Ethical Reflections

Vol. 7, Issue 2, No. 26, Summer 2026, pp. 1-25

Online ISSN: 2717-1159 / Print ISSN: 2676-4810

https://jer.znu.ac.ir

Analyzing the Three Dimensions of the "Responsibility Gap" in Artificial Intelligence: An Ethical Approach

Reza Jafari,[1] Mohammad Javad Fallah [2]

Article info

Original Article

DOI:
10.30470/er.2026.2080704.1504

Submission History
Received: 12/12/2025
Revised: 18/02/2026
Accepted: 28/04/2026
Published: 23/07/2026

Authors' Contribution
Reza Jafari: Collecting data, Study concept and design, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript for important intellectual content.
Mohammad Javad Fallah: statistical analysis, reading and confirming the last version of manuscript.

Conflict of Interests
The authors declare no conflict of interest.

Funding/Support
This research has not received any financial support from funding organizations.

Extraction
This article has not been extracted from thesis.

AI Usage
Artificial intelligence tools were used in the preparation of this article to a limited extent for the purpose of identifying relevant research sources.

Abstract

The question of responsibility in the age of artificial intelligence has become one of the central issues in the philosophy of technology and applied ethics. The increasing autonomy and complexity of intelligent systems have transformed the relationship between humans and machines, turning the concept of the "responsibility gap" into a major challenge in this field. Adopting an analytical-conceptual approach, this research revisits the theoretical and logical foundations of this concept and demonstrates that the responsibility gap, contrary to common perception, is not a simple and unified phenomenon. Rather, it is the result of conceptual conflation and the formation of three distinct discontinuities: the technical-metaphysical gap, which relates to the issue of agency and control; the organizational-structural gap, which arises from the unfair distribution of responsibility among institutions and individuals; and the epistemic gap, which is associated with the ambiguity and lack of transparency in the processes of learning systems. The research findings indicate that moving beyond the search for a single solution and shifting toward multi-level and integrative approaches is a necessary condition for solving the challenge of responsibility in intelligent systems.

Keywords: Artificial Intelligence Ethics; Responsibility Gap; Moral Responsibility; Accountability; Moral Agency; Epistemic Gap.

1. **Corresponding author**: Ph.D. in Islamic Ethics, Maaref Islamic University, Qom, Iran., Qom, Iran. Rezajafari128@gmail.com

2. Associate Professor, Department of Islamic Ethics, University of Islamic Sciences, Qom, Iran. fallah@maaref.ac.ir

Introduction

The rapid integration of Artificial Intelligence (AI) into various facets of human life—ranging from transportation and healthcare to judicial systems and financial markets—has confronted human society with a complex array of unprecedented ethical challenges. A quintessential example of this complexity is the fatal accident involving a self-driving Uber car in 2018. While a specific algorithmic error led to the tragic death of a pedestrian, the causal chain linking software developers, the safety driver, the operating company, and regulatory bodies was so deeply entangled that attributing moral and legal responsibility to a single entity became practically impossible.

This event, alongside numerous similar incidents, has catalyzed widespread debate surrounding the concept of the "Responsibility Gap." This notion, emphasizing the difficulty or sheer impossibility of ascribing responsibility to either human agents or the machine itself, has dominated the discourse in the philosophy of technology for nearly two decades, arguably leading to a theoretical and practical impasse. In attempting to resolve this deadlock, scholars have generally diverged into two main streams: one group seeks metaphysical solutions to bridge this gap by exploring the potential for artificial moral agency, while another group adopts a pragmatic approach, focusing directly on designing structural frameworks for distributing responsibility among human actors.

The primary problem addressed in this research is not merely to choose between these existing streams, but to critically analyze the very concept of the responsibility gap. The study posits that the theoretical ambiguity currently plaguing the field stems from an imprecise formulation of the problem itself. This research hypothesizes that the responsibility gap is not a single phenomenon but a conceptual umbrella covering at least three distinct discontinuities. Furthermore, this study aims to explore the capacity of religious—specifically Islamic—jurisprudential principles in addressing this modern challenge. The key question is whether traditional concepts such as attribution (Isnad), liability (Daman), and causation (Tasbib) can resolve the issue when applied to the autonomous nature of AI, or whether these analytical tools also face a deadlock.

Responsibility Gap

The Technical-Metaphysical Gap: The first and most recognized dimension of the responsibility gap originates directly from the nature of the technology itself. Pioneered by thinkers like Andreas Matthias, this gap pertains to the fundamental issues of agency, control, and knowledge. The core argument is that the autonomy and unpredictability of machine learning algorithms erode the two traditional preconditions for moral responsibility: human control and human knowledge. In modern deep learning systems, the programmer defines the initial architecture, but the final logic is distributed across millions of synaptic weights, creating a system that learns and adapts autonomously through its interaction with

the environment. From an ontological perspective, AI lacks the consciousness or volition required for true moral agency. Consequently, an ethical paradox emerges: the human creator lacks the necessary control to be held fully responsible, while the machine lacks the moral agency to bear the blame. From an Islamic jurisprudential (Fiqh) perspective, this autonomous nature disrupts the material relationship of attribution (Isnad) between the designer and the outcome. The traditional rule of the "precedence of the direct actor over the cause" (Mubashir over Sabab) falters because the AI acts as a direct actor without will, yet its computational complexity makes it exceptionally difficult to trace the act back to the original cause (the designer).

The Organizational-Structural Gap: Unlike the first dimension, the second gap is rooted not in the technology itself, but in the human structures and organizational frameworks that develop and deploy these technologies. This gap is best explained by the "Problem of Many Hands" in organizational theory. In complex socio-technical systems, the sheer number of human agents—programmers, data scientists, managers, users, and regulators—creates a dense web of causality where responsibility is dispersed to the point of vanishing. Madeleine Elish articulates this phenomenon through the concept of "Moral Crumple Zones," wherein responsibility for a systemic failure is unfairly concentrated on the nearest human operator (e.g., the backup driver), thereby protecting the broader technological and organizational structure from liability. Furthermore, this structural complexity facilitates psychological mechanisms of "Moral Disengagement," as theorized by Albert Bandura. In such environments, individuals experience the displacement and diffusion of responsibility, feeling that their personal contribution to any potential harm is negligible. In the context of Islamic jurisprudence, this presents a significant challenge in determining the primary cause of damage when multiple causes intersect, leading to a computational deadlock in assigning proportionate liability.

The Epistemic Gap: The third dimension concerns the profound inability to know, trace, and prove responsibility. Even if metaphysical responsibility exists and organizational structures are defined, the "Black Box" nature of AI creates an insurmountable epistemic barrier. This opacity manifests in two forms: technical opacity, stemming from the inherent complexity of deep neural networks, and commercial opacity, resulting from corporate secrecy and intellectual property protections. This lack of transparency disrupts the practical mechanisms of accountability. In legal and ethical analysis, the judgment of "custom" (Urf) is only valid when based on discoverable reality. The opacity of neural networks prevents the scientific discovery of the decision-making trajectory, rendering the attribution of responsibility empirically unverifiable. Furthermore, this lack of transparency creates a state of "Gharar" (uncertainty and risk resulting from ignorance). In Islamic jurisprudence, a deceived person acting in ignorance is not liable; responsibility shifts to the deceiver. However, in deep learning, even the designer is largely ignorant of the specific decision-making process. This creates a state of pervasive,

systemic ignorance, blurring the critical legal line between "fault" (Taqsīr) and "excusable ignorance" (Quṣūr), effectively allowing all actors to evade accountability.

Conclusion

The comprehensive analysis provided in this study reveals that overcoming the theoretical and practical deadlock of the responsibility gap requires a fundamental paradigm shift. We must move away from the search for a single, universal solution and instead adopt a multi-level, integrative approach tailored to the specific nature of each gap.

To address the Technical-Metaphysical Gap, the study suggests shifting from a paradigm of "responsibility based on intent" to one of "responsibility based on material attribution." Utilizing the principles of strict liability (Daman), accountability can be imposed based on the creation of risk, even in the absence of direct human control or malicious intent.

For the Organizational-Structural Gap, the solution lies in transitioning from the pursuit of a single culprit to a model of "Shared Risk Distribution." Drawing on jurisprudential capacities for distributing liability among multiple intersecting causes, all stakeholders—from designers to policymakers—should share in the compensation of damages based on their respective levels of benefit and influence.

Finally, regarding the Epistemic Gap, transparency must be elevated from a mere ethical recommendation to a mandatory precondition for system deployment. The "explainability" of AI should be regarded as the strict condition for removing "Gharar" (uncertainty). If a system's decision-making process remains uninterpretable, its deployment constitutes an action based on ignorance, and the opacity itself must be considered an independent fault (Taqsīr) that incurs liability. Ultimately, resolving the responsibility gap demands a synthesized framework of ethical design, legal reform, and social policy.

References

Alidoust, Abu al-Qasim; Bay, Husaynali. (1391 SH). «The Criterion of Liability». *Fiqh-e Ahl-e Bayt*, 18(70): 48-105. [In Persian].

Bandura, A. (1999). "Moral Disengagement in the Perpetration of Inhumanities". *Personality and Social Psychology Review*, 3(3): 193-209.

Bryson, J. J. (2018). "Patiency Is Not a Virtue: The Design of Intelligent Systems and Systems of Ethics". *Ethics and Information Technology*, 20(1): 15–26.

Busuioc, M. (2021). "Accountable Artificial Intelligence: Holding Algorithms to Account". *Public Administration Review*, 81(5): 825–836.

Canas, J. J. (2022). "AI and Ethics When Human Beings Collaborate with AI Agents". *Frontiers in Psychology*, 13: 836650.

Coeckelbergh, M. (2020). "Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability". *Science and Engineering Ethics*, 26: 2051–2068.

Da Silva, M. (2024). "Responsibility Gaps". *Philosophy Compass*, e70002.

Danesh, Javad. (1397 SH). *Mas'uliyyat-e Akhlaqi*. Qom: Pazhuheshgah-e 'Ulum va Farhang-e Islami. [In Persian].

Dastani, M.; Yazdanpanah, V. (2023). "Responsibility of AI Systems". *AI & Society*, 38(3): 859–872.

Elish, M. C. (2019). "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction". *Engaging Science, Technology, and Society*, 5: 40–60.

Fakhr al-Muhaqqiqin, Muhammad Ibn Hasan. (1387 SH). *Idah al-Fawa'id fi Sharh Ishkalat al-Qawa'id*. Qom: Isma'iliyan. [In Arabic].

Floridi, L.; Sanders, J. W. (2004). "On the Morality of Artificial Agents". *Minds and Machines*, 14(3): 349–379.

Fritz, A.; Brandt, W.; Gimpel, H.; Bayer, S. (2020). "Moral Agency Without Responsibility? Analysis of Three Ethical Models of Human-Computer Interaction in Times of Artificial Intelligence (AI)". *Ethics and Information Technology*, 22(1): 35–48.

Gorbanian, Saleh. (1399 SH). «Moral Agency in Artificial Intelligence (Robots)». Ethical Reflections, 1(1): 11-32. [In Persian].

