Ethical Reflections

Vol. 7, Issue 3, No. 27, Fall 2026, pp. 1-25
Online ISSN: 2717-1159 / Print ISSN: 2676-4810
https://jer.znu.ac.ir

Responsibility Gap in Autonomous Artificial Intelligence Systems: Reframing the Sparrow–Swoboda Debate through Verbeek's Approach

Soghra Salehi,[1] Mostafa Taqavi [2]

Article info

DOI: 10.30470/er.2026.2092018.1545

Original Article

Submission History
Received: 18/06/2026
Revised: 23/07/2026
Accepted: 07/08/2026
Published: 23/09/2026

Authors' Contribution
Soghra Salehi: Study concept and design, data collection, drafting of the manuscript, writing the final version of the manuscript.
Mostafa Taqavi: Providing suggestions and revision comments, reading and approving the final version of the 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 is derived from a dissertation with the following details (Title: Mediated Responsibility in the Age of Artificial Intelligence: A Verbeekian Approach to the Responsibility Gap. Degree: PhD. Student: Soghra Salehi. Supervisor: Mostafa Taqavi. Advisor: Hadi Samadi. University: Islamic Azad University, Science and Research Branch, Tehran, Iran).

AI Usage
Artificial intelligence tools were used in the preparation of this article to the extent of 10% for the purpose of literature review and research background (the percentage and manner of use should be specified.

Abstract

The emergence of autonomous artificial intelligence-based systems has introduced new challenges to the attribution of ethical responsibility, as in these systems ethical consequences do not result from the decision of a single human agent but rather emerge from the interaction among humans, algorithms, data, and technological environments. The concept of the "responsibility gap" represents one of the most significant formulations of this challenge, referring to the difficulty of identifying the responsible agent under conditions of reduced control and predictability in autonomous systems. In this context, the debate between Robert Sparrow and Torben Swoboda reflects two different approaches: Sparrow, emphasizing the lack of sufficient conditions for responsibility attribution, argues for the possibility of a responsibility gap. In contrast, Swoboda, by redefining the criteria of control and predictability, argues for the possibility of preserving human responsibility. Employing conceptual analysis and comparative examination, this study reconsiders the conceptual foundations of the Sparrow–Swoboda debate based on the perspectives of Peter Verbeek. The findings indicate that, despite their different conclusions, Sparrow and Swoboda's disagreement is grounded in shared assumptions concerning the separation of humans and technology, the instrumental view of technology, and a linear model of the relationship between action and outcome. From Verbeek's perspective, these assumptions overlook the role of technology in shaping perception, decision-making, and possibilities for action. Therefore, the responsibility gap cannot be regarded merely as a problem of finding or failing to find a specific human agent; rather, it should be reconsidered within the framework of human–technology relations and the concept of "mediated responsibility." This reframing does not attribute ethical responsibility to technology itself; instead, it emphasizes the necessity of understanding responsibility within a network of human and technological relations.

Keywords: Responsibility Gap; Artificial Intelligence; Autonomous Systems; Technological Mediation; Mediated Responsibility.

1. Ph.D. Student in Modern Philosophy, Department of Philosophy, SR. C., Islamic Azad University, Tehran, Iran. s.salehi761@iau.ir

2. **Corresponding author,** Associate Professor, Department of Philosophy of Science, Sharif University of Technology, Tehran, Iran. M_Taqavi@sharif.edu

Introduction

The expansion of autonomous artificial intelligence-based systems has turned the issue of responsibility attribution into one of the most significant topics in the ethics of technology. In these systems, decisions and outcomes result from the interaction of multiple layers, including design, data, algorithms, operational environments, and human interventions. Therefore, the direct attribution of consequences to a specific human agent faces difficulties that cannot be easily explained within traditional frameworks of responsibility. This situation is referred to in the philosophical and technological ethics literature as the "responsibility gap" (Matthias, 2004, p. 171).

A substantial part of the literature on the responsibility gap is based on two criteria of control and predictability and evaluates the possibility of responsibility attribution according to these criteria. However, as the autonomy of systems increases and their decision-making processes become more complex, the relationship between human agents and final outcomes becomes increasingly ambiguous. Consequently, the question arises as to whether responsibility can still be attributed to human agent(s).

Various perspectives have been proposed in response to this issue. Some theorists, including Sparrow, argue that in certain cases, the reduction of control and predictability leads to situations in which attributing responsibility to any human agent is not fully justified. In contrast, Svoboda maintains that even under conditions of technological complexity, it is possible to defend the attribution of responsibility to human actors by reconsidering the conditions of responsibility attribution. Accordingly, a debate has emerged between, on the one hand, doubts about the possibility of attributing responsibility and, on the other hand, the defense of the continuity of human responsibility.

Beyond these perspectives, Verbeek's approach provides a different conceptual framework for analyzing the relationship between humans, technology, and responsibility. Therefore, this article aims to compare Sparrow's and Svoboda's views in order to clarify the conceptual foundations of this debate and to explore the possibility of reconsidering it based on Verbeek's perspective. The main question of this research is: How can the Sparrow–Svoboda debate on the responsibility gap be reconsidered and reformulated based on Verbeek's approach? Accordingly, this study employs conceptual analysis and a comparative examination of existing arguments to investigate the foundations of this debate and examine the possibility of redefining the problem of responsibility within the framework of human–technology relations.

Sparrow and the Denial of Responsibility in Autonomous Systems

Sparrow argues that responsibility attribution in autonomous systems is impossible due to the loss of control, limited predictability, and absence of moral agency. Since programmers lack control, users cannot predict outcomes, and systems lack moral agency, no qualified agent remains responsible.

Swoboda and the Defense of Possibility of Responsibility in Autonomous Systems

Swoboda maintains that autonomy does not eliminate responsibility, since structural control and non-local predictability provide sufficient grounds for attributing responsibility to human agents.

A Verbeekian Reassessment of the Sparrow–Svoboda Debate

Verbeek transforms the debate by emphasizing technological mediation and hybrid agency. Accordingly, action emerges from human-technology interactions within a network of relations, leading to the concept of "mediated responsibility" shaped by mediations.

Conclusion

This article, focusing on the Sparrow–Swoboda debate concerning the possibility of attributing responsibility in autonomous systems, has shown that this disagreement, despite differences in their conclusions, is grounded in assumptions about the nature of action, the role of technology, and the criteria of responsibility. In both approaches, responsibility is analyzed within a framework in which control and predictability are considered conditions for responsibility attribution, and action is understood through a linear relationship between a human agent and an outcome. This shared framework limits the formulation of the problem, since it reduces the role of technology to instrumental mediation and does not consider its influence on the formation of possibilities of action. From this perspective, the disagreement between the two approaches does not concern their fundamental conceptual foundations, but rather remains at the level of evaluating whether the relationship between human action and outcomes can be preserved or disrupted under conditions of technological autonomy.

Nevertheless, the analysis of the Sparrow–Swoboda debate indicates that their disagreement is not merely about answering an existing problem; rather, the deeper issue concerns the explanatory limits of the traditional model of responsibility attribution when confronted with technologies. Sparrow, by emphasizing the rupture between human action and outcomes, considers the preservation of responsibility under conditions of technological autonomy problematic. In contrast, Swoboda attempts to demonstrate that technological complexity does not necessarily entail the disappearance of human responsibility. Despite this difference, both perspectives formulate the problem within a framework in which responsibility is analyzed based on the possibility of tracing actions to outcomes and identifying the human agent. Therefore, their disagreement concerns less the role of technology in shaping action and more the possibility of maintaining traditional criteria of responsibility attribution under new technological conditions.

In contrast, reconsidering the issue from Verbeek's perspective indicates that ethical action in autonomous systems cannot be reduced to the decision of a single agent, because the process of action

formation takes place within a network of human and technological relations. Within this framework, technology is not merely a tool for realizing intentions; rather, it constitutes part of the structure through which possibilities for action are created.

Accordingly, the issue of responsibility becomes, before being a question of "to whom responsibility should be attributed," a question of "how action is formed." This shift in analysis, rather than providing a final answer to the responsibility gap, suggests that the formulation of this problem within frameworks based on control and predictability requires reconsideration. However, Verbeek's perspective cannot be regarded as a final solution to the responsibility gap; rather, its significance lies in revealing the conceptual limitations within which this problem has been formulated. Verbeek changes the formulation of the problem by questioning the framework of the Sparrow–Swoboda debate. More precisely, what is formulated as the responsibility gap is, at a more fundamental level, the result of a particular understanding of action and technology. By reconsidering this understanding, a redefinition of the problem itself becomes possible. Accordingly, the responsibility gap is not merely a consequence of the absence of a human agent; rather, it indicates the limitations of a model of action that assumes humans and technology as separate domains. In contrast, Verbeek's approach rethinks responsibility at the level of the conditions of possibility of action within human–technology relations and as a mediated phenomenon.

This reconsideration does not amount to presenting a complete theory for determining ethical responsibility in all technological contexts. Verbeek's approach, rather than providing specific criteria for determining the share and boundaries of responsibility among human actors, offers a framework for re-understanding the relationship between action, technology, and responsibility. From this perspective, the concept of "mediated responsibility" can be regarded as an attempt to redefine the level of analysis of the problem of ethical responsibility. This redefinition indicates that the attribution of ethical responsibility in complex intelligent systems cannot be fully explained without considering the conditions under which action is formed within human–technology relations. Consequently, after this reconsideration, the central question is no longer whether human responsibility should be abandoned or attributed to technology; rather, it concerns how ethical responsibility should be formulated within the context of technological mediation.

References

Behkhooy, Mohsen. (1401 SH). *An Explanation and Critique of Verbeek's Theory of the Moralization of Technology*. Master's Thesis. Sharif University of Technology, Department of Philosophy of Science, [2022].

Burrell, Jenna. (2016). "How the Machine Thinks: Understanding Opacity in Machine Learning Algorithms". *Big Data & Society*. 3(1): 1–12. DOI: 10.1177/2053951715622512.

Fischer, John Martin, & Ravizza, Mark. (2000). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press, Second edition.

Floridi, Luciano. (2016). "Faultless Responsibility: On the Nature and Allocation of Moral Responsibility for Distributed Moral Actions". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 374(2083), Article 20160112. DOI: 10.1098/rsta.2016.0112.

Goetze, Theodore Simon. (2022). "Mind the Gap: Autonomous Systems, the Responsibility Gap, and Moral Entanglement". In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. New York: Association for Computing Machinery (ACM).

Jalilian, Mohammad Sobhan. (1396 SH). *Technical Artifacts and Human Beings: A Reflection on Verbeek's Thought*. Master's Thesis. Sharif University of Technology, Department of Philosophy of Science, [2017].

Kiener, Maximilian. (2024). "AI and Responsibility: No Gap, but Abundance". *Journal of Applied Philosophy*. Advance online publication. DOI: 10.1111/japp.12765.

Khalili, Mehdi. (1397 SH). "Can Technological Artifacts Be Good or Bad?". *Siasat-e Elm va Fanavari, Science and Technology Policy*. Fall, Vol. 10, No. 3, pp. 17–28, [2018].

Matthias, Andreas. (2004). "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata". *Ethics and Information Technology*. 6(3): 175–183. DOI: 10.1007/s10676-004-3422-1.

Mirzamohammadi, Abdolraoof; Taqavi, Mostafa. (1404 SH). "Technology Ethics: A Critical Evaluation of the Debate Between David Moreau and Joseph Pitt on the Value-Ladenness of Technology and Moral Responsibility". *Taammolat-e Akhlaghi, Ethical Reflections*. Summer, Vol. 6, No. 2, pp. 93–114, [2025].

