



Rereading the Philosophical-Hermeneutical Challenges and Strategies of Human Agency in the Age of Artificial Intelligence



ARTICLE INFO

Article Type

Original Research

Authors

Abdollahpour Sangchi F.

Department of Theoretical Foundations of Islam, Faculty of Islamic Thought and Education, University of Tehran, Tehran, Iran

Rahnamaei H.*

Department of Theoretical Foundations of Islam, Faculty of Islamic Thought and Education, University of Tehran, Tehran, Iran

Asgari Yazdi A.

Department of Theoretical Foundations of Islam, Faculty of Islamic Thought and Education, University of Tehran, Tehran, Iran

Rezaee M.

Department of Islamic Philosophy and Theology, Faculty of Theology and Islamic Studies, University of Mazandaran, Babolasar, Iran

How to cite this article

Abdollahpour Sangchi F, Rahnamaei H, Asgari Yazdi A, Rezaee M. Rereading the Philosophical-Hermeneutical Challenges and Strategies of Human Agency in the Age of Artificial Intelligence. Philosophical Thought. 2026;6(1):41-60.

*Correspondence

Address: Department of Theoretical Foundations of Islam, Faculty of Islamic Thought and Education, University of Tehran, Enghlab Street, Tehran, Iran. Postal Code: 1417864181

Phone: -

h-rahnamaei@ut.ac.ir

Article History

Received: November 3, 2025

Accepted: December 22, 2025

ePublished: February 4, 2026

ABSTRACT

The rapid expansion of artificial intelligence technologies, particularly large language models, has raised new questions in the domains of philosophy, hermeneutics, and human agency. At the epistemological level, the primary issue is the inability of AI systems to achieve “genuine hermeneutic understanding,” which is instead limited to the structural simulation of meaning, thereby risking a disconnection between meaning and truth. At the methodological level, the opacity and ambiguity of algorithmic mechanisms lead to a crisis of credibility and legitimacy in machine-generated interpretations. At the existential-ethical level, the potential threat to the human interpreter’s role and responsibility in the process of meaning-making comes to the fore. Based on this analysis, the article proposes theoretical, practical, and technological strategies to redefine and expand human hermeneutic agency in the age of artificial intelligence. It argues that a responsible and critical engagement with these technologies is not merely a technical necessity but a philosophical and ethical imperative—one that can redefine the relationship between humans, texts, and technology within the horizon of digital hermeneutics.

Keywords Philosophical Hermeneutics, Artificial Intelligence, Human Agency, Understanding, Ethics of Technology

CITATION LINKS

[Aalipour & Zanganeh, 2022] An investigation of theoretical foundations related to digital transformation and innovation in knowledge management systems (KMS) to understand its role in business management and sustainable development; [Alvand et al., 2023] A model of human resources development strategies in the digital age; [Amershi et al., 2014] Power to the people: The role of humans in interactive machine learning; [Brown et al., 2020] Language models are few-shot learners; [Capurro, 2010] Digital hermeneutics: An outline; [Daraei, 2024] Explainable artificial intelligence: A step towards transparency and trust in artificial intelligence; [Enayati et al., 2025] The role of artificial intelligence in identifying and reducing cognitive biases in the learning process; [Floridi, 2014] The fourth revolution: How the infosphere is reshaping human reality; [Floridi & Cowls, 2019] A unified framework of five principles for AI in society; [Floyd et al., 1992] Software development and reality construction; [Gadamer, 2004] Truth and method; [High-Level Expert Group on Artificial Intelligence, 2019] Ethics guidelines for trustworthy AI; [Jafari Tabrizi, 2012] Islamic mysticism; [Javadi Amoli, 2016] Tasnim commentary (Volume 46); [Jobin et al., 2019] The global landscape of ai ethics guidelines; [Lipton, 2017] The myths of model interpretability; [Marcus & Davis, 2019] Rebooting AI: Building artificial intelligence we can trust; [Moor, 2006] The dartmouth college artificial intelligence conference: The next fifty years; [Morley et al., 2019] From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices; [Motahari, 2015] Notes of professor Motahari (Volume 1); [Motalebi Karbakandi, 2013] Analysis of artificial intelligence and its functions in comparison with human thinking; [Noorani & Farazkish, 2025] Ethical considerations and bias in language models. Proceedings of the First National Conference on Artificial Intelligence and Emerging Research; [Ojjeabu, 2024] From automation to augmentation: The potential of artificial intelligence in biblical hermeneutics; [Pickering et al., 2017] The interplay between human and machine agency; [Romele et al., 2020] Digital hermeneutics: From interpreting with machines to interpretational machines; [Sadr Al-Din Shirazi, 2008] AL-MAZAHIR AL-ILAHYEH (Volume 3); [Sobhani Tabrizi, 2019] Hermeneutics; [Tabatabaei, 1995] TAFSIR AL-MIZAN; [Wang, 2021] Is artificial intelligence capable of understanding?;

نوع مقاله: پژوهشی اصیل

بازخوانی چالش‌های فلسفی-هرمنوتیک و راهبردهای عاملیت انسانی در عصر هوش مصنوعی

فاطمه عبدالله‌پور سنگچی

گروه مبانی نظری اسلام، دانشکده اندیشه و معارف اسلامی، دانشگاه تهران، تهران، ایران

حسین رهنمائی*

گروه مبانی نظری اسلام، دانشکده اندیشه و معارف اسلامی، دانشگاه تهران، تهران، ایران

علی عسگری یزدی

گروه مبانی نظری اسلام، دانشکده اندیشه و معارف اسلامی، دانشگاه تهران، تهران، ایران

مهران رضایی

گروه فلسفه و کلام اسلامی، دانشکده الهیات و معارف اسلامی، دانشگاه مازندران، بابل، ایران

چکیده

گسترش شتابان فناوری‌های هوش مصنوعی، به‌ویژه مدل‌های زبانی بزرگ، پرسش‌های تازه‌ای را در حوزه فلسفه، هرمنوتیک و عاملیت انسانی پدید آورده است. در سطح معرفت‌شناختی، مساله اصلی ناتوانی سامانه‌های هوش مصنوعی در تحقق «فهم هرمنوتیک اصیل» و محدود بودن آنها به شبیه‌سازی ساختاری معنا و در نتیجه خطر گسست میان معنا و حقیقت است. در سطح روش‌شناختی، ابهام و غیرشفافیت سازوکارهای الگوریتمی به بحران اعتبار و مشروعیت تفاسیر ماشینی می‌انجامد. در سطح وجودی-اخلاقی نیز، تهدید جایگاه مفسر انسانی و مسئولیت او در فرآیند معنابخشی مورد توجه قرار می‌گیرد. بر پایه این تحلیل، مقاله راهبردهایی نظری، عملی و فناورانه برای بازتعریف و بسط عاملیت هرمنوتیک انسان در عصر هوش مصنوعی ارائه می‌دهد و استدلال می‌کند که مواجهه مسئولانه و نقادانه با این فناوری‌ها نه صرفاً ضرورتی فنی، بلکه ضرورتی فلسفی و اخلاقی است؛ ضرورتی که می‌تواند به بازتعریف نسبت انسان، متن و فناوری در افق هرمنوتیک دیجیتال بینجامد.

کلیدواژگان: هرمنوتیک فلسفی، هوش مصنوعی، عاملیت انسانی، فهم، اخلاق فناوری

تاریخ دریافت: ۱۴۰۴/۰۸/۱۲

تاریخ پذیرش: ۱۴۰۴/۱۰/۰۱

تاریخ انتشار: ۱۴۰۴/۱۱/۱۵

*نویسنده مسئول: rahnamaei@ut.ac.ir

آدرس مکاتبه: تهران، دانشگاه تهران، خیابان انقلاب، دانشکده اندیشه و معارف اسلامی

تلفن محل کار: -

مقدمه

تحولات شتابان در حوزه هوش مصنوعی، به‌ویژه ظهور مدل‌های زبانی بزرگ، نه‌تنها شیوه‌های تولید و پردازش متن را دگرگون ساخته‌اند، بلکه پرسش‌هایی بنیادین درباره ماهیت فهم، تفسیر و جایگاه انسان در فرآیند معنابخشی برانگیخته‌اند. توانایی این سامانه‌ها در تولید متون منسجم و ظاهراً معنادار، این پرسش اساسی را پیش می‌کشد که آیا می‌توان خروجی‌های آنها را مصداق «فهم» دانست یا باید آنها را صرفاً نوعی بازنمایی آماری و شبیه‌سازی زبانی تلقی کرد. این پرسش، صرفاً فنی یا کاربردی نیست، بلکه مستقیماً به مبانی فلسفی فهم و تفسیر در سنت هرمنوتیک فلسفی بازمی‌گردد.

از منظر هرمنوتیک فلسفی، فهم امری صرفاً شناختی یا محاسباتی نیست، بلکه رویدادی وجودی است که در بستر تجربه زیسته، افق‌های تاریخی، پیش‌فهم‌ها و عاملیت مفسر انسانی تحقق می‌یابد. بر این اساس، تجربه و فهم همواره به سوژه‌ای تاریخ‌مند، بدن‌مند و مسئول گره خورده‌اند. در نتیجه، آنچه در سامانه‌های هوش مصنوعی رخ می‌دهد، نه تجربه به معنای دقیق فلسفی آن و نه تحقق فهم اصیل، بلکه شبیه‌سازی صوری آثار و نتایج فهم انسانی در سطح ساختارهای زبانی و آماری است. از این رو، اطلاق مفاهیمی چون «فهم» یا «تجربه»

به هوش مصنوعی، تنها در معنایی استعاری قابل دفاع است و فاقد حیث هستی‌شناختی مستقل در چارچوب هرمنوتیک فلسفی است.

در کنار این چالش معرفت‌شناختی، مسایل روش‌شناختی و اخلاقی نیز اهمیت می‌یابند. جعبه سیاه بودن مدل‌های یادگیری عمیق، تکرارناپذیری نتایج و بازتولید سوگیری‌های نهفته در داده‌ها، اعتبار تفاسیر الگوریتمی را با تردید مواجه می‌سازد. افزون بر این، واگذاری فرآیندهای تفسیری به سامانه‌های غیرانسانی، خطر تضعیف نقش مفسر انسانی و کمرنگ شدن مسئولیت اخلاقی او در تولید و داوری معنا را به همراه دارد. این چالش‌ها را می‌توان نشانه‌هایی از به‌حاشیه‌رفتن عاملیت هرمنوتیک انسان دانست؛ عاملیتی که در سنت فلسفی هرمنوتیک، شرط امکان فهم، معنا و مسئولیت تلقی می‌شود.

بر این اساس، مساله اصلی این مقاله نه ارزیابی کارآمدی فنی سامانه‌های هوش مصنوعی، بلکه بررسی فلسفی امکان یا عدم امکان فهم الگوریتمی و پیامدهای آن برای عاملیت هرمنوتیک انسان است. مقاله با تکیه بر مبانی هرمنوتیک فلسفی استدلال می‌کند که فهم و تفسیر، بدون عاملیت تفسیری انسان امکان‌پذیر نیست و راهبردهای فناورانه و اخلاقی در مواجهه با هوش مصنوعی تنها زمانی معنا دار خواهند بود که در خدمت حفظ و بازتعریف این عاملیت قرار گیرند. از این منظر، چالش اصلی عصر هوش مصنوعی نه جایگزینی انسان با ماشین، بلکه صیانت از نقش انسان به عنوان مفسر نهایی و مسئول معنا در افق هرمنوتیک دیجیتال است.