Gunkel, D. J. (2017). "Mind the Gap: Responsible Robotics and the Problem of Responsibility". *Ethics and Information Technology*, 22(4): 307–320.

Husayni 'Amili, Muhammad Javad. (1419 AH). *Miftah al-Karamah fi Sharh Qawa'id al-'Allamah*. Qom: Jami'ah-ye Mudarrisin-e Hawzah-ye 'Ilmiyyah-ye Qom (Mu'assasah al-Nashr al-Islami). [In Arabic].

Johnson, D. G.; Verdicchio, M. (2018). "AI, Agency and Responsibility: The VW Fraud Case and Beyond". *AI & Society*, 33(4): 517–526.

Kaplan, A.; Haenlein, M. (2019). "Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence". *Business Horizons*, 62(1): 15–25.

Khoei, Sayyid Abu al-Qasim. (1418 AH). *Mawsu'at al-Imam al-Khoei*. Qom: Mu'assasah Ihya' Athar al-Imam al-Khoei. [In Arabic].

Khomkhunsorn, T. (2025). "Meaningful Human Control and Responsibility Gaps in AI: No Culpability Gap, but Accountability and Active Responsibility Gap". *Journal of Integrative and Innovative Humanities*, 5(1): 35–57.

Kiener, M. (2022). "Can We Bridge AI's Responsibility Gap at Will?". *Ethical Theory and Moral Practice*, 25: 575–593.

Konigs, P. (2022). "Artificial Intelligence and Responsibility Gaps: What Is the Problem?". *Ethics and Information Technology*, 24(3): 1–11.

Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology*. New York: Viking.

Lancaster, C. M. et-al. (2024). "'It's Everybody's Role to Speak Up... But Not Everyone Will': Understanding AI Professionals' Perceptions of Accountability for AI Bias Mitigation". *ACM Journal on Responsible Computing*, 1(1): 1–30.

LeCun, Y.; Bengio, Y.; Hinton, G. (2015). "Deep Learning". *Nature*, 521(7553): 436–444.

Matthias, A. (2004). "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata". *Ethics and Information Technology*, 6(3): 175–183.

McGraw, D. (2020). "Ethical Responsibility in the Design of Artificial Intelligence (AI) Systems". *The Journal of Philosophy, Science & Law*, 20: 1–18.

Mesbah Yazdi, Muhammad Taqi. (1393 SH). *Falsafeh-ye Akhlaq*. Qom: Mu'assasah-ye Amuzheshi va Pazhuheshi-ye Imam Khomeini. [In Persian].

Mirzamohammadi, Abdolraoof; Taqavi, Mostafa. (1404 SH). «The Ethics of Technology: A Critical Assessment of the Debate Between David Morrow and Joseph C. Pitt on the Value-Ladenness of Technology and Moral Responsibility». *Ethical Reflections*, 6(2): 93-114. [In Persian].

Najafi, Muhammad Hasan. (1367 SH). *Jawahir al-Kalam fi Sharh Shara'i' al-Islam*. Beirut: Dar Ihya' al-Turath al-'Arabi. [In Arabic].

Novelli, C.; Taddeo, M.; Floridi, L. (2023). "Accountability in Artificial Intelligence: What It Is and How It Works". *AI & Society*.

Oimann, A.-K.; Tollon, F. (2025). "Responsibility Gaps and Technology: Old Wine in New Bottles?". *Journal of Applied Philosophy*, 42(1): 1–22.

Russell, S. J.; Norvig, P. (2010). *Artificial Intelligence: A Modern Approach* (3rd Ed.). New Jersey: Prentice Hall.

Sadeghi, Muhammad Hadi; Mirzaei, Muhammad. (1398 SH). «The Nature of Causation Relationship and the Criteria for Verify It». *Islamic Law and Jurisprudence Studies*, 11(21): 167-194. [In Persian].

Santoni de Sio, F.; Mecacci, G. (2021). "Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them". *Philosophy & Technology*, 34: 1057–1084.

Sparrow, R. (2007). "Killer Robots". *Journal of Applied Philosophy*, 24(1): 62–79.

Stahl, B. C. (2023). "Embedding Responsibility in Intelligent Systems: From AI Ethics to Responsible AI Ecosystems". *Ethics and Information Technology*, 25(3): 1–17.

Tigard, D. W. (2021). "There Is No Techno-Responsibility Gap". *Philosophy & Technology*, 33(3): 589–607.

Veluwenkamp, H. (2025). "What Responsibility Gaps Are and What They Should Be". *Ethics and Information Technology*, 27(14).

Verbeek, P.-P. (2015). "Beyond Interaction: A Short Introduction to Mediation Theory". *Interactions*, 22(3): 26–31.

تأملات اخلاقی

دورۀ هفتم، شمارۀ دوم (پیاپی ۲۶)، تابستان ۱۴۰۵، صفحات ۱-۲۵

شاپا الکترونیکی: ۱۱۵۹-۲۷۱۷/ شاپا چاپی: ۴۸۱۰-۲۶۷۶

https://jer.znu.ac.ir

تحلیل ابعاد سه‌گانه «شکاف مسئولیت» در هوش مصنوعی با رویکرد اخلاقی

رضا جعفری،[۱] محمدجواد فلاح[۲]

اطلاعات مقاله

DOI:
10.30470/er.2026.2080704.1504

مقاله پژوهشی

تاریخ دریافت: ۱۴۰۴/۰۹/۲۱

تاریخ اصلاح: ۱۴۰۴/۰۱۱/۲۹

تاریخ پذیرش: ۱۴۰۵/۰۲/۰۸

تاریخ انتشار: ۱۴۰۵/۰۵/۰۱

حمایت مالی و تعارض منافع

مقاله حامی مالی و تعارض منافع ندارد.

استخراج

این مقاله متخذ از رساله نیست.

استفاده از هوش مصنوعی

در نگارش این مقاله از هوش مصنوعی به میزان محدود برای یافتن منابع مرتبط با پژوهش استفاده شده است.

چکیده

پرسش از مسئولیت در عصر هوش مصنوعی به یکی از مسائل اصلی در فلسفۀ فناوری و اخلاق کاربردی تبدیل شده است. افزایش خودمختاری و پیچیدگی سامانه‌های هوشمند، رابطۀ میان انسان و ماشین را تغییر داده و مفهوم «شکاف مسئولیت» را به چالش مهمی در این حوزه بدل کرده است. این پژوهش با رویکردی تحلیلی-مفهومی، به بازنگری در مبانی نظری و منطقی این مفهوم می‌پردازد و نشان می‌دهد که شکاف مسئولیت، برخلاف تصور رایج، پدیده‌ای بسیط و واحد نیست، بلکه حاصل خلط مفهومی و شکل‌گیری سه گسست متمایز است: شکاف فنی-متافیزیکی که به موضوع عاملیت و کنترل مربوط می‌شود، شکاف سازمانی-ساختاری که از توزیع ناعادلانۀ مسئولیت در میان نهادها و افراد ناشی می‌شود و شکاف معرفتی که با ابهام و نبود شفافیت در فرایندهای سامانه‌های یادگیرنده ارتباط دارد. یافته‌های پژوهش نشان می‌دهد که عبور از جست‌وجوی یک راه‌حل واحد و حرکت به سوی رویکردهای چندسطحی و ترکیبی، شرط لازم برای حل چالش مسئولیت در سامانه‌های هوشمند است.

واژه‌های کلیدی: اخلاق هوش مصنوعی، شکاف مسئولیت، مسئولیت اخلاقی، پاسخگویی، عاملیت اخلاقی، شکاف معرفتی.

۱. نویسنده مسؤول: دانش آموخته دکترای تخصصی اخلاق اسلامی، دانشگاه معارف اسلامی، قم، ایران. Rezajafari128@gmail.com

۲ . دانشیار گروه اخلاق اسلامی، دانشگاه معارف اسلامی، قم، ایران. fallah@maaref.ac.ir

مقدمه

با نفوذ روزافزون هوش مصنوعی در شئون مختلف زندگی، از حمل‌ونقل و بهداشت تا قضاوت و امور مالی، جامعه انسانی با مجموعه‌ای از چالش‌های اخلاقی بی‌سابقه روبرو شده است. حادثه مرگبار خودروی خودران اوبر در سال ۲۰۱۸ نمونه‌ای برجسته از این پیچیدگی است؛ جایی که یک خطای الگوریتمی به مرگ یک انسان منجر شد، اما زنجیرهٔ علیت و تقصیر میان توسعه‌دهندگان، راننده پشتیبان، شرکت بهره‌بردار و نهادهای نظارتی چنان درهم‌تنیده بود که اسناد مسئولیت به یک فرد واحد را تقریباً غیر ممکن ساخت.

این رویداد و موارد مشابه، باعث شکل‌گیری بحث‌های گسترده‌ای پیرامون «شکاف مسئولیت» شد. این انگاره که بر عدم امکان اسناد مسئولیت به هر یک از عاملان انسانی یا ماشینی تأکید دارد، برای نزدیک به دو دهه بر مباحث این حوزه سایه افکنده و آن را به یک بن‌بست نظری کشانده است. پژوهشگران در تلاش برای حل این بن‌بست، عمدتاً در دو جریان حرکت کرده‌اند: گروهی به دنبال راه‌حل‌های متافیزیکی برای پر کردن این خلأ بوده‌اند و گروهی دیگر، با رویکردی عمل‌گرایانه، مستقیماً به طراحی چارچوب‌هایی برای توزیع مسئولیت پرداخته‌اند.

مسئله اصلی این پژوهش، نه انتخاب میان این دو جریان، بلکه تحلیل خود انگارهٔ شکاف مسئولیت است. این پژوهش می‌کوشد مشخص سازد که در این چالش، ما دقیقاً با چه سؤالاتی مواجهیم و چرا این شکاف رخ داده است. آیا شکاف مسئولیت یک مسئله واحد و یکپارچه است یا آنکه چالش‌های مجزا و متفاوتی در آن نهفته است که این پدیده را تا این حد پیچیده ساخته است؟ ما معتقدیم که ابهام نظری موجود، ناشی از عدم تبیین دقیق خود صورت‌بندی مسئله است. ازاین‌رو، پژوهش حاضر در صدد است تا به این پرسش اصلی پاسخ دهد که انگاره شکاف مسئولیت دقیقاً چیست و مؤلفه‌ها و ریشه‌های شکل‌گیری آن کدام‌اند؟ افزون بر این، پژوهش حاضر به واکاوی ظرفیت قواعد درون دینی در مواجهه با این چالش نیز می‌پردازد. پرسش کلیدی در این بخش آن است که آیا مفاهیم دقیق سنتی ما همچون استناد، ضمان و تسبیب، در برخورد با ماهیت خودمختار هوش مصنوعی، به سادگی قادر به حل مسئله هستند یا آنکه دستگاه تحلیلی ما نیز در این مسیر با بن‌بست یا شکاف مسئولیت روبرو خواهد شد؟

پژوهش حاضر از نوع کیفی و مبتنی بر روش کتابخانه‌ای و تحلیل محتوا است و مراحل انجام پژوهش در سه بخش اصلی سامان‌دهی شده است. گام نخست، به تحلیل و دسته‌بندی پیشینه پژوهش اختصاص دارد تا نشان دهد چگونه تلقی واحد و یکپارچه از شکاف مسئولیت، این بحث را به بن‌بست نظری کشانده است. در گام دوم، یک گونه‌شناسی تحلیلی از انگاره شکاف مسئولیت ارائه می‌گردد. در نهایت، گام نهایی پیامدهای نظری و عملی این گونه‌شناسی را بررسی می‌کند و نشان می‌دهد که این تفکیک مفهومی چگونه به هموار ساختن مسیر برای یافتن راه‌حل‌های عملی کمک می‌کند.