Santoni de Sio, Filippo, & Mecacci, Giulio. (2021). "Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them". *Philosophy & Technology*. 34(4): 1057–1084. DOI: 10.1007/s13347-021-00450-x.

Schur, Eva, Brouns, Anna, & Lee, Peter. (2025). "Ethical Analysis of the Responsibility Gap in Artificial Intelligence". *Ethics & Society*. 6(4): 1–10.

Sparrow, Robert. (2007). "Killer Robots". *Journal of Applied Philosophy*. 24(1): 62–77. DOI: 10.1111/j.1468-5930.2007.00346.x.

Sparrow, Robert. (2021). "Why Machines Cannot Be Moral". *AI & Society*. 36(3), 685–693. DOI: 10.1007/s00146-020-01132-6.

Sparrow, Robert, & Flenady, Gene. (2025). "The Testimony Gap: Machines and Reasons". *Minds & Machines*. 35(1): Article 12. DOI: 10.1007/s11023-025-09712-5.

Sparrow, Robert, & Flenady, Gene. (2026). "On Not Being a Robot: AI as a Threat to Rational Agency". *Inquiry: An Interdisciplinary Journal of Philosophy*. 1–20. DOI: 10.1080/0020174X.2026.2673086.

Swoboda, Torben. (2018). "Autonomous Weapon Systems - An Alleged Responsibility Gap". In *Philosophy and Theory of Artificial Intelligence 2017*, edited by Vincent C. Müller. Berlin: Springer Verlag. pp. 302–313. DOI: 10.1007/978-3-319-96448-5_32.

Swoboda, Torben & Lauwaert, Lode. (2025). "Can Artificial Intelligence Embody Moral Values?". *AI and Ethics*. 5(4): 3669–3680. DOI: 10.1007/s43681-025-00662-3.

Verbeek, Peter-Paul. (2005). *What Things Do: Philosophical Reflections on Technology, Agency and Design*. University Park, PA: Pennsylvania State University Press, First edition.

Verbeek, Peter-Paul. (2009). "Ambient Intelligence and Persuasive Technology: The Blurring Boundaries Between Human and Technology". *NanoEthics*. 3(3): 231–242. DOI: 10.1007/s11569-009-0077-8.

Verbeek, Peter-Paul. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: University of Chicago Press, First edition.

Verbeek, Peter-Paul. (2014). "Some Misunderstandings about the Moral Significance of Technology". In *The Moral Status of Technical Artefacts*, edited by Peter Kroes and Peter-Paul Verbeek. Dordrecht: Springer. pp. 75–88. DOI: 10.1007/978-94-007-7914-3.

تأملات اخلاقی

دورهٔ هفتم، شمارهٔ سوم (پیاپی ۲۷)، پاییز ۱۴۰۵، صفحات ۱-۲۵

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

https://jer.znu.ac.ir

شکاف مسئولیت در سامانه‌های هوش مصنوعی خودمختار، بازصورت‌بندی مناقشه اسپارو و اسووبودا با رویکرد فربیک

صغری صالحی،[۱] مصطفی تقوی[۲]

اطلاعات مقاله

DOI:
10.30470/er.2026.2092018.1545

مقاله پژوهشی

تاریخ دریافت: ۱۴۰۵/۰۳/۲۸
تاریخ اصلاح: ۱۴۰۵/۰۵/۰۱
تاریخ پذیرش: ۱۴۰۵/۰۵/۱۶
تاریخ انتشار: ۱۴۰۵/۰۷/۰۱

نقش نویسندگان:

صغری صالحی: ایده پرداز اصلی، جمع‌آوری داده، نگارش اولیه و نهایی مقاله.

مصطفی تقوی: ذکر پیشنهادات و نکات اصلاحی، خوانش و تأیید نهایی مقاله.

حمایت مالی و تعارض منافع

مقاله حامی مالی و تعارض منافع ندارد.

استخراج

این مقاله متخذ از رساله دکتری با این عنوان: «مسئولیت میانجی‌شده در عصر هوش مصنوعی: رویکردی فربیکی به شکاف مسئولیت»، دانشجو: صغری صالحی، استاد راهنما : مصطفی تقوی، استاد مشاور: هادی صمدی در دانشگاه آزاد اسلامی، واحد علوم و تحقیقات تهران است.

استفاده از هوش مصنوعی

در نگارش این مقاله از هوش مصنوعی به میزان ۱۰ درصد برای مرور ادبیات و پیشینه تحقیق استفاده شده است.

چکیده

ظهور سامانه‌های خودمختار مبتنی بر هوش مصنوعی، مسئله انتساب مسئولیت اخلاقی را با چالش‌های نوینی مواجه کرده است؛ زیرا در این سامانه‌ها، پیامدهای اخلاقی حاصل تصمیم یک عامل انسانی واحد نیست، بلکه برآمده از تعامل میان انسان، الگوریتم، داده و محیط فناورانه‌اند. مفهوم «شکاف مسئولیت» یکی از مهم‌ترین صورت‌بندی‌های این چالش است که به دشواری تعیین عامل مسئول در شرایط کاهش کنترل و پیش‌بینی‌پذیری سامانه‌های خودمختار اشاره دارد. در این زمینه، مناقشه میان رابرت اسپارو و توربن اسووبودا نشان‌دهنده دو رویکرد متفاوت است: اسپارو با تأکید بر فقدان شرایط کافی برای انتساب مسئولیت، از امکان شکل‌گیری خلأ مسئولیت دفاع می‌کند. در مقابل، اسووبودا با بازتعریف معیارهای کنترل و پیش‌بینی‌پذیری، امکان حفظ مسئولیت انسانی را مطرح می‌سازد. پژوهش حاضر با روش تحلیل مفهومی و بررسی تطبیقی، مبانی مفهومی مناقشه اسپارو و اسووبودا را براساس دیدگاه‌های پیتر فربیک بازخوانی می‌کند. یافته‌های پژوهش نشان می‌دهد که اختلاف اسپارو و اسووبودا، به‌رغم تفاوت در نتایج، بر مفروضات مشترکی درباره جدایی انسان و فناوری، نقش ابزاری فناوری و الگوی خطی رابطه کنش و پیامد استوار است. از منظر فربیک، این مفروضات نقش فناوری را در شکل‌گیری ادراک، تصمیم و امکان‌های کنش نادیده می‌گیرند. بنابراین، شکاف مسئولیت را نمی‌توان تنها مسئله‌ای درباره یافتن یا نیافتن یک عامل انسانی مشخص دانست، بلکه باید آن را در چارچوب روابط انسان-فناوری و مفهوم «مسئولیت میانجی‌شده» بازاندیشی کرد. این بازصورت‌بندی، بدون نسبت‌دادن مسئولیت اخلاقی به فناوری، بر ضرورت فهم مسئولیت در شبکه‌ای از روابط انسانی و فناورانه تأکید می‌کند.

واژه‌های کلیدی: شکاف مسئولیت، هوش مصنوعی، سامانه‌های خودمختار، میانجی‌گری فناورانه و مسئولیت میانجی‌شده.

۱. دانشجوی دکتری فلسفه عصر جدید، گروه فلسفه، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران. s.salehi761@iau.ir

۲ . **نویسنده مسئول،** دانشیار، گروه فلسفه علم، دانشگاه صنعتی شریف، تهران، ایران. m_Taqavi@sharif.edu

مقدمه

گسترش سامانه‌های خودمختار مبتنی بر هوش مصنوعی، مسئله انتساب مسئولیت را به یکی از مهم‌ترین مباحث اخلاق فناوری تبدیل کرده است. در این سامانه‌ها، تصمیم‌ها و پیامدها حاصل تعامل لایه‌های متعددی از طراحی، داده، الگوریتم، محیط عملیاتی و مداخله انسانی‌اند؛ ازاین‌رو، انتساب مستقیم پیامدها به یک عامل انسانی مشخص با دشواری‌هایی روبه‌رو می‌شود که در چارچوب‌های سنتی مسئولیت به‌آسانی قابل تبیین نیست. این وضعیت در ادبیات فلسفی و اخلاق فناوری با عنوان «شکاف مسئولیت» شناخته می‌شود (Matthias, 2004, p. 171).

بخش عمده‌ای از ادبیات شکاف مسئولیت بر دو معیار کنترل و پیش‌بینی‌پذیری بنا شده است و امکان انتساب مسئولیت را بر اساس این دو می‌سنجند. اما هرچه خودمختاری سامانه افزایش یابد و فرایند تصمیم‌گیری آن پیچیده‌تر شود، پیوند میان عامل انسانی و پیامد نهایی مبهم‌تر می‌شود. در نتیجه، این پرسش مطرح می‌شود که آیا همچنان می‌توان مسئولیت را به عامل یا عاملان انسانی نسبت داد یا خیر؟

در پاسخ به این مسئله، دیدگاه‌های متفاوتی مطرح شده‌اند. برخی نظریه‌پردازان، از جمله رابرت اسپارو[1] استدلال می‌کنند که در برخی موارد، کاهش کنترل و پیش‌بینی‌پذیری به ظهور موقعیت‌هایی می‌انجامد که انتساب مسئولیت به هیچ عامل انسانی به‌طور کامل موجه نیست. در مقابل، توربن اسووبودا[2] بر آن است که حتی در شرایط پیچیدگی فناورانه نیز می‌توان با بازفهم شرایط انتساب، از امکان مسئول دانستن کنش‌گران انسانی دفاع کرد. بدین‌ترتیب، مناقشه‌ای شکل می‌گیرد که در یک سوی آن تردید نسبت به امکان انتساب مسئولیت و در سوی دیگر دفاع از تداوم مسئولیت انسانی قرار دارد. فراتر از این دیدگاه‌ها، رویکرد پیتر فربیک[3] چارچوب مفهومی متفاوتی برای تحلیل نسبت انسان، فناوری و مسئولیت فراهم می‌کند. از این‌رو، هدف مقاله مقایسه دیدگاه‌های اسپارو و اسووبودا برای روشن شدن مبانی مفهومی این مناقشه و امکان بازخوانی آن برمبنای نگرش فربیک است. پرسش اصلی نیز این است که برمبنای رویکرد فربیک چگونه می‌توان مناقشه اسپارو و اسووبودا درباره شکاف مسئولیت را بازخوانی و صورت‌بندی کرد؟ بدین‌ترتیب، پژوهش حاضر با روش تحلیل مفهومی و بررسی تطبیقی استدلال‌های موجود، ضمن واکاوی مبانی این مناقشه، امکان بازتعریف مسئله مسئولیت در چارچوب روابط انسان- فناوری را بررسی می‌کند.

۲. پیشینه پژوهش

پژوهش‌های مرتبط با شکاف مسئولیت در سامانه‌های هوش مصنوعی و فناوری‌های خودمختار را می‌توان در چند محور اصلی دسته‌بندی کرد. در محور نخست، مطالعات فلسفی به تبیین مسئله مسئولیت در مواجهه با فناوری‌های پیچیده پرداخته‌اند. ماتیاس (۲۰۰۴) با طرح مفهوم شکاف مسئولیت نشان می‌دهد که در سامانه‌های خودآموز به‌دلیل کاهش پیش‌بینی‌پذیری و کنترل انسانی،

1. Robert Sparrow
2. Torben Swoboda
3. Peter Verbeek

مفاهیم سنتی مسئولیت دچار چالش می‌شوند. کوکلبرگ (۲۰۲۰) با تأکید بر نسبت میان کنترل، دانش و مسئولیت، چارچوبی رابطه‌ای برای فهم مسئولیت در فناوری‌های هوشمند ارائه می‌کند و نشان می‌دهد مسئولیت تنها به انتساب کنش به عامل انسانی محدود نیست. همچنین گوتزه (۲۰۲۲) استدلال می‌کند که راهکارهای حقوقی و نهادی موجود توان حل کامل مسئله شکاف مسئولیت را ندارند و نیازمند بازاندیشی نظری هستیم.