با بررسی پیشینه پژوهش‌ها در موضوع مورد نظر، عنوان مشابه این پژوهش یافت نشد؛ اما آثار علمی مرتبط با آن را می‌توان به عنوان پیشینه این تحقیق تلقی کرد [Motalebi Karbakandi, 2013; Wang, 2021; Ojjeabu, 2024]. نوآوری این مقاله در آن است که به جای تمرکز صرف بر چالش‌های فنی یا اخلاقی هوش مصنوعی، آنها را در چارچوب هرمنوتیک فلسفی بازخوانی می‌کند. این پژوهش نشان می‌دهد که بحران فهم در هوش مصنوعی نه صرفاً ناشی از محدودیت‌های تکنیکی، بلکه پیامدی هستی‌شناختی و معرفت‌شناختی است. همچنین مقاله با ارایه چهار سطح راهبردی (نظری، عملی، فناورانه و اخلاقی) برای بازتعریف عاملیت انسانی، گامی فراتر از نقد گذاشته و مسیر مشخصی برای مواجهه فعال و مسئولانه با هوش مصنوعی پیشنهاد می‌کند.

بازخوانی چالش‌ها در سه سطح بنیادین

چالش‌های فلسفی-هرمنوتیک هوش مصنوعی را می‌توان در سه سطح بنیادین بازخوانی کرد: معرفت‌شناختی، روش‌شناختی و وجودی-اخلاقی. این تمایز نه صرفاً یک طبقه‌بندی توصیفی، بلکه چارچوبی تحلیلی برای نشان دادن چگونگی تضعیف یا به‌حاشیه‌رفتن عاملیت هرمنوتیک انسان در مواجهه با تفسیر الگوریتمی است. بر این اساس، بازخوانی این چالش‌ها شالوده‌ای نظری فراهم می‌آورد تا در ادامه، امکان صورت‌بندی راهبردهایی فراهم شود که در خدمت حفظ و بازتعریف نقش انسان به عنوان مفسر نهایی معنا قرار گیرند.

چالش‌های معرفت‌شناختی؛ محدودیت‌های فهم و مساله افق زیسته

در سنت هرمنوتیک فلسفی، فهم صرفاً یک فرآیند شناختی یا پردازش اطلاعات نیست، بلکه رویدادی وجودی است که در بستر افق‌های تاریخی، پیش‌فهم‌ها و تجربه زیسته مفسر انسانی تحقق می‌یابد. فهم، در این معنا، همواره با نوعی عاملیت تفسیری گره خورده است که شامل موضع‌گیری، مسئولیت و مشارکت فعال در فرآیند معنابخشی است. سامانه‌های هوش مصنوعی، از جمله مدل‌های زبانی بزرگ، فاقد این زیرساخت‌های بنیادین‌اند و از این رو، تحقق فهم در معنای هرمنوتیک آن برای آنها ناممکن است.

در اینجا لازم است به صورت صریح به یک پرسش بنیادین پاسخ داده شود: آیا می‌توان برای سامانه‌های هوش مصنوعی از «تجربه»، حتی به صورت حداقلی سخن گفت؟ در سنت پدیدارشناسی و هرمنوتیک فلسفی، تجربه همواره امری زیسته، التفاتی و ریشه‌مند در وجود انسانی است؛ تجربه بدون سوژه‌ای آگاه، بدن‌مند و تاریخ‌مند، از حیث هستی‌شناختی فاقد معناست. بر این اساس، آنچه در سامانه‌های هوش مصنوعی رخ می‌دهد، نه تجربه و نه حتی «تجربه زیسته» به معنای دقیق فلسفی آن، بلکه صرفاً شبیه‌سازی ساختاری آثار و برون‌دادهای تجربه انسانی در سطح داده‌های آماری و الگوهای زبانی است. از این رو، اگر اصطلاح «تجربه زیسته» در توصیف عملکرد هوش مصنوعی به کار رود، صرفاً کارکردی استعاری دارد و نباید آن را به مثابه نوعی تجربه واقعی یا بدیل تجربه انسانی تلقی کرد.

در سطح عملیاتی، معنا در مدل‌های یادگیری ماشین عمدتاً به صورت الگوهای آماری و همبستگی‌های کمی میان داده‌ها بازنمایی می‌شود؛ رویکردی که، برخلاف هرمنوتیک انسانی، فاقد پیوند ذاتی با افق‌های فرهنگی، تاریخی و زیسته است [Capurro, 2010]. پیامد چنین رویکردی، تولید متونی است که از منظر صوری منسجم و کارکردی‌اند، اما از حیث هستی‌شناختی، واجد «فهم‌نمایی الگوریتمی» بوده و نه فهم اصیل [Moor, 2006]. به بیان دیگر، در سطح معرفت‌شناختی، مساله اصلی آن است که خروجی‌های هوش مصنوعی قادرند شکل و صورت تفسیر را تقلید کنند، بی‌آنکه در فرآیند دیالکتیکی میان متن و مفسر انسانی - که بنیان عاملیت هرمنوتیک و شرط امکان معناست - مشارکت داشته باشند. این گسست، خطر جدایی معنا از حقیقت و تضعیف نقش انسان به عنوان مرجع نهایی تفسیر را در پی دارد.

چالش‌های روش‌شناختی؛ ابهام، غیرشفافیت و بحران اعتبار در فرآیند تفسیر

سطح دوم چالش‌ها به روش و مکانیزم تولید تفسیر در سامانه‌های هوش مصنوعی بازمی‌گردد؛ سطحی که مستقیماً با مساله عاملیت هرمنوتیک انسان و امکان داوری مسئولانه درباره معنا در پیوند است. سامانه‌های هوش مصنوعی، به‌ویژه مدل‌های زبانی پیشرفته مانند GPT-4، قادرند متونی با انسجام بالا و ساختار زبانی پیچیده تولید کنند، اما فرآیند دستیابی به این خروجی‌ها برای کاربر انسانی شفاف، قابل مشاهده و قابل ارزیابی نیست. این عدم شفافیت، امکان مشارکت فعال مفسر انسانی در فرآیند تفسیر را به شدت محدود می‌سازد.

در این مدل‌ها، معنا از طریق وزن‌های شبکه‌های عصبی و سازوکارهای آماری بازنمایی می‌شود، بی‌آنکه این فرآیندها تابع قواعد، معیارها و روش‌شناسی سنتی هرمنوتیک باشند [Moor, 2006]. افزون بر این، «جعبه سیاه» بودن این سامانه‌ها، تکرارناپذیری نسبی پاسخ‌ها و وابستگی شدید به داده‌های آموزشی، موجب می‌شود که نتوان اعتبار و مشروعیت تفاسیر ماشینی را براساس معیارهای پذیرفته‌شده در تفسیر فلسفی یا دینی ارزیابی کرد [Brown et al., 2020]. در نتیجه، داوری درباره صحت، خطا یا مسئولیت تفسیری این خروجی‌ها با ابهام جدی مواجه می‌شود.

از این منظر، چالش روش‌شناختی صرفاً مساله‌ای فنی یا مهندسی نیست، بلکه به مساله‌ای فلسفی درباره امکان یا عدم امکان عاملیت تفسیری در سامانه‌های غیرانسانی بدل می‌گردد. تقلیل سنت‌های تفسیری به مدل‌های محاسباتی، بدون حضور سوژه‌ای مسئول و قادر به پاسخ‌گویی تفسیری، خطر گسست میان معنا، حقیقت و مسئولیت را در پی دارد. بدین ترتیب، بحران روش‌شناختی در تفسیر الگوریتمی را باید نشانه‌ای از به‌حاشیه‌رفتن عاملیت هرمنوتیک انسان دانست؛ عاملیتی که در سنت هرمنوتیک فلسفی، شرط اعتبار، پاسخ‌گویی و مشروعیت هرگونه تفسیر محسوب می‌شود.

چالش‌های وجودی-اخلاقی؛ تهدید جایگاه مفسر انسانی و مسئولیت معنابخشی

سطح سوم چالش‌ها مستقیماً به جایگاه وجودی و مسئولیت اخلاقی مفسر انسانی در فرآیند فهم و تفسیر بازمی‌گردد؛ سطحی که در آن، مساله عاملیت هرمنوتیک انسان به صورت صریح و تعیین‌کننده مطرح می‌شود. در سنت هرمنوتیک فلسفی، تفسیر صرفاً عملی شناختی یا زبانی نیست، بلکه کنشی وجودی است که با زیست جهان مفسر، تجربه‌های عاطفی، ایمان، نیت‌مندی و تعهد اخلاقی او پیوندی ناگسستنی دارد. از این منظر، معنا نه صرفاً «تولید» می‌شود و نه «محاسبه»، بلکه در افق زندگی انسانی و در نسبت با مسئولیت تفسیری پدیدار می‌گردد. سامانه‌های غیرانسانی، به دلیل فقدان این ابعاد وجودی، قادر به مشارکت واقعی در فرآیند عرضه و تصدیق معنا نیستند [Capurro, 2010].

با این حال، گسترش تفسیر الگوریتمی می‌تواند به تدریج مرز میان تاویل انسانی و تولید ماشینی متن را کم‌رنگ ساخته و به تضعیف جایگاه مفسر انسانی بینجامد. این وضعیت، صرفاً یک چالش معرفت‌شناختی نیست، بلکه پیامدهایی عمیق در سطح اخلاقی به همراه دارد؛ زیرا سامانه‌های یادگیری ماشین، به واسطه اتکای خود به داده‌های تاریخی و الگوهای آماری، در معرض بازتولید سوگیری‌های الگوریتمی و تثبیت معانی ناعادلانه یا تقلیل‌گرایانه قرار دارند [Floridi, 2014]. در چنین شرایطی، فقدان یک سوژه مسئول که بتواند نسبت به پیامدهای تفسیری پاسخ‌گو باشد، اعتبار اخلاقی تفسیر را به طور بنیادین مخدوش می‌سازد.

این چالش در بسترهای دینی و معنوی، اهمیتی دوچندان می‌یابد؛ چرا که تفسیر متون دینی نه تنها با فهم مفهومی، بلکه با ایمان، تعهد و حساسیت‌های فرهنگی و اعتقادی گره خورده است. تجربه معنوی، به مثابه بعدی اساسی از فهم دینی، مستلزم حضور عاطفی، دعا، زیست دینی و مشارکت وجودی مفسر است؛ اموری که در بستر صرفاً دیجیتال و الگوریتمی به سختی قابل تحقق‌اند. از این رو، واگذاری فرآیند تفسیر به سامانه‌های غیرانسانی، خطر تهی‌شدن معنا از بعد وجودی و تقلیل تفسیر به سطحی صرفاً کارکردی را در پی دارد.

در نتیجه، چالش وجودی-اخلاقی هوش مصنوعی را باید نه مساله‌ای حاشیه‌ای، بلکه نقطه کانونی مناقشه هرمنوتیک معاصر دانست؛ نقطه‌ای که در آن، حفظ عاملیت هرمنوتیک انسان به عنوان شرط امکان معنا، مسئولیت و حقیقت، به مساله‌ای بنیادین بدل می‌شود. این سطح از چالش نشان می‌دهد که بدون حضور مفسر انسانی به عنوان سوژه‌ای مسئول، هیچ سازوکار فناورانه‌ای نمی‌تواند مدعی تفسیر معتبر و اخلاقاً موجه باشد.

پیوند با راهبردها؛ از شناخت چالش تا کنش معنادار

سه سطح چالش معرفت‌شناختی، روش‌شناختی و وجودی-اخلاقی که در این مقاله تبیین شد، صرفاً توصیف وضعیت موجود نیستند، بلکه چارچوبی تحلیلی برای صورت‌بندی راهبردهای معنادار در مواجهه با هوش مصنوعی در عرصه تفسیر فراهم می‌آورند. از منظر هرمنوتیک فلسفی، راهبردها تنها زمانی موجه و اثربخش خواهند بود که در نسبت مستقیم با این سطوح چالش و با هدف صیانت از عاملیت هرمنوتیک انسان طراحی شوند.