اهمیت این پژوهش در ضرورت تحلیل دقیق این انگاره بنیادین است. تا زمانی که ابعاد و منشأهای این شکاف به صورت شفاف تحلیل نشوند، تلاش‌ها برای یافتن راه‌حل‌های عملی ممکن است ناکارآمد و بر پایه‌های مفهومی متزلزلی استوار باشند. این مقاله به دنبال آن است که با ایجاد درکی عمیق‌تر از چیستی مسئله، به انسجام‌بخشی این حوزه مطالعاتی کمک کرده و مبانی مستحکم‌تری برای پاسخ‌گویی به این چالش فراهم آورد.

۱. پیشینه پژوهش

مفهوم شکاف مسئولیت و پژوهش اختصاصی حول آن، در پژوهش‌های داخلی تاکنون به صورت جدی مطرح نشده است؛ گرچه در تحلیل مسئولیت در مباحث حقوقی و فقهی، پژوهش‌های متعددی انجام شده است؛ اما این مفهوم نزدیک به دو دهه در ادبیات اخلاق فناوری در غرب مطرح بوده و گفتمان گسترده‌ای را شکل داده است. پژوهش‌های مرتبط با این موضوع را می‌توان در سه دسته کلی دسته‌بندی کرد:

نخست، پژوهش‌های بنیادینی که به صورت‌بندی شکاف به مثابه یک بن‌بست متافیزیکی پرداخته‌اند. این جریان با کار پیشگامانه آندریاس ماتیاس در مقالهٔ «شکاف مسئولیت: اسناد مسئولیت به اعمال اُتوماتای یادگیرنده» (۲۰۰۴) آغاز شد. او استدلال کرد که خودمختاری و عدم پیش‌بینی‌پذیری ماشین‌های یادگیرنده، دو شرط سنتی مسئولیت یعنی کنترل و دانش را از بین برده و یک خلأ واقعی ایجاد می‌کند. این دیدگاه به سرعت توسط رابرت اسپارو در مقالهٔ «ربات‌های قاتل» (۲۰۰۷) به چالشی اخلاقی در سطح نظری انجامیده است. تفاوت این دسته از پژوهش‌ها با تحقیق حاضر در این است که آن‌ها شکاف مسئولیت را یک واقعیت متافیزیکی مسلم دانسته و پیامدهای آن را تحلیل می‌کنند، درحالی‌که پژوهش حاضر، خودِ این صورت‌بندی را زیر سؤال می‌برد. با این وجود، اهمیت این پژوهش‌ها برای مقاله حاضر در این است که این پژوهش‌ها نقطه شروع و بستر اصلی برای تحلیل یکی از مؤلفه‌های شکاف در این مقاله هستند و مورد استناد قرار می‌گیرند.

دوم، پژوهش‌هایی که بر ابعاد عملیاتی و انسانی شکاف متمرکز شده‌اند. این جریان، تحلیل را از یک بن‌بست متافیزیکی محض، به سمت چالش‌های انضمامی در ساختارهای انسانی سوق داده‌اند. برای نمونه، مقالهٔ «ناحیه‌های ضربه‌گیر اخلاقی: داستان‌های هشدارآمیز در تعامل انسان و ربات» (الیش، ۲۰۱۹) و همچنین تحلیل‌های «مشکل دست‌های بسیار» در آثاری چون «شکاف‌های مسئولیت» (دا سیلوا، ۲۰۲۴) و «هوش مصنوعی، اسناد مسئولیت، و توجیه رابطه‌ای توضیح‌پذیری» (کوکلبرگ، ۲۰۲۰)، نشان دادند که مشکل، گاهی فقدان مسئولیت نیست، بلکه توزیع ناعادلانه یا پراکندگی آن در ساختارهای پیچیده سازمانی است. از سوی دیگر، تحلیل‌های روان‌شناختی، مکانیسم‌های پنهانی را عواملی برای تشدید این شکاف دانسته‌اند؛ مانند «مهارت‌زدایی اخلاقی» و «حائل اخلاقی» است که توسط خوم‌خونسورن (۲۰۲۵) در مقاله «کنترل معنادار انسانی و شکاف مسئولیت در هوش مصنوعی...» به کار رفته‌اند. تفاوت این گروه از پژوهش‌ها با مقاله حاضر در سطح تمرکز است. آن‌ها نیز به‌درستی یکی از وجوه و مؤلفه‌های شکاف را بیان کرده‌اند، درحالی‌که جایگاه این پژوهش‌ها در پژوهش حاضر، بخشی از تحلیل جامع آن خواهد بود.

سوم، پژوهش‌هایی که به بازتعریف شکاف به مثابه یک چالش معرفتی پرداخته‌اند. این جریان سوم، کل صورت‌بندی مسئله را تغییر می‌دهند. متفکرانی چون پیتر کونیگز در مقاله «هوش مصنوعی و شکاف مسئولیت: مشکل چیست؟» (۲۰۲۲) و هرمان فِلُونِ‌کمپ در مقاله «شکاف‌های مسئولیت چه هستند و چه باید باشند» (۲۰۲۵) استدلال کردند که شکاف، اصولاً متافیزیکی نیست، بلکه معرفتی است. تفاوت این جریان با پژوهش حاضر در تقلیل‌گرایی آن است. این دیدگاه، کل پدیده پیچیده شکاف مسئولیت را به یک بُعد؛ یعنی چالش معرفتی فرومی‌کاهد و ابعاد دیگر را نادیده می‌گیرد. اهمیت این پژوهش‌ها نیز در شناسایی دقیق

بخشی از تحلیل جامع پژوهش حاضر است.

در این میان، پژوهش دسیو و مکاچی با عنوان «چهار شکاف مسئولیت در هوش مصنوعی: چرا اهمیت دارند و چگونه به آن‌ها بپردازیم» (Santoni de Sio & Mecacci, 2021) نزدیک‌ترین قرابت را با هدف این مقاله دارد. تفاوت اصلی آن پژوهش با مقاله حاضر در مبنای گونه‌شناسی است. تحلیل آن‌ها بر کارکردهای مختلف مسئولیت (شکاف مجازات، پاسخگویی اخلاقی، عمومی و فعال) متمرکز است، درحالی‌که پژوهش حاضر به دنبال ارائه یک گونه‌شناسی بنیادین‌تر بر اساس منشأ و علت شکل‌گیری هر شکاف است.

بنابراین، ادبیات پژوهشی موجود یک خلأ آشکار را نمایان می‌سازد. هر یک از سه جریان پژوهشی پیشین، عملاً تنها به یکی از ابعاد این شکاف پرداخته و در تحلیل خود، کل انگاره شکاف مسئولیت را به همان بُعد واحد تقلیل داده‌اند. از سوی دیگر، تلاش‌های اولیه برای گونه‌شناسی نیز بر کارکرد مسئولیت متمرکز بوده‌اند. آنچه در این میان مغفول مانده، یک تحلیل جامع است که نشان دهد شکاف مسئولیت نه یک مسئله واحد، بلکه چتری مفهومی برای چالش‌هایی با منشأهای کاملاً متفاوت است. این مقاله با هدف پر کردن این خلأ و ارائه یک گونه‌شناسی سه‌گانه مبتنی بر منشأ شکاف نگاشته شده است.

۲. مفاهیم کلیدی

۲-۱. مسئولیت اخلاقی

در این پژوهش، یک تمایز کلیدی که اغلب نادیده گرفته می‌شود، اهمیتی حیاتی دارد: تمایز میان «مسئولیت»[1] به معنای سنتی و فلسفی آن و «پاسخگویی»[2] به معنای یک سازوکار رویه‌ای. این تفکیک مفهومی در ادبیات اسلامی نیز با تعابیری همچون «مسئولیت ناظر به قابلیت» و «مسئولیت ناظر به پاسخگویی» مورد توجه قرار گرفته است (دانش، ۱۳۹۷، ص ۵۱).

«مسئولیت» که ریشه در اخلاق ارسطویی دارد، به این پرسش می‌پردازد که آیا یک عامل، به دلیل برخورداری از ویژگی‌های درونی مانند کنترل و دانش، شایستگی اخلاقی برای سرزنش یا تحسین را دارد یا خیر؟ (Coeckelbergh, 2020, p. 2052) تیگارد این سطح را «تفسیر توصیفی» از مسئولیت می‌نامد که بر وضعیت فاعل به عنوان منبع رویداد متمرکز است (Tigard, 2021, p. 115). در پژوهش‌های بومی این حوزه، عاملیت اخلاقی به معنای «توانایی تصمیم‌گیری و عمل به شیوه‌ای مسئولانه» تبیین شده است (قربانیان، ۱۳۹۹، ص ۳۰)؛ تعریفی که با مفهوم «اهلیت» نیز قرابت دارد، جایی که مسئولیت منوط به برخورداری از قابلیت‌هایی چون نیت، تأمل و اختیار دانسته می‌شود (دانش، ۱۳۹۷، ص ۵۱).

در مقابل، «پاسخگویی» مفهومی رویه‌ای و رابطه‌ای در نظر گرفته می‌شود و به این پرسش می‌پردازد که آیا سازوکارهای اجتماعی و اجرایی برای الزام یک عامل به گزارش‌دهی، توجیه و پذیرش عواقب عملش در برابر یک مرجع مشخص وجود دارد یا نه؟ (Novelli et-al., 2023, p. 1872) این مفهوم، خود دارای دو منظر است: از منظر ساختاری، این رابطه شامل چهار عنصر بنیادی است:

1. Responsibility
2. Accountability

فاعل، مرجع، مبنای هنجاری و پیامدها (Stahl, 2023, p. 3). از منظر فرایندی، این رابطه در یک عملیات سه‌مرحله‌ای محقق می‌شود: الزام به ارائه اطلاعات، توانایی ارائه توجیه، و آمادگی برای پذیرش پیامدها (Busuioc, 2021, p. 826). این رویکرد با تحلیل واژه «مسئولیت» در اندیشه اسلامی از ریشه سُؤال همسو است؛ به معنای «خواسته شدن چیزی از کسی» در جایی که سائل (خواهان) بتواند مسئول (خوانده) را مورد بازخواست قرار دهد (مصباح یزدی، ۱۳۹۳، ص ۱۵۶).