در محور دوم، پژوهش‌ها بر مدیریت یا کاهش شکاف مسئولیت از طریق سازوکارهای توسعه مسئولانه هوش مصنوعی تمرکز دارند. شور و همکاران (۲۰۲۵) بر هدایت، ردیابی و پاسخ‌گویی کنش‌گران انسانی در فرایند توسعه تأکید می‌کنند. کینر (۲۰۲۴) با مفهوم «فراوانی مسئولیت» تلاش می‌کند بحث را از فقدان مسئول به پراکندگی مسئولیت‌ها معطوف کند. با این‌حال، اغلب این رویکردها همچنان در چارچوب سنتی مسئولیت انسانی باقی می‌مانند.

در محور سوم، مطالعات فلسفه فناوری به نسبت انسان و فناوری پرداخته‌اند. در پژوهش‌های داخلی، جلیلیان (۱۳۹۶) و به‌خوی (۱۴۰۱) به نقد و بررسی نظریه فربیک پرداخته‌اند و خلیلی (۱۳۹۷) بر ارزش‌بار بودن مصنوعات فناورانه و نقش آن‌ها در شکل‌دهی به کنش انسانی تأکید کرده است. همچنین، میرزامحمدی و تقوی (۱۴۰۳) با بررسی مناقشه میان دیوید مورو و جوزف سی. پیت، به ارزیابی دو دیدگاه «ارزش‌بار بودن فناوری» و «خنثی بودن فناوری» پرداخته و نسبت این دو رویکرد را با مسئولیت اخلاقی تحلیل کرده‌اند. با وجود اهمیت این آثار، هیچ‌یک مسئله شکاف مسئولیت در سامانه‌های خودمختار را بررسی نکرده‌اند.

در مجموع، ادبیات موجود یا بر تبیین شرایط شکل‌گیری شکاف مسئولیت و ارائه راهکارهای نهادی متمرکز است یا مسئولیت را در چارچوب سنتی رابطه میان عامل انسانی، قصد، کنترل و پیامد تحلیل می‌کند. خلأ اصلی پژوهش‌ها در فقدان تحلیلی است که نشان دهد آیا شکاف مسئولیت فقط ناشی از ناتوانی انسان در کنترل فناوری است یا می‌توان با بازتعریف رابطه انسان و فناوری، صورت‌بندی متفاوتی از آن ارائه داد. همچنین، بازخوانی مناقشه اسپارو و اسووبودا از منظر نگرش‌های فربیک، مورد توجه قرار نگرفته است. این پژوهش با بهره‌گیری از دیدگاه‌های فربیک، درصدد بازصورت‌بندی این مناقشه و نشان‌دادن نقش فناوری به‌عنوان جزء سازنده کنش و مسئولیت است.

۳. شکاف مسئولیت در سامانه‌های خودمختار

ظهور سامانه‌های هوش مصنوعی خودمختار، مفهوم سنتی مسئولیت اخلاقی را با چالشی بنیادین مواجه کرده است؛ زیرا این سامانه‌ها رابطه کلاسیک میان فاعل، کنش و پیامد را دگرگون می‌کنند. در الگوی سنتی، کنش آگاهانه یک فاعل انسانی به‌طور مستقیم به پیامدها قابل انتساب است، اما در سامانه‌های یادگیری ماشین، تصمیم‌ها حاصل تعامل پیچیده داده‌ها، معماری الگوریتمی، یادگیری و محیط عملیاتی‌اند و به‌سادگی نمی‌توان آن‌ها را به اراده یا قصد یک عامل انسانی واحد تقلیل داد. در این چارچوب، «شکاف مسئولیت» به وضعیتی اشاره دارد که در آن پیامد اخلاقی رخ می‌دهد، اما انتساب آن به فاعل مشخص دشوار می‌شود؛ یعنی گسست میان وقوع رویداد و امکان پاسخ‌گویی. این گسست زمانی تشدید می‌شود که رفتار سامانه نه کاملاً قابل پیش‌بینی باشد و نه تحت کنترل مستقیم

انسان قرار گیرد (Matthias, 2004, pp. 175-181). یکی از عوامل اصلی این وضعیت، افزایش خودمختاری عملیاتی سامانه‌هاست. خودمختاری در این زمینه به‌معنای توانایی پردازش داده‌ها، یادگیری از تجربه و تولید خروجی بدون مداخله مستقیم و لحظه‌ای انسان است. در چنین شرایطی، نقش انسان از تصمیم‌گیرنده مستقیم به طراح و ناظر تغییر می‌یابد و فاصله‌ای میان محل طراحی و تحقق پیامد ایجاد می‌شود (Floridi, 2016, pp. 9-11).

افزایش خودمختاری با کاهش پیش‌بینی‌پذیری نیز همراه است؛ در سامانه‌های پیچیده یادگیری ماشین، امکان پیش‌بینی تصمیم خاص در موقعیتی معین محدود می‌شود و پیوند میان «دانستن»، «توانستن» و «مسئول بودن» تضعیف می‌گردد (Matthias, 2004, pp. 175-181). در کنار آن، مسئله عدم شفافیت الگوریتمی نیز مطرح است؛ به‌ویژه در شبکه‌های عمیق که حتی توسعه‌دهندگان نیز گاه قادر به توضیح دقیق چرایی تصمیم سامانه نیستند (Burrell, 2016, p. 10). این وضعیت «جعبه سیاه» امکان بازسازی زنجیره علّی مسئولیت را دشوار می‌کند. همچنین، این سامانه‌ها در شبکه‌ای از کنش‌گران انسانی و سازمانی عمل می‌کنند؛ از طراحی و تأمین داده تا استقرار و بهره‌برداری، که موجب توزیع عاملیت و شکل‌گیری وضعیت «دست‌های بسیار» می‌شود، به‌طوری‌که هر کنش‌گر تنها بخشی از فرایند را در اختیار دارد (Santoni de Sio & Mecacci, 2021, p. 1063). در نتیجه، انتساب کامل مسئولیت به یک عامل واحد دشوار می‌شود.

بنابراین، شکاف مسئولیت فقط ناشی از خطا یا نقص فنی نیست، بلکه حاصل برهم‌کنش خودمختاری فزاینده، کاهش پیش‌بینی‌پذیری، عدم شفافیت و پراکندگی عاملیت است. در این وضعیت، رابطه سنتی فاعل‌کنش و پیامد دچار اختلال می‌شود و مسئله اصلی نه تنها یافتن مقصر، بلکه بازاندیشی در امکان انتساب مسئولیت در ساختارهای فناورانه پیچیده است. این امر زمینه‌ساز مناقشه‌ای در ادبیات مسئولیت اخلاقی است که در ادامه بررسی می‌شود.

۴. اسپارو و انکار مسئولیت در سامانه‌های خودمختار

یکی از صورت‌بندی‌های بنیادین مسئله مسئولیت در سامانه‌های خودمختار، انکار آن است. رابرت اسپارو از برجسته‌ترین مدافعان این دیدگاه است که به استدلال او در این زمینه پرداخته می‌شود:

۴. ۱. سه‌گانه ناممکن مسئولیت

از نظر اسپارو در سامانه‌های خودمختار با یک «سه‌گانه مسئولیت» مواجه‌ایم؛ یعنی سه گزینه احتمالی برای انتساب مسئولیت وجود دارد، اما هرسه با چالش‌های فلسفی و اخلاقی مواجه‌اند. او چنین فرض می‌کند که سیستم تسلیحاتی خودمختاری، دسته‌ای از سربازان دشمن را تشخیص داده و هدف قرار می‌دهد. در حالی‌که برخی از آن‌ها اعلام کرده‌اند که می‌خواهند تسلیم شوند. برطبق قوانین اخلاق جنگ، سربازانی که قصد تسلیم دارند نباید مورد حمله قرار گیرند. با وجود این، سیستم خودمختار تصمیم می‌گیرد و آنها را می‌کشد. سؤال اسپارو چنین است: چه کسی را باید مسئول این جنایت جنگی دانست؟ او فرضیه‌هایش را در سه سطح سامانه (ربات)، برنامه‌نویسان و کاربران مطرح و استدلال می‌کند که انتساب مسئولیت در این سطح‌ها ممکن نیست:

۴. ۱. ۱. برنامه‌نویسان و شروط سه‌گانه مسئولیت

اسپارو ابتدا الگوی مسئولیت برنامه‌نویسان سیستم تسلیحاتی خودمختار را با شرط‌های کنترل، پیش‌بینی و آگاهی از نقص، نقض می‌کند؛ بدین‌صورت‌که برنامه‌نویسان بر روی اینکه سیستم خودمختار پس از آموزش و استقرار، چگونه تصمیم می‌گیرد، کنترل واقعی ندارند. او چنین می‌گوید: «هرچه یک سیستم خودمختارتر باشد، قابلیلت بیشتری دارد برای اینکه بتواند تصمیماتی را بگیرد که برخلاف آنچه برنامه‌نویسان پیش‌بینی کرده‌اند، باشد» (Sparrow, 2007, p. 67). با از بین رفتن کنترل پیشرفته و مداوم برنامه‌نویسان بر سیستم، چگونه می‌توان مدعی مسئولیت آنها شد؟ علاوه‌براین، سیستم خودمختار به‌ویژه سیستم‌های مبتنی بر یادگیری عمیق، ممکن است رفتاری انجام دهند که برای برنامه‌نویسان به‌طور کامل غیرمنتظره باشد. حتی اگر سازندگان سعی بر پیش‌بینی کنند که چه‌گونه سیستم رفتار خواهد کرد، اما پیچیدگی و خودمختاری سیستم، مانع آن است. به‌بیان دیگر، سیستم‌های مبتنی بر یادگیری عمیق از تجربه می‌آموزند و تصمیم‌های آن‌ها خارج از کنترل مستقیم برنامه‌نویسان اتخاذ می‌شود.

آگاهی از نقص، شرط دیگر مسئولیت‌پذیری است؛ بدین‌معناکه اگر برنامه‌نویسان می‌دانستند که سیستم دارای نقص خاصی است (مانند احتمال حمله به سربازان تسلیم‌شو)، و به‌رغم این آگاهی، سیستم را استقرار داده‌اند، آنگاه می‌توانند مسئول باشند. اما در فرض اسپارو، چنین خطایی را برنامه‌نویسان از قبل نمی‌شناختند و یا نمی‌توانستند شناسایی کنند (همان: ۶۷-۷۰): بنابراین، انتساب مسئولیت به آنها ناعادلانه است.

