Quarterly Journal of Virtual Space and Social Media
Vol. 2, Series. 7, No. 3, Autumn 2025/ 9 - 54
DOI: https://doi.org/10.22083/cssms.2025.538964.1067

A Systematic Review of Strategic Artificial Intelligence Governance Frameworks and Mechanisms in Studies from 2017-2025

Hassan Bashir, *Professor Department of Culture and Communication, Imam Sadiq University, Tehran, Iran. Email: bashir@isu.ac.ir*

Mohammad Javad Ravi, *PhD Student, Department of Culture and Communication, and researcher at the Roshd Center of Imam Sadeq University (Correspondence Author), Tehran, Iran. Email: mj.ravi@isu.ac.ir*

Extended Abstract

Introducion: Artificial intelligence, as one of the contemporary disruptive technologies, has transformed the structures of societies and industries. However, the rapid proliferation of this technology has raised questions about its management and oversight. AI governance, beyond technical aspects, requires attention to ethical, social, and legal dimensions in order to prevent negative consequences such as bias, privacy breaches, or discrimination. The swift advancement of artificial intelligence has challenged traditional regulatory frameworks that were designed for information and communication technologies. The speed of algorithm development has outpaced legislators' ability to enact up-to-date regulations, bringing challenges such as algorithmic bias, privacy violations, and potential misuse. Among the foundational and influential theories in this domain are Algorithmic Governance/Regulation and the artificial intelligence ecosystem theory. The lack of global consensus on definitions and approaches, coupled with policy fragmentation, has made the formulation of effective frameworks difficult.

Methods: Using a systematic protocol, we searched major databases (2017–2025) for "AI governance," retrieving 130 records. After de-duplication, screening, and criterion-based appraisal (English; direct governance focus; excluding first-generation symbolic AI), 63 studies remained. We extracted bibliographic/methodological data and qualitative evidence on ethics, policy, regulation, and operationalisation. Iterative thematic analysis-constant comparison, memoing, open-to-axial coding, return-to-text checks, and audit trails-ultimately produced 297 initial codes, consolidated into 16 organising codes and four overarching themes.

Results: Findings converge on four areas. (1) Ethics: prioritising transparency, explainability, bias mitigation, and meaningful human oversight to sustain trust. (2) Policies and outlooks: countries pursue divergent strategies-EU law-centric, US trust/ethics-led, and China industry-plus-social-stability-yet share goals to balance innovation with safety. (3) Coordination: persistent gaps remain in common definitions, interoperable standards, data-sharing mechanisms, and dispute-resolution, exacerbated by geopolitical frictions. (4) Governance domains: interlocking technical, ethical, and legal layers operate across national, sectoral, and organisational levels and throughout the system lifecycle, operationalised through risk classification, in-

dependent audits, model/data cards, incident reporting, procurement clauses, and other enforceable controls that translate principles into practice..

Discussion: Grounded in the review's evidence, we propose a "multi-layered governance based on risk and trust" model anchored in four pillars-ethics, policy, governance layers, and application. It operationalises values as measurable, enforceable controls; applies proportionately across technical, ethical, and legal layers; spans national, sectoral, organisational levels and the lifecycle. Practical levers include risk classification, independent audits, red-teaming, model/data cards, incident reporting, procurement clauses, and rollback protocols. The model reduces principle-to-practice gaps, improves accountability and interoperability, and preserves innovation under safety floors.

Keywords: AI governance, Systematic Review, AI ethics, Responsible development.

فصلنامه مطالعات فضای مجازی و رسانه‌های اجتماعی
سال ۲ / شماره ۳/ شماره مسلسل ۷ / پاییز ۱۴۰۴/ ۵۴ - ۹
Quarterly Journal of Virtual Space and Social Media
Vol. 2, Series. 7, No. 3, Autumn 2025/ 9 - 54
DOI: https://doi.org/10.22083/cssms.2025.538964.1067

مرور نظام‌مند چارچوب‌ها و سازوکارهای حکمرانی راهبردی هوش‌مصنوعی در مطالعات سال‌های ۲۰۱۷ـ۲۰۲۵

حسن بشیر[۱]، محمدجواد راوی[۲]

چکیده

هوش مصنوعی به‌عنوان یکی از فناوری‌های چالش‌برانگیز معاصر، ساختار جوامع و صنایع را دگرگون کرده است. بااین‌حال، گسترش پرشتاب این فناوری، پرسش‌هایی دربارهٔ مدیریت و نظارت بر آن مطرح ساخته است. پیشرفت سریع هوش مصنوعی، چارچوب‌های سنتی تنظیم‌گری را که برای فناوری‌های ارتباطی و اطلاعاتی طراحی شده بودند، به چالش کشیده است. سرعت توسعه الگوریتم‌ها از توانایی قانون‌گذاران برای وضع مقررات به‌روز پیشی گرفته و چالش‌هایی چون سوگیری الگوریتمی، نقض حریم خصوصی و سوءاستفاده‌های احتمالی به همراه داشته است. از جمله نظریه‌های بنیادین و اثرگذار در این قلمرو، می‌توان به طبقه‌بندی راسل و نورویگ و نظریه زیست‌بوم هوش مصنوعی اشاره کرد. نبود اجماع جهانی در تعاریف و رویکردها، همراه با پراکندگی سیاست‌ها، تدوین چارچوب‌های مؤثر را دشوار ساخته است. داده‌ها شامل تحلیل‌های کیفی و با جستجوی «حکمرانی هوش مصنوعی» در پایگاه‌های علمی گردآوری شد. تحلیل مضمون نتایج پژوهش‌ها، ۲۹۷ کد اولیه، ۱۶ کد سازمان‌دهنده و چهار مضمون فراگیر را استخراج کرد. یافته‌ها در چهار محور کلیدی اخلاق هوش مصنوعی با تأکید بر شفافیت و کاهش سوگیری در الگوریتم‌ها؛ سیاست‌ها و چشم‌اندازهای جهانی؛ چالش‌های هماهنگی بین‌المللی و وحدت در تعریف، توسعه و کاربست هوش مصنوعی و زمینه‌های حکمرانی هوش مصنوعی دسته‌بندی شدند. الگوی پیشنهادی این پژوهش، معماری چندلایه ریسک‌محور است که ارزش‌ها را به قیود اجرایی سنجش‌پذیر نگاشت می‌کند، لایه‌هایِ فنی، اخلاقی، حقوقی را در سطوح ملی، بخشی، سازمانی و چرخهٔ عمر پوشش داده، و در نهایت تلاش دارد شکاف مبنا تا عمل را کاهش دهد.

واژگان کلیدی

حکمرانی هوش مصنوعی، مرور نظام‌مند، اخلاق هوش مصنوعی، توسعهٔ مسئولانه.

تاریخ دریافت: ۱۴۰۴/۰۵/۱۲ تاریخ پذیرش: ۱۴۰۴/۰۸/۰۳

۱. استاد گروه فرهنگ و ارتباطات، دانشگاه امام صادق علیه السلام، تهران، ایران. bashir@isu.ac.ir

۲. دانشجوی دکتری فرهنگ و ارتباطات، پژوهشگر مرکز رشد دانشگاه امام صادق علیه السلام، تهران، ایران. (نویسنده مسئول) mj.ravi@isu.ac.ir

مقدمه

هوش مصنوعی در عصر حاضر به یکی از محرک‌های اصلی تحولات فناورانه بدل شده و نقشی بنیادین در دگرگونی ساختارهای اجتماعی، اقتصادی و صنعتی ایفا می‌کند. درک عمیق از ماهیت این فناوری از ابعاد نرم‌افزاری، سخت‌افزاری، برای حکمرانی مؤثر و آینده‌نگرانه کاملاً ضروری است؛ چرا که شناخت دقیق قابلیت‌ها، محدودیت‌ها و پیامدهای هوش مصنوعی، نه‌تنها به فهم واقع‌بینانه‌تر این فناوری منجر می‌شود، بلکه بستر تدوین راهبردهای مناسب برای مدیریت و هدایت آن در سطح جامعه را فراهم می‌سازد. بااین‌حال، توسعه شتابان هوش مصنوعی، علاوه بر ارتقای صنایع و خدمات، چالش‌های مهمی همچون سوگیری الگوریتمی، نقض حریم خصوصی و پیچیدگی مسئولیت‌های قانونی را نیز به وجود آورده است (UNESCO, 2021). اهمیت فزاینده هوش مصنوعی، نیاز به تدوین سیاست‌ها و چارچوب‌های حکمرانی هوشمندانه و چندبُعدی را به ضرورتی جدی تبدیل کرده است؛ زیرا غفلت از ابعاد اخلاقی و اجتماعی آن می‌تواند منجر به تعمیق نابرابری‌ها و کاهش اعتماد عمومی شود. تجربه جهانی نشان داده که نبود استانداردهای هماهنگ و پراکندگی سیاست‌ها، تحقق حکمرانی مؤثر را با دشواری روبه‌رو ساخته است. در پاسخ به این چالش‌ها، نهادهای بین‌المللی اقدام به تدوین اسناد و مقرراتی نظیر «پیشنهادنامه اخلاق هوش مصنوعی یونسکو» و «قانون هوش مصنوعی اتحادیه اروپا» کرده‌اند تا چارچوبی شفاف و پاسخگو برای توسعه مسئولانه این فناوری فراهم آید (European Parliament, 2024).

مسئله اساسی این مقاله، نبود درک جامع و قابل‌اتکا از وضعیت حکمرانی راهبردی هوش مصنوعی در سطح جهانی است؛ به‌ویژه در شرایطی که پیچیدگی‌های ذاتی این فناوری و سرعت تحولات آن، امکان شناخت عمیق و هستی‌شناختی را در برنامه‌های کوتاه‌مدت و میان‌مدت دشوار کرده است. این شکاف شناختی، باعث می‌شود سیاست‌گذاران و تصمیم‌گیرندگان داخلی برای تدوین خط‌مشی‌ها و پاسخگویی به چالش‌ها و فرصت‌های هوش مصنوعی، با ابهام و پراکندگی داده‌ها مواجه باشند. از سوی دیگر، اتکای صرف به مطالعات نظری یا سیاست‌گذاری‌های منفرد، قادر به ارائه تصویری جامع و راهبردی از الزامات حکمرانی هوش مصنوعی نیست. ازاین‌رو، انجام یک مطالعه فراتحلیلی نظام‌مند و مرور ادبیات معتبر بین‌المللی، ضرورتی انکارناپذیر است تا بتوان ضمن شناسایی چارچوب‌ها، الگوها و سازوکارهای اصلی حکمرانی هوش مصنوعی، مسیرهای بهینه برای سیاست‌گذاری و ارتقای ظرفیت‌های داخلی را تبیین کرد. این مقاله با تحلیل ۶۳ مطالعه منتخب خارجی (به زبان انگلیسی) منتشرشده بین

سال‌های ۲۰۱۷ تا ۲۰۲۵، از نشریات مختلف مرتبط با حکمرانی هوش مصنوعی در پی پاسخ به این پرسش کلیدی است که چارچوب‌ها و سازوکارهای حکمرانی راهبردی هوش مصنوعی چگونه در ادبیات علمی تعریف، ساختاربندی و ارزیابی شده‌اند و چه الگوهایی بیشترین تکرار و تأثیر را داشته‌اند. هدف نهایی پژوهش، فراهم‌آوردن بستر تصمیم‌سازی مبتنی بر شواهد و تقویت توان ملی برای حکمرانی اثربخش و مسئولانه هوش مصنوعی است. سؤالات پژوهش به شرح زیر است:

سؤال اصلی: چارچوب‌ها و سازوکارهای حکمرانی راهبردی هوش مصنوعی در مطالعات منتخب چگونه تعریف، ساختاربندی و ارزیابی شده‌اند؟

سؤال فرعی: کدام الگوهای حکمرانی بیشترین تکرار را دارند و ویژگی‌های مشترک آن‌ها چیست؟

پیشینه پژوهش

برای تبیین چارچوب مفهومی و جایگاه پژوهش، در این بخش پیشینه نظری و تجربی مرتبط مرور می‌شود. ابتدا مطالعات فارسی درباره حکمرانی هوش مصنوعی و حوزه‌های نزدیک (حکمرانی داده/دولت الکترونیک) و سپس پژوهش‌های بین‌المللی اخیر با رویکردهای فراتحلیل، فراترکیب و مرور نظام‌مند بررسی می‌گردد تا مؤلفه‌های کلیدی و شکاف‌های دانشی مشخص شود.

مقاله «ارائه مدل حاکمیت هوش مصنوعی برای حکمرانی دولت در جمهوری اسلامی ایران» با فراتحلیل نظام‌مند مطالعات ۲۰۰۰ تا ۲۰۲۴، ۱۲۷ پژوهش را خوشه‌بندی کرده و شش محور اصول حاکمیتی، فناوری‌های پیشرفته، مشارکت شهروندی، امنیت ـ دفاع، خدمات عمومی و اخلاق ـ حقوق را برجسته می‌کند و بر کارایی، شفافیت و چارچوب‌های اخلاقی تأکید دارد. بااین‌حال، تمرکز ملی، رویکرد عمدتاً توصیفی و فقدان معماری چندلایه و ابزار اجرایی از شکاف‌های آن است. در مقابل، مرور نظام‌مند حاضر با دامنه بین‌المللی ۲۰۱۷ تا ۲۰۲۵ با تمرکز بر مطالعات جدید و تلفیق فراتحلیل و تحلیل مضمون، فارغ از ارائه یک «مدل واحد»، به استخراج مؤلفه‌های کلیدی و صورت‌بندی معماری راهبردی چندسطحی می‌پردازد (ترابی و اقبال، ۱۴۰۳).

مقاله «فراتحلیل مطالعات دولت الکترونیک در ارتقای شاخص‌های حکمرانی خوب» (فراتحلیل و رگرسیون با CMA2) و بیشتر بر بدنه‌ای از منابع فارسی/

ملی در بازه ۱۳۹۰ـ۱۴۰۲ تکیه دارد؛ تمرکز آن بر سنجش اثر دولت الکترونیک بر شاخص‌های حکمرانی خوب (پاسخگویی، شفافیت، نظارت و...) است و به لایه‌ها و سطوح حکمرانی فناوری‌های نو کمتر می‌پردازد. در مقابل، پژوهش حاضر با رویکرد کیفی ـ ترکیبی، به محورهای کلیدی حکمرانی راهبردی هوش مصنوعی در مقیاس جهانی متمرکز است (مجدزاده و همکاران، ۱۴۰۲).

مقاله فراتحلیل کاربرد هوش مصنوعی در توسعهٔ کاربری‌های ترکیبی اراضی شهر تهران با مسئله ناکارآمدی الگوهای تفکیک کاربری در پاسخ به پیچیدگی شهری و ضرورت بهره‌گیری از هوش مصنوعی برای توسعه کاربری‌های ترکیبی پیش می‌رود. روش پژوهش، فراتحلیل و مرور نظام‌مند است که مطالعه بر ۲۱۰۷ سند و در ادامه به ۶۵ مطالعه واجد شرایط تقلیل یافت؛ نگاشت VOSviewer و سپس مدل‌یابی معادلات ساختاری با PLS-SEM/SmartPLS روی ۷ مؤلفه کاربست هوش مصنوعی و ۵ بعد توسعه. یافته‌ها حاکی از آن است که برازش قوی ($GoF=0.739$؛ $R^2=0.716$؛ $Q^2=0.506$)، و اثر غالب مدل‌سازی/شبیه‌سازی چیدمان فضایی ($t=15.858$)، سپس مشارکت ذی‌نفعان ($t=۵/۷۶۵$) و هم‌افزایی الگوریتم ($t=4.339$) در نتیجه ادغام روش‌های پیشرفته هوش مصنوعی برنامه‌ریزی داده‌محور را تقویت می‌کند؛ اما زیرساخت، کیفیت داده و عدالت الگوریتمی چالش‌اند و نیازمند سیاست‌گذاری حمایتی‌اند(آهنگری،۱۴۰۴).

مقاله «ساحت‌های چهارگانه حکمرانی هوش مصنوعی در آموزش‌وپرورش»، اشاره می‌کند برای به‌کارگیری ایمن و اثربخش هوش مصنوعی در آموزش‌وپرورش، الگوی حکمرانی منسجم لازم است. روش مقاله مطالعه کیفی با رویکرد فراترکیب و تکنیک هفت‌مرحله‌ای سندلوسکی و باروسو است در این راستا جست‌وجوی ۲۰۰۰ـ۲۰۲۴ (۲۰۹ منبع) و گزینش ۵۸ مطالعه برای ترکیب نهایی. یافته‌های پژوهش حکمرانی را در چهار ساحت تبیین می‌کند؛ خُرد (مشارکت و ارتباطات، اخلاق، مسئولیت‌پذیری، آموزش و یادگیری)، میانی (مدیریت دانش، مدیریت منابع انسانی و منابع)، کلان (انسجام، تمرکززدایی، انعطاف‌پذیری، کارایی، عدالت آموزشی، شفافیت) و فراکلان (قوانین و مقررات، ثبات، وفاق، مهارت‌های زندگی، دسترسی). در نتیجه استقرار حکمرانی کارآمد مستلزم توجه همزمان به هر چهار ساحت و ترجمه آن‌ها به سیاست‌ها و سازوکارهای اجرایی است (پورکریمی، علی‌اکبری، ۱۴۰۴).