اگرچه این دو مفهوم در گفتار روزمره اغلب به جای یکدیگر به کار می‌روند، اما تفکیک آن‌ها برای تحلیل دقیق چالش‌های هوش مصنوعی ضروری است، امری که محققان این حوزه بر آن تأکید دارند (Cañas, 2022, p. 2). خلاصه آن که مسئولیت به این پرسش می‌پردازد که آیا یک عامل، به دلیل برخورداری از ویژگی‌های درونی مانند کنترل و نیت، شایستگی اخلاقی برای سرزنش یا تحسین را دارد یا خیر. در مقابل، پاسخگویی به این پرسش می‌پردازد که آیا مکانیسم‌های اجتماعی و رویه‌ای برای الزام یک عامل به گزارش‌دهی و پذیرش عواقب عملش وجود دارد یا نه.

۲-۲. هوش مصنوعی

خاستگاه اصطلاح «هوش مصنوعی»[1] به میانه قرن بیستم بازمی‌گردد (Russell & Norvig, 2010, p. 17). با این حال، تعریف مدرن و عملیاتی آن که محور این پژوهش است بر «توانایی یک سیستم برای تفسیر صحیح داده‌های خارجی، یادگیری از این داده‌ها و استفاده از آن آموخته‌ها برای دستیابی به اهداف مشخص از طریق انطباق انعطاف‌پذیر» متمرکز است (Kaplan & Haenlein, 2019, p. 16). در این مقاله، تمرکز ما بر «هوش مصنوعی قوی»[2] یعنی سیستمی با خودآگاهی هم‌سطح انسان، نیست، بلکه بر «هوش مصنوعی ضعیف»[3] است؛ سیستم‌هایی که برای انجام وظایف مشخص و محدود طراحی شده‌اند و تمامی فناوری‌های کنونی را در بر می‌گیرند (Kurzweil, 2005, p. 203).

تحول مفهومی کلیدی که مستقیماً به بحث شکاف مسئولیت مرتبط است، گذار از «سیستم‌های خبره»[4] به «یادگیری ماشین»[5] است (McGraw, 2020, p. 2). سیستم‌های خبره که به رویکرد «بالا به پایین»[6] یا «نمادگرا»[7] تعلق دارند مانند برنامه شطرنج‌باز دیپ بلو، بر مجموعه‌ای از قوانین ثابت استوار بودند که مستقیماً توسط انسان برنامه‌ریزی می‌شدند. در مقابل، هوش مصنوعی مدرن از رویکرد «پایین به بالا»[8] یا «اتصال‌گرا»[9] مانند شبکه‌های عصبی استفاده می‌کند و با تحلیل حجم عظیمی از داده، به صورت خودکار

1. Artificial Intelligence
2. Strong AI
3. Weak AI
4. Expert Systems
5. Machine Learning
6. top-down
7. Symbolic
8. bottom-up
9. Connectionist

دانش و الگوهای رفتاری را استخراج می‌نماید (Kaplan & Haenlein, 2019, p. 18). این گذار از برنامه‌ریزی مبتنی بر قانون به یادگیری مبتنی بر داده، تمایز اصلی فناوری مورد بحث در این پژوهش است (Matthias, 2004, p. 178).

معماری‌های محاسباتی پیچیده‌ای چون «یادگیری عمیق»[1] پیامدهای مهمی برای بحث مسئولیت دارند. این مدل‌ها از چندین لایهٔ پردازشی پشت‌سرهم استفاده می‌کنند تا به صورت سلسله‌مراتبی، مفاهیم را بیاموزند؛ به این معنا که لایه‌های اولیه، الگوهای ساده مانند لبه‌ها در یک تصویر را کشف می‌کنند و لایه‌های بعدی، با ترکیب آن الگوهای ساده، مفاهیم پیچیده‌تر مانند چهره را می‌سازند. ویژگی کلیدی این روش‌ها آن است که این لایه‌ها توسط مهندسان انسانی طراحی نمی‌شوند، بلکه مستقیماً از داده‌ها آموخته می‌شوند (LeCun et-al., 2015, p. 436). دقیقاً همین ساختار پیچیده و خودآموخته که منطق درونی آن اغلب برای انسان قابل تفسیر نیست، منجر به پدیده «جعبه سیاه»[2] می‌شود (Coeckelbergh, 2020, p. 2059).

۳. انگارهٔ شکاف مسئولیت

۳-۱. مؤلفهٔ اول: شکاف فنی-متافیزیکی

نخستین و شناخته‌شده‌ترین بُعد انگارهٔ شکاف مسئولیت، چالشی است که مستقیماً از خود فناوری نشأت می‌گیرد. این بُعد از شکاف، برای نخستین بار به صورت منسجم توسط آندریاس ماتیاس ارائه شد. او استدلال کرد که ظهور ماشین‌های یادگیرنده و خودمختار، وضعیت جدیدی را ایجاد کرده است که در آن، سازنده یا کاربر دیگر قادر به پیش‌بینی کامل رفتار آینده ماشین نیست. از آنجا که کنترل و دانش، از پیش‌شرط‌های اصلی برای اسناد مسئولیت هستند، این عدم تسلط منجر به وضعیتی می‌شود که در آن، هیچ‌کس کنترل کافی بر اعمال ماشین ندارد تا بخواهد مسئولیت آن را بپذیرد (Matthias, 2004, p. 177).

این ادعای ماتیاس تکیه بر سه استدلال مشخص دارد که کاهش کنترل انسان بر این سیستم‌ها را تبیین می‌کند. استدلال اول او این است که نقش برنامه‌نویس در این سیستم‌های هوشمند، تغییر یافته و از یک کُدنویس سنتی به یک پدیدآورندهٔ سیستم‌های یادگیرنده تبدیل شده است. در مهندسی نرم‌افزارهای سنتی، برنامه‌نویس کنترل کاملی بر هر خط کد و نحوهٔ اجرای برنامه دارد و خطاها، مستقیماً خطاهای برنامه‌نویس تلقی می‌شوند؛ اما در ساختار سیستم‌های یادگیرنده، این کنترل مستقیم از بین می‌رود و دیگر از منطق نهایی یا جزئیات عملکرد آن اطلاع دقیقی ندارد؛ برای مثال، در شبکه‌های عصبی، برنامه‌نویس تنها معماری اولیه و پارامترهای یادگیری را تعریف می‌کند؛ دانش نهایی به صورت توزیع‌شده در میلیون‌ها وزن سیناپسی[3] ذخیره می‌شود که دیگر قابل بازرسی یا تفسیر مستقیم انسانی نیست و عملاً یک جعبه سیاه است (Matthias, 2004, p. 180). حتی این وضعیت در برخی برنامه‌نویسی‌ها شدیدتر هم می‌شود و برنامه‌نویس حتی معماری را هم تعریف نمی‌کند، بلکه الگوریتم خودش در جایگاه یک برنامه‌نویس عمل می‌کند و ماشینی را می‌سازد که خودش را برنامه‌نویسی می‌کند (Matthias, 2004, p. 181).

1. Deep Learning
2. Black Box
3. synaptic weights

استدلال دوم او مبتنی بر نقش تعیین‌کننده‌ای است که محیط بر عملکرد ماشین دارد. در سیستم‌های سنتی، رفتار ماشین تماماً توسط برنامه اولیه تعریف می‌شود، اما در سیستم‌های تطبیق‌پذیر، عملکرد آنها به‌شدت تحت‌تأثیر تعاملات بی‌واسطه با محیط عملیاتی و داده‌هایی است که از آن کسب می‌کند. در واقع، برنامه‌نویس، بخشی از کنترل خود را به محیط واگذار می‌کند (Matthias, 2004, p. 182). این مسئله در «عامل‌های خودمختار»[1]، چه نرم‌افزاری و چه فیزیکی، به اوج خود می‌رسد (Matthias, 2004, p. 181) و به خاطر اینکه محیط، وضعیتی پویا و غیرقابل‌پیش‌بینی دارد، برنامه‌نویس دیگر نمی‌تواند میزان یا نوع تأثیرات آن بر شکل‌گیری رفتار نهایی ماشین را کنترل کند.

سومین استدلال او، ضرورت یادگیری در حین عمل است که در سیستم‌های مدرن وجود دارد. سیستم‌های سنتی دارای فازهای مجزای آموزش و عملیات بودند، اما سیستم‌های مدرن برای حفظ کارآمدی خود در محیط‌های پویا مانند خودروهای خودران، باید دائماً در حال یادگیری و انطباق باشند. این امر مستلزم به‌کارگیری روش‌هایی مانند «یادگیری تقویتی»[2] است که این نوع یادگیری ذاتاً بر اساس آزمون‌وخطا طراحی شده است. در این پارادایم جدید، اشتباهات دیگر به عنوان یک نقص فنی یا خطای برنامه‌نویس شناسایی نمی‌شوند، بلکه ویژگی‌های اجتناب‌ناپذیر سیستم هستند و در واقع زمینهٔ یادگیری و رفتار بهینه آن سیستم را فراهم می‌آورد (Matthias, 2004, pp. 180, 183). این سه عامل فنی، مستقیماً دو شرط اصلی مسئولیت؛ یعنی کنترل و دانش را از بین می‌برند و هستهٔ اصلی شکاف را تشکیل می‌دهند.

این زوال کنترل را می‌توان بر مبنای تحلیل هستی‌شناختی و مبانی فقهی تبیین نمود. در تحلیل هستی‌شناختی، هوش مصنوعی نه به عنوان یک عامل دارای نفس و اراده، بلکه سیستمی با هوش محاسباتی و فاقد پشتوانهٔ روحی تحلیل می‌شود (Floridi & Sanders, 2004, p. 372)؛ دیدگاهی که با رویکردهای نوین فلسفه فناوری در «مصنوع» دانستن این سیستم‌ها همسو است (Bryson, 2018, p. 21). چالش فنی در اینجاست که الگوریتم‌های یادگیری عمیق، علی‌رغم فقدان قصد، به دلیل پیچیدگی محاسباتی، رفتاری شبه‌مختار بروز می‌دهند.

این وضعیت، قواعد فقهی را با چالش جدی مواجه می‌سازد. اگرچه در فقه، ضمان دایر مدار استناد تکوینی است (صادقی و میرزایی، ۱۳۹۸، ص ۱۷۲) و قصد شرط نیست (نجفی، ۱۳۶۷، ج ۴۳، ص ۵۰)، اما ماهیت یادگیرندهٔ سیستم، دقیقاً همین رابطهٔ مادی استناد میان طراح و نتیجه را مخدوش می‌کند. از این‌رو، قاعدهٔ «تقدم مباشر بر سبب» دچار توقف می‌شود؛ زیرا هوش مصنوعی وضعیتی دوگانه ایجاد می‌کند که از سویی مانند «مباشر مکرَه» فاقد اراده است که باید ضمان را به سبب (طراح) منتقل کرد و از سوی دیگر دارای «انتخاب‌گری محاسباتی» است که استناد فعل به طراح را دشوار می‌سازد.

تحلیل ارائه شده توسط ماتیاس، تبیین شکاف را به یک تناقض مفهومی جدی رساند. این تناقض از تقابل دو پیش‌فرض ناسازگار نشئت می‌گیرد؛ زیرا از یک سو، نمی‌توان یک برنامه‌نویس را به طور موجه مسئول اعمال یک ربات واقعاً خودمختار دانست و از سوی

1. Autonomous agents
2. Reinforcement learning

دیگر، مسئول دانستن یک ربات بی‌احساس، بی‌معنا و غیرقابل‌توجیه است (Kiener, 2022, p. 576).