۴. ۱. ۲. کاربران و از دست رفتن ارتباط علّی

اسپارو در مرحله دوم، امکان مسئول بودن کاربران را در مثال فرماندهان نظامی بررسی می‌کند. او ابتدا فرض می‌گیرد که فرمانده‌ها مسئول باشند؛ زیرا آنها تصمیم گرفته‌اند این سلاح را استقرار دهند. اما استدلالی متوازی برای این گروه نیز ارائه می‌دهد؛ بدین‌صورت‌که فرمانده‌ها نیز نمی‌توانند رفتار سیستم خودمختار را پیش‌بینی کنند. آنها ممکن است حمله به نقطه‌ای خاص را دستور دهند، اما اینکه سیستم چگونه تصمیم می‌گیرد که چه اهدافی را شامل حمله کند، از کنترل فرمانده خارج است. در واقع، ارتباط علّی میان «دستور فرمانده» و «رفتار نهایی سیستم» قطع شده است. به بیان دیگر، نوعی گسست علّی شکل می‌گیرد؛ بدین‌صورت که فرمانده ممکن است هدفی کلی را تعیین کند، اما نحوه انتخاب اهداف جزئی و تصمیم‌گیری عملیاتی در سطحی رخ می‌دهد که از کنترل و پیش‌بینی او خارج است. بنابراین، حتی اگر بتوان مسئولیت را در سطح تصمیم به استفاده از سیستم مطرح کرد، نمی‌توان آن را به پیامدهای خاص و پیش‌بینی‌ناپذیر عملکرد سیستم تعمیم داد. در نتیجه، ارتباط علّی کافی برای انتساب مسئولیت به کاربران(فرماندهان) نیز برقرار نمی‌شود. اسپارو این نکته را چنین بیان می‌کند: «ارتباط علّی میان تصمیم سیستم خودمختار و برنامه‌نویسان تنها به دلیل خودمختاری سیستم از بین رفته است. از آنجا که سیستم تصمیماتی می‌گیرد که برنامه‌نویسان نه پیش‌بینی کرده‌اند و نه کنترل می‌کنند، نمی‌توان ارتباط علّی بین آنها برقرار دانست» (Sparrow, 2007, pp. 70-71).

۴. ۱. ۳. سیستم خودمختار و مسئله رنج و معناداری

اسپارو در مرحله سوم، احتمال انتساب مسئولیت به خود سامانه خودمختار را بازگو می‌کند. این مرحله مستلزم آن است که سامانه واجد شرایطی مانند شایستگی مجازات و توانایی برای تنبیه یا پاداش باشد. همچنین بتواند احساس پشیمانی کند (Sparrow & Flenady, 2025, pp. 4-5). اما از نظر او، تحقق این شرط‌ها درباره سامانه‌های هوش مصنوعی خودمختار ممکن نیست؛ زیرا لازمه تنبیه اخلاقی آن است که موجود مورد نظر بتواند به‌نحوی معنادار رنج را تجربه کند. اسپارو در این‌باره چنین می‌گوید: «براساس یک نظریه معقول از تنبیه، اگر فرد الف تنبیه شود، آنگاه این تنبیه باید باعث رنج او شود» (Sparrow, 2007, p. 72). او تاکید می‌کند این شرط برای ماشین‌های خودمختار بی‌معناست؛ حتی اگر یک سامانه خاموش یا تخریب شود، نمی‌توان گفت که رنج برده است یا تجربه اخلاقی معناداری داشته است. در واقع، اسپارو به‌لحاظ فلسفی و روانی، اعمال تنبیه یا پاداش نسبت به یک سیستم هوشمند و از طریق آسیب فیزیکی یا ایجاد اختلال نرم‌افزاری را با شهود انسانی ناسازگار می‌داند (Sparrow, 2007, pp. 71-73). او این‌گونه توضیح می‌دهد: «اگر یکی از خانواده‌های قربانیان از ما بپرسد آیا ماشین خودمختار برای این جنایت به اندازه کافی تنبیه شد؟ چگونه می‌توان پاسخ داد: بله، ما روغن‌کاری‌اش را متوقف کردیم و اکنون رنج می‌برد. گرچه این پاسخ خنده‌آور است، اما مسئله‌ای عمیق را برملا می‌کند»(همان). بنابراین، سامانه خودمختار نیز نمی‌تواند، عامل مسئول باشد.

٤. ۱. ٤. نتیجه استدلال: انکار مسئولیت

براساس سه گزینه پیش‌فرضی فوق و دلالت‌های ارائه شده، اسپارو بدین نتیجه می‌رسد که در موقعیت سامانه‌های تسلیحاتی خودمختار، این گزینه‌ها برای انتساب مسئولیت، قابل دفاع نیستند. بنابراین، هیچ عامل مشخصی برای انتساب مسئولیت باقی نمی‌ماند و نوعی «خلأ مسئولیت» پدید می‌آید. اسپارو این وضعیت را تنها یک مسئله نظری نمی‌داند، بلکه آن را یک بحران اخلاقی تلقی می‌کند؛ زیرا اگر هیچ‌کس مسئول نباشد، بنیان‌های پاسخ‌گویی اخلاقی در استفاده از چنین سامانه‌هایی تضعیف می‌شود. از این‌رو، او در غیاب امکان انتساب مسئولیت نتیجه‌ای رادیکال را مطرح می‌کند که اگر کسی مسئول نخواهد بود، و اگر نمی‌توان کسی را برای تخلف‌های ممکن احتمالی مسئول دانست، بهتر است چنین سیستم‌هایی ساخته نشود (Sparrow, 2007, p. 75). در بیان دیگر چنین می‌گوید: « در جایی‌که پیامدها از حیث اخلاقی اهمیت بالایی داشته باشند، این امر می‌تواند دلیلی قوی برای عدم به‌کارگیری هوش مصنوعی باشد» (Sparrow, 2021, p. 692).

استدلال اسپارو را می‌توان چنین صورت‌بندی کرد:

- برای انتساب مسئولیت، باید میان عامل و پیامد رابطه‌ای مبتنی بر کنترل، پیش‌بینی‌پذیری و عاملیت اخلاقی برقرار باشد.
- در سامانه‌های خودمختار، برنامه‌نویسان کنترل کافی ندارند؛ کاربران قادر به پیش‌بینی پیامدهای خاص نیستند؛ سامانه نیز فاقد عاملیت اخلاقی است؛ زیرا رنج نمی‌برد.
- بنابراین، هیچ عامل واجد شرایط مسئولیت وجود ندارد (Sparrow, 2007, pp. 62-75).

نتیجه: در برخی سامانه‌های خودمختار، شکاف مسئولیت به‌طور ساختاری پدید می‌آید.

۵. اسووبودا و دفاع از امکان مسئولیت در سامانه‌های خودمختار

در برابر استدلال‌های انکاری اسپارو، دیدگاهی قرار دارد که می‌کوشد امکان انتساب مسئولیت به عامل انسانی را حفظ کند. در این رویکرد، مسئله اصلی نه نفی مسئولیت، بلکه بازتفسیر شرایط آن در بستر فناوری‌های پیچیده است. یکی از صورت‌بندی‌های مهم این موضع در آثار اسووبودا ارائه شده است. از منظر او، مسئولیت اخلاقی همواره در شرایطی اعمال می‌شود که درجاتی از عدم قطعیت، محدودیت معرفتی و فقدان کنترل کامل وجود دارد. بنابراین، اگر کنترل و پیش‌بینی‌پذیری کامل را شرط لازم مسئولیت بدانیم، دامنه مسئولیت اخلاقی به‌طور غیرواقع‌بینانه‌ای محدود خواهد شد. در حالی‌که در بسیاری از حوزه‌های زندگی اجتماعی، افراد مسئول پیامدهایی شناخته می‌شوند که نه به‌طور کامل تحت کنترل آن‌ها بوده‌اند و نه تمامی نتایج آن‌ها از پیش قابل پیش‌بینی بوده است. براین اساس، اسووبودا برای اثبات مسئولیت، میان «فقدان کامل کنترل» و «کنترل کافی برای انتساب مسئولیت» و همچنین پیش‌بینی محلی و غیرمحلی تمایز قائل می‌شود (Svoboda, 2018,p: 309-311).

۵. ۱. تمایز میان کنترل لحظه‌ای و کنترل ساختاری

نخستین گام در استدلال اسووبودا بازتعریف مفهوم کنترل است. او می‌پذیرد که طراح یا برنامه‌نویس بر همه خروجی‌های یک سامانه یادگیرنده کنترل مستقیم ندارد، اما کنترل‌ناپذیری را برای نفی مسئولیت کافی نمی‌داند. از نظر اسووبودا، کنترل اخلاقی لازم نیست به‌معنای کنترل لحظه‌ای بر کنش باشد، بلکه می‌تواند در سطحی ساختاری و طراحی‌شده اعمال شود. این سطح از کنترل شامل تصمیم‌هایی است که پیش از استقرار سامانه اتخاذ می‌شوند؛ از جمله آنها می‌توان به تعیین معماری یادگیری سامانه، تعیین محدوده و نحوه تداوم یا توقف فرایند یادگیریو امکان منجمدسازی مدل در مرحله‌ای مشخص از آموزش اشاره کرد. بنابراین، اگر طراح بتواند ساختار یادگیری را طراحی و محدود کند، آنگاه نوعی «کنترل راهنمایی‌کننده» بر رفتار کلی سامانه برقرار است. این سطح از کنترل برای انتساب مسئولیت کافی است، حتی اگر کنترل مستقیم بر رفتارهای جزئی وجود نداشته باشد (Svoboda, 2018, p. 309).

به‌بیان دیگر، اسووبودا میان کنترل بر رفتار آنی سیستم و کنترل بر چرخه حیات یادگیری تمایز قائل می‌شود. از نظر او، برنامه‌نویسان نمی‌توانند در هر لحظه مشخص کنند که شبکه عصبی به‌طور دقیق چه خروجی خواهد داد، اما آنها می‌توانند تعیین کنند که سیستم تا کجا و تا چه زمانی یادگیری را ادامه دهد؛ همچنین تصمیم بگیرند که آیا یادگیری در حین عمل مجاز باشد یا نباشد؛ علاوه‌براین، تصمیم بگیرند که فرایند آموزش در چه مرحله‌ای متوقف و مدل منجمد (ثابت) شود (همان). در واقع، طراح می‌تواند سامانه را در دو مرحله شکل دهد: در مرحله اول، شبکه براساس داده‌ها و هدف‌هایی که او تعیین کرده آموزش می‌بیند. در مرحله دوم، یادگیری متوقف می‌شود و سامانه تنها در وضعیت «استفاده» عمل می‌کند. اگر طراح، امکان یادگیری خودکار در حین استقرار را حذف یا به‌شدت محدود کند، آنگاه بخش بزرگی از آن چیزی که اسپارو «خودمختاری غیرقابل‌پیش‌بینی» می‌نامد، مهار می‌شود. بنابراین، این‌گونه نیست که طراح هیچ کنترل معناداری بر سیستم نداشته باشد؛ او بر طراحی، معماری یادگیری،داده‌های آموزشی و توقف یا استمرار

یادگیری کنترل دارد. برداشت اسووبودا از مفهوم کنترل، پیوند نزدیکی دارد با آنچه فیشر و راوِیزا (۲۰۰۰: ۱۸) «ضرورت قدرت» می‌نامند؛ بدین‌معنا که شخص دارای کنترل است اگر و تنها اگر، در صورتی‌که گزاره‌ای معین تحقق یابد، آن شخص قدرتِ آن را داشته باشد که به‌گونه‌ای عمل کند که آن گزاره تحقق نیابد.