مقالـه «حکمرانـی هـوش مصنوعـی: مضامیـن، شـکاف‌های دانـش و برنامه‌هـای آینـده باتوجه‌بـه ابعـاد اخـلاق هـوش مصنوعـی» بـا مـرور نظام‌منـد ادبیـات و به‌کارگیـری پروتـکل (PRISMA) بـه بررسـی حکمرانـی هـوش مصنوعـی می‌پـردازد. هـدف آن شناسـایی موضوعـات اصلی، شـکاف‌های دانشی و تنظیـم دسـتورکارهای آینـده اسـت. یافته‌هـا نشـان می‌دهـد حکمرانی بـر چهـار محـور فنـاوری، ذی‌نفعـان و زمینـه، مقـررات و فراینـدهـا اسـتوار اسـت. چالش‌هایـی چـون ضعـف در پیاده‌سـازی، ابهـام در اثربخشـی اصـول اخلاقـی و کمبـود راهکارهـای اجرایـی شناسـایی می‌شـود. پژوهـش گسـترش مطالعـات تجربـی، ایجـاد نهادهـای نظارتـی و ترویـج حکمرانـی مشـارکتی را پیشـنهاد می‌کنـد. چارچـوب مفهومـی تلفیقـی، پیش‌نیازهـا و پیامدهـای حکمرانی مسـئولانه را روشـن کـرده، تدویـن سیاسـت‌های اثربخـش و پیشـبرد تحقیقـات آتـی را ممکـن می‌سـازد (Papagiannidis et al, 2025).

مقاله «حکمرانـی هـوش مصنوعـی پیشـرفته: مـرور ادبیـات چالش‌هـا، گزینه‌هـا و پیشـنهادها» به بررسـی حکمرانی هـوش مصنوعی پیشـرفته از طریـق مـرور جامـع ادبیـات موجـود می‌پـردازد. هـدف آن ارائـه طبقه‌بنـدی روشـنی از پژوهش‌هـای ایـن حـوزه در سـه محـور اصلی شناسـایی چالش‌هـا، تحلیـل گزینه‌هـای حکمرانـی و تدویـن پیشـنهادهای سیاستی اسـت. یافته‌هـا نشـان می‌دهنـد کـه ایـن حـوزهٔ پژوهشـی نوظهـور حـول سـه خـط پژوهشـی متمایـز امـا مرتبـط شـکل گرفتـه اسـت:

۱. پژوهش‌هـای مسـئله‌یاب که بـه تحلیـل پارامترهـای فنـی، اسـتقرار و حکمرانـی هـوش مصنوعـی می‌پردازنـد؛
۲. پژوهش‌هـای گزینه‌یاب کـه بازیگـران کلیـدی، اهرم‌هـای حکمرانـی و راه‌هـای تأثیرگـذاری بـر آن‌هـا را بررسـی می‌کننـد؛
۳. پژوهش‌هـای تجویزی کـه بـه تدویـن سیاسـت‌های عملـی بـر اسـاس تحلیـل چالش‌هـا و گزینه‌هـا اختصـاص دارنـد. ایـن مطالعـه بـا سـازماندهی ادبیـات پراکنـده موجـود، بـه شفاف‌سـازی تحلیلی و راهبـردی ایـن حـوزه کمـک کـرده و بسـتری بـرای پژوهش‌هـای متمرکزتـر و سیاسـت‌گذاری اثربخش‌تـر در مواجهـه بـا چالش‌هـای هـوش مصنوعـی پیشـرفته فراهـم می‌کنـد (Maas, 2023).

مقاله چارچوب‌هـای حکمرانی هـوش مصنوعـی: ایـن مقالـه بـا مـرور نظام‌منـد ادبیـات، حکمرانـی هـوش مصنوعـی را بررسـی می‌کنـد. هـدف آن شناسـایی الگوهـای حکمرانـی، تحلیـل سـطوح اجـرا و ارائـه راهکارهـای مدیریـت ریسـک‌های نوظهـور اسـت. یافته‌هـا نشـان می‌دهـد انتخـاب و اجـرای چارچوب‌هـای مناسـب بـا چالش‌هایـی در چهـار محـور ذی‌نفعـان

پاسخگو، دامنه موضوعات، زمان‌بندی مداخلات و مکانیسم‌های اجرایی روبه‌روست. شکاف‌های اصلی شامل نبود سازوکار یکپارچه میان سطوح، کمبود معیارهای ارزیابی اثربخشی و فقدان راهنمای عملیاتی است. مطالعه توسعه چارچوب‌های انعطاف‌پذیر، ارزیابی مستمر و طراحی نظام چندلایه را پیشنهاد می‌کند. حکمرانی به‌عنوان فرایندی نظام‌مند برای مدیریت ریسک و همسوسازی توسعه با ارزش‌های اجتماعی تعریف می‌شود در سطوح مختلف حکمرانی (Batool et al, 2025).

مقاله «حکمرانی هوش مصنوعی: چارچوب‌های نظارتی و مسئولیت‌های اجتماعی» با تحلیل نظام‌مند ادبیات، به شناسایی خلأهای پژوهشی در حوزه اصول و راهبردهای نظارتی برای توسعه سیستم‌های هوش مصنوعی می‌پردازد. نتایج نشان می‌دهد که به‌رغم توجه روبه‌رشد به کاربردهای این فناوری، مطالعات نظام‌مند کافی در سه محور کلیدی وجود ندارد: چارچوب‌های حکمرانی پیشنهادی توسط نهادهای مختلف، پیوند میان حکمرانی و مسئولیت اجتماعی شرکتی، و ابعاد اساسی همچون پاسخگویی، شفافیت، تفسیرپذیری، انصاف، حریم خصوصی و امنیت سایبری. از جمله شکاف‌های شناسایی‌شده، می‌توان به نبود راهنمای عملی، ابهام در سازوکارهای نظارتی و کمبود استانداردهای اخلاقی جامع اشاره کرد. مقاله پیشنهاد می‌کند چارچوب‌های چندسطحی و سازوکارهای پویا تدوین شود و مسئولیت‌پذیری اجتماعی ذی‌نفعان تقویت گردد تا حکمرانی اخلاقی و مؤثر بر هوش مصنوعی تضمین شود (Camilleri, 2024).

مطالعه حاضر با رویکردی متمایز، از اخلاق‌گرایی صرف فراتر می‌رود و حکمرانی راهبردی چندلایه و ریسک‌محور را در چارچوب زیست‌بوم و حکمرانی چندسطحی صورت‌بندی می‌کند تا پیوندی عملی میان اصول هنجاری و سازوکارهای اجرایی برقرار شود. در سطح مسئله، هدف از «توصیف وضعیت» به حل سه‌گره اصلی بازتعریف می‌شود: شکاف اصل تا عمل، پراکندگی سیاستی، و فقدان تعریف/معماری مشترک؛ با تحدید دامنه به نسل دوم هوش مصنوعی شبکه‌محور برای هم‌سویی با واقعیت فنی امروز. در سطح روش، ترکیبی از مرور نظام‌مند بین‌المللی (۲۰۱۷ـ۲۰۲۵)، فراتحلیل و تحلیل مضمون با نمونه‌گیری هدفمند انگلیسی‌زبان، غربالگری چندمرحله‌ای و راهبرد دورانی کدگذاری به‌کار می‌رود؛ بی‌آنکه «یک مدل واحد» تحمیل شود، بلکه مؤلفه‌های معماری حکمرانی استخراج و منسجم می‌گردد.

جدول ۱. پیشینه پژوهش

مطالعه (سال)	دامنه/روش	تمرکز/یافتهٔ کلیدی	وجه تمایز با پژوهش حاضر
ترابی و اقبال، ۱۴۰۳	فراتحلیل نظام‌مند (۲۰۰۰ـ۲۰۲)	۶ محور حکمرانی دولت مبتنی بر هوش مصنوعی	مطالعه ملی و توصیفی
مجدزاده و همکاران، ۱۴۰۲	فراتحلیل + رگرسیون	دولت الکترونیک و شاخص‌های حکمرانی خوب	هوش مصنوعی به‌عنوان متغیر
آهنگری، ۱۴۰۴	مرور نظام‌مند PLS-SEM	کاربست هوش مصنوعی در برنامه‌ریزی شهری	بخشی/شهری و روش‌محور (PLS-SEM) با تمرکز بر کاربری اراضی
پورکریمی و علی‌اکبری، ۱۴۰۴	فراترکیب (۵۸ منبع)	۴ ساحت حکمرانی در آموزش‌وپرورش	بخشی (آموزش‌وپرورش) و داخلی با رویکرد فراترکیب ساحت‌محور
Papagiannidis et al., 2025	SLR (PRISMA)	شکاف اصل ـ عمل	اخلاق‌محور و هنجاری با تمرکز بر شکاف اصل ـ عمل
Maas, 2023	مرور ادبیات	مسائل/گزینه‌ها/پیشنهادها در AI پیشرفته	طبقه‌بندی مفهومی بدون بسط ابزارهای اجرایی
Batool et al., 2025	SLR	چالش اجرای چارچوب‌ها؛ چندلایه و ارزیابی مستمر	چارچوب‌محور با تأکید بر ارزیابی مستمر و فقدان سازوکار یکپارچه
Camilleri, 2024	مرور نظام‌مند	پیوند حکمرانی AI و مسئولیت اجتماعی	تمرکز بر مسئولیت اجتماعی شرکتی CSR
Jobin, Ienca & Vayena, 2019	نقشه‌برداری جهانی	همگرایی اصول اخلاق AI	اصول‌محور و توصیفی؛ غیرعملیاتی در سطح اجرا
Gasser & Almeida, 2017	مدل مفهومی	مدل لایه‌ای حکمرانی AI	مدل مفهومی و اولیه با جهت‌گیری کاملاً نظری

۲. ادبیات نظری

۲ـ۱. تعریف مفاهیم

۲ـ۱ـ۱. هوش مصنوعی: به طور کلی، هوش مصنوعی: «به سامانه‌ها یا ماشین‌هایی اطلاق می‌شود که توانایی انجام وظایفی را دارند که معمولاً مستلزم سطحی از هوشمندی مشابه انسان است»(کیسینجر و همکاران، ۱۴۰۳: ۴). این وظایف می‌تواند

فعالیت‌هایی مانند حل مسائل پیچیده، یادگیری از داده‌ها و تجربیات گذشته و حتی تولید محتوای خلاقانه را در برگیرد. از منظر دیگر، هوش مصنوعی را می‌توان :«عاملی عقلانی دانست که به‌گونه‌ای طراحی شده است تا در شرایط قطعی، به بهترین نتیجه ممکن دست یابد و در موقعیت‌های مبهم و نامطمئن، به مطلوب‌ترین نتیجه مورد انتظار برس» (Russell & Norvig, 2016)

۲ـ۱ـ۲. حکمرانی هوش مصنوعی: حکمرانی هوش مصنوعی مفهومی چندوجهی است با دو منظر اصلی:«از منظر توصیفی، مجموعه‌ای از سیاست‌ها، هنجارها، قوانین و نهادها را دربر می‌گیرد که توسعه و به‌کارگیری سامانه‌های هوشمند را سامان می‌دهد. از منظر هنجاری، چشم‌اندازی است که همین عوامل را به تصمیم‌گیری‌های مؤثر، ایمن، فراگیر، مشروع و منطبق با ارزش‌های انسانی هدایت می‌کند» (Black et al., 2024).

۲ـ۱ـ۳. زیست بوم هوش مصنوعی: زیست‌بوم هوش مصنوعی، «شبکه‌ای درهم‌تنیده از بازیگران (مانند شرکت‌های فناوری، دولت‌ها، دانشگاه‌ها و جامعه مدنی)، فناوری‌ها (مانند الگوریتم‌ها، زیرساخت‌های محاسباتی و داده‌ها) و فرآیندها (مانند توسعه، کاربرد و نظارت) است که با همکاری یکدیگر، ارزش‌های اقتصادی، اجتماعی و علمی تولید می‌کنند. به عنوان مثال، یک شرکت فناوری ممکن است الگوریتمی برای تشخیص پزشکی طراحی کند، اما موفقیت این الگوریتم به داده‌های باکیفیت (تأمین‌شده توسط بیمارستان‌ها)، قوانین نظارتی (وضع‌شده توسط دولت) و پذیرش اجتماعی (شکل‌گرفته توسط جامعه مدنی) وابسته است» (Black et al., 2024: 401). این تعاملات نشان‌دهنده پویایی و پیچیدگی زیست‌بوم هوش مصنوعی است که نیازمند یک رویکرد سیستمی برای حکمرانی است.

۲ـ۲. مبانی نظری پژوهش

باتوجه‌به نوظهور بودن این حوزه، پژوهشگران هنوز به چارچوب‌های نظری تثبیت‌شده‌ای دست نیافته‌اند و مطالعات در این زمینه ادامه دارد. بااین‌حال، در اینجا به برخی از چارچوب‌های نظری برجسته در این حوزه اشاره می‌کنیم. طبقه‌بندی راسل و نورویگ، به‌عنوان یکی از چارچوب‌های جامع و پذیرفته‌شده، سیستم‌های هوش مصنوعی را بر اساس قابلیت‌ها و درجات خودمختاری آن‌ها دسته‌بندی می‌کند (Russell & Norvig, 2016). این چارچوب در حوزه حکمرانی هوش مصنوعی نقشی محوری دارد، زیرا امکان درک دقیق‌تر انواع سیستم‌های هوش مصنوعی و شناسایی نیازهای نظارتی و اخلاقی مرتبط با هر دسته را فراهم می‌سازد. در ادامه، این طبقه‌بندی،

حکمرانی هوش مصنوعی را نه به‌مثابه مجموعه‌ای از «قواعد پراکنده» بلکه همچون یک معماری چندلایه، ریسک‌محور و اکوسیستم‌مدار صورت‌بندی می‌کند که در آن ارزش‌های هنجاری مطلوب و اثربخش، ایمنی، فراگیری، مشروعیت و سازگارپذیری به نحوی نظام‌مند به قیدهای طراحی و سازوکارهای اجرایی نگاشت می‌شوند. در این تلقی، «حکمرانی» صرفاً کنش دولت‌ها نیست؛ بلکه محصول برهم‌کنش و هم‌آرایی نهادهای دولتی، صنعت، جامعهٔ مدنی، اجتماعات تخصصی و نهادهای استانداردگذار در یک «سیستم از سیستم‌ها»ست. به طور هم‌زمان، این معماری دو وجه متمایز دارد: وجه توصیفی آن که به قواعد، هنجارها و نهادهای بالفعل اشاره می‌کند و وجه هنجاری آن که آرمان تصمیم‌های «خوب» و کارا، امن، فراگیر، مشروع و تطبیق‌پذیر را پی می‌گیرد (Dafoe, 2018).

این صورت‌بندی بر مبنای موضعی فرانظری استوار است: به‌جای تحمیل یک «مدل واحد» بر پدیده‌ای که هم از نظر فنی و هم نهادی متکثر و در حال تغییر است، بر استخراج مؤلفه‌ها، روابط و قیود عامی تمرکز می‌کند که هر سامانهٔ حکمرانی معتبر باید آن‌ها را برآورده سازد. از منظر معرفت‌شناختی، این چارچوب با «عدم قطعیت عمیق» و «کثرت ارزش‌ها» سازگار است و به‌جای پاسخ‌های محتوایی ثابت، «دستور زبان تحلیلی» برای طراحی و ارزیابی معماری ارائه می‌کند. همچنین بر واژگان مفهومی مشترک (قابلیت، ریسک، ظرفیت نهادی، مشروعیت)، روابط سازوکار محور (مقیاس‌پذیری اطلاعاتی، آثار شبکه‌ای، مسئلهٔ نمایندگی) و گزاره‌های پیونددهنده‌ای تکیه دارد که نسبت میان این عناصر را تبیین می‌کنند.