رابرت اسپارو با بیان دقیق خود این بن‌بست را به‌خوبی تحلیل کرد و نشان داد که چرا تلاش برای یافتن یک عامل مسئول در این چارچوب به شکست می‌انجامد. او استدلال می‌کند که در صورت وقوع یک خطای فاجعه‌بار، ما با سه گزینه روبرو هستیم و هیچ‌کدام نمی‌تواند پرسش مسئولیت را پاسخ دهد. این سه گزینه عبارت‌اند از اینکه از یک سو طراحان یا برنامه‌نویسان به دلیل خودمختاری و یادگیری ماشین، ارتباط علیتی‌شان با نتیجه قطع شده است و از سوی دیگر کاربر یا فرمانده عملیاتی نیز به دلیل عدم کنترل مستقیم بر تصمیمات ماشین، مسئول نیست و در نهایت، خود ماشین هم فاقد شرایط لازم برای عاملیت اخلاقی (مانند آگاهی، قصد، و قابلیت رنج‌بردن برای مجازات) است (Sparrow, 2007, pp. 69–73). در نتیجه این سیستم‌ها در یک وضعیت میانی قرار دارند؛ چرا که آن‌ها به اندازه‌ای خودمختار هستند که دیگران را از مسئولیت مبرا کنند، اما نه به اندازه‌ای که خودشان عامل اخلاقی مسئول شناخته شوند (Sparrow, 2007, p. 73). در این وضعیت ما با یک کنش و آسیب ناشی از آن مواجهیم، اما هیچ عامل اخلاقی مناسبی برای اسناد مسئولیت آن پیدا نمی‌کنیم.

بدین ترتیب، شکاف فنی–متافیزیکی به عنوان نخستین بُعد این چالش، خود را نشان می‌دهد. این شکاف از یک سو ریشه در زوال کنترل و دانش انسان در مواجهه با سیستم‌های یادگیرنده دارد و از سوی دیگر، ناشی از عدم قطعیت در مورد وضعیت و جایگاه عاملیت خود ماشین است.

۳-۲. مؤلفهٔ دوم: شکاف سازمانی–ساختاری

بُعد دوم شکاف مسئولیت، برخلاف بُعد نخست، مستقیماً از خود فناوری نشئت نمی‌گیرد، بلکه ریشه در «ساختارهای انسانی» و «چارچوب‌های سازمانی» دارد که این فناوری‌ها را در بر گرفته‌اند.

بهترین مفهوم برای تبیین این عامل، مفهوم «مشکل دست‌های بسیار»[1] در نظریه سازمان و فلسفه سیاسی است. این مفهوم که ریشه در اندیشهٔ متفکرانی چون دنیس تامپسون دارد، به وضعیتی اشاره دارد که در آن، تعداد زیاد عاملان انسانی در یک زنجیرهٔ علّی، یافتن یک فرد مشخص که مسئولیت یک شکست جمعی را بر عهده بگیرد، تقریباً غیرممکن می‌سازد (Lancaster et-al., 2024, p. 7).

هوش مصنوعی این مشکل قدیمی را در سطح بالاتری مطرح می‌کند؛ زیرا نه‌تنها تعداد دست‌های انسانی بیشتر شده، بلکه به تعبیر مارک کوکلبرگ، ما با «تعدد اجزای فنی»[2] نیز روبرو هستیم؛ یعنی انبوهی از مصنوعات فناورانه مانند حسگرهای خراب و الگوریتم‌های مبهم که آنها نیز به صورت علّی در وقوع آسیب نقش دارند (Coeckelbergh, 2020, p. 2052). این پیچیدگی، مسئولیت را در این شبکهٔ متراکم از عوامل دخیل، پراکنده کرده و عملاً غیرقابل‌دسترس می‌سازد.

اگرچه مشکل دست‌های بسیار گاهی با خود شکاف مسئولیت یکسان پنداشته می‌شود و در جایگاه تبیینی برای آن مطرح می‌شود،

1. Problem of Many Hands
2. Many Things

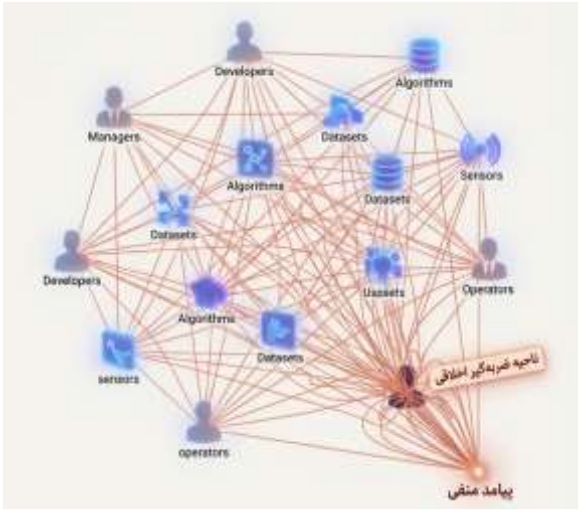
اما تحلیل دقیق‌تر نشان می‌دهد که این عاملی برای شکل‌گیری شکاف است. همان‌طور که دا سیلوا (Da Silva, 2024, p. 5) اشاره می‌کند، مهم است که این مؤلفه‌ها با دقت از هم متمایز شوند؛ زیرا شکاف سازمانی تنها یکی از وجوه چالش کلی است.

این پیچیدگی سازمانی را می‌توان در چارچوب دستگاه تحلیلی بومی، ذیل عناوین تسبیب و چالش تعیین قَرار ضَمان بازخوانی کرد. در فقه، وقتی خسارتی رخ می‌دهد، اصل اولی بر ضمان مباشر است؛ مگر آنکه سبب اقوی از مباشر باشد (حسینی عاملی، ۱۴۱۹، ج ۲۶، ص ۱۴۳؛ فخرالمحققین، ۱۳۸۷، ج ۴، ص ۶۸۸). در اکوسیستم هوش مصنوعی، ماشین نقش مباشر را ایفا می‌کند اما چون فاقد شعور است، علی‌القاعده بار مسئولیت باید به «سبب» (طراح، داده‌پرداز یا کاربر) منتقل شود. اما چالش اصلی در تطبیق مفهوم سبب نهفته است. در تعاریف فقهی، سبب چیزی است که «اگر نبود، تلف رخ نمی‌داد» (نجفی، ۱۳۶۷، ج ۳۷، ص ۴۸). در اینجا، هم طراح، هم داده‌پرداز و هم کاربر، همگی شرط لازم (سبب) هستند. فقه در مواجهه با اجتماع اسباب، به دنبال سببی است که تأثیر مادی یا عدوانی داشته باشد، اما ماهیت غیرخطی و یادگیرندهٔ سیستم، امکان تفکیک و وزن‌دهی به سهم هر یک را سلب می‌کند. اگرچه راهکارهای نوینی چون «توزیع مسئولیت بر اساس میزان تأثیر» مطرح شده است (علیدوست و بای، ۱۳۹۱، ص ۱۰۲)، اما شکاف سازمانی دقیقاً در مرحلهٔ سنجش میزان تأثیر رخ می‌دهد و عملاً تعیین سبب اقوی یا نقطه کانونی مسئولیت را با بن‌بست محاسباتی روبرو می‌سازد.

مادلین الیش با بیان یک تشبیه دقیق تحت عنوان «ناحیهٔ ضربه‌گیر اخلاقی»[1] اوج این پیچیدگی سازمانی و پیامد عملی آن را به‌خوبی نشان می‌دهد. او این مفهوم را با الهام از ناحیهٔ ضربه‌گیر در خودروها ابداع کرد که برای جذب نیروی تصادف و محافظت از سرنشینان طراحی شده است. او به طور مشابه، ناحیهٔ ضربه‌گیر اخلاقی را مطرح می‌کند و توضیح می‌دهد که چگونه مسئولیت شکست کل یک سیستم پیچیده، به صورت ناعادلانه بر عهدهٔ نزدیک‌ترین اپراتور انسانی متمرکز می‌شود، درحالی‌که آن فرد کنترل محدودی بر سیستم داشته و صرفاً نقش واسطه‌ای برای کاهش فشار مسئولیت بر ساختار فناورانه یافته است (Elish, 2019, p. 40).

او با بیان موارد عینی مشابه نشان می‌دهد که این مسئله‌ای دورازذهن نیست. یکی از این موارد، حادثهٔ نیروگاه هسته‌ای تری مایل آیلند[2] در سال ۱۹۷۹ است. در این حادثه هستهٔ رآکتور نسبتاً ذوب شد. آنچه در گزارش و روایت اولیهٔ رسانه‌ها در مورد این موضوع انتشار یافت، مقصر دانستن خطای انسانی بود. اما تحلیل‌های عمیق‌تر نشان داد که اپراتورها اطلاعات ناقصی را از سمت رابط کاربری معیوب اتاق کنترل دریافت می‌کردند. این نقص‌های در ابزارهای فنی و از سوی دیگر نادیده‌گرفتن آنها توسط مدیریت و عدم وجود فرهنگ‌سازمانی جهت به‌کارگیری قوانین ایمنی، زمینه‌ساز این فاجعه بودند. بااین‌حال در نهایت این اپراتورها بودند که به عنوان مقصر شناخته شدند و بار این شکست که بیشتر محصول یک سیستم معیوب بود، بر آنها تحمیل شد (Elish, 2019, p. 45).

1. Moral Crumple Zone
2. Three-Mile Island



تصویر شماره ۱: پراکندگی مسئولیت

این تحلیل‌ها نشان می‌دهند که بُعد دوم چالش مسئولیت، یک مشکل سازمانی-ساختاری است. این شکاف نشان می‌دهد که چالش اصلی، صرفاً فقدان یک عامل مسئول نیست، بلکه نحوهٔ طراحی ساختارهای انسانی است که در آن مسئولیت، یا در میان دست‌های بسیار آن‌قدر پراکنده می‌شود که در عمل هیچ‌کس پاسخگو نیست یا به‌صورت ناعادلانه به عامل انسانی نسبت داده می‌شود. این ساختارهای پیچیده، بستر روان‌شناختی‌ای را فراهم می‌کنند که رهایی از مسئولیت را برای عاملان انسانی تسهیل کرده و به آن مشروعیت می‌بخشد.

این بُعد روان‌شناختی پنهان شکاف سازمانی-ساختاری که به فعال‌شدن و تشدید آن کمک می‌کند، توسط آلبرت بندورا، روان‌شناس برجسته، در ضمن نظریهٔ «مکانیسم‌های رهایی از مسئولیت اخلاقی»[1] تبیین شده است. او در این نظریه این مسئله را بررسی می‌کند که چگونه افرادی که اخلاق‌مدار هستند، می‌توانند خودشان را برای انجام اعمال غیراخلاقی یا آسیب‌زا، قانع کنند، بدون اینکه درگیر عذاب وجدان و سرزنش درونی شوند؟ این نظریه توضیح می‌دهد که این امر، ناشی از عوامل شناختی است که باعث می‌شود شخص، کنترل اخلاقی درونی خودش را غیرفعال کند (Bandura, 1999, p. 197).