۵. ۲. تمایز میان پیش‌بینی محلی و غیرمحلی

گام دوم استدلال اسووبودا بازتفسیر مفهوم پیش‌بینی‌پذیری است. در دیدگاه او، خطای استدلال مخالفان در این‌است که پیش‌بینی‌پذیری را به سطح رفتارهای جزئی و لحظه‌ای تقلیل می‌دهند، اما او میان این دو سطح تمایز می‌گذارد؛ در رفتار محلی توجه به تصمیم‌های واحد و لحظه‌ای سامانه معطوف است ولی در رفتار غیرمحلی(ساختاری) الگوی کلی عملکرد سامانه در شرایط مختلف مدنظر است. ادعای اسووبودا این‌است که اگرچه رفتارهای محلی ممکن است غیرقابل پیش‌بینی باشند، اما رفتار غیرمحلی سامانه همچنان تا حد معقولی قابل پیش‌بینی است. او برای توضیح، از مثال شطرنج استفاده می‌کند: حرکت خاص فیل از خانه e۲ به b۵ نمونه‌ای از رفتار محلی است؛ اما «سبک بازی»، «راهبرد بُردن» و «الگوی کلی بازی» یک شطرنج‌باز، رفتار غیرمحلی اوست. به‌همین ترتیب، این‌که سامانه تسلیحاتی خودمختار به‌طور دقیق و در لحظه‌ای خاص روی کدام فرد یا شیء قفل و شلیک کند، یک رفتار محلی است، اما ساختار کلی تصمیم‌گیری سامانه؛ یعنی این‌که در چه شرایطی و به چه اهدافی حمله می‌کند یا خودداری می‌کند، مثالی از رفتار غیرمحلی است (Svoboda, 2018, pp. 310-311).

به‌بیان دیگر، می‌توان پیش‌بینی کرد که یک سامانه در مجموع چه نوع الگوهایی از طبقه‌بندی، هدف‌گیری یا خطا را ایجاد خواهد کرد، حتی اگر نتوان تعیین کرد در هر لحظه خاص چه تصمیمی خواهد گرفت. بنابراین، شرط پیش‌بینی‌پذیری برای مسئولیت نه در سطح جزئیات، بلکه در سطح الگوهای کلی رفتار باید ارزیابی شود. برمبنای این تمایز، اسووبودا نتیجه می‌گیرد که پیش‌بینی غیرمحلی برای انتساب مسئولیت کافی است؛ یعنی اگر طراح بداند که سامانه به‌طور کلی در چه شرایطی مستعد تولید چه نوع خطاهایی است، آنگاه نمی‌تواند نسبت به پیامدهای آن بی‌تفاوت باشد. در این‌معنا، ندانستن جزئیات دقیق هر کنش، دلیل موجهی برای نفی مسئولیت نیست؛ زیرا آنچه مهم‌تر است، نه پیش‌بینی تک‌رخدادها بلکه بررسی این است که طراحی سیستم چه نوع کنش‌ها و خطاهایی را ممکن می‌سازد. او برای تبیین این ادعا، از یک تمثیل استفاده می‌کند؛ بدین‌صورت که کارگری را در نظر بگیریم که از ساختمانی نیمه‌کاره مصالحی را به پایین پرتاب می‌کند و به‌طور تصادفی به عابری آسیب می‌زند. کارگر ممکن است ادعا کند که نمی‌دانسته کدام فرد در چه لحظه‌ای در محل حضور دارد. اما این ندانستن جزئی، مسئولیت او را منتفی نمی‌کند؛ زیرا کارگر می‌توانسته و باید می‌دانسته است که چنین رفتاری به‌طور کلی خطر آسیب به عابران را ایجاد می‌کند. بنابراین، مسئولیت نه بر مبنای پیش‌بینی فرد خاص، بلکه براساس آگاهی از الگوی کلی خطر تعیین می‌شود. در نتیجه، همان‌طورکه در این مثال پیش‌بینی جزئی شرط مسئولیت نیست، در سامانه‌های خودمختار نیز عدم امکان پیش‌بینی لحظه‌ای نمی‌تواند به‌تنهایی دلیل نفی مسئولیت باشد (Svoboda, 2018, p. 305).

براین اساس، اسووبودا نتیجه می‌گیرد که طراحان و برنامه‌نویسان همچنان می‌توانند مسئول پیامدهای سامانه‌های خودمختار

باشند، مشروط بر اینکه در طراحی ساختار سامانه نقش داشته باشند، نسبت به الگوهای کلی رفتار سامانه آگاه باشند و آگاهانه سامانه‌ای را مستقر کنند که دارای ظرفیت ایجاد آسیب است. بنابراین، برخلاف استدلال انکار مسئولیت، گسست میان طراح و پیامد به‌طور کامل برقرار نمی‌شود، بلکه مسئولیت در سطحی غیرمحلی (ساختاری) همچنان قابل دفاع باقی می‌ماند.

استدلال اسووبودا را می‌توان چنین صورت‌بندی کرد: اگر مسئولیت اخلاقی مستلزم کنترل و پیش‌بینی‌پذیری مطلق نباشد و اگر کنش‌گران انسانی همچنان در ایجاد شرایط امکان وقوع پیامدها نقش داشته باشند، آنگاه پیچیدگی و خودمختاری سامانه‌ها به‌تنهایی برای اثبات شکاف مسئولیت کافی نخواهد بود. بنابراین، در سامانه‌های خودمختار نیز می‌توان از تداوم مسئولیت انسانی دفاع کرد.

٦. وجوه اختلاف و اشتراک اسپارو و اسووبودا

چنانچه بررسی شد مناقشه میان اسپارو و اسووبودا دو نتیجه متضاد (انکار مسئولیت و دفاع از آن) را دربرداشت. اما این اختلاف، بر بستری از مفروضات مشترک شکل گرفته است که مورد بازاندیشی قرار نگرفته‌اند. به‌عبارت دیگر، اختلاف آن‌ها در نحوه تفسیر و اعمال اصول بنیادین است.

٦. ١. اختلاف در امکان انتساب مسئولیت

اسپارو و اسووبودا مسئله را از ظهور سامانه‌هایی آغاز می‌کنند که سطح کنترل مستقیم انسان بر فرایندهای تصمیم‌گیری را کاهش داده و پیش‌بینی پیامدهای آن‌ها را دشوار ساخته‌اند، اما از این وضعیت نتایج متفاوتی را مطرح می‌کنند. از نظر اسپارو، کاهش معنادار کنترل و پیش‌بینی‌پذیری، شرایط لازم برای انتساب مسئولیت را تضعیف می‌کند. ازاین‌رو، در برخی موارد ممکن است با وضعیتی مواجه شویم که در آن وقوع یک آسیب یا خطا روشن است، اما هیچ عامل انسانی واجد شرایط لازم برای پذیرش کامل مسئولیت نباشد. براین اساس، شکاف مسئولیت یک امکان واقعی و گاه اجتناب‌ناپذیر است. در مقابل، اسووبودا معتقد است که از کاهش کنترل یا افزایش پیچیدگی و پیش‌بینی‌ناپذیری نمی‌توان به عدم امکان انتساب مسئولیت رسید. حتی اگر سامانه‌های خودمختار کنترل مستقیم انسان را کاهش دهند، همچنان می‌توان مسئولیت را به افرادی نسبت داد که در طراحی، استقرار یا بهره‌برداری از آن‌ها نقش داشته‌اند. بنابراین، هسته اصلی مناقشه میان این دو رویکرد، فقدان مسئولیت و شرایط آن نیست، بلکه دشوارتر شدن فرایند انتساب و سطح بسندگی شرایط مسئولیت است که باید در چه سطحی تعریف شود.

٦. ٢. اشتراک در صورت‌بندی مسئله و مفروضات مفهومی

اسپارو و اسووبودا مسئله انتساب مسئولیت را در قالب یک پرسش مشترک صورت‌بندی می‌کنند. به‌بیان دیگر، درباره نحوه طرح مسئله توافق دارند؛ بدین‌صورت که بحث هر دو از این فرض آغاز می‌شود که انتساب مسئولیت تابع رابطه‌ای میان عامل انسانی و پیامد است. یعنی هر دو مسئله را در سطح یکسانی یعنی سطح شرایط انتساب مسئولیت به عاملان انسانی مطرح می‌کنند. بنابراین، مناقشه آن‌ها درون یک چارچوب مفهومی مشترک قرار دارد. علاوه‌براین، صورت‌بندی مشترک اسپارو و اسووبودا بر مسئله انتساب مسئولیت

بر مجموعه‌ای از مفروضات مفهومی استوار است.

نخستین فرض، جایگاه محوری کنترل و پیش‌بینی‌پذیری است؛ یعنی هر دو دیدگاه مسئولیت را به‌طور بنیادی وابسته به درجاتی از کنترل و امکان پیش‌بینی پیامدها می‌دانند. بنابراین، اختلاف آن‌ها در نقش محوری این مولفه‌ها نیست بلکه در تفسیر سطح مناسب وکفایت آنهاست؛ به‌طوری‌که از نظر اسپارو، شکاف مسئولیت نتیجه تضعیف این دو مؤلفه است ولی در دیدگاه اسووبودا مسئولیت حتی در صورت کاهش آن‌ها نیز قابل حفظ است.

دومین فرض، محوریت انسان در ساختار مسئولیت است؛ اسپارو هنگامی از شکاف مسئولیت سخن می‌گوید که نتوان مسئولیت را به عاملان انسانی نسبت داد (Sparrow, 2025, p. 1). اسووبودا نیز راه‌حل مسئله را در یافتن مبنایی برای حفظ مسئولیت انسانی جست‌وجو می‌کند (Svoboda, 2018). بنابراین، در هر دو رویکرد مسئولیت باید به انسان بازگردد و مسئله اصلی آن‌ها نیز این نیست که فناوری چه جایگاهی در ساختار مسئولیت دارد.

سومین فرض، تلقی ابزاری از فناوری است. در هر دو رویکرد، فناوری به‌عنوان واسطه‌ای برای تحقق اهداف انسانی در نظر گرفته می‌شود، حتی اگر این واسطه‌گری پیچیده و غیرشفاف باشد. فناوری در این چارچوب، نقش تعیین‌کننده در ساختار کنش ندارد، بلکه بیشتر به‌عنوان متغیری درون یک الگوی از پیش مفروض عمل می‌کند. به‌بیان دیگر، این مفروضه، به نحوه فهم فناوری مربوط می‌شود؛ در هر دو دیدگاه، فناوری در مقام ابزاری پیچیده ظاهر می‌شود که میان عامل انسانی و پیامد قرار گرفته است (Sparrow, 2026; Svoboda, 2025).

چهارمین فرض، پذیرش ضمنی نوعی الگوی خطی از رابطه کنش و پیامد است. در هر دو رویکرد، کنش در قالب رابطه‌ای خطی میان یک عامل(انسان یا سامانه) و یک پیامد مشخص فهم می‌شود. حتی در مواردی که پیچیدگی سامانه پذیرفته می‌شود، این پیچیدگی در قالب زنجیره‌ای از علل قابل تحلیل فرض می‌گردد. هردو می‌کوشند از طریق ردیابی زنجیره‌ای از کنش‌ها، مسئولیت را به عامل یا عاملان انسانی نسبت دهند. اختلاف در این است که اسپارو این زنجیره را در مواردی گسسته می‌داند و اسووبودا آن را همچنان قابل ردیابی تلقی می‌کند. اما در هر دو وضعیت، ساختار کلی تحلیل مبتنی بر جست‌وجوی منشأ انسانی پیامدهاست.