در سطح معرفت‌شناختی، این امر مستلزم بازاندیشی در چیستی فهم، تمایز آن از پردازش داده و بازتعریف معیارهای اعتبار و حقیقت در تفسیر است. در سطح روش‌شناختی، راهبردها باید معطوف به افزایش شفافیت، قابلیت بازبینی و امکان داوری انسانی درباره تفاسیر ماشینی باشند، به گونه‌ای که نقش مفسر انسانی به عنوان مرجع نهایی تفسیر حفظ گردد. در سطح وجودی-اخلاقی نیز، صیانت از جایگاه انسان به مثابه فاعل معنابخش و تضمین عدالت تفسیری، به عنوان شرط بنیادین هرگونه مداخله فناورانه، اهمیت می‌یابد [Floyd et al., 1992].

بر این اساس، راهبردهای پیشنهادی در ادامه مقاله نه صرفاً پاسخی فنی به مخاطرات هوش مصنوعی، بلکه تلاشی فلسفی برای بازتعریف نسبت انسان، متن و فناوری در افق هرمنوتیک دیجیتال‌اند. این راهبردها بر این پیش‌فرض استوارند که هوش مصنوعی فاقد تجربه زیسته و فهم اصیل بوده و تنها در سطح شبیه‌سازی معنا عمل می‌کند؛ پیش‌فرضی که مبنای دفاع از عاملیت هرمنوتیک انسان و محور اصلی استدلال مقاله را شکل می‌دهد. از این منظر، هدف نهایی مقاله نه مهار فناوری، بلکه تثبیت نقش انسان به عنوان مفسر مسئول و کانون معنا در عصر هوش مصنوعی است.

صیانت و بازتعریف عاملیت هرمنوتیک انسان در عصر هوش مصنوعی

در عصر هوش مصنوعی، مساله محوری دیگر صرفاً توسعه توانمندی‌های فناورانه نیست، بلکه چگونگی صیانت و بازتعریف عاملیت هرمنوتیک انسان در فرآیند فهم و تفسیر معناست. همان‌گونه که در بخش‌های پیشین نشان داده شد، سامانه‌های هوش مصنوعی - علی‌رغم توانایی در شبیه‌سازی ساختارهای زبانی و معنایی - فاقد تجربه زیسته، افق تاریخی و مسئولیت تفسیری‌اند و از این رو نمی‌توانند واجد عاملیت هرمنوتیک به معنای دقیق فلسفی باشند. بر این اساس، راهبردهای مواجهه با هوش مصنوعی باید نه در جهت اعطای نقش تفسیری مستقل به ماشین، بلکه در راستای تثبیت نقش انسان به عنوان مفسر نهایی و فاعل معنابخش صورت‌بندی شوند.

در این چارچوب، راهبردهای پیشنهادی را می‌توان به دو دسته کلی تقسیم کرد: راهبردهای نظری که معطوف به بازخوانی و تبیین مبانی معرفت‌شناختی، هستی‌شناختی و زبان‌شناختی فهم در سنت هرمنوتیک فلسفی‌اند و می‌کوشند تمایز میان فهم انسانی و شبیه‌سازی الگوریتمی معنا را به طور مفهومی روشن سازند؛ و راهبردهای عملی که ناظر به طراحی و ارزیابی سازوکارهایی هستند که امکان حفظ نقش انسان در فرآیند تفسیر و داوری معنا را در تعامل با سامانه‌های هوش مصنوعی فراهم می‌آورند. این تمایز امکان می‌دهد که از یک سو، بنیان‌های فلسفی بحث انسجام و عمق یابند و از سوی دیگر، این بنیان‌ها به صورت پاسخ‌هایی معنادار و مسئولانه در ساحت عمل ترجمه شوند.

راهبردهای نظری؛ چارچوب‌های مفهومی برای دفاع از فهم انسانی

راهبردهای نظری، در این مقاله، به مجموعه‌ای از چارچوب‌های مفهومی و تحلیلی اطلاق می‌شود که با تکیه بر هرمنوتیک فلسفی، امکان نقد پیش‌فرض‌های نهفته در تفسیر الگوریتمی و دفاع از فهم انسانی را فراهم می‌آورند. این راهبردها بر آن‌اند که نشان دهند فهم، برخلاف پردازش داده، مستلزم عاملیتی است که در افق تجربه زیسته، پیش‌فهم‌ها و مسئولیت تفسیری شکل می‌گیرد. از این منظر، نقش راهبردهای نظری نه ارایه تعاریف انتزاعی، بلکه فراهم‌ساختن مبنایی فلسفی برای تمایزگذاری میان تفسیر اصیل انسانی و فهم‌نمایی الگوریتمی است.

در حوزه‌ای که گاه از آن با عنوان «هرمنوتیک دیجیتال» یاد می‌شود، راهبردهای نظری تنها زمانی معنا دارند که به جای همسان‌سازی انسان و ماشین، به تحلیل انتقادی نسبت میان انسان، متن و فناوری بپردازند. این راهبردها مسیر را برای بازاندیشی مفاهیم کلیدی چون فهم، معنا، تفسیر و عاملیت هموار می‌سازند و نشان می‌دهند که چگونه می‌توان بدون نفی فناوری، از تقلیل معنا به محاسبه و از به‌حاشیه‌رفتن مفسر انسانی جلوگیری کرد.

در سطح نظری، هرمنوتیک دیجیتال بر این نکته تأکید می‌ورزد که انسان، به واسطه دانش تفسیری، تجربه زیسته و تعهدات ارزشی خود، قادر است افق‌های معنایی محدودشده در بستر فناوری‌های دیجیتال را بازسازی یا گسترش دهد. در غیاب این عاملیت انسانی، تفسیر داده‌های دیجیتال به فرآیندی تقلیل‌گرایانه بدل می‌شود که صرفاً به پردازش اطلاعات بسنده کرده و از درک معنای انسانی ناتوان است. از این منظر، به‌کارگیری فعال عاملیت هرمنوتیک شرط امکان هرگونه فهم معنادار در زیست‌جهان دیجیتال تلقی می‌گردد.

در این چارچوب، تمایز میان هرمنوتیک دیجیتال روش‌شناختی و هرمنوتیک هستی‌شناختی اهمیتی بنیادین می‌یابد. رویکرد روش‌شناختی عمدتاً بر ابزارها، الگوریتم‌ها و تکنیک‌های تفسیر دیجیتال تمرکز دارد، در حالی که رویکرد هستی‌شناختی به پرسش از جایگاه انسان به عنوان فاعل فهم، ظرفیت معناسازی و مسئولیت تفسیری او می‌پردازد [Romele et al., 2020]. تأکید مقاله حاضر بر این است که بدون این سطح هستی‌شناختی، هرمنوتیک دیجیتال به مجموعه‌ای از رویه‌های فنی فروکاسته می‌شود که فاقد عمق تفسیری و اخلاقی‌اند.

از این منظر، فناوری‌هایی چون کلان‌داده و اینترنت اشیا را نباید صرفاً زیرساخت‌هایی خنثی یا ابزارهای کارکردی تلقی کرد، بلکه آنها حامل الگوهای خاصی از معنا، قدرت و جهت‌مندی‌اند که در فرآیند تفسیر اجتماعی و انسانی شکل می‌گیرند. هرچند این فناوری‌ها در حوزه‌های اقتصادی و سازمانی به عنوان محرک‌های نوآوری و تحول معرفی می‌شوند [Aalipour & Zanganeh, 2022]، اما از منظر هرمنوتیک، اهمیت آنها در این است که بدون مداخله انتقادی و تفسیری انسان، می‌توانند افق‌های معنا را به منطق کارآمدی، محاسبه‌پذیری و کنترل تقلیل دهند. از این رو، راهبرد نظری هرمنوتیک دیجیتال باید معطوف به بازپس‌گیری نقش انسان در تفسیر این فناوری‌ها و جلوگیری از استحاله معنا در منطق صرفاً فنی باشد.

در بستر شتابان تحولات هوش مصنوعی، بازتعریف نقش مفسر انسانی دیگر صرفاً یک دغدغه نظری نیست، بلکه ضرورتی عملی و اخلاقی به شمار می‌آید. هرمنوتیک فلسفی، به‌ویژه در قرائت گادامر، با مفاهیمی چون «پیش‌فهم»، «گفت‌وگو» و «هم‌افقی» نشان می‌دهد که فهم همواره رویدادی انسانی است که در تعامل میان مفسر و متن و در افق تاریخی-فرهنگی خاصی تحقق می‌یابد [Gadamer, 2004]. از این منظر، مفسر انسانی در مواجهه با سامانه‌های هوش مصنوعی نه هم‌عرض ماشین، بلکه تنظیم‌کننده افق‌های معناست؛ یعنی فاعلی که با آگاهی از محدودیت‌های الگوریتمی، جهت‌گیری‌های معنایی برون‌دادهای ماشینی را ارزیابی، اصلاح و بازسازمان‌دهی می‌کند.

این تنظیم افق‌ها نباید به منزله اعطای عاملیت تفسیری به ماشین تلقی شود، بلکه بیانگر مسئولیتی انسانی در مواجهه با میانجی‌های فناورانه است. الگوریتم‌ها می‌توانند الگوهای زبانی را بازنمایی کنند، اما این مفسر انسانی است که با اتکا به تجربه زیسته، زمینه اجتماعی و تعهدات اخلاقی، تصمیم می‌گیرد کدام معنا معتبر، کدام تفسیر قابل قبول و کدام قرائت گمراه‌کننده است. بدین ترتیب، عاملیت هرمنوتیک همچنان به طور انحصاری به انسان تعلق دارد و فناوری صرفاً در مقام ابزار یا بستر ظهور معنا باقی می‌ماند.

از منظر هرمنوتیک دیجیتال، این مسئولیت تفسیری مستلزم آن است که مفسر انسانی نسبت به نحوه عملکرد سامانه‌های هوش مصنوعی، محدودیت‌ها و پیامدهای آنها آگاهی انتقادی داشته باشد. مفاهیمی چون خودکارآمدی و ادراک ریسک، در این چارچوب، نه به عنوان متغیرهای صرفاً روان‌شناختی، بلکه به مثابه شروط امکان ایفای مسئولیت تفسیری فهم می‌شوند؛ بدین معنا که تنها انسانی که توانایی مداخله، ارزیابی و مقاومت در برابر اقتدار ظاهری خروجی‌های ماشینی را دارد، می‌تواند نقش تنظیم‌گر افق‌های معنا را به‌درستی ایفا کند

[Pickering et al., 2017]. این راهبرد عملی، از این رو، در خدمت تقویت عاملیت هرمنوتیک انسان و جلوگیری از استحاله فهم در منطق خودکار فناوری قرار می‌گیرد.

راهبردهای عملی؛ ترجمه عملی دفاع از عاملیت هرمنوتیک انسان

راهبردهای عملی در این مقاله ناظر به آن دسته از سازوکارها و رویه‌هایی هستند که امکان تحقق عملی چارچوب‌های نظری هرمنوتیک را در تعامل انسان و هوش مصنوعی فراهم می‌آورند. مقصود از این راهبردها صرفاً بهینه‌سازی عملکرد سامانه‌های هوشمند نیست، بلکه تضمین آن است که فرآیند تفسیر و تصمیم‌گیری معنایی همچنان تحت نظارت و مسئولیت مفسر انسانی باقی بماند. از این منظر، هر مداخله عملی تنها در صورتی موجه تلقی می‌شود که به تقویت عاملیت هرمنوتیک انسان بینجامد، نه به جایگزینی آن با سازوکارهای خودکار.