برای تأمین کفایت تبیینی و تعمیم‌پذیری، مبانی نظری بر سه دیدگاه مکمل بنا شده است. نخست، دیدگاه «فناوری کاربری عام» که هوش مصنوعی را به‌منزلهٔ ورودی ارزشمند طیفی وسیع از فرایندها و محرک نوآوری‌های مکمل می‌فهمد (Garfinkel, 2024). پیامد نظری کلیدی این نگرش آن است که هرچه شدت کاربری عام و ظرفیت بهبود مستمر، بیشتر باشد، معماری حکمرانی باید تفکیک‌پذیرتر و سازگارپذیرتر طراحی شود و به‌جای قواعد ایستا از قیدهای طراحی قابل بازتنظیم بهره گیرد. دوم، دیدگاه «فناوری اطلاعات» که بر منطق تولید، فشرده‌سازی، انتقال و کنترل اطلاعات تأکید می‌کند و دو ویژگی برجسته را پیش می‌کشد: هزینهٔ نهایی بسیار پایین در مقیاس و آثار شبکه‌ای قوی. این ویژگی‌ها هم‌زمان به تمرکز ساختاری در تولید و نوعی «برابری مصرف» در دسترسی دامن می‌زنند؛ بنابراین، اصول حکمرانی باید شفافیت متناسب، ممیزی‌پذیری و میان‌عملیاتی‌پذیری را در کنار صیانت از امنیت و مالکیت فکری نهادینه کنند (Coyle et al, 2020). سوم،

دیدگاه «فناوری هوشمندی» که هوش مصنوعی را به‌منزلهٔ نوآوری در توان حلّ مسئله و تصمیم‌گیری می‌بیند و طیفی از «ابزار» تا «سامانه/عامل» را دربر می‌گیرد. هرچه از ابزارهای کمکی به سامانه‌ها و عامل‌های هدف‌جو نزدیک‌تر می‌شویم، مسائل سوگیری، هم‌سویی و کنترل اهمیتی فزاینده می‌یابند و اقتضا می‌کنند سازوکارهای فنی هم‌سویی با سازوکارهای نهادی پاسخ‌گویی، شفافیت متناسب و قابلیت بازخواست به‌صورت درهم‌تنیده طراحی شوند (Christian, 2020).

با اتکا به سه چشم‌انداز نظری یادشده، می‌توان مجموعه‌ای از «قیود فراگیر طراحی نهادی» را صورت‌بندی کرد که بی‌آنکه وارد مصادیق شوند، منطق معماری حکمرانی را هدایت می‌کنند. نخست، در چارچوب اقتصاد نهادی، نهادها برای داخلی‌سازی برون‌ریزها پدید می‌آیند؛ ازاین‌رو «برازش نهادی» یعنی هم‌ترازی قلمرو مسئله با دامنهٔ اختیار و ظرفیت نهاد شرط لازم کاهش ریسک‌های رابطه‌ای است، اما کافی نیست؛ الزامات ایمنی پیشینی و کانال‌های اعتراض و بازنگری نیز باید همراه آن برقرار باشند (Koremenos et al., 2001). دوم، تلقی هوش مصنوعی به‌مثابهٔ فناوری شدیداً دوکاربردی و منتشر، معادلهٔ دشوار شفافیت امنیت و مسئلهٔ راستی‌آزمایی را برجسته می‌کند؛ بدین‌معنا که سازوکارهای اعتمادسازی و نظارت باید شواهد کافی برای اطمینان از رفتار ایمن فراهم آورند، بی‌آنکه امنیت طرف‌ها را مخدوش کنند (Zaidi & Dafoe, 2021). سوم، در محیط‌هایی با بازده‌های نسبی بزرگ، فشارهای ساختاری به‌سوی کاستن از حاشیهٔ ایمنی و سایر ارزش‌ها میل می‌کند؛ بنابراین، تعریف «حدود غیرقابل مصالحه» و تعبیهٔ مکانیسم‌های همکاری و هماهنگی مشروط، پیش‌شرط پیشگیری از فرسایش تدریجی ارزش‌هاست (Askell et al., 2019).

در امتداد همین منطق، روایت نظری بر «تناسب ریسک» و «مقیاس‌پذیری حکمرانی» تأکید می‌گذارد. تناسب ریسک بدین معناست که شدت الزامات و سازوکارهای کنترلی باید تابع شدت ریسک نهفته در قابلیت‌ها و زمینه‌ها باشد و این تناسب از خلال نگاشت صریح ارزش‌های هنجاری به معیارها و کنترل‌های سنجش‌پذیر تحقق یابد. مقیاس‌پذیری حکمرانی نیز ایجاب می‌کند که قواعد و نهادها از اکنون معتبر باشند و در مواجهه با سطوح بالاتر توانمندی و تحول نیز پایدار و قابل تعمیم باقی بمانند؛ به عبارت دیگر، باید «قابلیت ارتقا» و «قابلیت بازگشت» در معماری پیش‌بینی شود تا اصلاحات کم‌هزینه و توقف امن در مواجهه با عدم‌قطعیت ممکن باشد (Gruetzemacher & Whittlestone, 2022).

نکتهٔ روش‌شناختی مهم در این چارچوب آن است که رابطهٔ میان کنترل‌های فنی و سازوکارهای نهادی، رابطه‌ای جانشین‌پذیر نیست؛ بلکه مکمل است. ارزیابی فنی، ممیزی رفتاری، پایش مداوم و ثبت قابل‌اتکا زمانی به افزایش مشروعیت و کارایی منتهی می‌شوند که خطوط مسئولیت، سازوکارهای اعتراض و حمایت از افشاگران و نظارت مستقل به‌عنوان اجزای درونی معماری تعریف شده باشند. به همین قیاس، شفافیت باید «متناسب با زمینه» باشد: افشای آنچه برای بازخواست عمومی و حرفه‌ای کفایت می‌کند، نه بیش و نه کم؛ زیرا شفافیت بی‌محابا همان‌قدر مسئله‌آفرین است که ابهام کامل (Coyle & Weller, 2020).

از این منظر، اصول طراحی که به‌صورت هنجاری و اجرایی قابل دفاع‌اند عبارت‌اند از: چندمرکزی و توزیع کارکردها میان سطوح و کنشگران با پیوندهای قاعده‌مند؛ تناسب ریسک و تدریجی‌سازی الزام‌ها؛ تعریف کف‌های ایمنی غیرقابل‌سازش؛ شفافیت متناسب و ممیزی‌پذیری؛ قابلیت اعتراض و بازنگری؛ نمایندگی معنادار ذی‌نفعان در طراحی و بازطراحی؛ ماژولاریتی و قابلیت بازگشت برای اصلاح کم‌هزینه؛ استانداردهای مشترک و میان‌عملیاتی‌پذیری برای کاهش قفل‌شدگی و تسهیل ممیزی در سطح اکوسیستم؛ هم‌سویی فنی و نهادی از رهگذر ادغام کنترل‌های فنی با ترتیبات حقوقی و سازمانی؛ و سرانجام، سازگارپذیری یادگیرنده از طریق بازنگری دوره‌ای مبتنی بر شواهد و سندباکس‌های تنظیمی (Amodei et al., 2016).

در سطح نظری پیونددهنده، چند گزارهٔ عمومی جهت‌دهندهٔ طراحی نیز برجسته است. نخست، با افزایش «قابلیت»، اگر شفافیت متناسب و ظرفیت نهادی هم‌پای آن رشد نکند، شکاف هم‌سویی با ارزش‌ها افزایش می‌یابد؛ دوم، افزایش مقیاس‌پذیری اطلاعاتی و آثار شبکه‌ای، در فقدان ممیزی مؤثر، خطر تمرکز بدون پاسخ‌گویی را تشدید می‌کند؛ سوم، برازش نهادی شرط لازم کاهش ریسک است؛ اما شرط کافی نیست و تنها در کنار کف‌های ایمنی پیشینی و کانال‌های اعتراض و بازنگری کفایت پیدا می‌کند؛ چهارم، هرجا پاداش‌های نسبی بزرگ وجود داشته باشد، حفظ ارزش‌ها محتاج قیود سخت و همکاری پذیرفتنی است (Dafoe, 2018). این گزاره‌ها، به‌صورت هنجاری، نقش «پل» میان تحلیل نظری و طراحی عملی را ایفا می‌کنند بی‌آنکه به مصداق‌گویی فروکاسته شوند.

در جمع‌بندی، مبانی نظری حاضر، حکمرانی هوش مصنوعی را به‌مثابهٔ معماری‌ای می‌بینند که از سه عدسی فناوری کاربری عام، «فناوری اطلاعات» و «فناوری هوشمندی»

تغذیه می‌کند؛ بر اقتصاد سیاسی نهادها برای داخلی‌سازی برون‌ریزها و برازش نهادی تکیه دارد؛ تجارت‌برگ شفافیت امنیت و دشواری‌های راستی‌آزمایی را در طراحی لحاظ می‌کند؛ فشارهای ساختاری رقابت را با کف‌های ایمنی و سازوکارهای همکاری مهارپذیر می‌سازد؛ و از رهگذر «دستور زبان تحلیلی»، ارزش‌های هنجاری را به قیدهای طراحی سنجش‌پذیر و مقیاس‌پذیر نگاشت می‌کند. بدین‌سان، روایت نظری، هم انتزاع لازم برای اجتناب از مصداق‌زدگی را حفظ می‌کند و هم بنیان‌های کافی برای ترجمهٔ هنجارها به معماری اجرایی چندلایه را فراهم می‌آورد.

۳. روش پژوهش

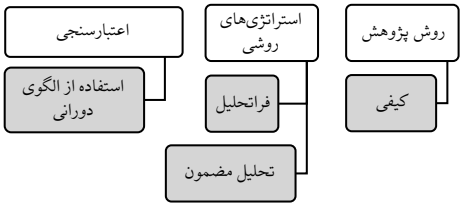
پژوهش حاضر با رویکرد کیفی و با تلفیق دو روش فراتحلیل و تحلیل مضمون انجام شده است. فراتحلیل زیرمجموعه‌ای از مرورهای سیستماتیک؛ روشی برای ترکیب سیستماتیک داده‌های کیفی و کمی مرتبط از چندین مطالعه انتخاب‌شده برای رسیدن به یک نتیجه واحد با قدرت است. فراتحلیل صرفاً یک تکنیک آماری نیست، بلکه روشی پژوهشی است که با بررسی نظام‌مند پژوهش‌های پیشین، فرضیه‌ها را بازتعریف کرده و از طریق ترکیب دقیق داده‌ها و یافته‌ها به نتیجه‌گیری‌های جامع‌تر دست می‌یابد (سهرابی‌فرد، ۱۳۸۵: ۱۶۹).

۱ـ۳. کاربست روش

در بخش فراتحلیل، با جستجوی کلیدواژه «حکمرانی هوش مصنوعی» در پایگاه‌های علمی، ابتدا ۱۳۰ مقاله مرتبط شناسایی شد. پس از ارزیابی دقیق‌تر موضوعات و بررسی چکیده‌های مقالات، ۶۳ مقاله علمی‌پژوهشی به زبان انگلیسی که به‌صورت مستقیم به ابعاد مختلف حکمرانی هوش مصنوعی پرداخته بودند، در این مطالعه، منابع برای تحلیل نهایی به روش «نمونه‌گیری هدفمند مبتنی بر معیار» برگزیده شدند. به‌منظور کاهش ناهمگنی پارادایمی و تفکیک روشن نسل‌ها، آثار متعلق به نسل نخست رویکرد نمادگرا و سامانه‌های قاعده‌محور و خودکار صرفاً برای زمینه‌سازی تاریخی مرور و از هستهٔ تحلیل کنار گذاشته شد (مفیدی، ۱۴۰۴). تمرکز اصلی بر نسل دوم، یعنی پارادایم پیوندگرا و شبکه‌محور، قرار گرفت؛ پارادایمی که اکنون میدان غالب پژوهش و صنعت را شکل می‌دهد و بخش عمدهٔ دستاوردهای عملی از بطن آن برآمده است (مفیدی، ۱۴۰۴). این تحدید دامنه، همگرایی مفهومی و مقایسه‌پذیری یافته‌ها را افزایش می‌دهد و با پرسش پژوهش دربارهٔ معماری‌های یادگیری داده‌محور

و دلالت‌های حکمرانی آن‌ها انطباق دارد. ازاین‌رو انتخاب پژوهش‌ها در بازهٔ ۸ ساله اخیر قرار گرفت تا این تمایز شکل گیرد و مرز پژوهش در نسبت با نسل اول هوش مصنوعی مشخص باشد. داده‌های دموگرافیک شامل حوزه تخصصی، محل نشر، تعداد نویسندگان، ترکیب جنسیتی، روش‌های پژوهش و نظریه‌های پرتکرار مقالات، استخراج شد و به‌عنوان پایه‌ای برای یافته‌های توصیفی پژوهش مورد استفاده قرار گرفت. در مرحله تحلیل مضمون، داده‌های کیفی مقالات بررسی شد که نتیجه آن استخراج ۲۹۷ کد پایه، ۱۶ کد سازمان‌دهنده و چهار کد فراگیر بود. این کدها در نهایت چارچوب تحلیلی پژوهش را شکل دادند.

استراتژی پژوهش در روند تحلیل داده‌ها، بدین صورت بود که از الگوی دورانی (فلیک، ۱۳۹۴: ۹۳) در تحلیل مضمون استفاده شد، به‌طوری که پژوهشگر تا پایان فرآیند نگارش، به‌صورت پیاپی بین سطوح مختلف کدها (پایه، سازمان‌دهنده و فراگیر) حرکت کرد تا به انسجام تحلیلی دست یابد. همچنین، به دلیل حجم بالای کدهای سازمان‌دهنده، این کدها در چندین مرحله با کدهای پایه مقایسه و بازبینی شدند. در نهایت، استخراج کدهای فراگیر نیز پس از چندین بار وارسی و اصلاح کدهای سازمان‌دهنده و پایه انجام پذیرفت تا دقت و اعتبار نتایج تضمین شود. این رویکرد نظام‌مند، امکان تحلیل عمیق و ساختاریافته موضوع حکمرانی هوش مصنوعی را فراهم کرد و به ارائه نتایجی روشن و مستدل منجر شد.



نمودار ۱. چارچوب روش پژوهش

۴. یافته‌های پژوهش

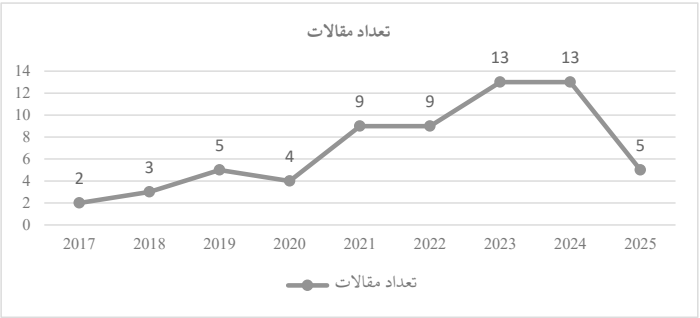
این پژوهش باهدف تبیین مؤلفه‌های کلیدی حکمرانی راهبردی هوش مصنوعی انجام شده است و در پی پاسخ به پرسش اصلی زیر است: «چارچوب‌ها و سازوکارهای حکمرانی راهبردی هوش مصنوعی در مطالعات خارجی به زبان انگلیسی منتشرشده بین سال‌های ۲۰۱۷ تا ۲۰۲۵ چگونه تعریف، ساختاربندی و ارزیابی شده‌اند؟» افزون بر

این، پژوهش حاضر به شناسایی مؤلفه‌های تکرارشونده در الگوهای حکمرانی هوش مصنوعی نیز می‌پردازد. برای نیل به این اهداف، رویکردی نظام‌مند اتخاذ شده است که زمینه درک جامع داده‌ها و یافته‌های مقالات منتخب را فراهم می‌آورد. روند انجام پژوهش نیز به شرح زیر پیگیری شده است:

- بخش نخست: توصیف ویژگی‌های مقالات موردمطالعه؛
- بخش دوم: بسط روش‌شناختی کدگذاری سه‌مرحله‌ای برای تحلیل نظام‌مند داده‌ها؛
- بخش سوم: واکاوی نتایج به‌دست‌آمده از مطالعات انتخابی.

۴ـ۱. یافته‌های توصیفی ـ آماری

بررسی روند انتشار مقالات در بازه زمانی ۲۰۱۷ تا ۲۰۲۵ نشان‌دهنده تغییرات قابل‌توجهی در تعداد مطالعات منتشرشده است. در سال‌های ابتدایی (۲۰۱۷ تا ۲۰۱۹)، رشد مقالات آرام و تدریجی بود، اما از ۲۰۲۰ به بعد شاهد افزایش چشمگیر تعداد پژوهش‌ها هستیم که به نظر می‌رسد تحت‌تأثیر پیشرفت‌های فناوری و اهمیت یافتن موضوع قرار گرفته است. اوج این روند صعودی در سال‌های ۲۰۲۲ تا ۲۰۲۵ مشاهده می‌شود که نشان‌دهنده اوج توجه علمی به این حوزه است. داده‌های سال ۲۰۲۵ نیز حاکی از تداوم این روند رو به رشد است. این الگو به‌وضوح نشان می‌دهد که موضوع مورد مطالعه در سال‌های اخیر به یکی از حوزه‌های پژوهشی پویا و پررونق تبدیل شده است[1] (نمودار شماره (۳).