در مجموع، این مفروضات نشان می‌دهد که مناقشه میان اسپارو و اسووبودا، در سطحی بنیادی‌تر درون یک صورت‌بندی مشترک از مسئله قرار دارد. به‌همین دلیل، اختلاف آن‌ها به‌جای آنکه به بازتعریف مسئله منجر شود، درون همان چارچوب باقی می‌ماند. این وضعیت زمینه را برای بازاندیشی در سطحی پیشینی‌تر فراهم می‌کند؛ سطحی که در آن نه پاسخ به شکاف مسئولیت، بلکه خود صورت‌بندی مسئله مورد پرسش قرار می‌گیرد. به‌بیان دیگر، تا زمانی‌که مسئله تنها در قالب انتساب مسئولیت به کنش‌گران انسانی فهم شود، اختلاف میان این دو ناگزیر در محدوده همان چارچوب باقی خواهد ماند. ازاین‌رو، برای ارزیابی حدود و امکان‌های این مناقشه، باید به رویکردهایی توجه کرد که مسئله مسئولیت را نه از نقطه پایانِ زنجیره انتساب، بلکه از نحوه شکل‌گیری رابطه میان انسان، فناوری و کنش مورد بررسی قرار می‌دهند. رویکرد پیتر فربیک را می‌توان یکی از مهم‌ترین تلاش‌ها در این جهت دانست.

۷. بازخوانی مناقشه از منظر فربیک

بررسی مناقشه میان اسپارو و اسووبودا نشان داد که هر دو دیدگاه، به‌رغم اختلاف در نتیجه‌گیری، بر مجموعه‌ای از مفروضات مشترک درباره ماهیت کنش و شرایط مسئولیت‌پذیری استوارند. در این چارچوب،کنش به‌مثابه رابطه‌ای میان یک عامل انسانی و پیامد قابل‌ردیابی آن، فهم می‌شود و فناوری فقط به‌عنوان واسطه‌ای در تحقق این رابطه محسوب می‌شود. اما اگر نقش فناوری در شکل‌دهی به امکان‌های ادراک و کنش در نظر گرفته شود، آنگاه صورت‌بندی مسئله تغییر می‌کند؛ یعنی پیش از آنکه پرسش از «چه کسی مسئول است» مطرح شود، باید پرسید «کنش چگونه و در چه شبکه‌ای از روابط انسانی- فناورانه ممکن شده است». رویکرد فربیک را می‌توان تلاشی برای این بازاندیشی دانست. اهمیت نظری دیدگاه او نیز در ارائه پاسخی جدید به این پرسش نیست، بلکه در به چالش کشیدن مفروضاتی است که خودِ پرسش بر آن‌ها بنا شده است.

۷. ۱. نقد دوگانه انسان/فناوری

برمبنای رویکرد فربیک، نخستین پیش‌فرض مناقشه اسپارو و اسووبودا در این است که انسان و فناوری به‌عنوان دو قلمرو مستقل از یکدیگر در نظر گرفته می‌شود. اما این جداسازی، تصویری ساده‌شده از رابطه انسان و فناوری ارائه می‌دهد؛ زیرا فناوری‌ها تنها ابزارهایی بیرونی در اختیار انسان نیستند و موجودیت‌هایی جدا از کنش انسانی نیز نمی‌باشند. در واقع، آنچه در موقعیت‌های فناورانه به‌عنوان کنش انسانی شناخته می‌شود، اغلب پیش از تحقق تصمیم نهایی، در فرایند تعامل انسان با فناوری شکل می‌گیرد. فناوری‌ها در این سطح فقط امکان اجرای کنش را فراهم نمی‌کنند، بلکه در نحوه ادراک موقعیت، تعریف مسئله و شکل‌گیری انتخاب نیز مداخله دارند (Verbeek, 2005, pp. 10-11; Verbeek, 2011, pp. 152-155).

بنابراین، اگر فناوری در ساختار ادراک و تصمیم دخالت داشته باشد، آنگاه نمی‌توان کنش را محصول اراده یک فاعل مستقل دانست که ابتدا تصمیم می‌گیرد و سپس فناوری را برای تحقق آن به‌کار می‌گیرد. در این صورت، فناوری بخشی از زمینه‌ای است که در آن خودِ فاعل، امکان‌های کنش و ساحت‌های تصمیم شکل می‌گیرند. نتیجه این استدلال آن است که دوگانه انسان/فناوری، به‌عنوان پیش فرض انتساب مسئولیت، نیازمند بازاندیشی است. این بازاندیشی به‌معنای انتقال مسئولیت از انسان به فناوری نیست؛ زیرا فربیک فناوری را همچون یک فاعل اخلاقی مستقل در نظر نمی‌گیرد. از نظر او، مسئولیت در موقعیت‌های فناورانه باید در بستری فهم شود که در آن، کنش انسانی از پیش با میانجی‌گری‌های فناورانه درهم‌تنیده است (Verbeek, 2011, p. 53).

از این منظر، نقد فربیک متوجه چارچوبی است که رابطه انسان و فناوری را رابطه‌ای بیرونی تصور می‌کند. اما او با عبور از دو تلقی کلاسیک فناوری - یعنی ابزارانگاری که فناوری را وابسته به اراده انسان می‌داند و ذات‌انگاری که آن را نیرویی مستقل و تعیین‌کننده تصور می‌کند - نشان می‌دهد که هر دو دیدگاه در یک پیش‌فرض مشترک یعنی جدایی اولیه انسان و فناوری سهیم‌اند (Verbeek, 2005, pp. 10-11). در حالی‌که فناوری‌ها در تجربه واقعی انسان، بخشی از شبکه‌ای هستند که در آن ادراک، تصمیم و کنش شکل می‌گیرد. به‌همین دلیل، مثال «انسان مسلح» در تحلیل فربیک اهمیت ویژه‌ای دارد. در این مثال، نه می‌توان اسلحه را ابزار منفعلی

دانست که تنها اراده انسانی را منتقل می‌کند و نه می‌توان آن را عامل مستقلی دانست که جدا از انسان عمل می‌کند. در واقع، آنچه رخ می‌دهد، شکل‌گیری نوعی کنش ترکیبی است که در آن ویژگی‌های فناوری و توانایی‌های انسانی در یک رابطه متقابل قرار می‌گیرند (Verbeek, 2005, pp. 154-155).

بنابراین، بازخوانی فربیکی از مناقشه مسئولیت در سامانه‌های خودمختار، صورت مسئله را از سطح انتساب به یک عامل مستقل به شبکه‌ای از روابط تغییر می‌دهد که در آن انسان و فناوری درهم‌تنیده‌اند. به بیان دیگر، مسئله فقط این نیست که مسئولیت در پایان زنجیره کنش به کدام انسان نسبت داده شود، بلکه باید بررسی کرد که این زنجیره چگونه از ابتدا در تعامل انسان و فناوری شکل گرفته است.

۷. ۲. میانجی‌گری فناورانه

دومین پیش فرض مناقشه اسپارو و اسووبودا مربوط به نگرش پسینی درباره فناوری است؛ بدین‌معنا که ابتدا یک کنش انسانی مفروض گرفته می‌شود و سپس بررسی می‌گردد که آیا پیامد حاصل را می‌توان به عامل انسانی نسبت داد یا نقش فناوری موجب گسست این زنجیره شده است. اما طبق رویکرد میانجی‌گرایانه فربیک، این پیش‌فرض نادرست است؛ زیرا از منظر او، فناوری تنها حلقه‌ای در زنجیره علّی میان انسان و پیامد نیست، بلکه عنصری است که پیش از تحقق کنش، در شکل‌گیری تجربه، ادراک موقعیت و امکان‌های تصمیم دخالت دارد. به‌بیان دیگر، فناوری تنها پس از آنکه انسان تصمیمی می‌گیرد وارد فرایند عمل نمی‌شود، بلکه در تعیین اینکه انسان چه چیزی را می‌بیند، چگونه موقعیت را تفسیر می‌کند و چه گزینه‌هایی را ممکن یا ناممکن می‌سازد، مشارکت دارد (Verbeek, 2011, pp. 14-15; Verbeek, 2005, pp. 159-160).

براین اساس، مفهوم میانجی‌گری فناورانه، فهم ابزارمحور فناوری را به فهم رابطه‌ای آن بدل می‌کند. در تلقی ابزاری، فناوری وسیله‌ای خنثی برای تحقق اهداف از پیش تعیین‌شده انسانی است؛ بنابراین منشأ اصلی کنش، همچنان در فاعل انسانی جست‌وجو می‌شود. اما در دید فربیک، این نوع تلقی، رابطه واقعی انسان و فناوری را نادیده می‌گیرد؛ زیرا فناوری‌ها نه فقط امکان اجرای یک کنش، بلکه شرایط ظهور همان کنش را نیز جهت می‌دهد. به‌عبارت دیگر، انسان همواره جهان را در وضعیتی میانجی‌گری‌شده تجربه می‌کند و این میانجی‌گری بخشی از نحوه شکل‌گیری تصمیم و کنش است (Verbeek, 2011, p. 97).

نتیجه ادعای فربیک این‌است که رابطه میان انسان، فناوری و پیامد را نمی‌توان همچون یک زنجیره خطی «تصمیم انسانی ← ابزار فناورانه ← نتیجه» فهم کرد؛ زیرا فناوری بخشی از فرایندی است که در آن خودِ تصمیم و ساحت‌های کنش شکل می‌گیرند. از این منظر، فناوری نه جایگزین عامل انسانی و نه مانعی در مسیر انتساب مسئولیت است؛ بلکه عنصری است که ساختار کنش را سامان‌دهی می‌کند. فربیک برای توضیح این نقش، میانجی‌گری فناورانه را در دو سطح ادراکی و کنشی مطرح می‌کند. در سطح ادراکی، فناوری‌ها فقط اطلاعات را منتقل نمی‌کنند، بلکه شیوه مواجهه انسان با جهان را شکل می‌دهند. آنچه انسان به‌عنوان واقعیت، داده یا موقعیت تصمیم‌گیری تجربه می‌کند، همواره از خلال سازوکارهای فناورانه قابل دسترس است. بنابراین، ادراک انسانی یک فرایند کاملاً مستقل و پیشافناورانه نیست؛ چراکه در شبکه‌ای از روابط میان انسان و مصنوعات شکل می‌گیرد (Verbeek, 2011, pp. 14-15).

در سطح کنشی نیز فناوری‌ها در ساختار امکان‌های عمل مشارکت دارند. این نقش به‌معنای تعیین جبری رفتار انسانی نیست؛ بلکه بدین‌معناست که فناوری‌ها ساحتی از امکان‌ها ایجاد می‌کنند که در آن برخی کنش‌ها تسهیل، برخی محدود و برخی دیگر دشوار می‌شوند. برای مثال، سرعت‌گیرها یا سامانه‌های کنترل ترافیک فقط ابزارهایی برای اجرای مقررات نیستند، بلکه با طراحی محیط، شیوه خاصی از رفتار را در کاربران شکل می‌دهند (Verbeek, 2009, p. 235). در این‌معنا، فناوری بخشی از زمینه‌ای است که کنش درون آن رخ می‌دهد.

