزشی، پاسخ‌گویی، شفافیت، مسئولیت انسانی، استقلال علمی، حکمرانی داده، سواد هوش مصنوعی، تنظیم‌گری

مشارکتی و صیانت از اصالت فرایند تعلیم و تربیت توجه کند. مضمون فراگیر استخراج شده یعنی «حکمرانی انسان‌محور و ارزش‌بنیاد هوش مصنوعی در دانشگاه» نشان داد که از دیدگاه مشارکت‌کنندگان، معیار اصلی ارزیابی فناوری در دانشگاه نباید صرفاً کارآمدی، سرعت یا کاهش هزینه باشد، بلکه فناوری باید در نسبت با غایات آموزشی، حقوق‌ذی‌نفعان، کرامت انسانی و رسالت معرفتی دانشگاه سنجیده شود. این نتیجه با دیدگاه‌هایی همسو است که بر استفاده مسئولانه و تربیت‌محور از هوش مصنوعی مولد تأکید کرده و نشان می‌دهند که بهره‌گیری مؤثر از این فناوری تنها زمانی مشروعیت آموزشی دارد که فرصت‌های یادگیری را توسعه دهد و در عین حال خطرهای اخلاقی، سوگیری، وابستگی و کاهش کیفیت یادگیری را کنترل کند (Abedin et al., 2026; Mhlongo, 2026). همچنین، مرور نظام‌مند اخلاق هوش مصنوعی در آموزش نشان داده است که اصولی مانند عدالت، شفافیت، پاسخ‌گویی، حریم خصوصی و عاملیت انسانی باید به‌صورت مرتبط و نه جدا از یکدیگر در سیاست‌گذاری آموزشی لحاظ شوند (Yan & Chau, 2025).

نخستین محور اصلی یافته‌ها، عدالت آموزشی بود. نتایج نشان داد که عدالت در استفاده دانشگاهی از هوش مصنوعی دارای ابعاد توزیعی، رویه‌ای و ترمیمی است. در بعد توزیعی، مشارکت‌کنندگان بر دسترسی برابر به ابزارهای هوشمند، زیرساخت مناسب، اینترنت، خدمات آموزشی و حمایت از گروه‌های کمتر برخوردار تأکید کردند. در بعد رویه‌ای، مسئله عدم تبعیض الگوریتمی، شفاف بودن معیارهای تصمیم‌گیری و ممیزی سامانه‌ها برجسته شد و در بعد ترمیمی نیز حق اعتراض، بازبینی و اصلاح تصمیمات الگوریتمی اهمیت یافت. این یافته بیانگر آن است که عدالت آموزشی صرفاً با فراهم کردن یک ابزار یکسان برای همه تحقق نمی‌یابد، زیرا تفاوت در سواد دیجیتال، سطح دسترسی، زبان، توان اقتصادی و مهارت استفاده می‌تواند حتی در شرایط برخورداری صوری برابر، به پیامدهای متفاوت منجر شود. این نتیجه با مطالعاتی همسو است که خطر بازتولید نابرابری و سوگیری در سامانه‌های آموزشی مبتنی بر هوش مصنوعی را برجسته کرده‌اند (Mokadi et al., 2026; Yan & Chau, 2025). همچنین، تأکید Che و Perkins بر حکمرانی فراگیر و حمایت از یادگیرندگان نشان می‌دهد که استفاده اخلاقی از هوش مصنوعی باید فراتر از کارایی سامانه، به آثار آن بر گروه‌های مختلف دانشجویان و میزان دسترسی عادلانه آنان توجه کند (Che & Perkins, 2026).

تأکید مشارکت‌کنندگان بر پیشگیری از سوگیری الگوریتمی نیز یکی از یافته‌های برجسته پژوهش بود. این نتیجه را می‌توان در پرتو ماهیت داده‌محور هوش مصنوعی تبیین کرد. سامانه‌های هوشمند بر اساس داده‌های پیشین آموزش می‌بینند و در صورتی که این داده‌ها منعکس‌کننده نابرابری‌های موجود باشند، فناوری می‌تواند همان الگوها را بازتولید یا حتی تقویت کند. این مسئله به‌ویژه در فرایندهایی مانند پذیرش، رتبه‌بندی، پیش‌بینی عملکرد، تشخیص تقلب و تخصیص حمایت آموزشی اهمیت دارد. Mokadi و همکاران نیز با تمرکز بر ابعاد اخلاقی هوش مصنوعی در علوم اجتماعی، به مسئله سوگیری، قدرت و عدم بی‌طرفی کامل فناوری اشاره کرده‌اند (Mokadi et al., 2026). از سوی دیگر، مسئله آمادگی نهادی نیز با عدالت پیوند مستقیم دارد، زیرا دانشگاهی که از نظر زیرساخت، سیاست، آموزش و منابع انسانی آماده نیست، نمی‌تواند تضمین کند که فناوری به‌صورت برابر و مسئولانه در اختیار همه گروه‌ها قرار گیرد. این یافته با پژوهش Mmari درباره آمادگی مؤسسات آموزش عالی برای پذیرش اخلاقی هوش مصنوعی و نیز مطالعه Sultana و همکاران درباره آمادگی هوش مصنوعی در سطح دانشگاه، صنعت و دولت هم‌راستا است (Mmari, 2026; Sultana et al., 2026).