در این چارچوب، طراحی سامانه‌های هوش مصنوعی باید به گونه‌ای صورت گیرد که امکان فهم، ارزیابی و مداخله انتقادی مفسر انسانی را حفظ کند. مفاهیمی چون شفافیت الگوریتمی، اخلاق‌محوری در طراحی و ارتقای سواد داده‌ای، در اینجا نه به عنوان توصیه‌های صرفاً فنی، بلکه به مثابه شروط عملی صیانت از مسئولیت تفسیری انسان مطرح می‌شوند. نهادینه‌سازی رویکرد «اخلاق از طریق طراحی» و گشودن امکان بازخورد انسانی مستمر، به مفسر اجازه می‌دهد که در برابر اقتدار ظاهری خروجی‌های ماشینی موضعی فعال و نقادانه اتخاذ کند [Alvand et al., 2023].

از منظر هرمنوتیک، فناوری‌های هوشمند دو مسیر متفاوت پیش روی مفسر انسانی قرار می‌دهند: از یک سو، خطر محدودسازی افق‌های معنا از طریق الگوریتم‌های فیلترکننده و داده‌های از پیش قالب‌بندی‌شده؛ و از سوی دیگر، امکان گسترش افق‌های تفسیری از طریق فراهم‌آوردن ابزارهای نوین برای تحلیل، مقایسه و بازخوانی معنا. راهبردهای عملی پیشنهادی در این مقاله بر این پیش‌فرض استوارند که تحقق مسیر دوم تنها در صورتی ممکن است که انسان جایگاه خود را به عنوان فاعل تفسیر حفظ کرده و نقش فناوری را در حد ابزار یا میانجی نگاه دارد.

چنان که فلوریدی نیز تأکید می‌کند، فناوری‌های هوشمند همزمان حامل ظرفیت‌های توانمندساز و محدودکننده‌اند [Floridi, 2014]. با این حال، آنچه اغلب در کاربردهای عملی مغفول می‌ماند، توجه به پیش‌فرض‌های اجتماعی، فرهنگی و ارزشی است که در طراحی و کاربست الگوریتم‌ها رسوب می‌کنند. از این رو، راهبرد عملی دفاع از عاملیت هرمنوتیک انسان مستلزم آن است که این پیش‌فرض‌ها به صورت آگاهانه و انتقادی مورد بازاندیشی قرار گیرند، تا تفسیر دیجیتال به ابزاری برای تثبیت سلطه‌های پنهان یا تقلیل معنا بدل نگردد.

در این چارچوب، سامانه‌های هوش مصنوعی نه به مثابه فاعلان تفسیری، بلکه به عنوان میانجی‌های فناورانه عمل می‌کنند که می‌توانند شرایط کنش و تفسیر انسانی را تسهیل یا محدود سازند. این تسهیل، در صورتی که تحت نظارت آگاهانه مفسر انسانی صورت گیرد، می‌تواند احساس توانمندی و خودکارآمدی او را در مواجهه با داده‌ها و خروجی‌های پیچیده تقویت کند؛ در مقابل، محدودیت‌های فنی، نهادی یا مقرراتی این سامانه‌ها ممکن است به تضعیف این خودکارآمدی و شکل‌گیری رویکردی محافظه‌کارانه در تفسیر بینجامد. از این رو، خودکارآمدی و مدیریت ریسک را باید نه به ماشین، بلکه به مثابه ظرفیت‌های انسانی در نظر گرفت که امکان حفظ فاصله انتقادی با خروجی‌های ماشینی را فراهم می‌آورند [Pickering et al., 2017].

بر این اساس، مفسر انسانی موظف است ضمن بهره‌گیری سنجیده از قابلیت‌های پشتیبان فناوری، مرزهای تفسیر انسانی را آگاهانه حفظ و حتی گسترش دهد و از واگذاری نقش داور معنایی به سامانه‌های خودکار پرهیز کند. در این معنا، نقش فناوری به ارایه ابزار و امکانات محدود می‌شود و مسئولیت نهایی معنا، اعتبار و پیامدهای تفسیری همچنان بر عهده انسان باقی می‌ماند. چنین موضعی امکان تعامل مسئولانه‌تر با فناوری را فراهم می‌سازد، بی‌آنکه به استحاله عاملیت هرمنوتیک انسان بینجامد.

در سطح عملی، تحقق این موضع مستلزم نهادینه‌سازی اخلاق در طراحی و به‌کارگیری نظام‌های داده‌محور، ارتقای سواد الگوریتمی مفسران انسانی و مشارکت فعال ذی‌نفعان در فرآیندهای تصمیم‌گیری تفسیری است. این اقدامات، نه به منظور افزایش استقلال ماشین، بلکه برای تقویت ظرفیت انسانی در بازتولید معنا و ارزش در افق عدالت، شفافیت و توسعه پایدار صورت می‌پذیرند [Aalipour & Zanganeh, 2022].

در نهایت به‌اختصار می‌توان گفت که مساله اصلی در مواجهه هرمنوتیک با هوش مصنوعی، نه ارتقای توان تفسیری ماشین، بلکه صیانت و بازتعریف عاملیت هرمنوتیک انسان است. تحلیل چالش‌ها در سطوح معرفت‌شناختی، روش‌شناختی و وجودی-اخلاقی روشن ساخت که سامانه‌های هوش مصنوعی، علی‌رغم شبیه‌سازی معنا، فاقد تجربه زیسته، افق تاریخی و مسئولیت تفسیری‌اند و از این رو نمی‌توانند واجد فهم اصیل باشند. بر این مبنا، راهبردهای نظری و عملی مقاله به گونه‌ای صورت‌بندی شدند که فهم انسانی همچنان کانون فرآیند تفسیر باقی بماند و فناوری در مقام میانجی، نه جایگزین، ایفای نقش کند. در افق هرمنوتیک دیجیتال، آینده فهم و معنا بیش از آنکه به توان محاسباتی ماشین وابسته باشد، به پایداری عاملیت تفسیری انسان و آگاهی انتقادی او نسبت به پیش‌فرض‌های نهفته در فناوری‌های هوشمند گره خورده است.

شفافیت الگوریتمی در خدمت عاملیت هرمنوتیک انسان

در بستر گسترش فزاینده کاربردهای هوش مصنوعی در حوزه‌های علمی، فرهنگی و دینی، مساله اصلی از منظر هرمنوتیک نه افزایش استقلال تفسیری ماشین، بلکه تضمین پاسخ‌گویی و شفافیت آن در برابر مفسر انسانی است. در چارچوب هرمنوتیک دیجیتال، استفاده از سامانه‌های «جعبه سیاه» در تحلیل متون معنادار، به‌ویژه متون دینی، می‌تواند به گسست میان معنا، زمینه و مسئولیت تفسیری بینجامد؛ گسستی که پیامدهای معرفتی و اخلاقی آن به مراتب فراتر از خطاهای فنی است. از این رو، راهبردهای فناورانه تنها در صورتی موجه‌اند که به تقویت نقش انسان به عنوان فاعل نهایی فهم و داور معنا منجر شوند.

بر این اساس، مفهوم «هوش مصنوعی توضیح‌پذیر» (Explainable Artificial Intelligence; XAI) در این مقاله نه به مثابه تلاشی برای اعطای فهم به ماشین، بلکه به عنوان سازوکاری برای قابل فهم و قابل نقد ساختن فرآیندهای الگوریتمی برای انسان تلقی می‌شود. شفافیت الگوریتمی، همراه با رویکردهای تعاملی نظیر سامانه‌های "انسان در چرخه" (Human-in-the-loop)، این امکان را فراهم می‌آورد که مفسر انسانی بتواند در فرآیند تولید، ارزیابی و اصلاح برون‌دادهای ماشینی مداخله کند و از واگذاری ناخواسته داور معنایی به سیستم‌های خودکار جلوگیری نماید. در این چارچوب، ابزارهای پردازش زبان طبیعی و مدل‌های ترکیبی تنها زمانی مشروعیت هرمنوتیک می‌یابند که در مقام میانجی‌های فناورانه عمل کرده و فهم انسانی را تقلیل یا جایگزین نکنند، بلکه امکان بسط و نظارت انتقادی بر آن را فراهم سازند.

راهبردهای فنی-فناورانه

این سطح به مجموعه‌ای از ابزارها و راهبردهای فنی می‌پردازد که هدف آنها شفاف‌سازی فرآیندهای تصمیم‌گیری الگوریتمی، افزایش تفسیرپذیری خروجی‌ها و ارتقای کیفیت پردازش زبان در تعامل انسان و ماشین است. تمرکز این راهبردها بر توسعه فناوری‌هایی است که به جای جایگزینی فهم انسانی، در خدمت تقویت نظارت، داوری و مسئولیت هرمنوتیک مفسر انسانی عمل می‌کنند. در این چارچوب، رویکردهایی همچون هوش مصنوعی توضیح‌پذیر، مدل‌های ترکیبی در پردازش زبان طبیعی و سامانه‌های تعاملی انسان در چرخه به عنوان ابزارهای کلیدی برای تحقق تعامل انسان‌محور و قابل اعتماد با سامانه‌های زبانی هوشمند بررسی می‌شوند.

هوش مصنوعی توضیح‌پذیر به مثابه شرط پاسخ‌گویی تفسیری

با گسترش حجم داده‌ها و افزایش توان محاسباتی، سامانه‌های هوش مصنوعی به طور فزاینده‌ای در فرآیندهای تصمیم‌گیری به کار گرفته می‌شوند؛ با این حال، از منظر هرمنوتیک، مساله اصلی نه انجام وظایفی شبیه به کنش انسانی، بلکه امکان پاسخ‌گویی این سامانه‌ها در برابر مفسر انسانی است. هوش مصنوعی توضیح‌پذیر در این چارچوب، نه به منزله برخورداری ماشین از فهم یا تجربه، بلکه به عنوان سازوکاری برای شفاف‌سازی فرآیندهای تصمیم‌سازی الگوریتمی برای انسان معنا می‌یابد. توضیح‌پذیری، شرطی است که امکان داوری انتقادی، ارزیابی زمینه‌ای و اصلاح تصمیم‌های ماشینی را برای مفسر انسانی فراهم می‌سازد و از استقرار غیرپاسخ‌گو در سطح الگوریتمی جلوگیری می‌کند.

از این منظر، اهمیت هوش مصنوعی توضیح‌پذیر در حوزه‌های حساس - از جمله تفسیر داده‌های معنادار و تصمیم‌های واجد پیامدهای اخلاقی - در آن است که امکان بازسازی زنجیره تصمیم‌گیری الگوریتمی به گونه‌ای قابل فهم برای انسان را فراهم می‌آورد. پژوهش‌های تجربی نشان داده‌اند که استفاده از مدل‌های توضیح‌پذیر می‌تواند قواعد تصمیم‌سازی را به شکلی شفاف آشکار سازد و امکان بازبینی و اصلاح آنها را مهیا کند [Daraei, 2024]. در افق هرمنوتیک دیجیتال، چنین قابلیت‌هایی نه صرفاً ابزاری برای افزایش اعتماد کاربر، بلکه شرط لازم برای حفظ مسئولیت تفسیری انسان و پیشگیری از تقلیل معنا به خروجی‌های خودکار محسوب می‌شوند. بدین ترتیب، هوش مصنوعی توضیح‌پذیر را باید نه گامی به سوی فهم ماشینی، بلکه ابزاری فناورانه برای تثبیت جایگاه انسان به عنوان فاعل نهایی معنا تلقی کرد.