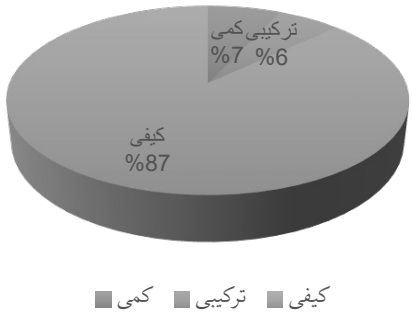


نمودار ۲. مقالات منتشر شده در هرسال

۱. از آنجا که این پژوهش در ابتدای سال ۲۰۲۵ نگارش شده است، تعداد مقالات مستخرج در این سال نسبت به سال‌های پیشین کمتر است.

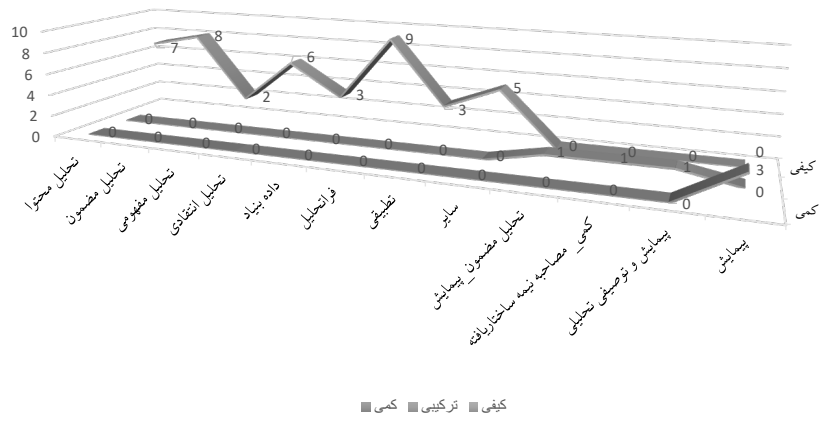
۴-۱-۱. روش شناسی مقالات منتخب

تحلیل روش‌شناسی مقالات منتخب نشان می‌دهد که ۸۷ درصد از پژوهش‌های منتخب از روش‌های کیفی، ۶ درصد از روش‌های کمی و ۷ درصد از روش‌های ترکیبی بهره برده‌اند. این ترجیح روش‌شناختی، حاکی از تناسب ذاتی رویکردهای کیفی با پیچیدگی‌های مفهومی و بافتاری حکمرانی هوش مصنوعی است، چراکه این روش‌ها امکان کاوش عمیق‌تر ابعاد اجتماعی، اخلاقی و حقوقی را فراهم می‌کنند. بااین‌حال، غلبه روش‌های کیفی می‌تواند نشان‌دهنده کم‌توجهی نسبی به سنجش‌های عینی و داده‌محور در این حوزه باشد. برای تکمیل یافته‌های موجود، ترکیب هوشمندانه‌تری از روش‌های کمی و کیفی پیشنهاد می‌شود تا هم عمق تحلیل و هم اعتبار نتایج افزایش یابد (نمودار شماره ۴).



نمودار ۳. روش های مورد استفاده در مقالات منتخب

در بخش دوم روش‌شناسی مقالات مورد مطالعه به مبحث استراتژی‌های روش‌شناختی می‌پردازیم که بر اساس اهداف پژوهشی انتخاب شده‌اند. در میان این روش‌ها، تحلیل مضمون، تحلیل محتوا، فراتحلیل و مطالعات تطبیقی به عنوان پرکاربردترین شیوه‌ها شناسایی شده‌اند که سهم قابل توجهی از مقالات را به خود اختصاص داده‌اند (نمودار شماره (۵).



نمودار ۴. استراتژی های پژوهش

غلبه این استراتژی‌های خاص نشان‌دهنده نیاز مبرم به رویکردهای تحلیلی عمیق در این حوزه پژوهشی است. تحلیل مضمون و محتوا امکان کاوش نظام‌مند مفاهیم کلیدی را فراهم می‌کنند، در حالی که فراتحلیل و مطالعات تطبیقی به سنتز یافته‌ها و شناسایی الگوهای کلان کمک می‌نمایند. این ترجیح روش‌شناختی احتمالاً ناشی از ماهیت میان‌رشته‌ای موضوع و ضرورت ترکیب دیدگاه‌های مختلف است. بااین‌حال، به نظر می‌رسد استفاده از روش‌های نوین مانند تحلیل شبکه‌ای یا مدلسازی پویا می‌تواند ابعاد جدیدی به پژوهش‌های آتی بیفزاید.

۴ـ۱ـ۲. نظریات پرتکرار در مقالات منتخب

در پژوهش‌های معاصر درباره حکمرانی هوش مصنوعی، پنج نظریه کلیدی به‌عنوان چارچوب‌های اصلی تحلیل و سیاست‌گذاری شناسایی شده‌اند:

۱. حکمرانی چندسطحی که تعاملات بین سطوح محلی تا بین‌المللی را در سیاست‌های دیجیتال و هوش مصنوعی تبیین می‌کند؛

۲. نظریه ذی‌نفعان که بر لحاظ منافع همه ذی‌نفعان در تصمیم‌گیری‌ها و کاربردهای اخلاق و مسئولیت اجتماعی تأکید دارد؛

۳. حکمرانی الگوریتمی که نقش الگوریتم‌ها در مدیریت و نظارت اجتماعی را بررسی می‌کند؛

۴. چارچوب مبتنی بر ریسک برای تنظیم فناوری‌ها بر پایه ارزیابی ریسک (مانند قانون اتحادیه اروپا)؛

۵. هوش مصنوعی مسئول با محور شفافیت و پاسخگویی، تأکید یونسکو و اصول اخلاقی نهادها. جدول شماره (۱).

جدول ۲. نظریات پرتکرار در مقالات مورد مطالعه

عنوان نظریه	شرح مختصر	تعداد
چارچوب مبتنی بر ریسک (Díaz-Rodríguez et al. 2023)	روشی برای تنظیم و مدیریت فناوری‌ها بر اساس ارزیابی و اولویت‌بندی ریسک‌های مرتبط با آن‌ها.	۷
هوش مصنوعی مسئول (Wu et al. 2024)	چارچوبی برای توسعه و استفاده از هوش مصنوعی با تأکید بر اصول اخلاقی، شفافیت، و پاسخگویی.	۶
حکمرانی الگوریتمی (Yigitcanlar. 2022)	استفاده از الگوریتم‌ها و سیستم‌های خودکار برای مدیریت و نظارت بر فرآیندهای اجتماعی و سیاسی.	۵
نظریه حکمرانی چندسطحی (Urs.2017)	چارچوبی برای تحلیل تعاملات و تصمیم‌گیری‌های مشترک بین سطوح مختلف حکمرانی، از محلی تا بین‌المللی.	۳
نظریه ذی‌نفعان (Cihon.2020)	نظریه‌ای که بر اهمیت در نظر گرفتن منافع و نقش تمامی ذی‌نفعان در تصمیم‌گیری‌های سازمانی تأکید دارد.	۲

۴ـ۲. یافته‌های تحلیلی ـ آماری

یافته‌های تحلیل مضمون ۶۳ مقاله منتخب، چارچوبی نظام‌مند برای بازنمایی محتوای مطالعات در حوزه حکمرانی هوش مصنوعی فراهم کرده است. بر این مبنا، محتوای پژوهش‌ها را می‌توان در چهار محور کلان طبقه‌بندی کرد: اخلاق هوش مصنوعی، سیاست‌ها و چشم‌اندازها، لایه‌های حکمرانی، و توسعه و کاربست. این محورها هر یک نمایانگر حوزه‌های گسترده‌تری هستند که طیف گسترده‌ای از مسائل، چالش‌ها و راهبردهای تحقیقاتی را در بر می‌گیرند و در عین حال ساختاری تحلیلی برای فهم هم‌پوشانی‌ها و تمایزات میان آنها فراهم می‌آورند.

در ستون نخست، محورهای فراگیر آمده و در مقابل هر یک، مضامین سازمان‌دهنده مرتبط به‌همراه شرح آنها فهرست شده است. این ساختار نشان می‌دهد که چگونه مضامین جزئی‌تر در تقاطع ارزش‌ها، سیاست‌ها، تنظیم‌گری و عملیاتی شدن فناوری، به شکل‌دهی فهمی یکپارچه از حکمرانی هوش مصنوعی کمک می‌کنند.

از آنجا که ارائه جدول کامل مضامین پایه به‌دلیل فشردگی تحلیلی و محدودیت فضا امکان‌پذیر نیست، تمرکز بر سطح ارتباط میان مضامین سازمان‌دهنده و فراگیر گذاشته شده است تا ساختار تحلیلی قابل تفسیر و بازتولید باقی بماند. این جدول نه صرفاً دسته‌بندی ایستا، بلکه چارچوبی برای خوانش تعاملی و داینامیک را پیش‌نهاد می‌دهد؛ به‌گونه‌ای که می‌توان هم‌پوشانی‌ها، تضادها و فضاهای میانجی میان ارزش‌ها، سیاست‌گذاری، سازوکارهای حکمرانی و اجرا را شناسایی کرد و از آن برای تبیین راهبردهای متوازن‌تر و مبتنی بر فهم ترکیبی بهره برد.

جدول ۳. تحلیل مضمون

کدهای مضامین پایه	مضمون سازمان‌دهنده	کد مضمون سازمان‌دهنده	مضمون فراگیر
A5, D6, I33, AZ4, BA1, BA2, BB4, BG6, BH1-BH3, B, BH7-BH8, BI1, I5, BI6-BI9, ...	اخلاق هوش مصنوعی: تعادل نوآوری، ایمنی و مدیریت ریسک‌های اخلاقی- اجتماعی	S2	اخلاق هوش مصنوعی
AD1, AD2, AD3	چالش‌های هماهنگی نهادی در اجرای اصول اخلاقی	S4	
AB1, AB2	ویژگی‌های کلیدی حکمرانی (پاسخگویی، شفافیت، توضیح‌پذیری)	S16	
Y1, Y2, Y3	سیاست‌های توسعه هوش مصنوعی با تمرکز بر مالکیت معنوی، حریم خصوصی و همکاری بین‌المللی	S3	سیاست‌ها و چشم‌اندازهای هوش مصنوعی
M1, M2, M3, M4, M5, M6, N1, N2, N3, N4, N5, N6, O1, O2, O3, O4	تدوین سیاست‌ها و اسناد مبتنی بر نوآوری، عدالت و رهبری جهانی	S5	
V5, W4, X1, X2, X3, X4	شکاف میان چشم‌اندازهای اسناد بالادستی و عرصه عمل	S6	
AP1, AP2, AP3, AP4, AP5	هشت محور اصلی در توسعه هوش مصنوعی (نوآوری، تولید، پژوهش، مدیریت عمومی، نیروی کار، آموزش و پرورش، تامین زیرساخت)	S8	

کدهای مضامین پایه	مضمون سازمان‌دهنده	کد مضمون سازمان‌دهنده	مضمون فراگیر
BA3	سه لایه (فنی، اخلاقی، حقوقی) در حکمرانی هوش مصنوعی	S9	لایه‌های حکمرانی هوش مصنوعی
B1, B2, B3, B4, B5	ابعاد حکمرانی هوش مصنوعی در سطوح، محورها و مراحل	S10	
C1, D10, D11, G1, G2, H5, I1, I2, K4, K5, L1, L2, M6, N6, O1, O2, R4, S4, Y2, Y3, AD3, AE1, AE6, AF1, AF2, AG1, BD4, BE4, BE6, BF13, BF4, BF6, BG1, G4, BI9, ...	حکمرانی در جهت کاهش خطر، حفظ تسلط انسان، جلوگیری از سوءاستفاده، ارتقا منافع عمومی، حفظ زنجیره نرم‌افزار و مشارکت عمومی	S11	
J2, AV1, AV2, AV3, AV4, AV5	تنظیم‌گری در طیف مقررات سخت و نرم	S12	
N1, N2, N3, N4, N5, N6	رویکرد کلان حکمرانی شامل، قانون جامع، توسعه اخلاقی و توجه بر زیرساخت‌های فنی	S13	
A1, A2, A4, A5, A6, B1-B5, C1, C2, BH4-BH8, BI1-BI3, BI7-BI9, ...	چالش‌های حکمرانی (کم‌توجهی به آموزش، جعبه سیاه الگوریتم‌ها)	S14	
AG4	زمینه‌های حکمرانی (فنی، اجتماعی، فرهنگی، اقتصادی، گفتمانی)	S15	
B1-B3-D6-L4-V1, T1, G4	فقدان تعریف مشترک و چالش‌های پیش‌بینی‌ناپذیری فناوری هوش مصنوعی	S1	توسعه و کاربست هوش مصنوعی
Y1, Y2, Y3	سیاست‌های توسعه هوش مصنوعی با تمرکز بر مالکیت معنوی، حریم خصوصی و همکاری بین‌المللی	S3	
D7, D11, K1, K2, R1-R4, T2, T3, U2-U4, W3, W4, X1AQ3, AQ4BH5, BI1, 3, BI7-BI9, ...	قراردادهای تأمین فناوری و خدمات دیجیتال در توسعه مسئولانه هوش مصنوعی	S7	
AP1, AP2, AP3, AP4, AP5	هشت محور اصلی در توسعه هوش مصنوعی	S8	

۴ـ۲ـ۱. اخلاق هوش مصنوعی

در زیر مجموعه مضمون فراگیر «اخلاق هوش مصنوعی»، به‌عنوان یکی از ستون‌های اصلی و چالشی حکمرانی هوش مصنوعی، نقشی حیاتی در ایجاد تعادل میان پیشرفت‌های علمی و مدیریت پیامدهای اخلاقی و اجتماعی آن ایفا می‌کند. این موضوع به عنوان چشم‌انداز حوزه هوش مصنوعی نه‌تنها به دنبال بهره‌برداری از ظرفیت‌های هوش مصنوعی است، بلکه تلاش دارد تا با شناسایی و کاهش خطرها، این فناوری را در مسیری هدایت کند که همسو با ارزش‌های انسانی و نیازهای جوامع باشد. در ادامه به محورهای اصلی «اخلاق هوش مصنوعی» بیان می‌شود:

۴ـ۲ـ۱ـ۱. چالش‌های اخلاقی هوش مصنوعی

هوش مصنوعی با وجود توانمندی‌های چشمگیر، با چالش‌هایی روبه‌روست که در صورت مدیریت نادرست می‌تواند پیامدهای گسترده‌ای ایجاد کند. مهم‌ترین آن‌ها شفافیت و پاسخگویی است؛ بسیاری از الگوریتم‌ها به‌صورت «جعبه‌سیاه» عمل می‌کنند و روند تصمیم‌گیری‌شان برای کاربران یا حتی توسعه‌دهندگان به‌سادگی قابل توضیح نیست (AB1,T2). این عدم شفافیت، پاسخگویی را دشوار کرده و نگرانی‌هایی درباره حریم خصوصی افراد (AB4)، آسیب‌پذیری در برابر حملات سایبری (AB5) و بروز سوگیری‌های ناخواسته در داده‌ها یا خروجی‌ها برمی‌انگیزد .(L4) چالش دیگر، سوگیری و تبعیض است؛ به‌ویژه در مدل‌های زبانی بزرگ و هوش مصنوعی مولد که گاه محتوایی ناعادلانه، تبعیض‌آمیز یا نادرست تولید می‌کنند .(AB3, AJ3) در حوزه استخدام و مدیریت منابع انسانی، این امر می‌تواند به نابرابری در دسترسی به فرصت‌های شغلی یا حذف برخی مهارت‌ها بینجامد (AJ2) و در خدمات عمومی تبعیض‌های نژادی یا جنسیتی را تشدید کرده، اعتماد عمومی را تضعیف کند. افزون بر این، خطرات ایمنی و تهدیدات وجودی ناشی از ضعف‌های فنی یا سوءاستفاده‌هایی مانند نظارت غیرمجاز نیز قابل اغماض نیست (AF1). همچنین، خودکارسازی گسترده می‌تواند بیکاری وسیع و آثار اجتماعی عمیق بر جای گذارد (V4). مجموع این چالش‌ها ضرورت تدوین چارچوب‌های اخلاقی محکم و نظام‌مند را یادآور می‌شود.