اهمیت این برداشت برای شکاف مسئولیت در سامانه‌های خودمختار این‌است که میانجی‌گری فناورانه، محل شکل‌گیری مسئله اخلاقی را تغییر می‌دهد؛ بدین‌صورت که اگر فناوری در ساختار ادراک و کنش دخالت داشته باشد، آنگاه نمی‌توان موقعیت اخلاقی را تنها پس از وقوع پیامد و در مرحله ارزیابی مسئولیت بررسی کرد؛ زیرا که فناوری‌ها پیش از آن، در تعیین اینکه چه چیزی دیده شود، چه مسئله‌ای به‌عنوان مسئله اخلاقی مطرح گردد و چه گزینه‌هایی در دسترس کنشگر قرار گیرد، نقش دارند (Verbeek, 2011, p. 64). از این منظر، شکاف مسئولیت در سامانه‌های خودمختار فقط ناشی از پیچیده‌تر شدن زنجیره علّی یا کاهش کنترل انسانی نیست، در واقع، نشانه تغییری عمیق‌تر در ساختار کنش است. سامانه‌های هوش مصنوعی نشان می‌دهند که تصمیم‌گیری انسانی و عملکرد فناورانه را نمی‌توان همیشه به دو مرحله جداگانه تقسیم کرد؛ زیرا الگوریتم‌ها، داده‌ها، رابط‌های کاربری و زیرساخت‌های فنی در شکل‌دهی به نحوه مشاهده، انتخاب و اقدام مشارکت می‌کنند. بنابراین، بازصورت‌بندی فربیکی از مناقشه مسئولیت به‌جای آنکه تنها به‌دنبال تعیین جایگاه فناوری در میان عوامل مسئول باشد، پرسش بنیادی‌تری مطرح می‌کند؛ بدین‌صورت که کنش مسئولانه در چه شبکه‌ای از روابط انسان- فناوری امکان‌پذیر می‌شود؟ پاسخ به این پرسش، زمینه را برای مفهوم عاملیت ترکیبی فراهم می‌کند؛ زیرا در نگرش فربیک، کنش محصول اراده یک فاعل واحد نیست، بلکه برآمده از تعامل میان عناصر انسانی و فناورانه است.

۷. ۳. عاملیت ترکیبی

سومین پیش‌فرض مناقشه اسپارو و اسووبودا، انسانی‌بودن عامل کنش است. مسئله مورد چالش نیز، قابلیت شناسایی و مسئولیت‌پذیری این عامل انسانی در مواجهه با سامانه‌های خودمختار است که زنجیره انتساب را دشوار، پیچیده یا مبهم می‌سازند. اما فربیک از منظر پساپدیدارشناختی، عاملیت را تنها ویژگی درونی یک فاعل انسانی مستقل نمی‌داند که ابتدا تصمیم می‌گیرد و سپس از فناوری برای تحقق تصمیم خود استفاده می‌کند. از نظر او، عاملیت و کنش‌گری درون شبکه‌ای از روابط میان انسان، مصنوعات فناورانه، داده‌ها و زیرساخت‌های مادی شکل می‌گیرد (Verbeek, 2011, pp. 107-109). براین‌اساس، مسئله اصلی آن نیست که آیا فناوری جایگزین انسان به‌عنوان عامل کنش شده است یا خیر؛ بلکه باید پرسید کنش چگونه در تعامل میان عناصر انسانی و فناورانه پدید می‌آید. به‌بیان دیگر، می‌توان چنین استدلال کرد که اگر فناوری در نحوه ادراک موقعیت، تعریف مسئله، انتخاب گزینه‌ها و اجرای عمل دخالت داشته باشد، آنگاه نمی‌توان کنش را محصول اراده واحد یک عامل دانست که فناوری تنها ابزار تحقق آن بوده است. در واقع، فناوری در این چارچوب نه عامل مستقلی در برابر انسان و نه ابزار منفعل در دست اوست، بلکه بخشی از شرایط امکان کنش است.

فربیک با الهام از دیدگاه‌های برونو لاتور می‌گوید که کنش‌های فناورانه از تعامل مجموعه‌ای از عناصر شکل می‌گیرند و هیچ عنصر واحدی به‌تنهایی دارای بسندگی کامل علّی برای ایجاد پیامد نیست (Verbeek, 2014, pp. 78-79). در چنین چارچوبی، عاملیت به انسان محدود نمی‌شود و به فناوری نیز معطوف نمی‌گردد، بلکه در شبکه‌ای از روابط میان آن‌ها تحقق می‌یابد. بدین‌ترتیب، عاملیت به‌صورت رابطه‌ای و توزیع‌شده مفهوم‌سازی می‌شود. این نوع مفهوم‌سازی، برای بازصورت‌بندی مناقشه شکاف مسئولیت دارای اهمیت است؛ زیرا مسئله مسئولیت را نمی‌توان تنها با جست‌وجوی یک عامل اصلی حل کرد. در واقع، اگر کنش در یک شبکه انسان- فناوری شکل گرفته باشد، پرسش از اینکه چه کسی به‌تنهایی منشأ کنش بوده است، بخشی از پیچیدگی واقعی موقعیت را نادیده گرفته است. بنابراین، شکاف مسئولیت در سامانه‌های خودمختار تنها ناشی از دشواری یافتن یک عامل انسانی در انتهای زنجیره علّی نیست، بلکه ریشه عمیق‌تری در این واقعیت دارد که خودِ ساختار کنش در فناوری‌های پیچیده دیگر مطابق الگوی ساده «فاعل ـ عمل ـ پیامد» سامان نمی‌یابد. چنانچه «انسان مسلح» در تحلیل فربیک نشان می‌دهد که ابزار فناورانه فقط امتداد اراده انسانی نیست؛ بلکه ترکیب انسان و فناوری نوع خاصی از امکان کنش را ایجاد می‌کند که نمی‌توان آن را تنها به انسان و یا به ابزار نسبت داد (Verbeek, 2005, pp. 154-155).

با وجود این، در نگرش فربیک مفهوم عاملیت ترکیبی به‌معنای انتساب مسئولیت اخلاقی به فناوری نیست. او میان مشارکت فناوری در شکل‌گیری کنش و جایگاه پاسخ‌گویی اخلاقی تمایز می‌گذارد؛ بدین‌صورت که فناوری‌ها در ایجاد و جهت‌دهی به کنش نقش دارند، اما در مقام سوژه اخلاقی قرار نمی‌گیرند؛ زیرا فاقد قابلیت پاسخ‌گویی اخلاقی‌اند. ازاین‌رو، مسئولیت همچنان به کنش‌گران انسانی مربوط است، اما این مسئولیت در بستری تحقق می‌یابد که از پیش توسط میانجی‌گری‌های فناورانه شکل گرفته است (Verbeek, 2011, p. 132).

در نتیجه، مفهوم‌سازی فربیک از عاملیت ترکیبی، سطح مناقشه اسپارو و اسووبودا را از مالکیت کنش به فرایند شکل‌گیری کنش تغییر می‌دهد. بدین‌ترتیب، عاملیت ترکیبی را می‌توان نقطه‌ای دانست که در آن، مسئله مسئولیت از مدل خطی و فردمحور فاصله می‌گیرد و به سوی فهم شبکه‌ای از کنش حرکت می‌کند؛ شبکه‌ای که در آن انسان‌ها و فناوری‌ها در ایجاد امکان‌های کنش مشارکت دارند، بدون آنکه این مشارکت به‌معنای یکسانی جایگاه اخلاقی آن‌ها باشد.

۷. ٤. مسئولیت میانجی‌شده

در صورت‌بندی اسپارو و اسووبودا از مسئله انتساب مسئولیت، بازگرداندن پیامدها به عامل انسانی مشخص محوریت دارد. اما مولفه‌های فکری فربیک مبنی بر نفی دوگانه انسان/فناوری، میانجی‌گری فناورانه و بازتعریف عاملیت به‌مثابه امری ترکیبی، صورت مسئله انتساب مسئولیت در سامانه‌های خودمختار را تغییر می‌دهد. بنابراین، چارچوب سنتیِ انتساب پیامد به یک فاعل مستقل انسانی برای این نوع سامانه‌ها راهگشا نخواهد بود؛ زیرا طبق نگرش او، مسئولیت اخلاقی زمانی معنا دارد که کنش‌گران در شکل‌گیری یک کنش مشارکت داشته باشند. حال اگر کنش تنها محصول اراده مستقل یک فرد نباشد، بلکه در تعامل با فناوری شکل گیرد، در بررسی

مسئله مسئولیت نیز باید ساختار واقعی شکل‌گیری کنش در نظر گرفته شود؛ زیرا نقش فناوری در شکل‌دهی به ادراک، تصمیم و امکان‌های کنش قابل حذف نیست. همچنین، فناوری فاعل اخلاقی مستقل نیست و نمی‌توان آن را مانند انسان مسئول دانست. بنابراین، مسئله مسئولیت نباید براساس یک رابطه خطی میان فاعل، کنش و پیامد فهم شود؛ چراکه فرض اساسی این مدل آن است که می‌توان یک نقطه آغاز مشخص برای کنش یافت و سپس پیامد را به آن عامل بازگرداند. در حالی‌که طبق دیدگاه‌های فربیک، کنش در سامانه‌های خودمختار را می‌توان محصول عاملیت ترکیبی و میانجی‌گری فناورانه دانست. بنابراین، مسئله انتساب مسئولیت در این سامانه‌ها نیز باید به صورت شبکه‌ای از روابطی فهم شوند که کنش را ممکن ساخته‌اند.

بازصورت‌بندی فربیک از مسئله مسئولیت در چارچوب روابط انسان- فناوری را می‌توان با عنوان مسئولیت میانجی شده نام برد؛ زیرا این مفهوم بیانگر آن است که مسئولیت انسانی همواره درون شرایطی تحقق می‌یابد که فناوری‌ها بخشی از ساختار آن را شکل داده‌اند و میانجیگری کرده‌اند. در واقع، در این سطح نیز انسان همچنان موضوع اصلی پاسخ‌گویی اخلاقی باقی می‌ماند، اما این پاسخ‌گویی از نسبت میان یک فرد واحد و پیامدی مستقل تغییر می‌کند و در بستری از میانجی‌گری‌های فناورانه، سازمانی و مادی تحقق پیدا می‌کند. در این چارچوب، فناوری نه حامل مسئولیت است و نه عنصری خنثی که تنها بعد از تصمیم انسانی وارد فرآیند کنش‌گری شود. بلکه فناوری‌ها در سطحی پیشینی، شرایط امکان کنش را شکل می‌دهند؛ بدین‌معنا که آن‌ها بر اینکه چه چیزی دیده شود، چه مسئله‌ای مطرح گردد، چه گزینه‌هایی در دسترس قرار گیرد و چه نوع تصمیم‌هایی امکان‌پذیر باشند، تأثیر می‌گذارند (Verbeek, 2011, p. 64).

اهمیت این صورت‌بندی برای سامانه‌های خودمختار هوش مصنوعی چنین است که در این سامانه‌ها، تصمیم‌ها تنها نتیجه اراده یک کاربر یا برنامه‌نویس نیست بلکه حاصل تعامل میان داده‌ها، معماری الگوریتمی، طراحی فنی، کاربران، نهادهای سازمانی و محیط اجتماعی‌اند. در چنین شرایطی، جست‌وجوی یک عامل واحد برای پیامدهای علّی و اخلاقی ممکن نیست و باعث نادیده‌انگاری ساختار واقعی ایجاد کنش می‌شود.

در مجموع، رهیافت فربیک را نباید به‌عنوان نظریه‌ای کامل برای تعیین چگونگی انتساب مسئولیت اخلاقی در همه موقعیت‌های فناورانه تلقی کرد. اهمیت این رهیافت در این سطح، ارائه یک الگوی اجرایی برای تعیین سهم دقیق هر یک از کنشگران نیست؛ درواقع، آشکار کردن محدودیت‌های چارچوب‌های مبتنی بر انتساب خطی مسئولیت و نشان دادن نقش روابط انسان-فناوری در شکل‌گیری کنش است. هرچند فربیک با تأکید بر عاملیت ترکیبی و توزیع مسئولیت میان کنشگران انسانی درگیر در فرایندهای فناورانه، مانند طراحان، کاربران و سیاست‌گذاران (Verbeek, 2011, p. 108)، چارچوبی گسترده‌تر برای فهم مسئولیت در بستر روابط انسان- فناوری ارائه می‌دهد، اما تعیین حدود و شرایط دقیق این مسئولیت‌ها همچنان نیازمند معیارهای تکمیلی اخلاقی و نهادی است. بنابراین، اهمیت رویکرد فربیک را باید در تغییر سطح تحلیل مسئولیت جست‌وجو کرد؛ بدین‌معنا که مسئله از پرسش درباره یافتن یک عامل مسئول، به بررسی چگونگی شکل‌گیری کنش و شرایطی معطوف می‌شود که در آن مسئولیت اخلاقی میان کنشگران انسانی معنا پیدا می‌کند.