با آنکه هوش مصنوعی قابل توضیح تلاشی معنادار برای شفاف‌سازی سازوکارهای درونی سامانه‌های یادگیری ماشین به شمار می‌آید، این شفافیت به‌ویژه در شبکه‌های عصبی عمیق با محدودیت‌های بنیادین مواجه است. هرچند این شبکه‌ها بر پایه پارامترهای عددی دقیق عمل می‌کنند، اما فقدان پیوند قابل ردیابی میان این پارامترها و معنای خروجی‌ها موجب می‌شود که تغییرات جزئی در داده‌ها به تفسیرهای به‌کلی نادرست بینجامد؛ خطاهایی که در برخی آزمایش‌ها به صورت تشخیص یک شیء معنادار به جای شیئی کاملاً نامرتبط بروز یافته‌اند و حتی می‌توانند در قالب حملات خصمانه مورد سوءاستفاده قرار گیرند [High-Level Expert Group on Artificial Intelligence, 2019]. از منظر هرمنوتیک، این خطاها صرفاً نقص فنی نیستند، بلکه نشانه فقدان فهم زمینه‌ای، افق معنایی و توان تشخیص تفاوت‌های معنادارند؛ عناصری که تنها در چارچوب عاملیت انسانی معنا می‌یابند.

بر همین اساس، اطلاق عنوان «جعبه سیاه» به مدل‌های پیچیده یادگیری ماشین، به‌ویژه شبکه‌های عصبی عمیق، دلالتی فراتر از دشواری فنی دارد و به مساله‌ای هرمنوتیک بدل می‌شود. هنگامی که فرآیند تصمیم‌گیری یک سامانه برای مفسر انسانی غیرقابل فهم باقی می‌ماند، امکان داوری انتقادی، مسئولیت‌پذیری و نسبت‌دادن

معنا از میان می‌رود. اهمیت این مساله در حوزه‌های حساسی چون پزشکی، حقوق و اقتصاد دوچندان است، زیرا در این زمینه‌ها خطای تفسیری می‌تواند پیامدهای جبران‌ناپذیری داشته باشد [Lipton, 2017]. هرچند لیپتون به درستی بر مشکل «جعبه سیاه» بودن مدل‌ها تاکید می‌کند، تمرکز او عمدتاً بر توصیف مساله باقی می‌ماند و کمتر به پیامدهای هرمنوتیک آن، یعنی ضرورت حفظ مفسر انسانی به عنوان فاعل نهایی فهم و داوری معنا، می‌پردازد. از این رو، چالش شفافیت الگوریتمی را باید نه صرفاً مساله‌ای فنی، بلکه شرطی اساسی برای صیانت از عاملیت هرمنوتیک انسان تلقی کرد.

اصل «توضیح‌پذیری» که به روشنی در آثار فلوریدی و کول مطرح شده است، از این جهت اهمیت بنیادین دارد که همچون گزاره‌ای شرطی، دو بعد به هم پیوسته معرفت‌شناختی و اخلاقی را دربر می‌گیرد. از بعد معرفت‌شناختی، توضیح‌پذیری به این معناست که اگر پرسش «این سامانه چگونه کار می‌کند؟» مطرح شود، آنگاه باید امکان ارایه پاسخی فهم‌پذیر برای مفسر انسانی وجود داشته باشد؛ پاسخی که نه صرفاً فنی، بلکه معنادار و قابل ادغام در افق تفسیری انسان باشد. از بعد اخلاقی، توضیح‌پذیری شرط امکان مسئولیت‌پذیری است؛ بدین معنا که اگر خطا یا پیامدی نامطلوب رخ دهد، آنگاه باید بتوان نسبت آن را به کنشگران انسانی و ساختارهای تصمیم‌گیری روشن ساخت. از این منظر، توضیح‌پذیری نه یک فضیلت افزوده، بلکه «شرط لازم» تحقق اصول اخلاقی بنیادینی چون عدالت، خیررسانی و صیانت از خودآیینی انسان به شمار می‌آید؛ زیرا در غیاب آن، این اصول در سطحی انتزاعی و غیرقابل اعمال باقی می‌مانند [Floridi & Cowls, 2019].

در همین راستا، برخی گزارش‌های سیاست‌گذارانه توضیح‌پذیری را به توانایی انسان در ردیابی و درک تصمیمات سامانه‌های هوشمند تعریف می‌کنند؛ مشروط بر آنکه توضیح ارایه‌شده متناسب با سطح تخصص مخاطب، به‌موقع و از حیث مفهومی روشن باشد. با این حال، در عمل تعارضی میان «دقت» مدل‌های پیچیده و «شفافیت» آنها پدید می‌آید که حل آن نیازمند داوری تفسیری و مداخله انسانی است [High-Level Expert Group on Artificial Intelligence, 2019]. هرچند این اسناد بر اهمیت توضیح‌پذیری تاکید می‌ورزند، توصیه‌های آنها عمدتاً کلی و سیاست‌محور باقی می‌ماند و کمتر به این نکته می‌پردازد که بدون حفظ عاملیت هرمنوتیک انسان به عنوان مفسر نهایی، حتی شفاف‌ترین توضیح‌ها نیز به داده‌هایی خام و فاقد معنا فروکاسته می‌شوند.

سیستم‌های تعاملی؛ طراحی ابزارهای هوش مصنوعی انسان در چرخه

سیستم‌های تعاملی مبتنی بر الگوی انسان در چرخه به گونه‌ای طراحی می‌شوند که انسان به صورت فعال در چرخه تصمیم‌گیری، یادگیری و ارزیابی سامانه‌های هوش مصنوعی حضور داشته باشد. در این رویکرد، کاربر صرفاً مصرف‌کننده خروجی مدل نیست، بلکه از طریق ارایه بازخورد، اصلاح ورودی‌ها و تنظیم چارچوب‌های معنایی، در جهت‌دهی به فرآیند یادگیری ماشین مشارکت می‌کند. چنین سامانه‌هایی، افزون بر افزایش دقت و کارایی فنی، امکان شفاف‌سازی تصمیمات الگوریتمی و اعمال نظارت تفسیری انسان را فراهم می‌سازند و از این طریق، بستر تحقق عاملیت انسانی در تعامل با فناوری را تقویت می‌کنند.

مدل‌های انسان در چرخه بر این پیش‌فرض استوارند که اتکای صرف به داده‌های آماری، به تولید خروجی‌هایی منجر می‌شود که اگرچه ممکن است از حیث محاسباتی دقیق باشند، اما فاقد عمق تفسیری و حساسیت زمینه‌ای‌اند. ورود بازخورد انسانی به چرخه یادگیری، این امکان را فراهم می‌آورد که مدل نه‌تنها الگوهای آماری، بلکه قیود معنایی، هنجاری و تفسیری را نیز لحاظ کند. از این منظر، همان‌گونه که آمرشی و همکاران نشان

می‌دهند، ترکیب توانمندی‌های انسان و ماشین به شکل‌گیری سامانه‌هایی می‌انجامد که هم کارآمدتر و هم از نظر تفسیری قابل اعتمادترند [Amershi et al., 2014].

در بستر هرمنوتیک دیجیتال، اهمیت این رویکرد دوچندان می‌شود؛ زیرا تفسیر متون دینی و فلسفی، فرآیندی صرفاً داده‌محور نیست، بلکه کنشی معنایی، تاریخی و چندلایه است که بدون مداخله آگاهانه مفسر انسانی ناقص باقی می‌ماند. در این چارچوب، می‌توان از نسبت عقل و وحی در سنت تفسیری اسلامی به مثابه تمثیلی روشنگر بهره گرفت؛ آن‌گونه که آیت‌الله جوادی آملی تأکید می‌کند، عقل برهانی و نقل معتبر هر دو منبع فهم‌اند، اما هیچ یک به تنهایی تمامی افق معنا را دربر نمی‌گیرند و نسبت آنها نسبتی تکمیلی و هدایت‌شده است [Javadi Amoli, 2016]. به طور مشابه، در هرمنوتیک دیجیتال نیز قدرت پردازش ماشینی و داوری تفسیری انسان باید در نسبتی تعاملی و غیرجان‌شینانه قرار گیرند؛ نسبتی که در آن، ماشین نقش ابزار پشتیبان را ایفا می‌کند و انسان همچنان فاعل نهایی معنا باقی می‌ماند. تنها در چنین چارچوبی است که فهم متن از سطح داده‌کاوی فراتر رفته و به افق معناشناسی و مسئولیت تفسیری ارتقا می‌یابد.

نه از آن رو که سامانه‌های هوش مصنوعی فاقد توان پردازش سریع و گسترده داده‌ها هستند، بلکه از آن جهت که بدون هدایت انسانی، فهم عمیق مفاهیم و معانی، به‌ویژه در متون مقدس به تدریج تهی می‌گردد. آنچه در غیاب مفسر انسانی رخ می‌دهد، نه فقدان محاسبه، بلکه فروکاست معنا به سطح همبستگی‌های آماری است. در این زمینه، تاملات علامه جعفری به مثابه پشتوانه‌ای فلسفی هشدار می‌دهد که هرگونه یک‌سویه‌گرایی چه ذهن‌گرایی محض و چه تجربه‌گرایی تقلیل‌گرایانه، به انسداد مسیر معرفت می‌انجامد. به تعبیر او، معرفت انسانی تنها در توازن میان واقعیت عینی، استدلال و خلاقیت ذهنی امکان‌پذیر است [Jafari Tabrizi, 2012]. این نکته نشان می‌دهد که اتکای صرف به داده‌های محسوس و آماری، بدون مداخله عقل تفسیری، به شناختی سطحی و فاقد عمق معنایی منجر می‌شود.

از منظر طراحی سامانه‌های هوش مصنوعی تعاملی، این تحلیل فلسفی دلالتی فناورانه دارد: اگر سیستم‌های انسان در چرخه صرفاً به منزله ابزار اصلاح خطای آماری تلقی شوند و نه سازوکاری برای اعمال داوری تفسیری انسان، آنگاه عاملیت انسانی به سطح نظارت حداقلی فروکاسته می‌شود. در مقابل، رویکرد انسان‌محور به هوش مصنوعی بر آن است که کرامت، قضاوت و نقش معنابخش انسان در مرکز طراحی و بهره‌برداری از سامانه‌ها قرار گیرد [High-Level Expert Group on Artificial Intelligence, 2019]. این نگاه با حکمت صدرایی نیز همخوان است؛ جایی که انسان، با حفظ وحدت هویتی خود، دارای مراتب حسی، خیالی و عقلی دانسته می‌شود و فهم، حاصل تعامل این مراتب است [Sadr Al-Din Shirazi, 2008]. نادیده‌گرفتن این مراتب در طراحی فناوری، تعامل انسان و ماشین را به رابطه‌ای صرفاً ابزاری تقلیل می‌دهد و خطر حذف ابعاد وجودی معنا را تشدید می‌کند.

از سوی دیگر، آمزشی نشان می‌دهد که یکی از چالش‌های اساسی سیستم‌های تعاملی، فاصله میان تخصص فنی توسعه‌دهندگان و دانش زمینه‌ای کاربران است [Amershi et al., 2014]. این امر شبیه همان چیزی است که آیت‌الله سبحانی در باب «تفسیر به رای» هشدار می‌دهد [Sobhani Tabrizi, 2019]:

«مفاهیمی که آیات قرآن بر آن دلالت ندارند بر آیه تحمیل نشود. تحمیل چنین معنی که برای آن در خود آیه و آیات دیگر شاهدهی نیست، تفسیر به باطن است که از نظر قواعد تفسیر کاملاً ممنوع است و گشودن این باب به روی مفسر یک نوع بازی با وحی الهی است.»