۴ـ۲ـ۱ـ۲. اصول بنیادین و چارچوب‌های اخلاقی

برای هدایت هوش مصنوعی در مسیری مسئولانه، اصول بنیادین و چارچوب‌های اخلاقی نقش راهنما را ایفا می‌کنند. یکی از این اصول کلیدی، چهارگانه اعتمادسازی است و بر اساس این اصول ساخته شده‌اند: نظارت انسانی، استحکام فنی و ایمنی،

حریم خصوصی و مدیریت داده‌ها، و شفافیت، عدالت و پاسخگویی (U2). این اصول در استانداردهای بین‌المللی مانند (ISO/IEC) و ابزار ارزیابی[1] (ALTAI) گنجانده شده‌اند و هدفشان ایجاد اعتماد در زیست‌بوم هوش مصنوعی است (AR3). نظارت انسانی به معنای حفظ کنترل انسان بر تصمیم‌گیری‌های حیاتی، استحکام فنی به ایمنی و قابلیت اطمینان سیستم‌ها، حریم خصوصی به حفاظت از داده‌های شخصی و شفافیت به ارائه اطلاعات روشن درباره عملکرد سیستم‌ها اشاره دارد.

عدالت و شمول نیز از دیگر اصول اساسی هستند که بر حذف سوگیری‌ها و تضمین برابری تأکید دارند (N6, AM4). در حوزه‌هایی مانند پزشکی یا قضاوت، حوزه‌هایی که تصمیمات هوش مصنوعی می‌توانند زندگی انسان‌ها را مستقیماً تحت تأثیر قرار دهند، وجود سوگیری می‌تواند عواقب جبران‌ناپذیری داشته باشد. برای دستیابی به این هدف، تنوع در طراحی و گفتگوهای چندجانبه با ذی‌نفعان مختلف به‌عنوان راه‌هایی مؤثر برای کاهش تبعیض پیشنهاد شده‌اند (AC3).

اصل انسان‌محوری نیز از جایگاه ویژه‌ای برخوردار است و بر این باور است که هوش مصنوعی باید در خدمت انسان و ارزش‌های او باشد (AK1-AK6). این اصل، حکمرانی پایدار را ترکیبی از سه دیدگاه کاربرمحور، جامعه‌محور و اجتماع‌محور[2] می‌داند. ابزارهایی مانند سکوهای تعامل شهروندی می‌توانند مشارکت عمومی را تقویت کنند و اطمینان دهند که توسعه هوش مصنوعی با نیازها و آرمان‌های جامعه هم‌راستا باشد (AK5).

۴ـ۲ـ۱ـ۳. راهکارهای عملیاتی و حکمرانی جامع

نتایج مقالات علاوه بر ارائه چالش‌های اخلاقی برای غلبه بر چالش‌های اخلاقی و کاهش ریسک‌ها، ارائه راهکارهای عملیاتی ضروری داده‌اند. این راهکارها می‌توانند در سطوح مختلف، از سیاست‌گذاری کلان تا فرآیندهای توسعه، پیاده‌سازی شوند (AQ1-AQ6). در سطح کلان، تشکیل هیئت‌های بازبینی الگوریتم با حضور متخصص فنی و غیرفنی، مانند متخصصان حقوقی و اخلاق‌شناسان، پیشنهاد شده است تا نظارت جامع‌تری بر سیستم‌ها اعمال شود (AL2). در مرحله توسعه نیز، ادغام اصول اخلاقی در تمامی مراحل، از جمع‌آوری داده‌ها تا ارزیابی عملکرد سیستم‌ها، از اهمیت بسزایی برخوردار است (BI1).

1. Assessment List for Trustworthy Artificial Intelligence

۲. این سطح گسترده‌تر از «جامعه محور» است و به کل ساختارهای اجتماعی، فرهنگی و نهادی توجه دارد.

در این راستا ابزارهای فناورانه و نظارتی نیز می‌توانند به بهبود شفافیت و ایمنی کمک کنند (BI2-BI9). برای نمونه، استفاده از Model Cards برای ارائه گزارش‌های شفاف درباره عملکرد مدل‌ها و ارزیابی ریسک‌ها از جمله ابزارهای پیشنهادی است (AQ6). بااین‌حال، چالش‌هایی مانند ناکافی بودن بلوغ این ابزارها در محیط‌های عملیاتی (BI3) و نیاز به یکسان‌سازی اصطلاحات و استانداردها همچنان باقی است (BI8).

مقررات نرم، مانند کدهای اخلاقی و استانداردهای داوطلبانه، رویکردی انعطاف‌پذیر برای تنظیم هوش مصنوعی ارائه می‌دهند (AV1-AV5). اصول یونسکو و همکاری‌های عمومی‌ـ‌خصوصی نمونه‌هایی از این رویکرد هستند که تلاش دارند تعادلی میان نوآوری و مسئولیت‌پذیری ایجاد کنند. بااین‌حال، ضعف در اجرا و دشواری جلب اعتماد عمومی از جمله موانع این روش به شمار می‌روند (AV3).

در این راستا، اخلاق در هوش مصنوعی، حوزه‌ای چندوجهی است که نیازمند نگاهی جامع به مسائل فنی، اجتماعی و اخلاقی است. شفافیت، عدالت و انسان‌محوری باید به‌عنوان اصول راهنما در تمامی مراحل توسعه و استفاده از این فناوری مدنظر قرار گیرند. ابزارهای عملیاتی و مقررات نرم می‌توانند به کاهش ریسک‌های اخلاقی کمک کنند، اما موفقیت آن‌ها در گرو اجرا و نظارت دقیق و مؤثر است.

نمودار ۵. چارچوب اخلاقی حکمرانی هوش مصنوعی

۴ـ۲ـ۲. سیاست‌ها و چشم‌اندازهای هوش مصنوعی

محور فراگیر دوم، یعنی «سیاست‌ها و چشم‌اندازهای هوش مصنوعی»، اصلی مهم در نحوه تدوین و پیاده سازی حکمرانی است. محور «سیاست‌ها و چشم‌اندازهای هوش مصنوعی» بر تدوین چارچوب‌هایی متمرکز است که توسعه و کاربرد این فناوری را به شکلی مسئولانه و هم‌راستا با ارزش‌های انسانی هدایت کند. محورهای اصلی این بخش به بررسی رویکردهای مختلف کشورها در تدوین سیاست‌های ملی و بین‌المللی، چالش‌های هماهنگی جهانی و اصول بنیادین در این حوزه می‌پردازد.

۴ـ۲ـ۲ـ۱. تدوین سیاست‌ها و اسناد مبتنی بر نوآوری، عدالت و رهبری جهانی

کشورها بر اساس ظرفیت‌ها، اولویت‌ها و شرایط فرهنگی و ژئوپلیتیکی خود، سیاست‌ها و چشم‌اندازهای متفاوتی را برای توسعه هوش مصنوعی دنبال می‌کنند. این تفاوت‌ها در اسناد ملی و مطالعات تطبیقی به‌وضوح قابل‌مشاهده است. برای مثال، چین بر تقویت رهبری خود در حوزه سخت‌افزار و پیشرفت صنعتی تمرکز دارد و هدفش تبدیل‌شدن به قطب اصلی تولید فناوری‌های پیشرفته است (M2, M4). در مقابل، هند توسعه کاربردهای هوش مصنوعی را با تأکید بر رفع نیازهای گروه‌های مختلف اجتماعی و اقتصادی در اولویت قرار داده است (AM2). ایالات متحده نیز رویکردی دوجانبه اتخاذ کرده و ضمن توجه به توسعه اخلاق‌محور، رهبری در زیرساخت‌ها و نرم‌افزار را دنبال می‌کند تا جایگاه خود را در عرصه جهانی حفظ کند (M1, M5).

تفاوت‌های موجود صرفاً به جنبه‌های فنی محدود نمی‌شود، بلکه متأثر از ارزش‌های فرهنگی و رقابت‌های ژئوپلیتیکی است. آمریکا با محوریت ارزش‌هایی مانند آزادی فردی و اخلاق پروتستانی سیاستگذاری می‌کند (M1, M3, M5)، درحالیکه چین هماهنگی اجتماعی و توسعه صنعتی را به عنوان اصول راهبردی خود دنبال می‌نماید (M2, M4, M5). تحریم‌های آمریکا علیه چین نمونه‌ای از تأثیر این رقابت‌ها بر جهت‌گیری سیاست‌هاست (M6).

در سطح حکمرانی هوش مصنوعی نیز شاهد تنوع رویکردها هستیم:

- *اتحادیه اروپا: تدوین چارچوبی قانونم حور و الزام‌آور* (N1)
- *آمریکا: تأکید بر توسعه اخلاق‌محور و اعتمادسازی* (N1)
- *چین: پیوند توسعه فناوری با ثبات اجتماعی بر مبنای عدالت* (N1)

این اسناد بر سه محور کلیدی تأکید دارند:

۱. *تعادل بین نوآوری و ایمنی*(S2)

۲. *تقویت همکاری‌های بین‌المللی* (N5)

۳. *رعایت اصولی مانند شفافیت و عدالت* (N6)

۴ـ۲ـ۲ـ۲. چالش‌های هماهنگی جهانی

مطالعات حکمرانی در ذیل عنوان سیاست‌ها و چشم‌اندازهای هوش مصنوعی در سطح جهانی به موضوع چالش‌های هماهنگی سیاست‌های هوش مصنوعی پرداخته‌اند. تعاملات و همکاری‌ها در سطح جهانی با موانع متعددی روبه‌روست که از جمله آن‌ها می‌توان به رقابت‌های ژئوپلیتیکی و تفاوت‌های فرهنگی اشاره کرد. این عوامل بر تدوین سیاست‌های یکپارچه تأثیر می‌گذارند و گاه همکاری بین‌المللی را دشوار می‌سازند (M1,M6). برای مثال، نبود تعریف مشترک از «هوش مصنوعی سیاست‌پذیر» و پراکندگی سیاست‌ها در کشورهایی مانند هند، موانعی جدی در مسیر همگرایی جهانی ایجاد کرده است (AD1-AD3, AO3, AM2). رقابت‌های ژئوپلیتیکی نیز، مانند تنش‌های میان آمریکا و چین، این چالش‌ها را تشدید می‌کند (AT4).

برای غلبه بر این موانع، پیشنهادهایی مانند ایجاد سازوکارهای حل اختلاف و اشتراک‌گذاری داده‌ها مطرح شده است. این سازوکارها می‌توانند زمینه‌ساز هم‌افزایی بیشتر میان کشورها شوند و به تدوین سیاست‌هایی هماهنگ‌تر کمک کنند (AN5). همچنین، تقویت گفت‌وگوهای چندجانبه و مشارکت ذی‌نفعان مختلف، از جمله دولت‌ها، شرکت‌های خصوصی و جامعه مدنی، می‌تواند به کاهش شکاف‌های موجود یاری رساند (O1, O2).

از سوی دیگر برخی مطالعات به الزامات و بایسته‌های سیاست‌ها و چشم‌اندازهای هوش مصنوعی پرداخته‌اند. تحلیل مضمون نتایج این مطالعات نشان می‌دهد که سیاست‌های و چشم‌اندازهای کلی هوش مصنوعی نیازمند رویکردی چندوجهی است که ارزش‌های فرهنگی، الزامات فنی و نیاز به همکاری بین‌المللی را در خود جای دهد. تفاوت‌های ژئوپلیتیکی و فرهنگی، اگرچه موانعی در مسیر سیاست‌گذاری جهانی ایجاد می‌کنند، اما اصول مشترکی مانند شفافیت، عدالت و پاسخگویی می‌توانند به‌عنوان پایه‌ای برای اجماع عمل کنند (N6, O3). این اصول نه‌تنها اعتماد عمومی را تقویت می‌کنند، بلکه به کاهش خطرهای ناشی از توسعه بی‌رویه هوش مصنوعی کمک می‌کنند.

همکاری بین‌المللی، به‌ویژه از طریق سازوکارهای حل اختلاف و تبادل دانش، می‌تواند به همگرایی سیاست‌ها منجر شود و زمینه را برای استفاده مسئولانه از این فناوری فراهم کند (N5). درعین‌حال، کشورها باید باتوجه‌به ظرفیت‌ها و اولویت‌های خود، سیاست‌هایی را تدوین کنند که نوآوری را با ایمنی و اخلاق درآمیزد. برای مثال،

تمرکز چین بر سخت‌افزار، هند بر کاربردها و آمریکا بر زیرساخت‌ها، نشان‌دهنده این تنوع راهبردی است که در نهایت باید در چارچوبی جهانی هماهنگ شود.

نمودار ۶. سیاست ها و چشم اندازی های هوش مصنوعی

۴ـ۲ـ۳. لایه‌های حکمرانی هوش مصنوعی

محور فراگیر سوم در مرحله اول به بیان «لایه‌های اصلی حکمرانی هوش مصنوعی» می‌پردازد. ازاین‌رو برای دستیابی به حکمرانی مؤثر در حوزه هوش مصنوعی، ضروری است که چارچوبی چندلایه طراحی شود که لایه‌های «فنی، اخلاقی و حقوقی» را به‌صورت هم‌افزا در خود جای دهد. این لایه‌ها نه‌تنها مکمل یکدیگرند، بلکه با همکاری و تعامل، زمینه‌ساز توسعه و کاربرد مسئولانه این فناوری می‌شوند.

- لایه فنی: این لایه بر طراحی الگوریتم‌هایی متمرکز است که هم مسئولانه عمل کنند و هم از شفافیت برخوردار باشند. یکی از موانع اصلی در این حوزه، پدیده «جعبه‌سیاه» است که فرایند تصمیم‌گیری سیستم‌های هوش مصنوعی را

غیرقابل‌درک می‌سازد (B1,D1,T2). برای غلبه بر این مشکل، رعایت اصول طراحی مسئولانه الگوریتم‌ها، از جمله شفافیت و توضیح‌پذیری فنی، اهمیت بسزایی دارد (B2, D7, J1). این اصول به توسعه‌دهندگان امکان می‌دهد سیستم‌هایی خلق کنند که عملکردشان برای کاربران قابل‌فهم و قابل‌اتکا باشد.

- لایه اخلاقی: به دلیل اهمیت این لایه، در بخش اول یافته‌ها به طور مفصل بررسی شد، لایه اخلاقی بر ارزش‌های انسانی مانند انصاف، کاهش سوگیری و احترام به عدالت و خودمختاری تأکید دارد (B3, N6, AB3, U1, AK4). سوگیری در الگوریتم‌ها می‌تواند به تبعیض و نابرابری منجر شود، به‌ویژه در حوزه‌هایی مانند سلامت و نظام قضایی که حساسیت بالایی دارند. ازاین‌رو، شناسایی و رفع سوگیری‌ها از طریق ابزارهایی نظیر تنوع در مخاطبان و گفتگوهای چندجانبه ضروری است (AC3).
- لایه حقوقی: بر ایجاد چارچوب‌های تنظیم‌گری و نهادهای مسئول برای نظارت بر کاربرد هوش مصنوعی و اجرای قوانین تمرکز دارد (B4, J2, AI2, F1, P1) چارچوب‌های حقوقی باید به‌گونه‌ای طراحی شوند که ضمن تضمین مسئولیت‌پذیری، انعطاف‌پذیری لازم برای سازگاری با پیشرفت‌های سریع فناوری را نیز داشته باشند.

این سه‌لایه با یکدیگر هم‌افزایی دارند و از طریق پیوند قواعد و ابزارهای فناورانه (W1) و تأکید بر هوش مصنوعی قابل‌اعتماد (AR2)، به کاهش مخاطرات اجتماعی و اخلاقی کمک می‌کنند (C3, AF1). همکاری میان توسعه‌دهندگان، سیاست‌گذاران و جامعه مدنی برای موفقیت این چارچوب ضروری است (J4, I2).

۴ـ۲ـ۳ـ۲. سطوح، محورها و مراحل حکمرانی هوش مصنوعی

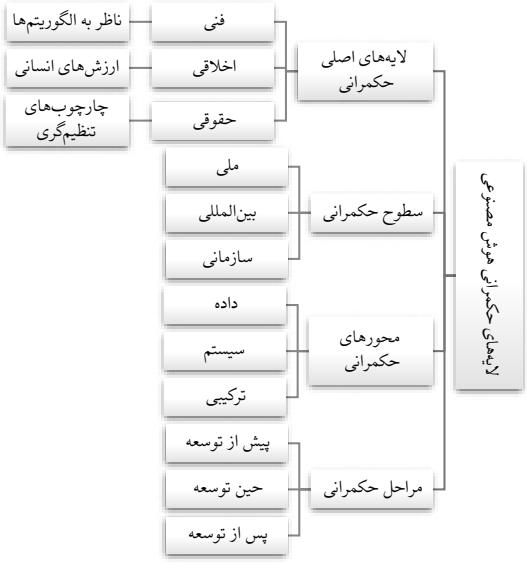
بخش دوم این محور فراگیر به سطوح، محورها و مراحل حکمرانی هوش مصنوعی می‌پردازد. حکمرانی هوش مصنوعی تنها به لایه‌های اصلی محدود نمی‌شود، بلکه نیازمند درنظرگرفتن سطوح، محورها و مراحل مختلفی است که به‌صورت یکپارچه عمل کنند تا این فناوری به شکلی مسئولانه مدیریت شود.