محور دوم یافته‌ها، پاسخ‌گویی، شفافیت و مسئولیت انسانی بود. یافته‌ها نشان داد که مشارکت‌کنندگان به‌طور جدی با شکل‌گیری «خلأ مسئولیت» در تصمیمات مبتنی بر هوش مصنوعی مخالفت داشتند و معتقد بودند که حتی در صورت استفاده گسترده از سامانه‌های خودکار،

مسئولیت نهایی باید نزد انسان و نهاد دانشگاه باقی بماند. این نتیجه از منظر اخلاقی اهمیت اساسی دارد؛ زیرا واگذاری مسئولیت به الگوریتم می‌تواند رابطه میان تصمیم و پاسخ‌گویی را تضعیف کند. در یک نظام دانشگاهی، تصمیمات مرتبط با نمره، پذیرش، تشخیص تخلف، ارزیابی علمی یا تخصیص فرصت‌ها آثار مستقیم بر حقوق و آینده افراد دارند و بنابراین باید امکان انتساب مسئولیت وجود داشته باشد. **Fraidan** نیز در بحث ارزیابی زبانی مبتنی بر هوش مصنوعی بر اهمیت شفافیت رویه‌ای و پاسخ‌گویی حقوقی تأکید کرده و نشان داده است که مشروعیت تصمیمات فناورانه بدون امکان توضیح و اعتراض تضعیف می‌شود (**Fraidan, 2025**). همچنین، مطالعات مربوط به سیاست‌های دانشگاهی هوش مصنوعی بر ضرورت تدوین ساختارهای روشن مسئولیت، قواعد استفاده و سازوکارهای نظارتی تأکید دارند (**Basil et al., 2026**; **Joshi et al., 2026**).

یافته مهم دیگر در این محور، مفهوم «نظارت انسانی معنادار» بود. نتایج نشان داد که حضور یک فرد در انتهای فرایند تصمیم‌گیری به‌تنهایی برای اخلاقی تلقی شدن نظام کافی نیست، بلکه ناظر انسانی باید دانش، اختیار، اطلاعات و زمان کافی برای بررسی و در صورت لزوم رد یا اصلاح خروجی سامانه داشته باشد. این یافته اهمیت تمایز میان نظارت صوری و نظارت واقعی را برجسته می‌کند. اگر نقش انسان تنها تأیید خروجی الگوریتم باشد، در عمل مسئولیت به فناوری منتقل شده است، حتی اگر از نظر اداری یک فرد امضاکننده نهایی باشد. این نتیجه با رویکردهای جدید در زمینه استفاده مسئولانه از هوش مصنوعی همسو است که بر حفظ عاملیت و قضاوت حرفه‌ای انسان تأکید دارند (**Abedin et al., 2026**; **McGinty, 2025**). از سوی دیگر، **Razi** و همکاران نیز نشان داده‌اند که استفاده از هوش مصنوعی مولد همواره با مصالحه میان سرعت، استقلال، اصالت و کیفیت همراه است و بنابراین تصمیم‌گیری حرفه‌ای نباید به خروجی سامانه تقلیل یابد (**Razi et al., 2025**).

حق اعتراض و قابلیت بازبینی نیز از نتایج مهم این پژوهش بود. مشارکت‌کنندگان تأکید کردند که دانشجو یا استاد باید بدانند در چه مواردی هوش مصنوعی بر یک تصمیم اثر گذاشته و چگونه می‌تواند درخواست بازبینی انسانی ارائه کند. این موضوع، پاسخ‌گویی را از سطح صرفاً سازمانی به سطح حقوق فردی گسترش می‌دهد. در واقع، شفافیت زمانی معنادار است که بتوان بر اساس اطلاعات ارائه‌شده اقدام کرد. سیاست‌هایی که صرفاً اعلام می‌کنند از هوش مصنوعی استفاده می‌شود، اما مسیر اعتراض یا اصلاح را مشخص نمی‌کنند، از منظر اخلاقی ناکافی‌اند. مطالعات مربوط به سیاست‌های دانشگاهی نشان داده‌اند که بسیاری از دانشگاه‌ها در حال حرکت به سمت تدوین دستورالعمل‌هایی هستند که استفاده از هوش مصنوعی را از حالت مبهم خارج کرده و حدود مسئولیت کاربر و نهاد را مشخص کنند (**Basil et al., 2026**; **Toquero, 2025**). یافته‌های **Joshi** و همکاران نیز نشان می‌دهد که تفاوت میان مؤسسات از نظر وضوح و انسجام سیاست‌های هوش مصنوعی قابل توجه است و وجود سیاست روشن می‌تواند نقش مهمی در کاهش ابهام اخلاقی ایفا کند (**Joshi et al., 2026**).

محور سوم یافته‌ها، استقلال علمی و عاملیت انسانی بود. نتایج نشان داد که مشارکت‌کنندگان استفاده از هوش مصنوعی را تا زمانی مشروع می‌دانند که فناوری جایگزین قضاوت حرفه‌ای استاد و فعالیت شناختی دانشجو نشود. این یافته از منظر فلسفه تعلیم و تربیت بسیار مهم است، زیرا هدف آموزش دانشگاهی صرفاً تولید سریع پاسخ یا انجام کار نیست، بلکه رشد توانایی استدلال، تحلیل، قضاوت، نقد و مسئولیت‌معرفتی است. اگر دانشجو در فرایند یادگیری به استفاده منفعلانه از خروجی هوش مصنوعی عادت کند، ممکن است بخشی از تجربه شناختی لازم برای رشد استقلال فکری از میان برود. **Razi** و همکاران نیز تأکید کرده‌اند که هوش مصنوعی مولد نباید به‌عنوان ابزاری بی‌محدودیت و