بندورا هشت عامل برای این مسئله شناسایی کرده که دو مورد از مهم‌ترین و مرتبط‌ترین آنها با بحث ما، «جابه‌جایی مسئولیت»[2] و «پخش‌کردن مسئولیت»[3] است. در جابه‌جایی مسئولیت، افراد با نسبت‌دادن مسئولیت اعمال خود به یک مرجع یا اقتدار بیرونی، از پیامدهای آن فرار می‌کنند. آنها کارهای خودشان را نتیجهٔ دستورات و الزامات خارجی وانمود می‌کنند و خود را در قبال آن مسئول نمی‌دانند. در پخش‌کردن مسئولیت نیز به‌خاطر دخیل بودن عوامل متعدد در شکل‌گیری یک عمل، هر فرد سهم خودش را کوچک می‌شمارد و به همین جهت احساس مسئولیت او در قبال این عمل نیز کاهش پیدا می‌کند (Bandura, 1999, pp. 197-198).

1. Mechanisms of Moral Disengagement
2. Displacement of Responsibility
3. Diffusion of Responsibility

گسترش سیستم‌های فنی پیچیده و خودمختار هوشمند، بستر روان‌شناختی مساعدی را برای این دو عامل فراهم کرده است. وقتی نوع عملکرد این سیستم‌ها شفاف نیست، در صورت بروز یک پیامد منفی، عاملان انسانی درگیر با استفاده از جابه‌جایی مسئولیت خود ماشین‌ها را مرجع تصمیم‌گیر معرفی می‌کنند و عهدهٔ خود را از مسئولیت خالی می‌کنند. همین‌طور حضور عوامل متعدد در بستر شکل‌گیری و به‌کارگیری این سیستم‌ها (که پیش‌تر با عنوان مشکل دست‌های بسیار بررسی شد)، موجب می‌شود که هر فرد در این زنجیره، نقش خود را ناچیز بداند و از بار اخلاقی آن شانه خالی کند.

این زمینهٔ روان‌شناختی در جدیدترین تحلیل‌ها به عنوان هسته اصلی شکاف شناسایی شده است. برخی استدلال می‌کنند که این شکاف زمانی رخ می‌دهد که عاملان انسانی به دلیل عوامل روان‌شناختی، دیگر به طور تجربی، خروجی‌های سیستم را متعلق به خود نمی‌دانند (Khomkhunsorn, 2025, p. 2). این امر ناشی از پدیده‌هایی مانند «اتکای بیش از حد به اتوماسیون»[1] و «حائل اخلاقی»[2] است که نوعی فاصله‌گذاری روان‌شناختی ایجاد می‌کند که به افراد اجازه می‌دهد خود را از نظر اخلاقی از اعمالشان جدا کنند (Khomkhunsorn, 2025, p. 2).

این وضعیت، مستقیماً به چیزی منجر می‌شود که برخی آن را «مهارت‌زدایی اخلاقی»[3] نامیده‌اند؛ زیرا اتکای مستمر به سیستم‌های هوشمند، فرصت‌های لازم برای تمرین و تقویت مهارت‌های اخلاقی انسان را از بین می‌برد. در نتیجه، این عدم درگیری خود شکاف مسئولیتی را ایجاد می‌کند که در آن، انسان‌ها حتی اگر اسماً مسئول باشند، انگیزه و توانایی درونی برای اقدام پیشگیرانه و مسئولانه را از دست می‌دهند (Khomkhunsorn, 2025, p. 9).

ازاین‌رو شکاف سازمانی-ساختاری تنها یک چالش طراحی سازمانی نیست، بلکه در درون خود یک چالش روان‌شناختی عمیق را دارد که این پیچیدگی ساختار سازمانی، زمینه‌ای برای رهایی از مسئولیت ایجاد می‌کند.

۳-۳. مؤلفهٔ سوم: شکاف معرفتی

بُعد سوم و نهایی این چالش، «شکاف معرفتی» است. این شکاف، برخلاف دو بُعد پیشین به بودن یا نبودن مسئولیت یا نحوهٔ توزیع آن نمی‌پردازد، بلکه موضوع آن دانستن و توانایی اثبات آن است. این مسئله زمانی نمود می‌یابد که بپذیریم حتی اگر مسئولیت متافیزیکی وجود داشته باشد و ساختار سازمانی هم شفاف باشد، ما به دلیل ماهیت خود فناوری، قادر به ردیابی و اسناد صحیح مسئولیت نیستیم.

این شکاف معرفتی که از پیچیدگی فنی هوش مصنوعی نشأت می‌گیرد، باعث شکست مکانیسم‌های عملی پاسخگویی می‌شود. این اختلال به دلیل پدیدهٔ «جعبه سیاه» رخ می‌دهد. جعبه سیاه به معنای یک مانع اطلاعاتی است که خود به دو شکل ظاهر می‌شود. نخست، «جعبه سیاه فنی» است که از پیچیدگی ذاتی مدل‌های یادگیری عمیق مانند شبکه‌های عصبی نشئت می‌گیرد که حتی طراحان

1. Automation Bias
2. Moral Buffer
3. Moral Deskilling

سیستم نیز به دلیل تعداد بالای پارامترها و محاسبات پنهان، قادر به درک کامل منطق درونی تصمیم‌گیری‌های مدل نیستند (Fritz et-al., 2020, p. 39). دوم، «جعبه سیاه تجاری» است که در آن، شرکت‌ها به دلیل حفظ اسرار تجاری و مالکیت معنوی، مانع از افشای الگوریتم‌ها و داده‌های آموزشی خود می‌شوند؛ همان‌طور که در مورد الگوریتم‌های تجاری پیش‌بینی تکرار جرم این سیاست مشاهده می‌شود (Fritz et-al., 2020, p. 37). این دو مانع با ابهامی که ایجاد می‌کند، توانایی ناظران برای شناسایی و ردیابی در زنجیرهٔ علیت را از بین می‌برد و در نتیجه امکان اسناد مسئولیت را با مشکل مواجه می‌کند.

پیامد مستقیم این جعبه سیاه، ایجاد ابهام در «احراز استناد» است. اهمیت این شکاف زمانی آشکارتر می‌شود که بدانیم در تحلیل‌های دقیق حقوقی، اگرچه مرجع تشخیص استناد «عرف» است (علیدوست و بای، ۱۳۹۱، ص ۱۱)، اما داوری عرف تنها زمانی معتبر است که منطبق با «واقعیات موجود و برگرفته از طرق علمی» باشد و ارتکازات مبتنی بر تسامح در اینجا راه ندارد (صادقی و میرزایی، ۱۳۹۸، ص ۱۸۰). پدیده جعبه سیاه دقیقاً همین رکن کشف علمی را هدف قرار می‌دهد. وقتی ساختار شبکه عصبی مانع از ردیابی منطقی خط سیر تصمیم می‌شود، ابزار علم برای کشف حقیقت استناد ناکارآمد می‌گردد. در نتیجه، حتی اگر عرف بخواهد به صورت مسامحه‌آمیز طراح را مسئول بداند، این اسناد به دلیل فقدان پشتوانهٔ احراز علمی و مادی، از منظر مبانی دقیق استناد، مخدوش خواهد بود.

از سوی دیگر، این عدم شفافیت منجر به شکل‌گیری وضعیتی می‌شود که می‌توان آن را تولید غَرَر و ریسک ناشی از جهل نامید؛ وضعیتی که مصداق قاعدهٔ «المغرور یرجع الی من غرّه» است (خویی، ۱۴۱۸، ج ۳۷، ص ۸۶). چالش اصلی در اینجاست که قواعد فقهی تصریح دارند که «مغرور» (کسی که جاهلانه و بر اثر اعتماد عمل کرده) ضامن نیست و مسئولیت بر عهده «غارّ» (فریب‌دهنده) است. اما در یادگیری عمیق، حتی طراح سیستم (غارّ احتمالی) نیز نسبت به فرایند تصمیم‌گیری جاهل است. اینجا با وضعیتی روبرو هستیم که جهل فراگیر بر کل سیستم حاکم است.

توسعهٔ فنی این چالش در آنجاست که مبنای احراز مسئولیت در نظام اخلاقی و فقهی، تمایز میان «جاهل قاصر» و «جاهل مُقصِّر» است؛ تمایزی که دایر مدار قدرت بر تحصیل علم است (خویی، ۱۴۱۸، ج ۱، ص ۱۶۰). اما ویژگی ذاتی یادگیری عمیق و پیچیدگی لایه‌های پنهان، با ایجاد مانع، قدرت پیش‌بینی را حتی از متخصصان سلب می‌کند. در نتیجه، خط‌کشی میان «قصور» و «تقصیر» ناممکن شده و راه برای تمسک به جهل ناشی از قصور و گریز از پاسخگویی برای همگان هموار می‌گردد.

این مانع توسط برخی متفکران به عنوان یک محور اصلی در بحث شکاف مسئولیت تفسیر شده است و آن‌ها استدلال می‌کنند که شاید مشکل اصلاً متافیزیکی نبوده است، بلکه از ابتدا حقیقت این چالش شکاف مسئولیت ناشی از این شکاف معرفتی؛ یعنی ناتوانی در یافتن مسئول بوده است.

پیتر کونیگز مدافع این دیدگاه، استدلال می‌کند که مشکل اصلی، یک شکاف متافیزیکی نیست، بلکه یک شکاف معرفتی است. از نظر او، این شکاف، یک چالش عملی در تعیین دقیق مسئولیت است که مستقیماً از ماهیت فنی خود سیستم ناشی می‌شود. از منظر این دیدگاه، مسئولیت وجود دارد، اما به دلیل پیچیدگی سیستم و ابهام فنی آن، پنهان شده و اثبات آن بسیار دشوار است

.(Königs, 2022, p. 10)

این نوع تحلیل، در جدیدترین تحلیل‌های مفهومی این حوزه نیز بازتاب یافته است. فِلُوِن‌کمپ در مقاله‌ای با عنوان «شکاف‌های مسئولیت چه هستند و چه باید باشند»، استدلال می‌کند که مفهوم شکاف مسئولیت به معنای شکاف متافیزیکی صحیح نیست و باید با مفهوم دقیق‌تر شکاف مسئولیت معرفتی جایگزین شود. او این شکاف را دقیقاً ناشی از ساختار تصمیم‌گیری مبهم می‌داند که شناسایی عامل قابل سرزنش را دشوار یا غیرممکن می‌سازد (Veluwenkamp, 2025, p. 12).