- سطوح حکمرانی: این سطوح شامل سه دسته ملی، بین‌المللی و سازمانی است. در سطح ملی، کشورها سیاست‌هایی را بر اساس نیازها و اولویت‌های داخلی خود تدوین می‌کنند، درحالی‌که در سطح بین‌المللی، همکاری و هماهنگی برای مواجهه با چالش‌های جهانی ضروری است (N5, AD3). در سطح سازمانی

نیز، شرکت‌ها و نهادها موظف‌اند استانداردهای داخلی را برای توسعه مسئولانه هوش مصنوعی پیاده‌سازی کنند.

- محورهای حکمرانی: این محورها به سه بخش داده، سیستم و ترکیبی تقسیم می‌شوند. حکمرانی داده بر کیفیت، امنیت و حفظ حریم خصوصی داده‌ها تمرکز دارد (J1, L2)، درحالی‌که حکمرانی سیستم بر شفافیت و توضیح‌پذیری الگوریتم‌ها تأکید می‌کند (K3, AB2). رویکرد ترکیبی نیز این دو را ادغام می‌کند تا چارچوبی جامع ارائه دهد.
- مراحل حکمرانی: این مراحل شامل پیش از توسعه، حین توسعه و پس از توسعه است. در مرحله پیش از توسعه، ارزیابی ریسک‌ها و تدوین اصول اخلاقی از اهمیت برخوردار است (AQ6). در حین توسعه، نظارت انسانی و بازرسی‌های اخلاقی ضروری است (J3, K4). پس از توسعه نیز، ممیزی‌های دوره‌ای و اصلاح رفتار سیستم‌ها برای تضمین عملکرد مسئولانه لازم است (L3, AO4).

این چارچوب چندبعدی با تمرکز بر رفع شکاف اطلاعاتی ناشی از جعبه سیاه (B1, T2) و تقویت شفافیت فنی (D7, B2)، به توسعه مسئولانه هوش مصنوعی کمک می‌کند. چالش‌هایی مانند پیش‌بینی‌ناپذیری ریسک‌ها (G4 و AF2) و رقابت‌های ژئوپلیتیکی (M6, AD2) نیازمند همکاری بین‌المللی و رویکردهای تطبیقی است (AF3, N5).



نمودار ۷. ابعاد مختلف حکمرانی هوش مصنوعی

۴ـ۲ـ۴. توسعه و کاربست هوش مصنوعی

محور فراگیر چهارم به توسعه و کاربست هوش مصنوعی اشاره دارد. یکی از محورهای اساسی در حکمرانی این فناوری به شمار می‌رود که بر طراحی، پیاده‌سازی و استفاده مسئولانه و پایدار از سیستم‌های هوش مصنوعی تمرکز دارد. این محور به بررسی ابزارها، قراردادها و زمینه‌های متنوع حکمرانی می‌پردازد که برای تضمین رشد اخلاقی و ایمن این فناوری ضروری هستند. در این راستا، قراردادهای تأمین فناوری و خدمات دیجیتال، زمینه‌های فنی، اجتماعی، فرهنگی، اقتصادی و گفتمانی، و همچنین چالش‌های پیاده‌سازی موردتوجه قرار گرفته‌اند تا چارچوبی جامع برای توسعه هوش مصنوعی ارائه شود.

۴ـ۲ـ۴ـ۱. قراردادهای تأمین فناوری و خدمات دیجیتال

قراردادهای تأمین فناوری و خدمات دیجیتال ابزارهایی کلیدی برای توسعه مسئولانه هوش مصنوعی هستند. این قراردادها باهدف ایجاد شفافیت، پاسخگویی و کاهش ریسک‌های اخلاقی و اجتماعی طراحی می‌شوند و به استفاده ایمن و اخلاقی از این فناوری کمک می‌کنند.

- توضیح‌پذیری الگوریتم‌ها: این قراردادها باید تضمین کنند که الگوریتم‌های به‌کاررفته در سیستم‌های هوش مصنوعی قابل‌فهم و شفاف باشند. این ویژگی به کاربران امکان می‌دهد تا منطق تصمیم‌گیری سیستم را درک کرده و در صورت نیاز آن را مورد پرسش قرار دهند (D7, AB2).
- کنترل کاربر بر داده‌ها: قراردادها باید حقوق کاربران را در دسترسی، اصلاح و حذف داده‌های شخصی‌شان تأمین کنند تا حریم خصوصی و خودمختاری آن‌ها حفظ شود (D8, AB4).
- نظارت مستقل و زنجیره تأمین باکیفیت: برای جلب‌اعتماد عمومی، قراردادها باید مفادی برای نظارت مستقل بر عملکرد سیستم‌ها و اطمینان از کیفیت زنجیره تأمین داشته باشند (D10, D11).

این قراردادها چارچوبی حقوقی فراهم می‌کنند که به‌عنوان خط مقدم دفاع در برابر سوءاستفاده از هوش مصنوعی عمل کرده و با انعطاف‌پذیری و اجرای سریع، ریسک‌های اخلاقی و اجتماعی را کاهش می‌دهند (AP5 - AP1).

۴ـ۲ـ۴ـ۲. زمینه‌های توسعه: فنی، اجتماعی، فرهنگی، اقتصادی و گفتمانی

توسعه مسئولانه هوش مصنوعی نیازمند توجه به زمینه‌های متعددی است که بر

طراحی، اجرا و کاربرد این فناوری اثر می‌گذارند. این زمینه‌ها شامل جنبه‌های فنی، اجتماعی، فرهنگی، اقتصادی و گفتمانی هستند که هر یک به‌نوبهٔ خود در شکل‌دهی به حکمرانی هوش مصنوعی نقش دارند.

- زمینه فنی: این زمینه بر طراحی الگوریتم‌ها و سیستم‌های هوش مصنوعی تمرکز دارد و بر شفافیت، توضیح‌پذیری و حذف سوگیری تأکید می‌کند (B2, D7, J1, AB2). سیستم‌هایی که از این اصول پیروی کنند، اعتماد کاربران را جلب کرده و از سوءتفاهم‌ها جلوگیری می‌کنند.
- زمینه اجتماعی: مشارکت عمومی و اعتماد جامعه به هوش مصنوعی از اهمیت ویژه‌ای برخوردار است. این زمینه بر تعامل با ذی‌نفعان و هم‌راستایی فناوری با نیازهای جامعه تأکید دارد (D11, AS3, AK5).
- زمینه فرهنگی: ارزش‌ها و هنجارهای فرهنگی در پذیرش هوش مصنوعی نقش کلیدی دارند. توجه به تفاوت‌های فرهنگی در طراحی سیستم‌ها و سیاست‌ها برای جلوگیری از تبعیض ضروری است (M1, M4, AA1, N6).
- زمینه اقتصادی: نوآوری و منابع اقتصادی از عوامل اصلی توسعه هوش مصنوعی هستند. سرمایه‌گذاری در زیرساخت‌ها و تقویت زیست‌بوم نوآوری در این زمینه حیاتی است (E2, BA3 و Y1).
- زمینه گفتمانی: گفت‌وگوهای عمومی و هنجارسازی در شکل‌دهی به ادراکات و پذیرش هوش مصنوعی تأثیرگذارند. تقویت گفتگوهای شفاف و چندجانبه در این زمینه ضروری است (BH5 و AZ5).

این زمینه‌ها به‌صورت متقابل بر یکدیگر اثر می‌گذارند و هماهنگی آن‌ها برای توسعه مسئولانه هوش مصنوعی لازم است.

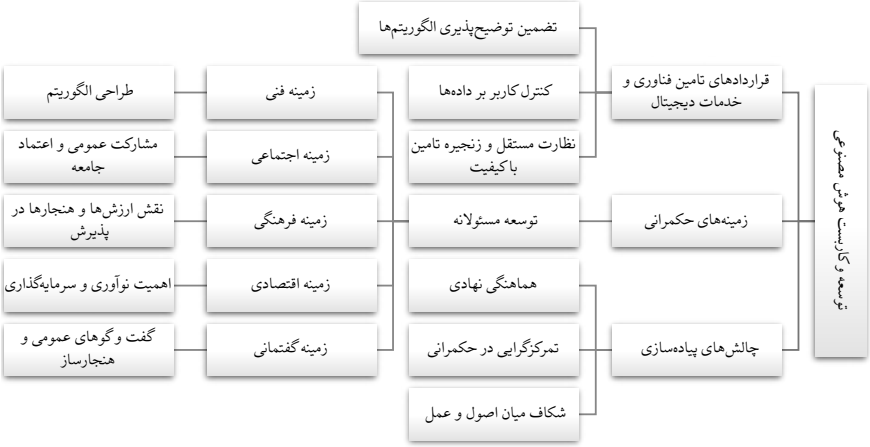
۴ـ۲ـ۴ـ۳. چالش‌های پیاده‌سازی و راه‌حل‌های پیشنهادی

تحلیل نتایج مقالات نشان می‌دهد، پیاده‌سازی چارچوب‌های حکمرانی هوش مصنوعی با موانع متعددی مواجه است که از جمله می‌توان به هماهنگی نهادی، تمرکزگرایی و شکاف میان اصول و عمل اشاره کرد. این چالش‌ها نیازمند راه‌حل‌های عملی و خلاقانه هستند.

- هماهنگی نهادی: پراکندگی سیاست‌ها و نبود تعریف مشترک از هوش مصنوعی، همکاری‌های میان نهادی را دشوار کرده است (AD1 و AD3). پیشنهاد می‌شود با ایجاد سازوکارهای حل اختلاف و اشتراک‌گذاری داده‌ها، این مشکل برطرف شود (AN5).

- تمرکزگرایی در حکمرانی: نهادهای متمرکز اگرچه کارآمدند، اما در برابر نفوذ گروه‌های خاص آسیب‌پذیرند (AE1 تا AE6 و AU1 تا AU5). در مقابل، ساختارهای غیرمتمرکز انعطاف‌پذیرند، اما ممکن است فاقد قدرت اجرایی باشند. یافتن تعادل میان کارایی و انعطاف‌پذیری از طریق رویکردهای نوآورانه ضروری است.
- شکاف میان اصول و عمل: نبود چارچوب‌های قانونی عملیاتی، کمبود آموزش و ابهام در اصول اخلاقی، اجرای مؤثر را دشوار می‌کند (W4, BI4, BF4 و AO3). توسعه «بانک سؤال ارزیابی ریسک» و ممیزی اخلاقی در فرایندها از راه‌حل‌های پیشنهادی است (AQ6 و AO4).

این چالش‌ها با رویکردهای تطبیقی و همکاری بین‌المللی قابل‌حل هستند تا حکمرانی هوش مصنوعی به شکلی پایدار اجرا شود (N5 و AV5). توسعه و کاربست هوش مصنوعی نیازمند چارچوبی جامع است که قراردادهای تأمین فناوری را با زمینه‌های حکمرانی تلفیق کند. شفافیت، پاسخگویی و مشارکت عمومی از طریق این قراردادها و نظارت مستقل تقویت می‌شوند. زمینه‌های فنی، اجتماعی، فرهنگی، اقتصادی و گفتمانی باید به‌صورت هماهنگ در نظر گرفته شوند تا هوش مصنوعی با ارزش‌های انسانی هم‌سو باشد.



نمودار ۸. توسعه و کاربست هوش مصنوعی

بحث و نتیجه‌گیری

مقاله حاضر باهدف صورت‌بندی منسجم مسئله و ترسیم نقشه‌ای شواهد بنیاد برای حکمرانی هوش مصنوعی، یک مرور نظام‌مند بر ادبیات انجام داده تا ظرفیت تصمیم‌سازی را تقویت و مسیر مواجهه مسئولانه با فرصت‌ها و تهدیدهای این فناوری را روشن کند. مسئله اصلی آن است که شتاب توسعه، شکاف‌هایی میان اصول هنجاری و سازوکارهای اجرایی پدید آورده و پراکندگی سیاست‌ها و نبود اجماع مفهومی، طراحی چارچوب‌های کارآمد حکمرانی را دشوار ساخته است. این مطالعه می‌کوشد با جمع‌بندی منظم یافته‌های معتبر، از سطح توصیف پراکنده فراتر رود و بنیانی برای اقدام عملی فراهم آورد.

پژوهش پیرامون دو پرسش سامان یافته است: نخست، چارچوب‌ها و سازوکارهای حکمرانی راهبردی هوش مصنوعی در مطالعات ۲۰۱۷ تا ۲۰۲۵ چگونه تعریف، ساختاربندی و ارزیابی می‌شوند؛ دوم، کدام الگوهای حکمرانی پرتکرارند و چه ویژگی‌های مشترکی دارند. برای پاسخ، رویکردی کیفی با تلفیق فراتحلیل و تحلیل مضمون برگزیده شد. از ۱۳۰ مقاله مرتبط، ۶۳ مطالعه که مستقیماً به ابعاد حکمرانی پرداخته بودند به‌صورت هدفمند انتخاب گردید. داده‌های دموگرافیک (حوزه، محل نشر، ترکیب نویسندگان، روش‌ها و نظریه‌های پرتکرار) استخراج شد و سپس با فرایندی دورانی، ۲۹۷ مضمون پایه به ۱۶ مضمون سازمان‌دهنده و در نهایت به ۴ مضمون فراگیر تنقیح گردید تا انسجام و دقت تحلیل تضمین شود.

هسته تحلیلی پژوهش بر ۱۶ مضمون سازمان‌دهنده استوار است که از فقدان تعریف مشترک و پیش‌بینی‌ناپذیری فناوری تا ویژگی‌های کلیدی حکمرانی مانند پاسخگویی، شفافیت و توضیح‌پذیری را دربر می‌گیرد. تنقیح این کدها به چهار مضمون فراگیر انجامیده است. نخست، «اخلاق هوش مصنوعی» به‌مثابه ستون حکمرانی و نقطه تلاقی شتاب نوآوری با ارزش‌های انسانی: چالش‌هایی چون جعبه سیاهی الگوریتم‌ها، ابهام در مسئولیت‌پذیری، نقض حریم خصوصی و آسیب‌پذیری‌های امنیتی، همراه با سوگیری‌ها و تبعیض‌های ساختاری، اعتماد عمومی را تهدید می‌کند. در برابر، بسته‌ای از اصول برای اعتمادسازی پیشنهاد می‌شود: نظارت انسانی بر تصمیم‌های حیاتی، استحکام فنی و ایمنی، حکمرانی داده و حریم خصوصی، و شفافیت، عدالت و پاسخگویی؛ همراه با انسان‌محوری، توجه به تنوع در طراحی و گفت‌وگوی چندجانبه با ذی‌نفعان. برای کاهش شکاف «اصل تا عمل»، ادغام اخلاق در چرخه توسعه،

هیئت‌های بازبینی الگوریتم و گزارش‌های شفاف مانند «کارت‌های مدل» و بهره‌گیری از مقررات نرم در کنار مقررات سخت توصیه می‌شود.

مضمون دوم «سیاست‌ها و چشم‌اندازها» است که تنوع رویکرد ملی را، از قانون‌محور و الزام‌آور تا اخلاق‌محور و اعتمادساز یا صنعتی‌محور و ثبات‌آفرین، به تصویر می‌کشد. باوجود تفاوت‌ها، سه محور مشترک برجسته است: تعادل میان نوآوری و ایمنی، تقویت همکاری‌های بین‌المللی و پاسداشت شفافیت و عدالت. موانع هماهنگی جهانی (رقابت‌های ژئوپلیتیکی، تفاوت‌های فرهنگی و فقدان تعریف مشترک از «هوش مصنوعی سیاست‌پذیر») با سازوکارهای حل اختلاف، اشتراک‌گذاری داده و تبادل دانش و گفت‌وگوهای چندجانبه با مشارکت واقعی دولت‌ها، بخش خصوصی و جامعه مدنی قابل مدیریت‌اند تا اجماع حداقلی بر اصول مشترک فراهم شود و همگرایی تدریجی شکل گیرد.