نتیجه‌گیری

این مقاله با تمرکز بر مناقشه اسپارو و اسووبودا درباره امکان انتساب مسئولیت در سامانه‌های خودمختار، نشان داد که این اختلاف، به‌رغم تفاوت در نتیجه‌گیری، بر مجموعه‌ای از مفروضات مشترک درباره ماهیت کنش، نقش فناوری و معیارهای مسئولیت استوار است. در هر دو رویکرد، مسئولیت در چارچوبی تحلیل می‌شود که در آن کنترل و پیش‌بینی‌پذیری به‌عنوان شروط بنیادین انتساب مسئولیت در نظر گرفته می‌شوند و کنش در قالب رابطه‌ای خطی میان عامل انسانی و پیامد فهم می‌شود. این چارچوب مشترک، امکان صورت‌بندی مسئله را محدود می‌کند؛ زیرا نقش فناوری در آن به سطح واسطه‌گری ابزاری تقلیل می‌یابد و تأثیر آن بر شکل‌گیری خودِ امکان‌های کنش مورد توجه قرار نمی‌گیرد. از این منظر، اختلاف میان دو دیدگاه، نه در سطح بنیان‌های مفهومی، بلکه در نحوه ارزیابی امکان حفظ یا گسست رابطه میان کنش انسانی و پیامد در شرایط خودمختاری فناورانه باقی می‌ماند.

با وجود این، تحلیل مناقشه اسپارو و اسووبودا نتیجه می‌دهد که اختلاف آنها فقط درباره پاسخ به یک مسئله موجود نیست؛ بلکه مسئله عمیق‌تر، حدود توان توضیحی الگوی سنتی انتساب مسئولیت در مواجهه با فناوری‌های پیچیده است. اسپارو با تأکید بر گسست میان کنش انسانی و پیامد، امکان حفظ مسئولیت در شرایط خودمختاری فناورانه را با چالش مواجه می‌داند. در مقابل، اسووبودا تلاش می‌کند نشان دهد که پیچیدگی فناوری الزاماً به‌معنای از میان رفتن امکان مسئولیت انسانی نیست. با وجود این تفاوت، هر دو دیدگاه مسئله را در چارچوبی مطرح می‌کنند که در آن مسئولیت بر مبنای امکان ردیابی کنش تا پیامد و شناسایی عامل انسانی دخیل در آن تحلیل می‌شود. بنابراین، اختلاف آنها بیش از آنکه به اصل نقش فناوری در شکل‌گیری کنش مربوط باشد، به امکان حفظ معیارهای سنتی انتساب مسئولیت در شرایط جدید فناورانه مربوط می‌شود.

در مقابل، بازخوانی مسئله از منظر فربیک بیانگر آن است که کنش اخلاقی در سامانه‌های خودمختار را نمی‌توان تنها به تصمیم یک عامل واحد تقلیل داد؛ زیرا فرایند شکل‌گیری کنش درون شبکه‌ای از روابط انسانی و فناورانه تحقق می‌یابد. در این چارچوب، فناوری فقط ابزار تحقق اراده انسانی نیست، بلکه بخشی از ساختار ایجاد امکان‌های کنش است.

براین‌اساس، مسئله مسئولیت پیش از آنکه مسئله‌ای درباره «انتساب به چه کسی» باشد، به مسئله‌ای درباره «چگونگی شکل‌گیری کنش» تبدیل می‌شود. این جابه‌جایی سطح تحلیل، به‌جای ارائه یک پاسخ نهایی به مسئله شکاف مسئولیت، حاکی از آن است که خود صورت‌بندی این مسئله در چارچوب‌های مبتنی بر کنترل و پیش‌بینی‌پذیری، نیازمند بازنگری است. البته دیدگاه‌های فربیک را نمی‌توان راه‌حلی نهایی برای مسئله شکاف مسئولیت دانست؛ بلکه اهمیت آنها در آشکارکردن محدودیت‌های مفهومی‌ای است که این مسئله در درون آنها شکل گرفته است. در واقع، فربیک صورت‌بندی مسئله را تغییر می‌دهد؛ یعنی چارچوب مناقشه اسپارو و اسووبودا را مورد پرسش قرار می‌دهد. به‌بیان دقیق‌تر، آنچه به‌عنوان شکاف مسئولیت صورت‌بندی می‌شود، در سطحی بنیادی‌تر، حاصل یک نحوه خاص از فهم کنش و فناوری است که با بازاندیشی در آن، بازتعریف خود مسئله ممکن می‌شود. بدین‌ترتیب، شکاف مسئولیت فقط پیامد فقدان یک عامل انسانی مشخص نیست، بلکه نشانه محدودیت الگویی از کنش است که انسان و فناوری را به‌صورت دو قلمرو جداگانه فرض می‌کند. در حالی‌که رویکرد فربیک، مسئولیت را در سطح عمیق‌تر، یعنی سطح شرایط امکان کنش

در روابط انسان- فناوری و به‌صورت میانجی‌شده بازاندیشی می‌کند.

درعین‌حال، این بازخوانی به‌معنای ارائه نظریه‌ای کامل برای تعیین مسئولیت اخلاقی در همه موقعیت‌های فناورانه نیست. رهیافت فربیک بیش از آنکه معیارهای مشخصی برای تعیین سهم و حدود مسئولیت کنشگران انسانی ارائه دهد، چارچوبی برای فهم دوباره رابطه میان کنش، فناوری و مسئولیت فراهم می‌آورد. از این منظر، می‌توان مفهوم «مسئولیت میانجی‌شده» را به‌عنوان تلاشی برای بازتعریف سطح تحلیل مسئله مسئولیت اخلاقی محسوب کرد؛ این بازتعریف بر این نکته دلالت دارد که انتساب مسئولیت اخلاقی در سامانه‌های پیچیده هوشمند، بدون توجه به شرایط شکل‌گیری کنش در روابط انسان- فناوری، تبیین کاملی نخواهد داشت. در نتیجه، پس از این بازخوانی، پرسش اصلی درباره کنار گذاشتن مسئولیت انسانی یا نسبت دادن آن به فناوری نیست؛ بلکه به نحوه صورت‌بندی مسئولیت اخلاقی در بستر میانجی‌گری فناورانه بازمی‌گردد.

منابع

به‌خوی، محسن. (۱۴۰۱). *شرح و نقد نظریهٔ اخلاقی‌سازی تکنولوژی فربیک*. پایان‌نامه کارشناسی‌ارشد. دانشگاه صنعتی شریف، گروه فلسفه علم.

جلیلیان، محمدسبحان. (۱۳۹۶). *مصنوعات تکنیکی و انسان: تاملی بر اندیشه‌های فربیک*. پایان‌نامه کارشناسی‌ارشد. دانشگاه صنعتی شریف، گروه فلسفه علم.

خلیلی، مهدی.(۱۳۹۷). «آیا مصنوعات فناورانه می‌توانند خوب یا بد باشند؟». *سیاست علم و فناوری*. پائیز. سال دهم، شماره سوم، صص ۱۷–۲۸.

میرزامحمدی، عبدالرئوف، و تقوی، مصطفی. (۱۴۰۴). «اخلاق فناوری: ارزیابی انتقادی مناقشه میان دیوید مورو و ژوزف پیت درباره ارزش‌باری فناوری و مسئولیت اخلاقی». *تأملات اخلاقی*. تابستان. سال ششم، شماره دوم، صص ۹۳–۱۱۴.

Burrell, Jenna. (2016). "How the Machine Thinks: Understanding Opacity in Machine Learning Algorithms". *Big Data & Society*. 3(1): 1–12. DOI: 10.1177/2053951715622512.

Fischer, John Martin, & Ravizza, Mark. (2000). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press, Second edition.

Floridi, Luciano. (2016). "Faultless Responsibility: On the Nature and Allocation of Moral Responsibility for Distributed Moral Actions". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 374 (2083), Article 20160112. DOI: 10.1098/rsta.2016.0112.

Goetze, Theodore Simon. (2022). "Mind the Gap: Autonomous Systems, the Responsibility Gap, and Moral Entanglement". In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)*. New York: Association for Computing Machinery (ACM).

Kiener, Maximilian. (2024). "AI and Responsibility: No Gap, but Abundance". *Journal of Applied Philosophy*. Advance online publication. DOI: 10.1111/japp.12765.

Matthias, Andreas. (2004)."The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata". *Ethics and Information Technology*. 6(3): 175–183. DOI: 10.1007/s10676-004-3422-1.

Santoni de Sio, Filippo, & Mecacci, Giulio. (2021). "Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them". *Philosophy & Technology*. 34(4): 1057–1084. DOI: 10.1007/s13347-021-00450-x.

Schur, Eva, Brouns, Anna, & Lee, Peter. (2025). "Ethical Analysis of the Responsibility Gap in Artificial Intelligence". *Ethics & Society*. 6(4): 1–10.

Sparrow, Robert. (2007). "Killer Robots". *Journal of Applied Philosophy*. 24(1): 62–77. DOI: 10.1111/j.1468-5930.2007.00346.x.

Sparrow, Robert. (2021). "Why Machines Cannot Be Moral". *AI & Society*. 36(3), 685–693. DOI: 10.1007/s00146-020-01132-6.

Sparrow, Robert, & Flenady, Gene. (2025). "The Testimony Gap: Machines and Reasons". *Minds & Machines*. 35(1): Article 12. DOI: 10.1007/s11023-025-09712-5.

Sparrow, Robert, & Flenady, Gene. (2026). "On Not Being a Robot: AI as a Threat to Rational Agency". *Inquiry: An Interdisciplinary Journal of Philosophy*. 1–20. DOI: 10.1080/0020174X.2026.2673086.

Swoboda, Torben. (2018). "Autonomous Weapon Systems - An Alleged Responsibility Gap". In *Philosophy and Theory of Artificial Intelligence 2017*, edited by Vincent C. Müller. Berlin: Springer Verlag. pp. 302–313. DOI: 10.1007/978-3-319-96448-5_32.

Swoboda, Torben & Lauwaert, Lode. (2025). "Can Artificial Intelligence Embody Moral Values?". *AI and Ethics*. 5(4), 3669–3680. DOI: 10.1007/s43681-025-00662-3.

Verbeek, Peter-Paul. (2005). *What Things Do: Philosophical Reflections on Technology, Agency and Design*. University Park, PA: Pennsylvania State University Press, First edition.

Verbeek, Peter-Paul. (2009). "Ambient Intelligence and Persuasive Technology: The Blurring Boundaries Between Human and Technology". *NanoEthics*. 3(3): 231–242. DOI: 10.1007/s11569-009-0077-8.

Verbeek, Peter-Paul. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago: University of Chicago Press, First edition.

Verbeek, Peter-Paul. (2014). "Some Misunderstandings about the Moral Significance of Technology". In *The Moral Status of Technical Artefacts*, edited by Peter Kroes and Peter-Paul Verbeek. Dordrecht: Springer. pp. 75–88. DOI: 10.1007/978-94-007-7914-3.