همه‌کاره تلقی شود و استفاده از آن با مبادله‌هایی میان کارآمدی و اصالت یادگیری همراه است (Razi et al., 2025). همچنین، McGinty بر ضرورت حفظ اصالت و تمامیت علمی در استفاده اعضای هیئت علمی از هوش مصنوعی تأکید کرده است (McGinty, 2025). یافته مربوط به عاملیت دانشجو نیز با مطالعات حوزه تمامیت علمی همسو بود. مشارکت‌کنندگان معتقد بودند که سیاست‌های صرفاً منع‌کننده نمی‌توانند پاسخ کاملی به مسئله هوش مصنوعی باشند و دانشگاه باید به‌جای تمرکز انحصاری بر کشف تقلب، شیوه‌های ارزشیابی و آموزش را بازطراحی کند. این نتیجه با یافته‌های Güneş و Kaban درباره چالش‌های اخلاقی و تمامیت علمی و همچنین مرور Kangwa و همکاران درباره ایجاد تعادل میان استفاده مؤثر از هوش مصنوعی مولد و حفظ تمامیت دانشگاهی هم‌راستا است (Güneş & Kaban, 2025; Kangwa et al., 2025). مرور Jamil و همکاران نیز نشان داده است که ادغام ChatGPT در آموزش عالی با طیفی از معضلات اخلاقی همراه است که با ممنوعیت ساده قابل حل نیست و نیازمند چارچوب‌های روشن مسئولیت و بازطراحی فرایندهای یادگیری است (Jamil et al., 2026).

یکی دیگر از یافته‌های مهم، استقلال نهادی دانشگاه در برابر پلتفرم‌ها و شرکت‌های فناوری بود. مشارکت‌کنندگان به خطر وابستگی فناورانه، قفل‌شدگی پلتفرمی، انتقال داده و کاهش کنترل دانشگاه بر فرایندهای تصمیم‌گیری اشاره کردند. این نتیجه نشان می‌دهد که استقلال علمی تنها به آزادی فردی استاد محدود نیست، بلکه دارای بعد ساختاری و نهادی نیز هست. اگر دانشگاه به سامانه‌ای وابسته شود که منطق تصمیم، نحوه پردازش داده یا شرایط استفاده آن خارج از کنترل دانشگاه باشد، بخشی از استقلال نهادی عملاً به بیرون منتقل می‌شود. Souli و Pierrakeas در بررسی روندهای مدیریت دانشگاهی و ادغام فناوری‌های جدید نیز نشان داده‌اند که تحول دیجیتال دانشگاه نیازمند بازنگری در ساختارهای مدیریت، کنترل و تصمیم‌گیری است (Souli & Pierrakeas, 2025). همچنین، Shi و Zhang در بحث حکمرانی مشارکتی آموزش پزشکی در عصر آموزش هوشمند بر ضرورت بازسازی ساختارهای مدیریتی و هماهنگ کردن فناوری با اهداف نظام تربیت حرفه‌ای تأکید کرده‌اند (Zhang & Shi, 2026).

حکمرانی داده و حریم خصوصی نیز در یافته‌های پژوهش جایگاه مهمی داشت. مشارکت‌کنندگان بر حداقل‌گرایی در گردآوری داده، تعیین هدف روشن، محدودیت در استفاده ثانویه و حفاظت از داده‌های حساس تأکید کردند. این نتیجه قابل تبیین است؛ زیرا دانشگاه‌ها حجم قابل توجهی از داده‌های تحصیلی، رفتاری، پژوهشی و هویتی را در اختیار دارند و ادغام سامانه‌های هوش مصنوعی می‌تواند انگیزه و ظرفیت بیشتری برای گردآوری و تحلیل این داده‌ها ایجاد کند. اگر این فرایند بدون قواعد روشن انجام شود، محیط دانشگاه می‌تواند از فضای یادگیری به فضای نظارت دائمی تبدیل شود. مرورهای اخلاق هوش مصنوعی در آموزش نیز حریم خصوصی و حفاظت از داده را از جمله نگرانی‌های اصلی معرفی کرده‌اند (Mhlongo, 2026; Yan & Chau, 2025). از این منظر، حکمرانی داده نه یک مسئله فرعی، بلکه بخشی از معماری اصلی حکمرانی اخلاقی هوش مصنوعی است.

نتایج همچنین نشان داد که سواد هوش مصنوعی و ظرفیت اخلاقی پیش‌شرط اجرای موفق سیاست‌های حکمرانی است. وجود دستورالعمل بدون توانایی کاربران برای درک محدودیت‌ها، خطاها، سوگیری‌ها و قابلیت‌های واقعی فناوری نمی‌تواند رفتار مسئولانه ایجاد کند. این یافته با مطالعاتی همسو است که بر ضرورت آموزش اعضای هیئت علمی و دانشجویان در زمینه استفاده مسئولانه از فناوری تأکید کرده‌اند (Abedin et al., 2026; McGinty, 2025). در این چارچوب، سواد هوش مصنوعی باید فراتر از مهارت فنی در نوشتن دستور یا استفاده از ابزار باشد و ابعاد انتقادی، اخلاقی و معرفتی را نیز شامل شود. دانشجو و استاد باید بتوانند درباره اعتبار خروجی، منشأ اطلاعات،

احتمال خطا، سوگیری و حدود قابل قبول استفاده قضاوت کنند. این موضوع در حوزه‌های تخصصی نیز اهمیت دارد. Li و همکاران در زمینه آموزش شیمی بر ضرورت هماهنگ کردن ادغام هوش مصنوعی با راهنماهای جهانی و اهداف توسعه پایدار تأکید کرده‌اند و نشان داده‌اند که استفاده مسئولانه از فناوری باید در پیوند با اهداف گسترده‌تر آموزشی و اجتماعی تعریف شود (Li et al., 2025).