مضمون سوم «معماری حکمرانی» را در چهار بعد هم‌بسته می‌چینند. در بعد لایه‌ها، سه‌لایه فنی، اخلاقی و حقوقی باید هم‌افزا عمل کنند: شفافیت و توضیح‌پذیری برای مهار جعبه سیاهی، انصاف و کاهش سوگیری به‌عنوان هسته هنجاری، و چارچوب‌های حقوقی انعطاف‌پذیر برای تضمین مسئولیت‌پذیری و انطباق با تحول سریع. در بعد سطوح، هماهنگی میان سطح ملی، بین‌المللی و سازمانی ضروری است. در بعد محورها، حکمرانی داده (کیفیت، امنیت، حریم خصوصی) در کنار حکمرانی سیستم (توضیح‌پذیری و شفافیت الگوریتمی) و رویکرد تلفیقی معنا می‌یابد. در بعد مراحل، چرخه کامل عمر (پیش از توسعه با ارزیابی ریسک و تعیین اصول؛ حین توسعه با نظارت انسانی و بازرسی اخلاقی؛ و پس از توسعه با ممیزی دوره‌ای و اصلاح) باید به‌صورت منسجم پوشش داده شود تا گذر از شعار به اجرا ممکن گردد و رویکردهای تطبیقی و همکاری‌های فرامرزی تقویت شود.

مضمون چهارم به «توسعه و کاربست» می‌پردازد و «قراردادهای تأمین فناوری و خدمات دیجیتال» را ریل اجرایی حکمرانی می‌داند. الزام توضیح‌پذیری الگوریتم‌ها، تثبیت حقوق داده‌ای کاربران در دسترسی، اصلاح و حذف، و پیش‌بینی نظارت مستقل و تضمین کیفیت زنجیره تأمین، ریسک‌های اخلاقی ــ اجتماعی را از ابتدا مهار و اعتماد را تقویت می‌کند. تحقق توسعه مسئولانه نیازمند حساسیت هم‌زمان به زمینه‌های فنی، اجتماعی، فرهنگی، اقتصادی و گفتمانی است. در اجرا، سه گره‌گاه اصلی تشخیص داده می‌شود: هماهنگی نهادی در سایه پراکندگی سیاست‌ها و فقدان تعریف مشترک؛

دوگانه کارایی/نفوذپذیری در ساختارهای متمرکز در برابر انعطاف/کاهش ضمانت اجرا در ساختارهای غیرمتمرکز؛ و شکاف «وضعیت مطلوب و موجود» ناشی از کمبود راهنمای عملیاتی، آموزش و وضوح مفهومی. پاسخ‌ها شامل سازوکارهای حل اختلاف و اشتراک‌گذاری داده، «بانک سؤال ارزیابی ریسک»، ممیزی اخلاقی مستمر و ترکیب سنجیده مقررات سخت و نرم است.

برآیند این مضامین، یک چارچوب حکمرانی شواهدبنیاد است: ستون فقرات سه‌لایه فنی، اخلاقی، حقوقی که بر سه سطح ملی، بین‌المللی، سازمانی امتداد می‌یابد و در سه محور حکمرانی داده، حکمرانی سیستم و رویکرد تلفیقی عمل می‌کند؛ چرخه عمر توسعه نیز از پیشینی تا پسینی پوشش می‌یابد. درون این معماری، ابزارهای عملیاتی برای پیوند «وضعیت موجود» به «وضعیت مطلوب» جای‌گذاری می‌شود: قراردادهای تأمین، هیئت‌های بازبینی الگوریتم، ممیزی‌های اخلاقی و ابزارهای گزارش‌دهی مدل، در کنار سازوکارهای حل اختلاف و تبادل دانش. این چارچوب یک فرایند یادگیرنده مبتنی بر پایش مستمر و بازنگری دوره‌ای است که می‌تواند ریسک‌ها را مهار، اعتماد را تقویت و ظرفیت نوآوری را در خدمت منافع عمومی و ارزش‌های انسانی قرار دهد. نتیجه آنکه حکمرانی راهبردی زمانی کارآمد است که در معماری‌ای چندلایه و چندسطحی متجسد شود و با ابزارهای شفاف، شکاف مزمن «وضعیت مطلوب و موجود» را کاهش دهد؛ همان مسیری که این مقاله، بر پایه ادبیات مرورشده، به‌روشنی ترسیم کرده است.

نمودار ۹. محورهای اصلی چارچوب حکمرانی راهبردی هوش مصنوعی

بحث و پیشنهاد

باتکیه‌بر یافته‌های مرور نظام‌مند، «مدل پیشنهادی حکمرانی هوش مصنوعی برای ایران» را می‌توان به‌صورت معماری چندلایه، ریسک‌محور و اعتمادساز صورت‌بندی کرد؛ معماری‌ای که از یک‌سو ارزش‌های هنجاری (ایمنی، عدالت، مشروعیت، فراگیری و سازگاری) را به قیدهای طراحی سنجش‌پذیر نگاشت می‌کند و از سوی دیگر، فاصلهٔ مزمن «مبنا تا عمل» را با ابزارهای اجرایی پر می‌کند. در این الگو، سه‌لایهٔ فنی، اخلاقی، حقوقی به‌مثابهٔ ستون فقرات، بر سه سطح ملی، بخشی، سازمانی امتداد می‌یابند و چرخهٔ عمر سامانه (پیشینی/حین توسعه/پسینی) را به طور یکپارچه پوشش می‌دهند. منطق راهبر مدل «تناسب ریسک» است: هرچه ریسک کاربرد و بستر بالاتر، الزامات سخت‌گیرانه‌تر؛ و «مقیاس‌پذیری حکمرانی» شرط بقا و ارتقا در برابر تحولات سریع فناوری.

در سطح ملی، پیشنهاد می‌شود «شورای ملی حکمرانی هوش مصنوعی» ذیل ساختار عالی سیاست‌گذاری دیجیتال کشور (با دبیرخانهٔ چابک) تشکیل شود تا پراکندگی نهادی کاهش یابد و «مرکز ارزیابی ریسک و ایمنی الگوریتمی» و «ثبت ملّی سامانه‌های پرریسک» را مستقر کند. این سطح باید سه سند پایه را مصوب و عملیاتی کند: ۱. «سند ملی حکمرانی داده» (تعریف حقوق داده‌ای، الزامات ناشناس‌سازی و استانداردسازی، اشتراک امن و مرزهای دسترسی عمومی و حاکمیتی)، ۲. «نظام طبقه‌بندی ریسک کاربردها» (هم‌ارز ساده‌شدهٔ رویکرد اتحادیه اروپا، بومی‌سازی‌شده برای ایران) و ۳. راهنمای ارزش‌های محوری همچون ایمنی پیشینی، منع تبعیض، حفاظت از کودکان، حریم خصوصی و امنیت کاربران. باتوجه‌به محدودیت‌های تحریمی و زیرساختی، ایجاد «صندوق محاسباتی ملّی» و «ابر بومی قابل حسابرسی» برای تأمین محاسبات ایمن ارزیابی و ممیزی و یک مرکز پاسخ‌گویی به رخدادهای هوش مصنوعی ضروری است.

در سطح بخشی، رگولاتورهای تخصصی حوزه‌های حساس سلامت، مالی، حمل‌ونقل، آموزش، رسانه و زیرساخت‌های حیاتی باید سندباکس‌های تنظیمی را راه‌اندازی کنند تا نوآوری کنترل‌شده و یادگیری مقرراتی رخ دهد. هر رگولاتور، ضمیمهٔ بخشی طبقه‌بندی ریسک و بستهٔ الزامات فنی و اخلاقی خود را منتشر کند: ارزیابی اثرات داده و حریم خصوصی، الزامات توضیح‌پذیری و قابلیت بازخواست، آزمون‌های پیش از بهره‌برداری و ممیزی‌های پسینی، و نیز الگوهای قراردادی الزام‌آور برای تأمین‌کنندگان بندهای شفافیت، حقوق داده‌ای کاربر، گزارش‌دهی رخداد و دسترسی

ممیز به لاگ‌ها. بدین ترتیب، پیوند «حکمرانی داده» و «حکمرانی سیستم» در هر بخش به‌صورت عملیاتی برقرار می‌شود.

در سطح سازمانی، وزارتخانه‌ها، نهادهای عمومی و کسب‌وکارهای بزرگ، ایجاد هیئت‌های بازبینی الگوریتم بین‌رشته‌ای الزامی باشد تا از مرحلهٔ طراحی تا استقرار، کنترل‌های اخلاقی و فنی را اعمال کنند. سازمان‌ها باید «کارت مدل» و «دفترچهٔ داده» برای سامانه‌های پرریسک منتشر کنند، کنترل نسخه و ثبت تصمیم‌های مهم را برقرار سازند، سازوکار افشاگران و رسیدگی به شکایات را به‌صورت روشن مستقر کنند و ممیزی‌های دوره‌ای مستقل را بپذیرند. برای جلب‌اعتماد عمومی، «اتاق شیشه‌ای الگوریتم‌ها» (پنجره‌های شفافیت متناسب با ریسک) در کنار داشبوردهای سنجش اعتماد، ایمنی، نوآوری و شمول، به‌صورت منظم گزارش‌دهی شوند.

کنترل‌های چرخه عمر باید یکپارچه و مرحله‌بندی‌شده اعمال شوند. پیش از توسعه: طبقه‌بندی ریسک، ارزیابی اثرات، طراحی مسئولانه و معیارهای آزمون. حین توسعه: نظارت انسانی مؤثر، کنترل داده و نسخه، آزمون تقابلی و پایش سوگیری. پس از بهره‌برداری: ممیزی‌های دوره‌ای، ثبت و گزارش رخداد، به‌روزرسانی اجباری و امکان بازگشت امن در صورت بروز خطر. «قراردادهای تأمین فناوری» به‌عنوان ریل اجرایی حکمرانی، باید شروط دسترسی ممیز به مستندات، SLA[1] های ایمنی و حریم خصوصی و سازوکار حل اختلاف را تصریح کنند.

برای هم‌افزایی با جهان و کاهش اصطکاک مبادلاتی، رویکرد «هم‌ارزی کارکردی» پیشنهاد می‌شود: همسوسازی حداقلی با هنجارهای معتبر بین‌المللی (مانند الزامات شفافیت و ارزیابی ریسک) بدون وابستگی حقوقی کامل؛ پذیرش متقابل معیارهای ارزیابی و نتیجهٔ ممیزی، به‌ویژه با کشورهای منطقه؛ و سرمایه‌گذاری هدفمند روی استانداردهای باز و ابزارهای متن‌باز قابل‌ممیزی. این رویکرد ضمن حفظ استقلال سیاستی، امکان تعامل‌پذیری فنی و نهادی و جذب سرمایه و دانش را افزایش می‌دهد.

در بُعد اجتماعی، تقویت «مشارکت معنادار ذی‌نفعان» ضروری است: شوراهای مشورتی شهروندی برای کاربردهای پرریسک، گفت‌وگوهای چندجانبه با صنوف و انجمن‌های تخصصی، و برنامهٔ ارتقای «سواد داده و هوش مصنوعی» برای مدیران و عموم. این لایه اعتمادساز، هزینه‌های اجرای مقررات را کاهش می‌دهد و مشروعیت مداخلات را بالا می‌برد. هم‌زمان، سیاست‌های فعال سرمایه انسانی (آموزش

1. Service Level Agreemen

میان‌رشته‌ای، حمایت از پژوهش ایمنی و تفسیرپذیری، و برنامه‌های نگهداشت و بازگشت نخبگان) باید به‌عنوان بازوی توانمندساز حکمرانی پیش برود.

جمع‌بندی این‌که، مدل متناسب با زمینهٔ ایران، همان معماری چندلایهٔ ریسک‌محور و اعتمادساز است که با «یک مرکز هماهنگی ملّی»، «رگولاتوری بخشی سندباکسی»، «حاکمیت دادهٔ شفاف و امن»، «ابزارهای اجرایی قرارداد ممیزی گزارش‌دهی»، و «توانمندسازی اجتماعی و زیرساختی» عمل می‌کند. این مدل هم‌زمان چابک و پاسخ‌گوست: برای شرایط تحریم و محدودیت محاسباتی راه‌حل دارد، با الگوهای جهانی هم‌ارزی کارکردی برقرار می‌کند، و شکاف میان اصول و اجرا را با سازوکارهای عینی کم‌رنگ می‌کند. هوش مصنوعی، به‌جای منبعی از ریسک‌های تجمیعی، به پیشران قابل‌اعتماد رفاه، عدالت و تاب‌آوری ملی بدل شود.

تعارض منافع

تعارض منافع ندارم.

منابع و مأخذ

آهنگری، نوید (۱۴۰۴). فراتحلیل کاربرد هوش مصنوعی در توسعهٔ کاربری‌های ترکیبی اراضی شهر تهران، اقتصاد و برنامه‌ریزی شهری. DOI:10.22034/uep.2025.537527.1680

پورکریمی، جواد و زهرا علی‌اکبری (۱۴۰۴). ساحت‌های چهارگانه حکمرانی هوش مصنوعی آموزش: مطالعه‌ای فراترکیب. پیشرفت‌های نوین در مدیریت آموزشی. DOI:10. 2034/njournal.2025.501570.1008

ترابی، محمدامین و حدیث اقبال (۱۴۰۳). ارائه مدل حاکمیت هوش مصنوعی برای حکمرانی دولت در جمهوری اسلامی ایران، دولت پژوهی ایران معاصر.

سهرابی‌فرد، نسرین (۱۳۸۵). مروری بر مبانی فراتحلیل، فصلنامه روان‌شناسان ایرانی، ۳(۱۰)، ۱.

فلیک، اووه (۱۳۹۴). درآمدی بر تحقیق کیفی، ترجمه هادی جلیلی، تهران: نشر نی.

کیسینجر، هنری؛ اشمیت، اریک؛ دانیل هوتنلوچر (۱۴۰۳). عصر هوش مصنوعی و آینده ما انسان‌ها، چاپ اول، تهران: نشر کتاب پارسه.

مجدزاده، سید انوشیروان، سلاجقه، سنجر، نیک‌پور، امین، و محمدجلال، کمالی. (۱۴۰۲). فراتحلیل مطالعات دولت الکترونیک در ارتقای شاخص‌های حکمرانی خوب. پژوهش‌های مدیریت راهبردی، ۲۹(۸۹)، ۷۷. ـ۱۰۸. Dor:20.1001.1.22285067.1402.29.89.3.4

مفیدی‌یزدی، حسین (۱۴۰۲ ـ ۱۴۰۴). مسئولیت‌پذیری هوش مصنوعی، درس فقه هوش مصنوعی. fiqhci.com

Agrawal, A., Gans, J., & Goldfarb, A. (2019). Economic policy for artificial intelligence. *Innovation policy and the economy*, *19*(1), 139-159. https//:doi.org10.1086/699935/.

Ahangari, Navid. (2025). Meta-analysis of Artificial Intelligence Applications in the Development of Mixed Land Uses in Tehran City. *Urban Economics and Planning*, **[In Persian]**. https://www.juep.net/article-229669.html.

Allen, D., Hubbard, S., Lim, W., Stanger, A., Wagman, S., Zalesne, K., & Omoakhalen, O. (2025). A roadmap for governing AI: technology governance and power-sharing liberalism. *AI and Ethics*, *5*(3), 3355-3377. https://doi.org/10.1007/s43681-024-00635-y.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). *Concrete Problems in AI Safety*. arXiv:1606.06565. https://doi.org/10.48550/arXiv.1606.06565.

Andrews, P., de Sousa, T., Haefele, B., Beard, M., Wigan, M., Palia, A.,... & Jacquet, A. (2022). A Trust Framework for Government Use of Artificial Intelligence and Automated Decision Making. *arXiv preprint arXiv:2208.10087*. https://doi.org/10.48550/arXiv.2208.10087.

Askell, A., Brundage, M., & Hadfield, G. (2019). The Role of Cooperation in Responsible AI Development. arXiv:1907.04534.https://doi.org/10.48550/arXiv.1907.04534.

Attard-Frost, B., Brandusescu, A., & Lyons, K. (2024). The governance of artificial intelligence in Canada: Findings and opportunities from a review of 84 AI governance initiatives. *Government Information Quarterly*, *41*(2), 101929. https://doi.org/10.1016/j.giq.2024.101929.

Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: a systematic literature review. *AI and Ethics*, 1-15. https://doi.org/10.1007/s43681-024-00653-w

Birkstedt, T., Minkkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: themes, knowledge gaps and future agendas. *Internet Research*, *33*(7), 133-167. https://doi.org/10.1108/INTR-01-2022-0042.

Birkstedt, T., Minkkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: themes, knowledge gaps and future agendas. *Internet Research*, *33*(7), 133-167. https://doi.org/10.1108/INTR-01-2022-0042.