بنابراین، برای کارایی و اعتمادپذیری سامانه‌های هوش مصنوعی در هرمنوتیک دیجیتال، باید کاربر نه صرفاً به عنوان تامین‌کننده داده، بلکه به عنوان شریکی واقعی در شکل‌دهی معنا حضور یابد. این همان چیزی است که در فلسفه اسلامی نیز به عنوان «تفاعل عقل و وحی» مطرح می‌شود؛ «اصولاً بعضی حقایق است که با هدایت حس و تجربه و یا هدایت عقل، بشر به‌آسانی به آنها پی می‌برد و ایمان می‌آورد و آنها همانهاست که بشر در زندگی ساده حسی و عقلی خود با آنها مواجه است، اما بعضی حقایق است که مافوق این دو است (همان‌طور که حقایق عقلی مافوق حقایق حسی است) و لهذا احتیاج به هدایت وحی است» [Motahari, 2015].

در نهایت، همان‌گونه که کمیسیون اروپا تأکید می‌کند، هوش مصنوعی نباید خود را به جای انسان جا بزند و باید مرز خود را با فهم انسانی روشن سازد [High-Level Expert Group on Artificial Intelligence, 2019]. این سخن را می‌توان با آموزه اسلامی چنین تبیین کرد: «ماشین ابزار است، نه صاحب معنا؛ معنا صرفاً در قلمرو انسان برخوردار از عقل و روح شکل می‌گیرد و هر کوششی برای جایگزینی کامل انسان با ماشین، نقیض حقیقت انسانی است.

از سوی دیگر، آمرشی و همکاران نشان می‌دهند که یکی از چالش‌های بنیادین در طراحی سیستم‌های تعاملی انسان در چرخه، شکاف میان دانش فنی توسعه‌دهندگان و دانش زمینه‌ای و تفسیری کاربران است [Amershi et al., 2014]. این شکاف، اگر به‌درستی مدیریت نشود، می‌تواند به تحمیل چارچوب‌های ازپیش‌تعیین‌شده الگوریتمی بر فرآیند معنابخشی بینجامد. این وضعیت، از حیث ساختاری، قابل مقایسه با هشدار است که در سنت تفسیر دینی ذیل عنوان «تفسیر به رای» مطرح شده است؛ یعنی تحمیل معنایی که ریشه‌ای در متن و شواهد درون‌متنی ندارند. آیت‌الله سبحانی با تأکید بر همین خطر، هرگونه تحمیل معنا بدون شاهد متنی را نوعی اخلال در فرآیند معتبر تفسیر می‌داند [Sobhani Tabrizi, 2019]. این قیاس، نه برای همسان‌سازی دو حوزه، بلکه برای روشن‌ساختن یک الگوی مشترک خطا در فرآیندهای تفسیری انسانی و فناورانه طرح می‌شود.

بر این اساس، برای ارتقای کارایی و اعتمادپذیری سامانه‌های هوش مصنوعی در هرمنوتیک دیجیتال، کاربر انسانی نباید صرفاً به عنوان منبع داده یا ناظر نهایی تلقی شود، بلکه باید به مثابه فاعل تفسیری و شریک فعال در شکل‌دهی معنا حضور داشته باشد. این رویکرد با آنچه در فلسفه اسلامی ذیل «تفاعل عقل و وحی» مطرح می‌شود همخوانی ساختاری دارد؛ بدین معنا که برخی سطوح معنا از دسترس محاسبه صرف یا استدلال ابزاری فراتر می‌روند و نیازمند افقی هستند که تنها از طریق هدایت انسانی و تفسیر عقلانی-وجودی گشوده می‌شود [Motahari, 2015]. در هرمنوتیک دیجیتال نیز، بدون این مداخله انسانی، تفسیر به سطح بازتولید الگوهای زبانی تقلیل می‌یابد.

در نهایت، همان‌گونه که گزارش کمیسیون اروپا تصریح می‌کند، سامانه‌های هوش مصنوعی نباید خود را جایگزین فهم انسانی معرفی کنند، بلکه باید حدود و نقش ابزاری خود را به‌روشنی حفظ نمایند [High-Level Expert Group on Artificial Intelligence, 2019]. این اصل را می‌توان در قالب یک تز فلسفی چنین صورت‌بندی کرد: ماشین می‌تواند میانجی معنا باشد، اما هرگز خاستگاه آن نیست. معنا تنها در افق عاملیت انسانی، با برخورداری از عقل، تاریخ‌مندی و مسئولیت اخلاقی تحقق می‌یابد و هر طراحی فناورانه‌ای که این افق را نادیده بگیرد، در نهایت به تضعیف فهم و فروکاست معنا خواهد انجامید.

راهبردهای فلسفی-هرمنوتیک؛ عاملیت ترکیبی در افق تفسیر انسانی

این سطح از راهبردها به بررسی نحوه تعامل فهم انسانی با سامانه‌های زبانی می‌پردازد و می‌کوشد چارچوب مفهومی «عاملیت ترکیبی» را تبیین نماید؛ چارچوبی که در آن، هوش مصنوعی نه به مثابه فاعل مستقل تفسیر، بلکه به عنوان ابزار و میانجی در خدمت عاملیت هرمنوتیک انسان عمل می‌کند. مقصود از عاملیت ترکیبی، هم‌افزایی میان توان محاسباتی ماشین و ظرفیت تفسیری، تاریخی و ارزشی انسان است، به گونه‌ای که مسئولیت نهایی معنا همچنان بر عهده مفسر انسانی باقی بماند.

رویکردهای ترکیبی در پردازش زبان طبیعی؛ پیوند یادگیری آماری و قواعد معنایی

رویکردهای ترکیبی در با هدف تلفیق قدرت یادگیری آماری با ساختارهای منطقی و قواعد معنایی زبان توسعه یافته‌اند. این رویکردها واکنشی انتقادی به محدودیت‌های مدل‌های صرفاً آماری‌اند؛ مدل‌هایی که، علی‌رغم موفقیت در پیش‌بینی الگوهای زبانی، فاقد درک زمینه‌ای، تاریخی و مفهومی معنا هستند. در چنین مدل‌هایی، معنا عمدتاً به همبستگی‌های آماری فروکاسته می‌شود و توان تفسیر مفاهیم چندلایه و متون هنجاری به شدت محدود می‌گردد.

از همین رو، رویکردهای ترکیبی می‌کوشند با افزودن قواعد معنایی، دانش پیشین و ساختارهای منطقی به مدل‌های آماری، این کاستی را جبران نمایند. نتیجه چنین تلفیقی، شکل‌گیری سامانه‌هایی است که نه تنها از نظر دقت محاسباتی کارآمدترند، بلکه از حیث تفسیرپذیری و توضیح‌دادن تصمیم‌ها نیز به فهم انسانی نزدیک‌تر می‌شوند. به تعبیر مارکوس و دیویس، اتکای صرف به یادگیری آماری، بدون بهره‌گیری از دانش نمادین و معنایی، به سامانه‌هایی می‌انجامد که در مواجهه با معناهای عمیق، شکننده و ناتوان‌اند [Marcus & Davis, 2019].

در افق هرمنوتیک دیجیتال، اهمیت این رویکرد دوچندان می‌شود؛ زیرا تفسیر متون مقدس و فلسفی، نه یک مساله صرفاً آماری، بلکه فرآیندی معنایی، تاریخی و هنجاری است. رویکردهای ترکیبی می‌توانند با حفظ شفافیت و امکان توضیح، ابزارهایی فراهم سازند که مفسر انسانی از آنها برای گسترش افق فهم خود بهره‌گیرد، بی‌آنکه جایگاه تفسیری او به ماشین واگذار شود. به بیان دیگر، این مدل‌ها زمانی از منظر فلسفی قابل دفاع‌اند که نقش آنها در حد «پشتیبانی از تفسیر انسانی» باقی بماند و نه جایگزینی آن؛ زیرا معنا، حتی در بستر فناوری، همچنان در افق عاملیت انسانی تحقق می‌یابد.

مارکوس (استاد روان‌شناسی و علوم عصبی) و دیویس (استاد علوم کامپیوتر) با تأکید بر ضرورت عبور از داده‌محوری صرف، استدلال می‌کنند که هوش مصنوعی قابل اعتماد تنها زمانی امکان‌پذیر است که توانایی درک قواعد معنایی و ساختارهای شناختی در طراحی مدل‌ها لحاظ شود. به نظر آنان، اگرچه سامانه‌های یادگیرنده معاصر در برخی حوزه‌ها به عملکردی «فرا انسانی» دست یافته‌اند، اما همچنان در عرصه‌های بنیادینی چون فهم زبان طبیعی، درک جهان، انعطاف‌پذیری در موقعیت‌های نو، یادگیری سریع بدون داده‌های انبوه و استدلال در شرایط ناقص یا متناقض، به طور معناداری از انسان عقب‌ترند. از این منظر، اتکای کامل به الگوی «لوح سفید» و یادگیری صرفاً داده‌محور، به خطایی بنیادین می‌انجامد؛ زیرا نادیده‌گرفتن منطق، ساختارهای شناختی و تجربه انباشته انسانی، امکان دستیابی به فهم عمیق زبان و جهان را محدود می‌سازد [Marcus & Davis, 2019].

این تحلیل، در افق هرمنوتیک دیجیتال، دلالتی فراتر از یک نقد فنی دارد. تفسیر متون مقدس، به دلیل لایه‌مندی معنایی، تاریخ‌مندی و پیوند با افق‌های اعتقادی و وجودی، قابل فروکاست به همبستگی‌های آماری

نیست. اگر فرآیند تفسیر صرفاً به الگوهای داده‌ای واگذار شود، معنا به سطح بازنمایی‌های سطحی تقلیل می‌یابد. در مقابل، بهره‌گیری از رویکردهای ترکیبی که در آنها مدل‌های آماری در خدمت چارچوب‌های مفهومی و داوری تفسیری انسان قرار می‌گیرند، می‌تواند امکان تفسیری زمینه‌مندتر، شفاف‌تر و قابل اعتمادتر را فراهم سازد. بدین ترتیب، اهمیت این رویکردها نه در نزدیک‌کردن ماشین به فهم انسانی، بلکه در حفظ و تقویت عاملیت هرمنوتیک انسان در مواجهه با فناوری‌های زبانی هوشمند نهفته است.

در نهایت به‌اختصار می‌توان گفت، راهبردهای فناورانه بررسی‌شده در این مقاله نشان می‌دهند که توسعه هوش مصنوعی در حوزه تفسیر، تنها زمانی از منظر هرمنوتیک و اخلاقی قابل دفاع است که در خدمت تقویت عاملیت تفسیری انسان قرار گیرد. رویکردهایی همچون هوش مصنوعی توضیح‌پذیر، سامانه‌های تعاملی انسان در چرخه و مدل‌های ترکیبی در پردازش زبان طبیعی، نه به منزله جایگزین فهم انسانی، بلکه به عنوان ابزارهایی برای شفاف‌سازی، بازبینی و هدایت فرآیندهای تفسیری معنا می‌یابند. تحلیل این راهبردها روشن ساخت که اتکای صرف به مدل‌های داده‌محور، بدون مداخله عقل تفسیری و زمینه‌مند انسان، به فروکاست معنا و تضعیف مسئولیت هرمنوتیک می‌انجامد. از این رو، ارزش فناورانه این رویکردها نه در افزایش خودمختاری ماشین، بلکه در امکان‌پذیر ساختن تعامل مسئولانه، قابل توضیح و انسان‌محور با سامانه‌های زبانی هوشمند نهفته است؛ تعاملی که در آن، فناوری نقش میانجی دارد و انسان همچنان فاعل نهایی معنا باقی می‌ماند.