یافته مربوط به تنظیم‌گری مشارکتی و ایجاد کمیته‌های چندرشته‌ای نیز با ادبیات موجود همخوان است. مشارکت‌کنندگان بر این باور بودند که تصمیم درباره استفاده از هوش مصنوعی نباید صرفاً به واحد فناوری اطلاعات واگذار شود، بلکه متخصصان آموزش، اخلاق، حقوق، فناوری، اعضای هیئت علمی و نمایندگان دانشجویان باید در فرایند سیاست‌گذاری مشارکت داشته باشند. این نتیجه با مفهوم «سرپرستی دانشگاهی» در مطالعه Che و Perkins همسو است که حکمرانی یادگیری مبتنی بر هوش مصنوعی را نیازمند ساختارهای اخلاقی و مشارکت چندذی‌نفعی می‌داند (Che & Perkins, 2026). همچنین، Sultana و همکاران نشان داده‌اند که آمادگی هوش مصنوعی به همکاری میان دانشگاه، صنعت و دولت وابسته است و نمی‌توان آن را صرفاً مسئله‌ای درون‌سازمانی در نظر گرفت (Sultana et al., 2026). بنابراین، حکمرانی مؤثر به شبکه‌ای از نقش‌ها، مسئولیت‌ها و تعاملات نیاز دارد.

در مجموع، نتایج پژوهش حاضر نشان می‌دهد که حکمرانی اخلاقی هوش مصنوعی در دانشگاه‌ها باید بر سه ستون اصلی عدالت آموزشی، پاسخ‌گویی و استقلال علمی استوار شود و این سه ستون از طریق حکمرانی داده، سواد هوش مصنوعی، نظارت انسانی، مشارکت ذی‌نفعان و ارزیابی مستمر حمایت شوند. این چارچوب با جهت‌گیری کلی ادبیات اخیر همسو است که از رویکردهای ساده ممنوعیت یا پذیرش نامحدود فاصله گرفته و بر استفاده مشروط، مسئولانه و ارزش‌محور از هوش مصنوعی تأکید دارد (Rao & Suhasini, 2025; Toquero, 2025). همچنین، یافته‌ها نشان می‌دهد که مسئله اصلی آینده آموزش عالی احتمالاً این نخواهد بود که آیا دانشگاه‌ها از هوش مصنوعی استفاده خواهند کرد یا خیر، بلکه این خواهد بود که چه نوع استفاده‌ای مشروع، عادلانه، قابل توضیح و سازگار با رسالت تربیتی دانشگاه است. از این منظر، حکمرانی اخلاقی باید فناوری را در خدمت اهداف تعلیم و تربیت قرار دهد، نه اینکه اهداف دانشگاه را با منطق فناوری بازتعریف کند.

از جمله محدودیت‌های پژوهش حاضر آن است که مشارکت‌کنندگان از دانشگاه‌ها و نهادهای آموزش عالی شهر تهران انتخاب شدند و اگرچه تنوع تخصصی مناسبی در میان آنان وجود داشت، ممکن است دیدگاه‌های آنان بازتاب کامل شرایط دانشگاه‌های سایر استان‌ها، دانشگاه‌های کوچک‌تر، مؤسسات غیردولتی یا مراکز آموزش عالی با سطح متفاوتی از زیرساخت دیجیتال نباشد. ماهیت کیفی پژوهش نیز به این معناست که یافته‌ها با هدف تعمیم آماری تولید نشده‌اند و بیشتر برای تبیین عمیق ابعاد مفهومی و نهادی مسئله به کار می‌روند. علاوه بر این، سرعت تحول فناوری‌های هوش مصنوعی سبب می‌شود برخی مصادیق و چالش‌های اجرایی در فاصله زمانی کوتاهی تغییر کنند. همچنین، مطالعه حاضر بر دیدگاه خبرگان و متخصصان متمرکز بود و تجربه مستقیم دانشجویان، کارکنان اداری و ارائه‌دهندگان فناوری به صورت مستقل مورد بررسی قرار نگرفت. از این رو، چارچوب حاصل باید در بسترهای متنوع‌تر و با مشارکت گروه‌های گسترده‌تر آزمون و تکمیل شود.

برای پژوهش‌های آینده پیشنهاد می‌شود چارچوب استخراج‌شده در این مطالعه با روش‌های کمی یا آمیخته اعتبارسنجی شود و اهمیت نسبی ابعاد و مؤلفه‌های آن در میان گروه‌های مختلف دانشگاهی مورد سنجش قرار گیرد. همچنین، انجام پژوهش‌های تطبیقی میان دانشگاه‌های دولتی و غیردولتی، دانشگاه‌های علوم پزشکی و غیرپزشکی، دانشگاه‌های بزرگ و کوچک و نیز نظام‌های آموزش عالی کشورهای مختلف می‌تواند به شناسایی عوامل زمینه‌ای مؤثر بر حکمرانی اخلاقی کمک کند. بررسی دیدگاه مستقل دانشجویان، استادان، مدیران، متخصصان فناوری و سیاست‌گذاران نیز می‌تواند اختلاف برداشت میان ذی‌نفعان را آشکار سازد. انجام مطالعات طولی برای ارزیابی تغییر سیاست‌ها و

پیامدهای استفاده از هوش مصنوعی در طول زمان، طراحی و آزمون ابزار سنجش بلوغ حکمرانی اخلاقی هوش مصنوعی و بررسی رابطه این بلوغ با متغیرهایی مانند اعتماد، عدالت ادراک شده، کیفیت یادگیری و تمامیت علمی نیز می تواند مسیرهای مهمی برای توسعه این حوزه فراهم کند.