۴. واسازی انگاره واحد: خلط مفهومی و مسیر پیش رو

تحلیل‌های تطبیقی ارائه شده در سه بخش پیشین نشان داد که انگاره «شکاف مسئولیت»، برخلاف تلقی رایج که آن را یک مسئله واحد می‌داند، پدیده‌ای چندوجهی است. بن‌بست کنونی در این بحث تا حد زیادی ناشی از عدم تفکیک نظام‌مند و خلط این چالش‌های مجزا با یکدیگر بوده است. این پژوهش، با واسازی این انگاره، نشان داد که «شکاف مسئولیت» در واقع چتری مفهومی برای حداقل سه چالش مجزا است: ۱. چالش فنی-متافیزیکی (زوال کنترل و استناد)، ۲. چالش سازمانی-ساختاری (تعدد اسباب و عدم تعین)، ۳. چالش معرفتی (ابهام و جهل). صورت‌بندی دقیق این چالش‌ها به مثابه سه مؤلفه مجزا، به جای یک شکاف واحد، می‌تواند مسیر را برای ارائه راهکاری جامع و چندلایه جهت خروج از بن‌بست فعلی فراهم آورد.

ضرورت چنین رویکردی در تحلیل این شکاف، در دیگر پژوهش‌هایی در این حوزه نیز که به سبکی دیگر به این تحلیل پرداخته‌اند، تأیید می‌شود. برخی در تحلیل خود، این شکاف را نه مفهومی واحد، بلکه در چهار بُعد مجزا تحلیل می‌کنند: شکاف مجازات، شکاف پاسخگویی، شکاف پاسخگویی عمومی و شکاف پاسخگویی فعال (Santoni de Sio & Mecacci, 2021, p. 1059).

این تحلیل، یافته‌های پژوهش حاضر را به شکلی دقیق تأیید می‌کند. شکاف مجازات نزد آن‌ها که به دشواری یافتن مقصر برای سرزنش در یک سیستم پیچیده می‌پردازد (Santoni de Sio & Mecacci, 2021, p. 1060)، مستقیماً با آنچه ما به عنوان شکاف سازمانی-ساختاری شناسایی کردیم، همسو است. به همین ترتیب، شکاف پاسخگویی عمومی نزد آنان که صراحتاً به جعبه سیاه فنی ارجاع داده می‌شود (Santoni de Sio & Mecacci, 2021, p. 1064)، توصیف دقیقی از آن چیزی است که ما شکاف معرفتی نامیده‌ایم. در نهایت، شکاف پاسخگویی اخلاقی آنان که به آگاهی عامل از ابعاد اخلاقی عملش که با زوال کنترل و دانش از بین می‌رود، بازمی‌گردد (Santoni de Sio & Mecacci, 2021, p. 1062)، بر شکاف فنی-متافیزیکی ما قابل انطباق است. این همسویی نشان می‌دهد که حرکت از یک نگاه یکپارچه به شکاف به سمت یک تحلیل چندبعدی، گامی ضروری و معتبر برای حل بن‌بست نظری موجود است.

در راستای همین رویکرد، فِلُوِن‌کمپ نیز پیشنهاد می‌کند که بخش دیگری از این چالش، نه یک شکاف به معنای خلأ، بلکه یک عدم هم‌ترازی کنترل؛ به معنای تخصیص نادرست کنترل به عاملان اشتباهی در یک سیستم است (Veluwenkamp, 2025, p. 12). این مفهوم به طور دقیق، هر دو بُعد فنی و سازمانی شناسایی‌شده در این مقاله را پوشش می‌دهد.

تحلیل ارائه شده توسط برخی پژوهش‌های جدید در این موضوع، به ما کمک می‌کند تا ریشه‌های فلسفی این خلط مفهومی را درک کنیم. آن‌ها استدلال می‌کنند که بحث جدید شکاف مسئولیت در هوش مصنوعی، در واقع تصویری جدید از یک مناقشه قدیمی در نظریه مسئولیت در نظریهٔ استراوسون بازمی‌گردد (Oimann & Tollon, 2025, p. 4).

اویمان و تولون شکاف بنیادین در نظریه مسئولیت را حول دو دیدگاه متضاد درباره «تقدم ذاتی» توضیح می‌دهند:

دیدگاه اول، «مستقل از پاسخ»[1] است که موضع سنتی و شهودی‌تر محسوب می‌شود. در این دیدگاه، مسئول بودن یک واقعیت عینی و مستقل از واکنش‌های ماست. به عبارت دیگر، مسئولیت یک ظرفیت است که یک فرد یا دارد یا ندارد و یک فرد زمانی مسئول است که دارای آن ظرفیت‌های مشخص مانند کنترل، دانش یا اراده آزاد باشد. واکنش‌های اجتماعی ما مانند سرزنش نیز صرفاً ابزارهایی هستند که در صورت موجه بودن، این واقعیت عینی از پیش موجود را به درستی شناسایی و ردیابی می‌کنند.

در مقابل، دیدگاه «وابسته به پاسخ»[2] که ریشه در کار استراوسون دارد، این اولویت را معکوس می‌کند. طبق این دیدگاه، مسئول بودن یک واقعیت متافیزیکی نیست که ما کشف کنیم، بلکه توسط اعمال اجتماعی ما؛ یعنی مسئول دانستن یکدیگر، ساخته و برقرار می‌شود. در این نگاه، مسئولیت اساساً یک رویهٔ اجتماعی و فرایند پویا و انعطاف‌پذیر است، نه یک ظرفیت فردی (Oimann & Tollon, 2025, p. 6; Tigard, 2021, p. 598).

این چارچوب تحلیلی به ما در تحلیل چرایی خلط ابعاد شکاف تاکنون کمک می‌کند. دیدگاه مستقل از پاسخ، مستقیماً با دو بُعد از شکاف درگیر است. این دیدگاه با «شکاف فنی-متافیزیکی» درگیر است؛ زیرا ظرفیت کنترل و دانش از بین رفته است و با «شکاف معرفتی» درگیر است؛ زیرا شناسایی آن واقعیت عینی به دلیل پدیده جعبه سیاه غیرممکن شده است. در مقابل، دیدگاه «وابسته به پاسخ»، مستقیماً با «شکاف سازمانی-ساختاری» درگیر است؛ زیرا در این موقعیت رویه‌های اجتماعی ما در اسناد مسئولیت شکست می‌خورند.

نتیجهٔ این خلط مفهومی و عدم تحلیل جامع از شکاف را می‌توان در پاسخ‌هایی که برای این شکاف ارائه شده ملاحظه نمود. دیوید گانکل سه رویکرد اصلی نظری در پاسخ به این چالش را شناسایی می‌کند: نخست، «ابزارانگاری»[3] است که فناوری را ابزار خنثی و مسئولیت را منحصراً نزد انسان می‌داند. این رویکرد در توضیح سیستم‌های یادگیرنده که از کنترل طراح خارج می‌شوند، شکست می‌خورد. دوم، «اخلاق ماشین»[4] که با اعطای نوعی عاملیت اخلاقی به ماشین سعی در پر کردن شکاف دارد. این رویکرد نیز ناکافی است زیرا اسناد مسئولیت به یک ماشین فاقد قصد و آگاهی، کارکردهای اصلی سرزنش را بی‌معنا می‌سازد. سوم، «مسئولیت ترکیبی یا توزیع‌شده»[5] است که مسئولیت را در شبکه‌ای از عاملان انسانی و غیرانسانی توزیع می‌کند. این رویکرد نیز در عمل با خطر

1. Response-Independence
2. Response-Dependence
3. Instrumentalism
4. Machine Ethics
5. Hybrid/Distributed Responsibility

پخش‌شدن مسئولیت تا حد از بین رفتن کامل آن مواجه است (Gunkel, 2017, pp. 314-318).

اکنون با تحلیلی که از مؤلفه‌های شکاف ارائه شد، مشخص می‌شود که بن‌بست بحث «شکاف مسئولیت»، ناشی از تلاش برای یافتن یک راه‌حل واحد برای این سه مسئلهٔ کاملاً متفاوت بوده است و هر کدام با پاسخ به یک قسمت از آن به دنبال حل کل مسئله بوده‌اند. اکنون با واسازی این انگاره، می‌توانیم نظریات رقیب را از حالت ایدهٔ محوری خارج کرده و هر یک را پاسخ به بخشی از این شکاف، بدانیم و در جایگاه دقیق خود قرار دهیم. بنابراین، راه‌حل، جستجوی یک پاسخ واحد برای این مسئله نیست، بلکه نیازمند سه دسته راه‌حل مجزا و هدفمند است که هر یک، یکی از ابعاد سه‌گانه این چالش را پوشش دهد.

پاسخ به شکاف فنی-متافیزیکی: این چالش ماهیتی مفهومی دارد. در ادبیات غربی، راهکارهایی چون «عاملیت ترکیبی» (Verbeek, 2015) یا «عاملیت سه‌گانه» جانسون (Johnson & Verdicchio, 2018) پیشنهاد شده است. تحلیل این مسئله با قواعد دینی نشان داد که راهکار دقیق‌تر، گذار از «مسئولیت مبتنی بر قصد» به «مسئولیت مبتنی بر استناد مادی» است؛ یعنی استفاده از مبانی ضمان که حتی بدون نیت انسانی، صرف استناد تکوینی برای جبران خسارت کافی باشد.

پاسخ به شکاف سازمانی-ساختاری: این چالش ماهیتی مدیریتی و حقوقی دارد. در ادبیات موضوع، راهکارهایی مبتنی بر شفافیت سازمانی مانند اجرای شرط «پیگیری» (Santoni de Sio & Mecacci, 2021, p. 1074) و یا «اسناد پاسخگویی و هماهنگی وظایف» در تیم‌های میان‌رشته‌ای (Dastani & Yazdanpanah, 2023, p. 869) مطرح شده است. همچنین رویکردهای انتقادی بومی نیز بر این نکته تأکید دارند که برای خروج از این بن‌بست، باید سطوح مسئولیت را فراتر از فرد، در لایه‌های «اجتماعی» و «نهادی» بازتعریف کرد تا نهادهای سیاست‌گذار نیز در جبران خسارت سهیم باشند (میرزامحمدی و تقوی، ۱۴۰۴، ص ۱۱۱). هم‌راستا با این رویکردها ولی با مبنایی متفاوت، ظرفیت‌های فقهی نشان می‌دهد که برون‌رفت از بن‌بست تعیین مقصر در شبکه‌های پیچیده، در گرو گذار به مدل‌های «توزیع اشتراکی ریسک» (توزیع ضمان بر اساس میزان بهره‌مندی و تأثیر) است تا از لوث شدن مسئولیت در میان اسباب متعدد جلوگیری شود.

پاسخ به شکاف معرفتی: این چالش ماهیتی فنی و نظارتی دارد. راهکارهای فنی موجود بر اجرای شرط «ردیابی» (Santoni de Sio & Mecacci, 2021, p. 1075) و توسعه ابزارهایی نظیر «هوش مصنوعی قابل‌توضیح» (Dastani & Yazdanpanah, 2023, p. 870) تأکید دارند. تحلیل بر اساس قواعد فقهی نیز مؤید این است که شفافیت باید فراتر از یک ابزار فنی، به عنوان شرط لازم «رفع غرر» و پیش‌شرط مشروعیت بکارگیری سیستم لحاظ گردد؛ به گونه‌ای که عدم شفافیت، خود به عنوان یک قصور مستقل و موجب ضمان تلقی شود.