راهبردهای اخلاقی؛ طراحی چارچوب‌های هنجاری و پاسخ‌گویی

در بستر تحول‌آفرین هوش مصنوعی، به‌ویژه در حوزه مدل‌های زبانی، طراحی راهبردهای اخلاقی نه یک ملاحظه جانبی، بلکه شرط بنیادین امکان تفسیر مسئولانه است. از منظر هرمنوتیک، اخلاق صرفاً به معنای محدودسازی فناوری نیست، بلکه چارچوبی هنجاری برای صیانت از عاملیت انسانی، عدالت معرفتی و پاسخ‌گویی در فرآیند معنابخشی به شمار می‌آید. چنین چارچوبی، از یک سو، مستلزم شناسایی و کاهش سوگیری‌های نهفته در مدل‌هاست؛ زیرا بی‌توجهی به این مساله می‌تواند به بازتولید تبعیض‌های اجتماعی و تحمیل افق‌های تفسیری محدود منجر شود. از سوی دیگر، رعایت کثرت‌گرایی تفسیری اهمیت اساسی دارد؛ چرا که فهم متن، به‌ویژه در حوزه‌های دینی و فلسفی، امری تک‌ساحتی نیست و هر چارچوب اخلاقی معتبر باید امکان هم‌زیستی و گفت‌وگوی افق‌های معنایی متنوع را فراهم آورد. افزون بر این، در مواجهه با خطاهای تفسیری، طراحی نظامی شفاف و قابل پیگیری برای پاسخ‌گویی ضروری است؛ نظامی که نه تنها امکان اصلاح و جبران پیامدهای خطا را فراهم می‌سازد، بلکه اعتماد عمومی به کاربردهای هوش مصنوعی را نیز تقویت می‌کند. در مجموع، راهبردهای اخلاقی زمانی موجه‌اند که توسعه مدل‌های زبانی را در جهت ارزش‌های انسانی، مسئولیت اجتماعی و حفظ نقش محوری مفسر انسانی هدایت نمایند.

شناسایی سوگیری در مدل‌های زبانی

یکی از مهم‌ترین چالش‌های اخلاقی در توسعه مدل‌های زبانی، مساله سوگیری است؛ سوگیری‌هایی که اغلب ریشه در داده‌های آموزشی، ساختارهای اجتماعی مسلط و پیش‌فرض‌های فرهنگی دارند و می‌توانند به بازتولید کلیشه‌های ناعادلانه و تفسیرهای جهت‌دار بینجامند. از آنجا که مدل‌های زبانی قادرند الگوهای نادرست را همانند الگوهای معتبر بیاموزند، شناسایی سوگیری نه صرفاً یک اقدام فنی، بلکه پیش‌شرطی اخلاقی برای حفظ عدالت معرفتی است. تحلیل نحوه تولید پاسخ‌ها، ردیابی الگوهای زبانی تکرارشونده، حسابرسی داده‌ها و به‌کارگیری آزمون‌های حساسیت، از جمله ابزارهایی هستند که امکان آشکارسازی این سوگیری‌های نهفته را

فراهم می‌سازند. چنین شناسایی‌ای زمینه‌ساز مداخلات اصلاحی و تولید خروجی‌هایی منصفانه‌تر، شفاف‌تر و قابل اعتمادتر خواهد بود [Enayati et al., 2025].

مدل‌های زبانی پیشرفته، همچون دیگر سامانه‌های هوش مصنوعی، در معرض یادگیری سوگیری‌های نژادی، جنسیتی و فرهنگی از داده‌های آموزشی خود قرار دارند. از منظر تحلیلی، می‌توان این مساله را با منطق صوری مقایسه کرد: همان‌گونه که در استدلال منطقی، اگر مقدمه‌ها مخدوش باشند، نتیجه نیز الزاماً دچار خطا خواهد شد، در سامانه‌های داده‌محور نیز داده‌های آلوده یا نامتوازن به تولید خروجی‌های تبعیض‌آمیز می‌انجامد. از این رو، شناسایی و کاهش سوگیری‌ها نه صرفاً یک اقدام فنی، بلکه ضرورتی اخلاقی برای حفظ عدالت معرفتی و اعتبار تفسیری سامانه‌های زبانی است [Floridi & Cows, 2019]. با این حال، باید توجه داشت که مساله سوگیری به درستی یا نادرستی صوری داده‌ها محدود نمی‌شود، بلکه ریشه در نابرابری‌های ساختاری و زمینه‌های تاریخی دارد که در داده‌ها بازتاب یافته و از طریق الگوریتم‌ها بازتولید می‌شوند.

بر این اساس، شناسایی سوگیری در مدل‌های زبانی بزرگ مستلزم تحلیلی چندلایه است؛ تحلیلی که هم داده‌های آموزشی، هم معماری مدل و هم الگوهای خروجی تولیدشده را دربر گیرد. آزمون‌های سیستماتیک با داده‌های هدفمند، تحلیل تفاوت پاسخ‌ها نسبت به گروه‌های اجتماعی گوناگون و ارزیابی عملکرد مدل در زبان‌ها و زمینه‌های کم‌منبع، می‌تواند الگوهای تبعیض‌آمیز پنهان را آشکار سازد. چنین رویکردی، افزون بر شناسایی منشا سوگیری، مبنایی برای طراحی مداخلات اصلاحی فراهم می‌آورد و امکان هدایت سامانه‌های زبانی به سوی تفسیری منصفانه‌تر، شفاف‌تر و مسئولانه‌تر را تقویت می‌کند [Noorani & Farazkish, 2025]. در افق هرمنوتیک، این مداخلات تنها زمانی معنادار خواهند بود که در خدمت حفظ عاملیت تفسیری انسان و جلوگیری از تحمیل افق‌های معنایی ناعادلانه قرار گیرند.

رعایت کثرت‌گرایی تفسیری

زبان و تفسیر به طور ذاتی واجد لایه‌مندی و تکثر معنایی‌اند؛ بدین معنا که فهم متون و پدیده‌ها الزاماً به یک قرائت یگانه فروکاستنی نیست. از منظر هرمنوتیک، معنا در تعامل میان متن، زمینه و افق مفسر شکل می‌گیرد و همین امر امکان ظهور قرائت‌های متنوع را فراهم می‌سازد. بر این اساس، هر چارچوب اخلاقی در حوزه هوش مصنوعی به‌ویژه در کاربردهای زبانی، باید پذیرای کثرت‌گرایی تفسیری باشد تا از تحمیل یک افق معنایی مسلط و حذف دیدگاه‌های فرهنگی، زبانی و معرفتی دیگر جلوگیری شود. رعایت این اصل، افزون بر افزایش مشروعیت اجتماعی فناوری، امکان هم‌افزایی افق‌های معنا و ارتقای کیفیت زیست‌جهان انسانی را فراهم می‌آورد [Morley et al., 2019].

با این حال، کثرت‌گرایی تفسیری به معنای پذیرش بی‌قید و شرط همه قرائت‌ها نیست. در حوزه‌های حساس و هنجاری مانند عدالت، سلامت یا تصمیم‌های حیاتی، ضرورت انسجام، پاسخ‌گویی و قابلیت الزام عملی ایجاب می‌کند که تکثر تفسیری در چارچوب معیارهای عقلانی، اخلاقی و نهادی مهار شود. در غیر این صورت، خطر نسبی‌گرایی مفرط و تضعیف مسئولیت اخلاقی پدید می‌آید. از این منظر، کثرت‌گرایی باید به مثابه اصلی «مشروط» فهم شود؛ اصلی که امکان تنوع معنا را می‌پذیرد، اما آن را تحت نظارت داوری انسانی و هنجارهای مشترک سامان می‌دهد.

این صورت‌بندی با سنت هرمنوتیک تفسیر متون مقدس نیز همخوان است. علامه طباطبایی با تأکید بر لایه‌مندی طولی معانی قرآن، نشان می‌دهد که تکثر معنا نه به معنای استعمال لفظ در معانی متعارض، بلکه ناظر به تفاوت افق‌ها و مراتب فهم است؛ به گونه‌ای که هر معنا در سطح خاصی از ادراک انسانی تحقق می‌یابد

[Tabatabaei, 1995]. این تلقی می‌تواند به عنوان پشتوانه‌ای فلسفی برای اخلاق هوش مصنوعی به کار رود: تکثر معنا پذیرفته می‌شود، اما معنا همواره در افق عاملیت انسانی، عقلانی و مسئولانه قوام می‌یابد، نه در خودمختاری الگوریتمی.

رویکردهای ترکیبی که ابزارهای فنی، سازوکارهای حسابرسی و فرآیندهای آموزش و مشارکت اجتماعی را به صورت همزمان به کار می‌گیرند، از حیث تحلیلی می‌توانند به مثابه یک ساختار شرطی هم‌بسته فهم شوند؛ بدین معنا که اعتبار اخلاقی و تفسیری سامانه تنها در صورتی تحقق می‌یابد که مجموعه‌ای از شروط به طور توأمان برقرار باشند. چنین ترکیبی، از یک سو، به شفافیت و پاسخ‌گویی تصمیمات الگوریتمی یاری می‌رساند و از سوی دیگر، امکان بازخوانی انتقادی و تفسیر متنوع عملکرد سیستم را برای کاربران انسانی فراهم می‌سازد [Morley et al., 2019]. از این منظر، رویکردهای ترکیبی می‌توانند پیوندی عملی میان اصل توضیح‌پذیری و ضرورت کثرت‌گرایی تفسیری برقرار کنند، بی‌آنکه یکی فدای دیگری شود.

افزون بر این، بررسی دستورالعمل‌های اخلاقی هوش مصنوعی نشان می‌دهد که اصولی همچون شفافیت، عدالت، مسئولیت و صیانت از حریم خصوصی، اگرچه در اغلب اسناد به صورت مشترک تکرار می‌شوند، اما معانی و شیوه‌های تحقق آنها یکسان نیست. این تفاوت‌ها نشان می‌دهد که اخلاق هوش مصنوعی نه مجموعه‌ای از قواعد صلب و یکنواخت، بلکه شبکه‌ای از ارزش‌های هنجاری است که تفسیر و اجرای آنها وابسته به زمینه‌های فرهنگی، نهادی و نوع ذی‌نفعان است [Jobin et al., 2019]. بر این اساس، کثرت‌گرایی تفسیری در اخلاق هوش مصنوعی نه نشانه ضعف هنجاری، بلکه شرط امکان پاسخ‌گویی معنادار و حفظ عاملیت انسانی در مواجهه با فناوری‌های هوشمند به شمار می‌آید؛ مشروط بر آنکه این تکثر در چارچوب معیارهای عقلانی و اخلاقی مشترک سامان یابد.

در نهایت به‌اختصار می‌توان گفت، راهبردهای اخلاقی مطرح‌شده در این مقاله نشان می‌دهند که مواجهه مسئولانه با مدل‌های زبانی هوش مصنوعی مستلزم طراحی چارچوب‌های هنجاری‌ای است که همزمان شفافیت، عدالت معرفتی، پاسخ‌گویی و صیانت از عاملیت انسانی را تضمین کنند. تحلیل مساله سوگیری، کثرت‌گرایی تفسیری و تنوع قرائت‌های اخلاقی آشکار ساخت که اخلاق هوش مصنوعی نه مجموعه‌ای از قواعد صلب و یکنواخت، بلکه ساختاری زمینه‌مند و تفسیرپذیر است که تنها در پیوند با داوری انسانی معنا می‌یابد. از این رو، راهبردهای اخلاقی زمانی موجه و کارآمد خواهند بود که از یک سو، امکان شناسایی و اصلاح خطاها و سوگیری‌های الگوریتمی را فراهم سازند و از سوی دیگر، از فروکاست معنا به الگوهای داده‌محور جلوگیری کرده و مسئولیت نهایی تفسیر و تصمیم را بر عهده انسان نگه دارند. در افق هرمنوتیک دیجیتال، اخلاق نه محدودکننده فهم، بلکه شرط امکان فهم مسئولانه و عادلانه در تعامل انسان و فناوری است.