در سطح عملی، پیشنهاد می شود دانشگاه ها چارچوب رسمی حکمرانی هوش مصنوعی را به عنوان بخشی از نظام سیاست گذاری آموزشی و پژوهشی خود تدوین کنند و در آن نقش ها، حدود مسئولیت و فرایندهای تصمیم گیری را به روشنی مشخص سازند. تشکیل کمیته دائمی و چندرشته ای برای ارزیابی سامانه های هوش مصنوعی، انجام ارزیابی اخلاقی پیش از استقرار، ممیزی دوره ای الگوریتم ها و ایجاد سازوکار مشخص برای اعتراض و بازبینی تصمیمات ضروری است. دانشگاه ها همچنین باید برای دانشجویان، اعضای هیئت علمی و مدیران دوره های مستمر سواد هوش مصنوعی و اخلاق فناوری برگزار کنند و سیاست های استفاده از ابزارهای مولد را به صورت شفاف، قابل فهم و متناسب با رشته های مختلف تنظیم نمایند. در فرایند خرید یا عقد قرارداد با ارائه دهندگان فناوری نیز باید مالکیت داده، نحوه پردازش، امنیت، حق ممیزی، امکان انتقال داده و شرایط خروج از پلتفرم به طور دقیق مشخص شود. در نهایت، هرگونه استفاده از هوش مصنوعی باید بر اصل حفظ نظارت انسانی، امکان مداخله و اصلاح، حمایت از استقلال علمی و تقویت اهداف تربیتی دانشگاه استوار باشد تا فناوری در نقش ابزار پشتیبان رشد انسانی و علمی باقی بماند.

تشکر و قدردانی

از تمامی کسانی که در انجام این پژوهش ما را همراهی کردند تشکر و قدردانی به عمل می آید.

مشارکت نویسندگان

در نگارش این مقاله تمامی نویسندگان نقش یکسانی ایفا کردند.

تعارض منافع

در انجام مطالعه حاضر، هیچ گونه تضاد منافی وجود ندارد.

موازین اخلاقی

در تمامی مراحل پژوهش حاضر اصول اخلاقی مرتبط با نشر و انجام پژوهش رعایت گردیده است.

Extended Abstract

Introduction

The rapid expansion of artificial intelligence (AI), particularly generative AI, has transformed the operational, pedagogical, and epistemic landscape of higher education. Universities are increasingly integrating AI-enabled systems into teaching, assessment, academic advising, research support, administrative decision-making, and student services. These technologies offer substantial opportunities for personalized learning, rapid feedback, content generation, data-informed decision-making, and improved institutional efficiency. However, their adoption also raises fundamental ethical questions concerning educational justice, accountability, transparency, academic integrity, privacy, human agency, and academic autonomy. Accordingly, the central challenge facing universities is no longer whether AI should be used, but how its use should be governed in ways that remain consistent with the educational, ethical, and epistemic missions of higher education (Abedin et al., 2026; Mhlongo, 2026; Yan & Chau, 2025).

From the perspective of the philosophy of education, AI governance cannot be reduced to technical regulation or risk management. Higher education is not merely a system for transmitting information or increasing instructional efficiency; it is also expected to cultivate critical reasoning, intellectual autonomy, responsible judgment, epistemic integrity, and meaningful relationships between teachers and learners. The growing use of generative AI in writing, assessment, research, and knowledge production therefore creates a potential tension between technological efficiency and the formative purposes of education. While AI can support teaching and learning, excessive reliance on automated systems may weaken independent thinking, reduce students' active cognitive engagement, and transform faculty members from autonomous academic decision-makers into passive users of algorithmic recommendations. This concern is consistent with research emphasizing that generative AI should not be treated as a universal solution and that its use involves important trade-offs between efficiency, authenticity, autonomy, and educational quality (McGinty, 2025; Razi et al., 2025).

Educational justice represents one of the most important ethical dimensions of AI integration in universities. Although AI may improve access to learning resources and individualized educational support, its benefits may not be distributed equally among students. Differences in access to advanced tools, digital infrastructure, financial resources, language proficiency, digital literacy, and prior technological experience can produce unequal educational outcomes even when access is formally available to all. Moreover, AI systems may reproduce existing social, cultural, linguistic, or institutional biases when trained on historically unbalanced datasets. Such risks are particularly important when AI is used for admissions, student assessment, academic risk prediction, plagiarism detection, resource allocation, or recommendations that influence educational opportunities. Previous research has therefore emphasized fairness, inclusion, bias mitigation, and equitable access as foundational requirements for ethically responsible AI in education (Che & Perkins, 2026; Mokadi et al., 2026; Yan & Chau, 2025).

Accountability is another critical issue. As AI systems become increasingly involved in educational and administrative decision-making, responsibility can become fragmented among software developers, technology providers, university administrators, faculty members, and end users. This creates the possibility of an accountability gap in which no actor assumes clear responsibility for errors, discriminatory outputs, privacy violations, or harmful educational consequences. Procedural transparency and legal accountability have been identified as essential for maintaining legitimacy in AI-mediated assessment and decision-making (Fraidan, 2025). At the institutional level, universities are beginning to develop policies and guidelines governing generative AI; however, considerable variation remains in the comprehensiveness, clarity, enforceability, and ethical maturity of these policies (Basil et al., 2026; Joshi et al., 2026; Toquero, 2025). This suggests that universities require governance models that move beyond fragmented rules toward integrated systems of responsibility, oversight, explanation, appeal, and institutional learning.