نوع شکاف	ماهیت چالش	دسته راه‌حل‌ها
فنی-متافیزیکی	مفهومی و حقوقی	گذار از مسئولیت مبتنی بر قصد به **ضمان مبتنی مبتنی بر خطر و استناد مادی**
سازمانی-ساختاری	مدیریتی و نظارتی	طراحی **ساختارهای پاسخگویی شفاف** و حرکت به سوی مدل‌های **توزیع اشتراکی ریسک**
معرفتی	فنی و رویه‌ای	الزام به **شفافیت فنی (XAI)** و در نظر گرفتن عدم شفافیت به مثابه **قصور مستقل و موجب ضمان**

تصویر شماره ۲: راه‌حل چندلایه

نتیجه‌گیری

برآیند تحلیل‌های این پژوهش آشکار می‌سازد که گام نخست در مواجهه با شکاف مسئولیت، واسازی این انگاره و تفکیک ابعاد درهم‌تنیدهٔ آن است؛ اقدامی که پیش از هر چیز، چرایی دشواری اسناد را در این بستر نوین تبیین می‌کند. از سوی دیگر، این پژوهش نشان داد که نمی‌توان انتظار داشت ظرفیت‌های موجود در قواعد مسئولیت و ضمان، با تکیه بر صورت‌بندی‌های کلاسیک، به سادگی بر این چالش فائق آیند؛ بلکه کارآمدی این قواعد در گرو بازخوانی مبنایی و تطبیق ساختاری آن‌ها با ویژگی‌های ماهوی و کارکردی این فناوری‌های نوظهور است.

در همین راستا، تحلیل بُعد فنی–متافیزیکی نشان داد که اصرار بر جستجوی عاملیت انسانی یا نیت مجرمانه در ماشین، یک خطای راهبردی است. تحلیل هستی‌شناختی هوش محاسباتی اثبات کرد که ما با گسست در رابطه مادی استناد مواجهیم. نتیجهٔ این گسست آن است که برای پر کردن این خلأ، باید از پارادایم مسئولیت مبتنی بر قصد و تقصیر عبور کرده و به سمت پارادایم مسئولیت مبتنی بر استناد عرفی توسعه‌یافته یا ضمان ناشی از خطر حرکت کنند؛ چرا که در غیر این صورت، ماهیت نوظهور الگوریتم‌ها، همواره راه گریز از مسئولیت را باز خواهد گذاشت.

همچنین، ضرورت تطبیق ساختاری در بُعد سازمانی–ساختاری آشکار ساخت که قاعدهٔ انتقال ضمان به سبب اقوی، در مواجهه با شبکه‌های پیچیدهٔ هوش مصنوعی کارایی خود را از دست می‌دهد و نتیجه‌ای جز لوث شدن خون یا مال؛ به معنای هدر رفتن حق در میان انبوهی از اسباب ناقص ندارد. از این رو، راهکار برون‌رفت، گذار از مدل تعیین مقصر واحد به مدل توزیع نسبی ضمان است؛ مدلی که در آن تمامی ذی‌نفعان (از طراح تا کاربر) نه بر اساس تقصیر فردی که احراز آن ناممکن است، بلکه بر اساس میزان بهره‌مندی و دخالت، در جبران خسارت سهیم شوند.

نهایتاً در بُعد معرفتی، بررسی پدیدهٔ جعبه سیاه روشن نمود که این مانع، صرفاً یک چالش فنی نیست، بلکه با ایجاد وضعیت غَرَر، مشروعیت اخلاقی به‌کارگیری سیستم را مخدوش می‌کند. نتیجهٔ حاصله آن است که شفافیت دیگر یک فضیلت اخلاقی مستحب نیست، بلکه باید به عنوان پیش‌شرط استناد و مبنای رفع ضمان تلقی گردد. تا زمانی که فرایند تصمیم‌گیری سیستم تفسیرپذیر نشود،

هرگونه اعتماد به آن مصداق اقدام بر امر مجهول بوده و رافع مسئولیت کاربر یا طراح نخواهد بود.

در نهایت، این پژوهش با عبور از تقلیل‌گرایی تک‌بعدی، استدلال می‌کند که برون‌رفت از بن‌بست شکاف مسئولیت، در گرو استقرار یک نظام مسئولیت چندلایه است. چنین نظامی قادر خواهد بود با اتخاذ رویکردی جامع، به ابعاد گوناگون این معضل پاسخ دهد؛ زیرا همان‌گونه که تبیین‌های تک‌بعدی در تحلیل ماهیت شکاف ناتوان بودند، راه‌حل‌های تک ساحتی نیز در مقام عمل کارآمد نخواهند بود.

منابع

حسینی عاملی، محمدجواد. (۱۴۱۹). *مفتاح الکرامة فی شرح قواعد العلامة*. قم: جامعه مدرسین حوزه علمیه قم (موسسه النشر الاسلامی)، چاپ اول.

خویی، سید ابوالقاسم. (۱۴۱۸). *موسوعة الإمام الخوئی*. قم: مؤسسة إحیاء آثار الامام الخوئی، چاپ اول.

دانش، جواد. (۱۳۹۷). *مسئولیت اخلاقی*. قم: پژوهشگاه علوم و فرهنگ اسلامی، چاپ اول.

صاحب جواهر، محمدحسن. (۱۳۶۷). *جواهر الکلام فی شرح شرائع الإسلام*. بیروت: دار إحیاء التراث العربی، چاپ هفتم.

صادقی، محمدهادی؛ میرزایی، محمد. (۱۳۹۸). «ماهیت رابطه استناد و معیار احراز آن». *مطالعات فقه و حقوق اسلامی*، ۱۱(۲۱): ۱۶۷-۱۹۴.

علیدوست، ابوالقاسم؛ بای، حسینعلی. (۱۳۹۱). «معیار ضمان». *فقه اهل بیت*، ۱۸(۷۰): ۴۸-۱۰۵.

فخرالمحققین، محمد بن حسن. (۱۳۸۷). *إیضاح الفوائد فی شرح إشکالات القواعد*. قم: اسماعیلیان، چاپ اول.

قربانیان، صالح. (۱۳۹۹). «عاملیت اخلاقی در مصنوعات هوشمند (ربات‌ها)». *تأملات اخلاقی*، ۱(۱): ۱۱-۳۲.

مصباح یزدی، محمدتقی. (۱۳۹۳). *فلسفه اخلاق*. قم: موسسه آموزشی و پژوهشی امام خمینی (ره)، چاپ اول.

میرزامحمدی، عبدالرئوف؛ تقوی، مصطفی. (۱۴۰۴). «اخلاق فناوری: ارزیابی انتقادی مناقشه میان دیوید مورو و ژوزف سی. پیت دربارهٔ ارزش‌باری فناوری و مسئولیت اخلاقی». *تأملات اخلاقی*، ۶(۲): ۹۳-۱۱۴.

Bandura, A. (1999). "Moral Disengagement in the Perpetration of Inhumanities". *Personality and Social Psychology Review*, 3(3): 193-209.

Bryson, J. J. (2018). "Patiency Is Not a Virtue: The Design of Intelligent Systems and Systems of Ethics". *Ethics and Information Technology*, 20(1): 15–26.

Busuioc, M. (2021). "Accountable Artificial Intelligence: Holding Algorithms to Account". *Public Administration Review*, 81(5): 825–836.

Cañas, J. J. (2022). "AI and Ethics When Human Beings Collaborate with AI Agents". *Frontiers in Psychology*, 13: 836650.

Coeckelbergh, M. (2020). "Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability". *Science and Engineering Ethics*, 26: 2051–2068.

Da Silva, M. (2024). "Responsibility Gaps". *Philosophy Compass*, e70002.

Dastani, M.; Yazdanpanah, V. (2023). "Responsibility of AI Systems". *AI & Society*, 38(3): 859–872.

Elish, M. C. (2019). "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction". *Engaging Science, Technology, and Society*, 5: 40–60.

Floridi, L.; Sanders, J. W. (2004). "On the Morality of Artificial Agents". *Minds and Machines*, 14(3): 349–379.

Fritz, A.; Brandt, W.; Gimpel, H.; Bayer, S. (2020). "Moral Agency Without Responsibility? Analysis of Three Ethical Models of Human-Computer Interaction in Times of Artificial Intelligence (AI)". *Ethics and Information Technology*, 22(1): 35–48.

Gunkel, D. J. (2017). "Mind the Gap: Responsible Robotics and the Problem of Responsibility". *Ethics and Information Technology*, 22(4): 307–320.

Johnson, D. G.; Verdicchio, M. (2018). "AI, Agency and Responsibility: The VW Fraud Case and Beyond". *AI & Society*, 33(4): 517–526.

Kaplan, A.; Haenlein, M. (2019). "Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence". *Business Horizons*, 62(1): 15–25.

Khomkhunsorn, T. (2025). "Meaningful Human Control and Responsibility Gaps in AI: No Culpability Gap, but Accountability and Active Responsibility Gap". *Journal of Integrative and Innovative Humanities*, 5(1): 35–57.

Kiener, M. (2022). "Can We Bridge AI's Responsibility Gap at Will?". *Ethical Theory and Moral Practice*, 25: 575–593.

Königs, P. (2022). "Artificial Intelligence and Responsibility Gaps: What Is the Problem?". *Ethics and Information Technology*, 24(3): 1–11.

Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology*. Viking.

Lancaster, C. M. et-al. (2024). "'It's Everybody's Role to Speak Up... But Not Everyone Will": Understanding AI Professionals' Perceptions of Accountability for AI Bias Mitigation". *ACM Journal on Responsible Computing*, 1(1): 1–30.

LeCun, Y.; Bengio, Y.; Hinton, G. (2015). "Deep Learning". *Nature*, 521(7553): 436–444.

Matthias, A. (2004). "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata". *Ethics and Information Technology*, 6(3): 175–183.

McGraw, D. (2020). "Ethical Responsibility in the Design of Artificial Intelligence (AI) Systems". *The Journal of Philosophy, Science & Law*, 20: 1–18.

Novelli, C.; Taddeo, M.; Floridi, L. (2023). "Accountability in Artificial Intelligence: What It Is and How It Works". *AI & Society*.

Oimann, A.-K.; Tollon, F. (2025). "Responsibility Gaps and Technology: Old Wine in New Bottles?". *Journal of Applied Philosophy*, 42(1): 1–22.

Russell, S. J.; Norvig, P. (2010). *Artificial Intelligence: A Modern Approach* (3rd Ed.). Prentice Hall.

Santoni de Sio, F.; Mecacci, G. (2021). "Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them". *Philosophy & Technology*, 34: 1057–1084.

Sparrow, R. (2007). "Killer Robots". *Journal of Applied Philosophy*, 24(1): 62–79.

Stahl, B. C. (2023). "Embedding Responsibility in Intelligent Systems: From AI Ethics to Responsible AI Ecosystems". *Ethics and Information Technology*, 25(3): 1–17.

Tigard, D. W. (2021). "There Is No Techno-Responsibility Gap". *Philosophy & Technology*, 33(3): 589–607.

Veluwenkamp, H. (2025). "What Responsibility Gaps Are and What They Should Be". *Ethics and Information Technology*, 27(14).

Verbeek, P.-P. (2015). "Beyond Interaction: A Short Introduction to Mediation Theory". *Interactions*, 22(3): 26–31.