نتیجه‌گیری

این پژوهش نشان داد که ورود هوش مصنوعی، به‌ویژه مدل‌های زبانی پیشرفته، به عرصه تفسیر و تولید معنا، سه ساحت بنیادین معرفت‌شناختی، روش‌شناختی و وجودی-اخلاقی را به طور همزمان با چالش مواجه می‌سازد. این چالش‌ها صرفاً ماهیتی فنی یا مهندسی ندارند، بلکه ریشه در پرسش‌های عمیق فلسفی درباره ماهیت فهم، معنا و نسبت انسان با متن دارند. از این رو، مواجهه با آنها مستلزم رویکردی فلسفی-هرمنوتیک است که بتواند از سطح کارکردی تولید متن فراتر رفته و به بنیان‌های معناشناختی و هنجاری فهم بپردازد.

در سطح معرفت‌شناختی، تحلیل‌ها نشان داد که خروجی‌های هوش مصنوعی، هرچند می‌توانند صورت و ساختار ظاهری تفسیر انسانی را بازتولید کنند، فاقد افق‌های تاریخی، پیش‌فهم‌ها و زیست‌جهان انسانی‌اند. بر این اساس، آنچه مدل‌های زبانی تولید می‌کنند نه «فهم اصیل»، بلکه نوعی «فهم‌نما» است که بدون اتکا به تجربه زیسته، قصد معنابخش و مسئولیت‌تفسیری شکل می‌گیرد. تمایز میان فهم انسانی و فهم‌نمای ماشینی، نقشی کلیدی در جلوگیری از خلط معنا، حقیقت و اعتبار تفسیری ایفا می‌کند و می‌تواند معیاری راهنما برای ارزیابی انتقادی متون تولیدشده توسط ماشین باشد.

از منظر روش‌شناختی، غیرشفافیت سازوکارهای یادگیری ماشین و ماهیت جعبه سیاه بسیاری از مدل‌های زبانی، بحران اعتماد و اعتبار تفاسیر ماشینی را تشدید می‌کند. وابستگی شدید به داده‌های آموزشی، محدودیت در تکرارپذیری و فقدان امکان توضیح معنادار فرآیند تصمیم‌گیری، نشان می‌دهد که مدل‌های صرفاً آماری نمی‌توانند به‌تنهایی معیارهای کلاسیک هرمنوتیک را برآورده سازند. از این رو، ضرورت توسعه راهبردهای فنی-فناورانه‌ای همچون هوش مصنوعی توضیح‌پذیر، رویکردهای ترکیبی آماری-معنایی در پردازش زبان طبیعی و سامانه‌های تعاملی انسان در چرخه برجسته می‌شود؛ راهبردهایی که شفافیت، تفسیرپذیری و نقش فعال انسان را در چرخه فهم تقویت می‌کنند.

در سطح وجودی و اخلاقی، یافته‌های پژوهش نشان داد که تضعیف جایگاه مفسر انسانی و جایگزینی او با سامانه‌های خودکار، خطری جدی برای معنابخشی انسانی و عدالت معرفتی به شمار می‌آید. فروکاستن معنا به خروجی‌های الگوریتمی، همراه با بازتولید سوگیری‌های ساختاری و تاریخی نهفته در داده‌ها، می‌تواند به تحریف فهم در حوزه‌های حساس دینی و فرهنگی بینجامد. از این رو، طراحی و التزام به راهبردهای اخلاقی همچون شناسایی و کاهش سوگیری، رعایت کثرت‌گرایی تفسیری و ایجاد نظام‌های روشن پاسخ‌گویی، شرطی اساسی برای اعتمادپذیری و مشروعیت اجتماعی هوش مصنوعی در حوزه تفسیر است.

بررسی‌ها همچنین نشان داد که راه برون‌رفت از این وضعیت، نه در طرد فناوری، بلکه در بازتعریف «عاملیت هرمنوتیک» در عصر دیجیتال نهفته است. در این چارچوب، انسان باید نه به عنوان مصرف‌کننده منفعل خروجی‌های ماشینی، بلکه به مثابه تنظیم‌کننده افق‌های معنا و شریک تفسیری فعال حضور یابد. این عاملیت ترکیبی، امکان بهره‌گیری از توان پردازشی ماشین را در کنار حفظ مرجعیت تفسیری، عقلانی و اخلاقی انسان فراهم می‌سازد.

بر این اساس، راهبردهای پیشنهادی پژوهش در سه سطح به‌هم‌پیوسته صورت‌بندی می‌شوند: نخست، در سطح فلسفی-هرمنوتیک، بازاندیشی در مبانی فهم، معنا و تفسیر در بستر فناوری‌های زبانی؛ دوم، در سطح فنی-فناورانه، توسعه سامانه‌های توضیح‌پذیر، تعاملی و ترکیبی که شفافیت و کنترل انسانی را تقویت می‌کنند؛ و سوم، در سطح اخلاقی-هنجاری، طراحی چارچوب‌هایی برای پاسخ‌گویی، عدالت معرفتی و پذیرش کثرت‌گرایی تفسیری، همراه با ارتقای سواد دیجیتال و هرمنوتیک پژوهشگران علوم انسانی.

در نهایت، نتیجه اصلی پژوهش آن است که مواجهه با هوش مصنوعی در حوزه تفسیر نباید به سطح کارآمدی فنی یا نوآوری فناورانه فروکاسته شود، بلکه باید به فرصتی برای بازاندیشی نسبت انسان، متن و فناوری بدل گردد. تحقق این امر مستلزم تلفیق نقد فلسفی، طراحی فناورانه توضیح‌پذیر و چارچوب‌های اخلاقی پاسخ‌گو است. چنین رویکردی می‌تواند افق «هرمنوتیک دیجیتال» را بگشاید؛ افقی که در آن، ماشین ابزار فهم باقی می‌ماند و انسان، همچنان محور معنا، مسئولیت و داوری تفسیری است.

تشکر و قدردانی: موردی برای گزارش وجود ندارد.

تاییدیه اخلاقی: موردی برای گزارش وجود ندارد.

تعارض منافع: موردی برای گزارش وجود ندارد.

سهم نویسندگان: فاطمه عبدالله‌پور سنگچی (نویسنده اول)، ایده‌پردازی پژوهش/گردآوری داده‌ها/تحلیل مطالب (۴۰٪)؛ حسین رهنمائی (نویسنده دوم)، مشارکت در تدوین چارچوب نظری/نظارت علمی بر پژوهش (۲۰٪)؛ علی عسگری یزدی (نویسنده سوم)، مشارکت در تحلیل داده‌ها/ارزیابی علمی محتوا (۲۰٪)؛ مهران رضایی (نویسنده چهارم)، مشارکت در توسعه مباحث نظری/ارزیابی انتقادی محتوا (۴۰٪)

منابع مالی: مقاله مستخرج از رساله دکتری نویسنده اول است.

منابع

- Aalipour N, Zanganeh M (2022). An investigation of theoretical foundations related to digital transformation and innovation in knowledge management systems (KMS) to understand its role in business management and sustainable development. Proceedings of the Third International Conference on Management, Accounting, Economics and Banking in the Third Millennium. Melbourne: Arvin Alborz. [Persian]
- Alvand F, Afsharnejad A, Nobari A, Majidi M (2023). A model of human resources development strategies in the digital age. Journal of Innovation Management in Defensive Organizations. 6(4):225-248. [Persian]
- Amershi S, Cakmak M, Knox WB, Kulesza T (2014). Power to the people: The role of humans in interactive machine learning. AI Magazine. 35(4):105-120.
- Brown T, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al (2020). Language models are few-shot learners. Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: NIPS'20. p. 1877-1901.
- Capurro R (2010). Digital hermeneutics: An outline. AI & Society. 25(1):35-42.
- Daraei M (2024). Explainable artificial intelligence: A step towards transparency and trust in artificial intelligence. In Proceedings of the 6th International Conference on Computer Engineering, Electrical Engineering and Technology, Hamedan, Iran. [Persian]
- Enayati Z, Ramezani SN, Raisi M, Heydari Z (2025). The role of artificial intelligence in identifying and reducing cognitive biases in the learning process. Proceedings of the First International Conference on Technological Innovations in Education. Ramshir: Binesh Higher Education Institute. [Persian]
- Floridi L (2014). The fourth revolution: How the infosphere is reshaping human reality. Oxford: Oxford University Press.
- Floridi L, Cowls J (2019). A unified framework of five principles for AI in society. Harvard Data Science Review. 1(1).
- Floyd C, Züllighoven H, Budde R, Keil-Slawik R, editors (1992). Software development and reality construction. Berlin: Springer-Verlag.
- Gadamer HG (2004). Truth and method. Weinsheimer J, Marshall DG, translators. New York: Continuum.
- High-Level Expert Group on Artificial Intelligence (2019). Ethics guidelines for trustworthy AI. Brussels: European Commission.
- Jafari Tabrizi MT (2012). Islamic mysticism. Tehran: Allameh Jafari Works Compilation and Publishing Institute. [Persian]
- Javadi Amoli A (2016). Tasnim commentary (Volume 46). Qom: Israa Publishing Center. [Persian]
- Jobin A, Ienca M, Vayena E (2019). The global landscape of ai ethics guidelines. Nature Machine Intelligence. 1(9):389-399.
- Lipton ZC (2017). The mythos of model interpretability. Communications of the ACM. 61(10):36-43.
- Marcus G, Davis E (2019). Rebooting AI: Building artificial intelligence we can trust. New York: Pantheon Books.
- Moor JH (2006). The dartmouth college artificial intelligence conference: The next fifty years. AI Magazine. 27(4):87-91.
- Morley J, Floridi L, Kinsey L, Elhalal A (2019). 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.
- Motahari M (2015). Notes of professor Motahari (Volume 1). Tehran: Sadra. [Persian]

- Motalebi Karbakandi H (2013). Analysis of artificial intelligence and its functions in comparison with human thinking [dissertation]. Qom: University of Qom, Faculty of Theology and Islamic Studies. [Persian]
- Noorani A, Farazkish R (2025). Ethical considerations and bias in language models. Proceedings of the First National Conference on Artificial Intelligence and Emerging Research. Tehran: Islamic Azad University South Tehran Branch. [Persian]
- Ojieabu SO (2024). From automation to augmentation: The potential of artificial intelligence in biblical hermeneutics. Crowther Journal of Arts and Humanities. 1(4).
- Pickering JB, Engen V, Walland P (2017). The interplay between human and machine agency. Proceedings of the 19th International Conference on Human-Computer Interaction. Vancouver: Springer. p. 47-59.
- Romele A, Severo M, Furia P (2020). Digital hermeneutics: From interpreting with machines to interpretational machines. AI & Society. 35(1):73-86.
- Sadr Al-Din Shirazi M (2008). AL-MAZAHIR AL-ILAHYEH (Volume 3). Tabibiyani H, translator. Tehran: Amir Kabir. [Persian]
- Sobhani Tabrizi J (2019). Hermeneutics. Qom: Tohid. [Persian]
- Tabatabaei MH (1995). TAFSIR AL-MIZAN (Volume 3). Qom: Islamic Publications Office. [Persian]
- Wang J (2021). Is artificial intelligence capable of understanding? An analysis based on philosophical hermeneutics. Cultures of Science. 4(3):135-146.