Academic integrity has also become a central concern in the use of generative AI. Traditional distinctions between independent work, acceptable assistance, and academic misconduct have become increasingly difficult to maintain. Policies based exclusively on prohibition or detection may be inadequate because AI detection systems themselves are imperfect and because the educational challenge extends beyond identifying unauthorized use. Universities increasingly need to reconsider assessment design, disclosure requirements, expectations of authorship, and the development of responsible AI literacy among students and faculty. Studies have highlighted the need to balance

effective use of generative AI with the preservation of academic integrity, authenticity, and responsible academic practice (Güneş & Kaban, 2025; Jamil et al., 2026; Kangwa et al., 2025).

Academic autonomy constitutes a further dimension of ethical governance. Universities have traditionally protected the professional discretion of faculty, freedom of inquiry, diversity of academic perspectives, and institutional independence. However, increasing reliance on commercial AI platforms may create new forms of technological and epistemic dependence. Algorithmic recommendation systems may influence which sources are prioritized, how research questions are framed, and how academic decisions are made. In addition, dependence on external vendors may limit institutional control over data, platform rules, system updates, and the underlying logic of automated recommendations. The broader digital transformation of university administration has already demonstrated the importance of aligning technological innovation with institutional governance and decision-making structures (Souli & Pierrakeas, 2025). Similarly, the restructuring of educational governance in smart learning environments requires attention to institutional coordination, human oversight, and professional autonomy (Zhang & Shi, 2026).

Institutional preparedness is therefore a major prerequisite for ethical AI adoption. Universities need not only technical infrastructure but also clear policies, trained personnel, ethical review mechanisms, data governance procedures, and organizational capacity for ongoing monitoring. Research on institutional readiness has shown that universities differ significantly in their ability to adopt AI responsibly and that effective implementation requires coordination among governance, infrastructure, human resources, and ethical standards (Mmari, 2026; Sultana et al., 2026). Furthermore, disciplinary contexts matter. AI integration in fields such as science and professional education requires adaptation to disciplinary goals, global ethical guidance, and broader social responsibilities (Li et al., 2025). Thus, ethical AI governance should be sufficiently general to protect common academic values while remaining flexible enough to respond to disciplinary and institutional variation.

Overall, the literature indicates a need for an integrated framework that links educational justice, accountability, academic autonomy, data governance, AI literacy, and institutional oversight within a philosophy-of-education perspective. Existing research has often addressed these issues separately, focusing on academic integrity, policy development, institutional readiness, responsible use, or transparency. What remains insufficiently developed is a unified conceptual framework that explains how these dimensions interact and how universities can institutionalize them in governance structures. Therefore, the present study aimed to develop a framework for the ethical governance of artificial intelligence in universities, with particular emphasis on educational justice, accountability, and academic autonomy within the philosophy of education.

Methods and Materials

This study employed a qualitative research design based on thematic analysis. The qualitative approach was selected because the research problem involved normative, philosophical, institutional, and interpretive dimensions that required in-depth exploration of expert perspectives rather than statistical measurement. Participants were selected from universities and higher education institutions in Tehran using purposive criterion-based sampling. The sample included faculty members and specialists in philosophy of education, AI and educational technology, technology ethics, higher education management and policy, as well as university administrators with relevant experience in educational, research, or digital transformation policy. Eligibility criteria included appropriate academic or professional qualifications, at least five years of relevant experience, demonstrated involvement in teaching, research, management, policymaking, or

professional practice related to higher education, AI, ethics, or educational technology, and willingness to participate. Sampling continued until conceptual saturation was reached, resulting in a final sample of 18 participants.

Data were collected through in-depth semi-structured interviews. The interview guide was developed based on the objectives of the study and the major conceptual domains of AI governance in higher education. Questions explored participants' understanding of ethical AI governance, fairness in access and algorithmic decision-making, transparency, accountability, human responsibility, privacy, academic freedom, faculty autonomy, student agency, appropriate limits of AI use in teaching and assessment, institutional oversight, and mechanisms for ethical governance. Interviews were conducted individually and lasted approximately 45 to 75 minutes. With participants' consent, interviews were recorded and transcribed verbatim. Field notes and analytic memos were also maintained throughout data collection.

Data analysis was conducted concurrently with data collection using thematic analysis. Each transcript was read repeatedly to achieve immersion in the data. Meaning units relevant to the research question were identified and assigned initial codes. Similar codes were then grouped into preliminary categories and themes. Through constant comparison, themes were reviewed, merged, separated, and refined to achieve internal coherence and conceptual distinction. Relationships among themes were subsequently examined to construct a higher-order governance framework. MAXQDA software was used to organize transcripts, codes, categories, and thematic relationships. Trustworthiness was enhanced through prolonged engagement with the data, peer review of coding decisions, review of selected interpretations by participants, documentation of analytic decisions, and attention to credibility, dependability, confirmability, and transferability.

Findings

The 18 participants included 11 men and 7 women. Their professional backgrounds represented five major areas: philosophy of education, AI and educational technology, ethics and professional responsibility, higher education management and policy, and university administration. Participants had between 5 and more than 15 years of relevant professional experience, with an average experience of approximately 15 years. This diversity allowed the phenomenon to be examined from philosophical, pedagogical, technological, managerial, and ethical perspectives.

The thematic analysis identified one overarching theme, seven major themes, and multiple subthemes. The overarching theme was "human-centered and value-based governance of artificial intelligence in universities." This theme reflected the shared view that the legitimacy of AI use should not be judged exclusively by efficiency, speed, or automation, but by its alignment with educational purposes, human dignity, academic values, and institutional responsibility.