Budhwar, P., Chowdhury, S., Wood, G., Aguinis, H., Bamber, G. J., Beltran, J. R.,... & Varma, A. (2023). Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. *Human Resource Management Journal*, *33*(3), 606-659. https://doi.org/10.1111/1748-8583.12524.

Bullock, Justin B. (Ed.) (2024): The Oxford handbook of AI governance. New York: Oxford University Press.

Camilleri, M. A. (2024). Artificial intelligence governance: Ethical considerations and implications for social responsibility. *Expert systems*, *41*(7), e13406.

Camilleri, M. A. (2024). Artificial intelligence governance: Ethical considerations and implications for social responsibility. *Expert systems*, *41*(7), e13406. https://doi.org/10.1111/exsy.13406.

Camilleri, M. A. (2024). Artificial intelligence governance: Ethical considerations and implications for social responsibility. *Expert systems*, *41*(7), e13406. https://doi.org/10.1111/exsy.13406.

Cancela-Outeda, C. (2024). The EU's AI act: A framework for collaborative governance. *Internet of Things*, *27*, 101291. https://doi.org/10.1016/j.iot.2024.101291.

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International journal of educational technology in higher education*, *20*(1), 38. https://doi.org/10.1186/s41239-023-00408-3

Cheng, J., & Zeng, J. (2023). Shaping AI's future? China in global AI governance. *Journal of Contemporary China*, *32*(143), 794-810. https://doi.org/10.1080/10670564.2022.2107391.

Chiarelli, P. (2016). Artificial and Biological Intelligence: Hardware vs Wetware. American Journal of Applied Psychology, 5(6), 98-103. doi: 10.11648/j.ajap.20160506.19.

Christian, B. (2020). The Alignment Problem: Machine Learning and Human Values. W. W. Norton.

Cihon, P., Maas, M. M., & Kemp, L. (2020). Fragmentation and the Future: Investigating Architectures for International AI Governance. Global Policy. https://doi.org/10.1111/1758-5899.12890.

Cihon, P., Maas, M. M., & Kemp, L. (2020). Fragmentation and the future: Investigating architectures for international AI governance. *Global Policy*, *11*(5), 545-556. https://doi.org/10.1111/1758-5899.12890.

Cihon, P., Maas, M. M., & Kemp, L. (2020, February). Should artificial intelligence governance be centralised? Design lessons from history. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 228-234). https://dl.acm.org/doi/abs/10.1145/3375627.3375857.

Coglianese, C. (2023). *Procurement and artificial intelligence* (Public Law and Legal Theory Research Paper Series No. 23-33). University of Pennsylvania Carey Law School.

Coyle, D., & Weller, A. (2020). "Explaining" machine learning reveals policy challenges. Science, 368(6498), 1433-1434. DOI: 10.1126/science.aba9647.

Dafoe, A. (2018). AI governance: a research agenda. Governance of AI Program, Future of Humanity Institute, University of Oxford: Oxford, UK, 1442, 1443.

David, A., Yigitcanlar, T., Desouza, K., Li, R. Y. M., Cheong, P. H., Mehmood, R., & Corchado, J. (2024). Understanding local government responsible AI strategy: An international municipal policy document analysis. *Cities*, *155*, 105502. https://cdn.governance.ai/GovAI-Research-Agenda.pdf.

de Almeida, P. G. R., & dos Santos Junior, C. D. (2025). Artificial intelligence governance: Understanding how public organizations implement it. *Government information quarterly*, *42*(1), 102003. https://doi.org/10.1016/j.cities.2024.105502.

Díaz-Rodríguez, N., Coeckelbergh, M., Herrera-Viedma, E., Herrera, F., & Del Ser, J. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. University of Granada. https://doi.org/10.1016/j.inffus.2023.101896.

Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., De Prado, M. L., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, *99*, 101896. https://doi.org/10.1016/j.inffus.2023.101896

European Parliament and Council. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence* (Artificial Intelligence Act). Official Journal of the European Union.

Flick, Uwe. (2015). *An Introduction to Qualitative Research*. (Trans. Hadi Jalili). Tehran: Ney Publishing, **[In Persian].**

Garfinkel, B. (2024). General purpose technologies. In J. Bullock et al. (Eds.) , The Oxford Handbook of AI Governance. Oxford University Press.

Gasser, U., & Almeida, V. A. (2017). A layered model for AI governance. *IEEE Internet Computing*, *21*(6), 58-62. DOI: 10.1109/MIC.2017.4180835

Gritsenko, D., Markham, A., Pötzsch, H., & Wijermars, M. (2022). Algorithms, contexts, governance: An introduction to the special issue. *new media & society*, *24*(4), 835-844. https://doi.org/10.1177/14614448221079037

Gruetzemacher, R., & Whittlestone, J. (2022). Transformative AI: Definitions, thresholds, and development scenarios. (Working paper / preprint).

Hadley, E., Blatecky, A., & Comfort, M. (2024). Investigating algorithm review boards for organizational responsible artificial intelligence governance. *AI and Ethics*, 1-11.

Hassan, M., Borycki, E. M., & Kushniruk, A. W. (2025, March). Artificial intelligence governance framework for healthcare. In *Healthcare Management Forum* (Vol. 38, No. 2, pp. 125-130). Sage CA: Los Angeles, CA: SAGE Publications. DOI: 10.1177/08404704241291226.

Herrera-Poyatos, A., Del Ser, J., de Prado, M. L., Wang, F. Y., Herrera-Viedma, E., & Herrera, F. (2025). Responsible Artificial Intelligence Systems: A Roadmap to Society's Trust through Trustworthy AI, Auditability, Accountability, and Governance. *https://doi.org/10.48550/arXiv.2503.04739.*

Hildebrandt, M. (2018). Algorithmic regulation and the rule of law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2128), 20170355. https://doi.org/10.1098/rsta.2017.0355

Hine, E., & Floridi, L. (2024). Artificial intelligence with American values and Chinese characteristics: a comparative analysis of American and Chinese governmental AI policies. *Ai & society*, *39*(1), 257-278. https://doi.org/10.1007/s00146-022-01499-8.

Hooghe, L., & Marks, G. (2003). Unraveling the central state, but how? Types of multi-level governance. *American Political Science Review, 97*(2), 233-243

Housley, W., & Smith, S. (2023). Algorithmic governance: Technology, power, and knowledge. In *The SAGE handbook of digital society* (pp. 439-457). DOI10.4135/9781529783193:.

Jobin, A. (2019). Artificial Intelligence: the global landscape of ethics guidelines. *arXiv preprint arXiv:1906.11668*.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, *1*(9), 389-399. https://arxiv.org/pdf/1906.11668.

Kissinger, Henry; Schmidt, Eric; & Huttenlocher, Daniel. (2024). *The Age of Artificial Intelligence and Our Human Future.* 1st ed. Tehran: Parseh Publishing, **[In Persian**].

Koremenos, B., Lipson, C., & Snidal, D. (2001). The rational design of international institutions. *International Organization, 55*(4), 761-799. DOI: https://doi.org/10.1162/002081801317193592.

Krafft, P. M., Young, M., Katell, M., Huang, K., & Bugingo, G. (2020, February). Defining AI in policy versus practice. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 72-78). doi/abs/10.1145/3375627.3375835.

Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications policy*, *44*(6), 101976. https://doi.org/10.1016/j.telpol.2020.101976.

Loi, M., & Spielkamp, M. (2021, July). Towards accountability in the use of artificial intelligence for public administrations. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 757-766). doi/abs/10.1145/3461702.3462631.

Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., & Jacquet, A. (2024). Responsible AI pattern catalogue: A collection of best practices for AI governance and engineering. *ACM Computing Surveys*, *56*(7), 1-35. https://doi.org/10.1145/3626234

Maas, M. M. (2023). Advanced AI Governance: A literature review of problems, options, and proposals. *AI Foundations Report*, *4*. Maas, Matthijs M., Advanced AI Governance: A Literature Review of Problems, Options, and Proposals (November 10, 2023). AI Foundations Report 4, Available at SSRN: http://dx.doi.org/10.2139/ssrn.4629460

Majdzadeh, Seyed Anoushirvan; Salajegheh, Sanjar; Nikpour, Amin; & Kamali Mohammadjalal, Kamal. (2023). Meta-Analysis of E-Government Studies in Enhancing Good Governance Indicators. *Strategic Management Research*, 29(89), 77-108, **[In Persian]**.

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics*, *2*(4), 603-609. https://doi.org/10.1007/s43681-022-00143-x

Marchant, G. E., & Gutierrez, C. I. (2022). Soft law 2.0: an agile and effective governance approach for artificial intelligence. *Minn. JL Sci. & Tech.*, *24*, 375.

Marda, V. (2018). Artificial intelligence policy in India: a framework for engaging the limits of data-driven decision-making. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180087. https://doi.org/10.1098/rsta.2018.0087.

Minkkinen, M., & Mäntymäki, M. (2023). Discerning between the “easy” and “hard” problems of AI governance. *IEEE Transactions on Technology and Society*, *4*(2), 188-194.

Mofidi-Yazdi, Hossein. (2023-2025). Artificial Intelligence Responsibility. *Fiqh of Artificial Intelligence Course*. Retrieved from, **[In Persian]**.

Mökander, J., & Floridi, L. (2023). Operationalising AI governance through ethics-based auditing: an industry case study. *AI and Ethics*, *3*(2), 451-468. https://doi.org/10.1007/s43681-022-00171-7

Molloy, B. T. (2021). Project Governance for Defense Applications of Artificial Intelligence. *Prism*, *9*(3), 106-121. https://www.jstor.org/stable/48640749?seq=1

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and engineering ethics*, *26*(4), 2141-2168. https://doi.org/10.1126/science.132.3429.741

Papagiannidis, E., Enholm, I. M., Dremel, C., Mikalef, P., & Krogstie, J. (2023). Toward AI governance: Identifying best practices and potential barriers and outcomes. *Information Systems Frontiers*, *25*(1), 123-141. https://doi.org/10.1007/s10796-022-10251-y

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, *34*(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

Perry, B., & Uuk, R. (2019). AI governance and the policymaking process: key considerations for reducing AI risk. *Big data and cognitive computing*, *3*(2), 26. https://doi.org/10.3390/bdcc3020026.

Pourkarimi, Javad, & Ali-Akbari, Zahra. (2025). The Four Dimensions of Artificial Intelligence Governance in Education: A Meta-Synthesis Study. *Modern Advances in Educational Management*, **[In Persian]**.

Roberts, H., Cowls, J., Morley, J., Taddeo, M., Wang, V., & Floridi, L. (2021). The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. In *Ethics, governance, and policies in artificial intelligence* (pp. 47-79).

Roberts, H., Hine, E., Taddeo, M., & Floridi, L. (2024). Global AI governance: barriers and pathways forward. *International Affairs*, *100*(3), 1275-1286. https://doi.org/10.1093/ia/iiae073.

Robles, P., & Mallinson, D. J. (2025). Artificial intelligence technology, public trust, and effective governance. *Review of Policy Research*, *42*(1), 11-28. https://doi.org/10.1111/ropr.12555.

Roche, C., Wall, P. J., & Lewis, D. (2023). Ethics and diversity in artificial intelligence policies, strategies and initiatives. *AI and Ethics*, *3*(4), 1095-1115. https://doi.org/10.1007/978-3-030-20680-2.

Russell, Stuart J.; Norvig, Peter (2016): Artificial intelligence. A modern approach. 3. edition. Global edition. Upper Saddle River: Pearson (Prentice Hall series in artificial intelligence).

Schiff, D. (2022). Education for AI, not AI for education: The role of education and ethics in national AI policy strategies. *International Journal of Artificial Intelligence in Education*, *32*(3), 527-563. https://doi.org/10.1007/s40593-021-00270-2

Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (2020, February). What's next for ai ethics, policy, and governance? a global overview. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 153-158). https://doi.org/10.1145/3375627.33758

Schneider, J., Abraham, R., Meske, C., & Vom Brocke, J. (2023). Artificial intelligence governance for businesses. *Information Systems Management*, *40*(3), 229-249. https://doi.org/10.1080/10580530.2022.2085825.

Sigfrids, A., Leikas, J., Salo-Pöntinen, H., & Koskimies, E. (2023). Human-centricity in AI governance: A systemic approach. *Frontiers in artificial intelligence*, *6*, 976887. https://doi.org/10.3389/frai.2023.976887

Socol, A., & Iuga, I. C. (2024). Addressing brain drain and strengthening governance for advancing government readiness in artificial intelligence (AI). *Kybernetes*, *53*(13), 47-71. https://doi.org/10.1108/K-03-2024-0629.

Sohrabifard, Nasrin. (2006). A Review of the Fundamentals of Meta-Analysis. *Iranian Psychologists Quarterly*, 3(10), 1-10, **[In Persian]**.

Stix, C. (2021). The ghost of AI governance past, present and future: AI governance in the European Union. *arXiv preprint arXiv:2107.14099*.https://doi.org/10.48550/arXiv.2107.14099

Straub, V. J., Morgan, D., Bright, J., & Margetts, H. (2023). Artificial intelligence in government: Concepts, standards, and a unified framework. *Government Information Quarterly*, *40*(4), 101881. https://doi.org/10.1016/j.giq.2023.101881.

Taeihagh, A. (2021). Governance of artificial intelligence. *Policy and society*, *40*(2), 137-157. https://doi.org/10.1080/14494035.2021.1928377

Tallberg, J., Erman, E., Furendal, M., Geith, J., Klamberg, M., & Lundgren, M. (2023). The global governance of artificial intelligence: Next steps for empirical and normative research. *International Studies Review*, *25*(3), viad040. https://doi.org/10.1093/isr/viad040.

Torabi, Mohammad-Amin, & Eghbal, Hadis. (2024). A Model of Artificial Intelligence Governance for Government Administration in the Islamic Republic of Iran. *Contemporary Iranian Government Studies*, **[In Persian].**

Turk, Ž. (2024). Regulating artificial intelligence: A technology-independent approach. *European View*, *23*(1), 87-93. https://doi.org/10.1177/17816858241242890

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W. G. (2021). Framing governance for a contested emerging technology: insights from AI policy. *Policy and Society*, *40*(2), 158-177. https://doi.org/10.1080/14494035.2020.1855800

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. Adopted at the 41st session of the General Conference of UNESCO, November 2021.

Wang, W., & Siau, K. (2018). Artificial intelligence: a study on governance, policies, and regulations.

Wodi, A. (2024). Artificial Intelligence (AI) Governance: An Overview. *Available at SSRN 4840769*. dx.doi.org/10.2139/ssrn.4840769

Wu, C., Zhang, H., & Carroll, J. M. (2024). AI governance in higher education: Case studies of guidance at big ten universities. *arXiv preprint arXiv:2409.02017*. https://doi.org/10.48550/arXiv.2409.02017.

Wu, Y., & Shan, S. (2021). Application of Artificial Intelligence to Social Governance Capabilities under Public Health Emergencies. *Mathematical problems in engineering*, *2021*(1), 6630483. https://doi.org/10.1155/2021/6630483.

Xue, J. H. (2024). Polycentric theory diffusion and AI governance. In *Global digital development* (pp. 243-263).

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance, 12*(4), 505-523. https://doi.org/10.1111/rego.12158

Yigitcanlar, T., Agdas, D., & Degirmenci, K. (2022). Artificial intelligence in local governments: Perceptions of city managers on prospects, constraints and choices. Queensland University of Technology. https://doi.org/10.1007/s00146-022-01450-x.

Yigitcanlar, T., Agdas, D., & Degirmenci, K. (2023). Artificial intelligence in local governments: perceptions of city managers on prospects, constraints and choices. *Ai & Society*, *38*(3), 1135-1150. https://doi.org/10.1007/s00146-022-01450-x.

Yigitcanlar, T., David, A., Li, W., Fookes, C., Bibri, S. E., & Ye, X. (2024). Unlocking artificial intelligence adoption in local governments: Best practice lessons from real-world implementations. *Smart Cities*, *7*(4), 1576-1625. https://doi.org/10.3390/smartcities7040064.

Yigitcanlar, T., Senadheera, S., Marasinghe, R., Bibri, S. E., Sanchez, T. W., Cugurullo, F., & Sieber, R. (2024). The Local Government and Artificial Intelligence Nexus: A Five-Decade Scientometric Analysis on Evolution, State-of-The-Art, and Emerging Trends. *State-of-The-Art, and Emerging Trends*. https://doi.org/10.1016/j.cities.2024.105151.

Zaidi, S., & Dafoe, A. (2021). AI and arms control: Lessons and challenges for governing emerging technologies. (Report / working paper).

Zambonelli, F., Salim, F., Loke, S. W., De Meuter, W., & Kanhere, S. (2018). Algorithmic governance in smart cities: The conundrum and the potential of pervasive computing solutions. *IEEE Technology and Society Magazine*, *37*(2), 80-87.