The first major theme was educational justice. Participants emphasized equitable access to AI tools, reduction of digital inequalities, mitigation of algorithmic bias, inclusive design, fairness in assessment, and targeted support for vulnerable or disadvantaged student groups. Educational justice was interpreted in distributive, procedural, and restorative terms. Distributive justice referred to equitable access to technological resources and opportunities. Procedural justice concerned transparency and fairness in algorithmically supported decisions. Restorative justice referred to the right to challenge, review, and correct harmful or unfair decisions.

The second major theme was accountability, transparency, and human responsibility. Participants strongly emphasized that AI should not create an accountability gap. Final responsibility for important educational and administrative decisions should remain with identifiable human actors and institutional units. Meaningful human oversight was distinguished from nominal oversight; a human

reviewer should possess the knowledge, authority, information, and time necessary to question or override an AI-generated recommendation. Transparency, explainability, auditability, documentation, and formal appeal mechanisms were identified as essential components of accountable governance.

The third major theme was academic autonomy and human agency. Participants argued that AI should support rather than replace faculty judgment and student intellectual activity. Faculty should retain authority over curriculum design, teaching methods, assessment, and scholarly decision-making. Students should remain active participants in reasoning, writing, inquiry, and knowledge construction. Excessive dependence on generative AI was viewed as a potential threat to critical thinking, intellectual independence, and authentic learning.

The fourth major theme concerned data governance and privacy. Participants emphasized data minimization, purpose limitation, secure storage, informed use, restricted secondary use, and clear rules regarding ownership and access. They warned that universities should not become environments of continuous behavioral surveillance and that sensitive educational data should not be processed without clear necessity and safeguards.

The fifth major theme was AI literacy and ethical capacity. Participants stressed that technical skills alone were insufficient. Students, faculty, and administrators needed critical AI literacy that included understanding limitations, bias, hallucination, data risks, authorship concerns, and ethical responsibility.

The sixth major theme was participatory regulation and institutional oversight. Participants proposed multidisciplinary governance structures involving educational, legal, ethical, technological, administrative, faculty, and student representation. Pre-deployment ethical assessment, periodic auditing, continuous policy review, and documented institutional procedures were considered necessary.

The seventh major theme was preservation of the authenticity of education. Participants emphasized the primacy of educational purposes, the importance of the faculty-student relationship, prevention of cognitive dependency, preservation of critical thinking, and clear distinctions between AI assistance and substitution of human intellectual work.

Overall, the final framework consisted of three interrelated layers. The first included normative and philosophical principles such as justice, human dignity, academic freedom, responsibility, and educational authenticity. The second included procedural mechanisms such as transparency, explainability, oversight, appeal, auditing, and data protection. The third included institutional mechanisms such as governance committees, AI literacy programs, procurement standards, periodic review, and pre-deployment assessment.

Discussion and Conclusion

The findings demonstrate that ethical AI governance in universities is fundamentally a governance problem rather than merely a technology-management problem. Universities cannot ensure ethical use simply by selecting technically reliable systems or issuing isolated usage rules. Ethical governance requires an integrated architecture that connects philosophical principles, procedural safeguards, and institutional capacity. The overarching theme of human-centered and value-based governance indicates that AI must remain subordinate to the educational mission of the university rather than becoming a force that redefines that mission according to technological efficiency.

Educational justice emerged as a foundational dimension because AI can simultaneously expand opportunity and reproduce inequality. The findings suggest that formal equality of access is insufficient. Universities must consider whether different groups are able to benefit from AI to a

comparable degree and whether automated systems generate unequal outcomes. This requires monitoring not only who has access to AI, but also how different groups experience its consequences.

Accountability was equally central. The results indicate that universities should never allow AI systems to become de facto decision-makers in high-stakes academic contexts without identifiable human responsibility. Human oversight must be substantive and capable of altering outcomes. This principle is especially important for assessment, academic misconduct decisions, admissions, resource allocation, and other processes with significant consequences for individuals.

The findings on academic autonomy extend the governance debate beyond student misuse of generative AI. They show that AI can also affect faculty independence, institutional self-governance, and epistemic diversity. Universities therefore need to evaluate not only the quality of AI outputs but also the structural implications of long-term dependence on commercial platforms and opaque recommendation systems.

The study also demonstrates that ethical governance requires organizational learning. Because AI systems evolve rapidly, universities need governance structures that can monitor emerging risks, revise policies, and incorporate feedback from users. Static guidelines are unlikely to remain adequate. Ethical governance should therefore be continuous, participatory, and adaptive.

In conclusion, the study proposes that ethical AI governance in universities should be built around three core pillars: educational justice, accountability, and academic autonomy. These pillars should be supported by data governance, meaningful human oversight, critical AI literacy, participatory institutional regulation, and ongoing ethical evaluation. The central principle is that AI should function as a tool under human direction, responsibility, and judgment rather than as a substitute for human agency. A university can therefore be considered ethically prepared for AI not when it adopts the most advanced technologies, but when it can explain why those technologies are used, who is responsible for their consequences, how affected individuals can challenge decisions, and how academic freedom, fairness, and educational purposes are protected.

References

- Abedin, M. Z., Hayajneh, A., & Raahemi, B. (2026). Pedagogical Use of Responsible Generative AI in Higher Education; Opportunities and Challenges: A Systematic Literature Review. *Ai in Education*, 2(2), 11. <https://doi.org/10.3390/aieduc2020011>
- Basil, M., Soltani, A., & Mukhalalati, B. (2026). Generative Artificial Intelligence Policies and Guidelines Across Top Ranked Universities: Insights From AI Ecological Education Policy Framework for AI Integration and University Policy Development Framework for GAI. <https://doi.org/10.21203/rs.3.rs-9788323/v1>
- Che, C., & Perkins, B. (2026). Academic Stewardship Through AI-Enabled Learning Governance: Ethical Systems for Inclusive Learner Support in Higher Education. <https://doi.org/10.21203/rs.3.rs-9759894/v1>
- Fraidan, A. A. (2025). Procedural Transparency and Legal Accountability to Sustain AI-Mediated Language Assessment in Saudi Arabia. *Sage Open*, 15(4). <https://doi.org/10.1177/21582440251396113>
- Güneş, A., & Kaban, A. L. (2025). A Delphi Study on Ethical Challenges and Ensuring Academic Integrity Regarding <sc>AI</Sc> Research in Higher Education. *Higher Education Quarterly*, 79(4). <https://doi.org/10.1111/hequ.70057>
- Jamil, B., Shaoan, M. M. R., Zeyuan, P. Y., Irfan, S., Arif, M., & Golam Ali Azgar Hossain, M. (2026). Revisiting the Research on ChatGPT Integration in Higher Education: A Systematic Review on Ethical Dilemmas. *Sage Open*, 16(2). <https://doi.org/10.1177/21582440261431568>
- Joshi, A., McGoldrick, B., Mittal, N., Roy, S., Zeba, Z., Ofori, M. A., Dockery, S., Jha, N., Auerbach, J., Cadet, J., Kruger, J. S., Leeman, R., Mitchell, K., Molter, G., Naidech, A. M., Nezami, E., Orimaye, S. O., Queen, C., Shah, U., . . . Magana, L. (2026). Investigating the Use of Generative AI Policies Among ASPPH Member Schools and Programs of Public Health. *Frontiers in Public Health*, 14. <https://doi.org/10.3389/fpubh.2026.1796810>

- Kangwa, D., Msafiri, M. M., & Fute, A. (2025). Exploring the Factors That Promote a Balance Between Academic Integrity and the Effective Use of <scp>GenAI</Scp> Tools in Higher Education: A Systematic Review. *Journal of Computer Assisted Learning*, 41(5). <https://doi.org/10.1111/jcal.70109>
- Li, Y., Tolosa, L., Rivas, F., & Márquez, R. (2025). Integrating AI in Chemical Education: Navigating UNESCO Global Guidelines, Emerging Trends, and Its Intersection With Sustainable Development Goals. <https://doi.org/10.26434/chemrxiv-2025-wz4n9-v2>
- McGinty, J. (2025). Ethical Frameworks of Artificial Intelligence for Faculty: Upholding Academic Integrity and Authenticity. *New Directions for Adult and Continuing Education*, 2025(188), 15-23. <https://doi.org/10.1002/ace.70012>
- Mhlongo, T. (2026). Navigating Ethical Complexities in Educational AI: A Systematic Review of Generative Chatbot Integration in Teaching and Learning. <https://doi.org/10.21203/rs.3.rs-9419663/v1>
- Mmari, W. D. (2026). Institutional Preparedness for the Adoption of Ethical Artificial Intelligence in Tanzanian Higher Learning Institutions. <https://doi.org/10.21203/rs.3.rs-10619789/v1>
- Mokadi, T., Jarrar, O., & Yousef, A. (2026). Social Science in the Age of AI: Unveiling Opportunities, Confronting Biases, and Charting Ethical Pathways. *Philosophies*, 11(2), 52. <https://doi.org/10.3390/philosophies11020052>
- Rao, G. T., & Suhasini, N. (2025). Integrating Artificial Intelligence in Higher Education to Enhance Teaching and Learning. *Computer Applications in Engineering Education*, 33(6). <https://doi.org/10.1002/cae.70085>
- Razi, A., Bouzoubaa, L., Pessianzadeh, A., Seberger, J. S., & Rezapour, R. (2025). Not a Swiss Army Knife: Academics' Perceptions of Trade-Offs Around Generative Artificial Intelligence Use. <https://doi.org/10.21203/rs.3.rs-5686084/v1>
- Souli, S. P., & Pierrakeas, C. (2025). Contemporary Trends in University Administration With the Integration of Digital/New Technologies. *Administrative Sciences*, 15(11), 437. <https://doi.org/10.3390/admsci15110437>
- Sultana, S., Samad, R., Haque, M., Sultana, Z., Jahangir, Z., Hossain, B. M. M., Noman, R. M., & Ahmed, S. I. (2026). Bangladesh AI Readiness: Perspectives From the Academia, Industry, and Government. <https://doi.org/10.48550/arxiv.2601.12934>
- Toquero, C. M. D. (2025). Trends, Trajectories and Tomorrows Of <scp>AI</Scp> Regulations in Philippine Higher Educational Institutions. *Higher Education Quarterly*, 80(1). <https://doi.org/10.1111/hequ.70090>
- Yan, Y., & Chau, T. (2025). A Systematic Review of AI Ethics in Education. *Journal of Global Information Management*, 33(1), 1-50. <https://doi.org/10.4018/jgim.386381>
- Zhang, X. L., & Shi, P. (2026). The Digital Transformation of Medical Education Collaborative Governance in the Era of Smart Education: An Innovative Management Restructuring of Medical Talent Cultivation Systems. *Frontiers in Medicine*, 13. <https://doi.org/10.3389/fmed.2026.1781836>