Interdisciplinary Studies in Ethics
JISE

Comparative Analysis of two Ethical Approaches—Optimistic and Pessimistic—toward AI Development, with a Focus on the Problem of the Responsibility Gap

Massoud Toossi Saeidi

Institute for Science and Technology Studies, Shahid Beheshti University, Tehran, Iran. m_tousi@sbu.ac

Abstract

Introduction: This paper examines two contrasting ethical approaches to the development of artificial intelligence (AI): the optimistic and the pessimistic. Both approaches aim to analyze the ethical and human-centered dimensions of AI, yet they differ fundamentally in their assumptions and conclusions. The optimistic approach emphasizes AI's potential to enhance human life and argues that ethical concerns are often based on speculative or non-specialist assumptions. In contrast, the pessimistic approach deems unrestricted AI development ethically unjustifiable due to unpredictable consequences, algorithmic bias, and the erosion of human decision-making capacity. The focal point of this paper is the "responsibility gap"—a dilemma that complicates the attribution of negative outcomes of AI systems to any specific individual or institution, raising profound questions about moral and legal accountability. The central question addressed is: which of the two approaches offers a more reasonable response to the responsibility gap?

Findings: The optimistic approach is grounded in three core arguments:

- The benefits of AI development outweigh its harms, and depriving societies of these benefits is ethically unjustifiable.
- Pessimistic concerns often stem from non-expert perceptions, whereas specialists tend to offer more balanced and optimistic assessments.
- Philosophical assumptions underlying pessimistic views—such as the claim that robots lack human-like qualities—remain unresolved and cannot serve as a decisive basis for restricting AI development.

Conversely, the pessimistic approach draws on empirical evidence of AI's problematic effects:

- AI systems may possess unethical tendencies such as deception and malicious intent.
- AI development leads to undesirable consequences like institutionalized inequality and diminished human autonomy, which cannot be ethically offset by potential benefits.

Cite this article: Toossi Saeidi, M.(2025). Comparative Analysis of two Ethical Approaches—Optimistic and Pessimistic—toward AI Development, with a Focus on the Problem of the Responsibility Gap. *Interdisciplinary Studies in Ethics*, 1(2), p. 7-24. https://doi.org/10.48308/jiethics.2026.241786.1030

Received: 2025/05/02 ; **Received in revised form:** 2025/06/06 ; **Accepted:** 2025/07/10 ; **Published online:** 2025/10/02
Article type: Research Article jiethics.sbu.ac.ir

● Ethical considerations should extend beyond normative human life to include potential harm to nature and ecosystems, threatening the very foundation of human existence.

Regarding the responsibility gap, pessimistic thinkers such as Matthias and Sparrow argue that autonomous systems make it impossible to assign moral responsibility, especially in sensitive domains like warfare. Optimists like Danaher, however, view the gap as an opportunity to reduce the psychological burden of tragic human decisions, presenting it as a potential ethical advantage.

Discussion: The paper offers an independent analysis that distinguishes moral accountability from moral worth, arguing that the responsibility gap in AI is no more complex than that found among humans. Epistemic uncertainty and lack of full control are inherent to all moral agents, and the emergence of intelligent entities is not fundamentally different from the birth of new human beings.

Thus, ethical pessimism that rejects AI development due to the responsibility gap suffers from internal contradiction—if blameworthiness is a condition for moral legitimacy, then human reproduction itself would be ethically suspect. Accordingly, a combined neutral-optimistic approach to the responsibility gap is logically superior to absolute pessimism. This conclusion demonstrates the overall implausibility of the pessimistic approach and supports the preference for optimistic and neutral perspectives.

Keywords: The Responsibility Gap Challenge, AI Ethics, Optimistic Approach, Pessimistic Approach, Rationality.

References

Ahmad, S. F.; Han, H.; Alam M. M.; Rehmat, M.; Irshad, M.; Arraño-Muñoz, M.; & Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications*, 10(1), 1–14. https://doi.org/10.1057/s41599-023-01787-8

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications*, 10(1), 1–12. https://doi.org/10.1057/s41599-023-02079-x

Danaher, J. (2019a). The Philosophical Case for Robot Friendship. *Journal of Posthuman Studies*, 3(1), 5–24. https://doi.org/10.5325/jpoststud.3.1.0005

Danaher, J. (2019b). The Philosophical Case for Robot Friendship. *Journal of Posthuman Studies*, 3(1), 5–24. https://doi.org/10.5325/jpoststud.3.1.0005

Danaher, J. (2022). Tragic choices and the virtue of techno-responsibility gaps. *Philosophy & Technology*, 35(2), 26. https://doi.org/10.1007/s13347-022-00519-1

Ferlito, B., Segers, S., De Proost, M., & Mertes, H. (2024). Responsibility Gap(s) Due to the Introduction of AI in Healthcare: An Ubuntu-Inspired Approach. *Science and Engineering Ethics*, 30(4), 34. https://doi.org/10.1007/s11948-024-00501-4

Ferrara, E. (2024). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci*, 6(1), 3. https://doi.org/10.3390/sci6010003

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 175–183. https://doi.org/10.1007/s10676-004-3422-1

Narayanan, A., & Kapoor, S. (2025a). *AI as normal technology*. Knight First Amend. Inst. https://thedocs.worldbank.org/en/doc/d6e33a074ac9269e4511e5d44db2f9ac-0050022025/original/AI-as-Normal-Technology-Narayanan-Kapoor-Final.pdf

Narayanan, A., & Kapoor, S. (2025b). *AI as normal technology: An alternative to the vision of AI as a potential superintelligence*. Knight First Amendment Institute, Columbia University, https://Kfai-Documents.S3.Amazonaws.Com/Documents/C3cac5a2a7/AI-as-Normal-Technology—Narayanan—Kapoor.Pdf

Prentice, R. (2025). *Techno-Optimist or AI Doomer? Consequentialism and the Ethics of AI.* Ethics Unwrapped. https://ethicsunwrapped.utexas.edu/techno-optimist-or-ai-doomer-consequentialism-and-the-ethics-of-ai

Schwaller, F. (2025). Will AI improve your life? Here's what 4,000 researchers think. *Nature*, 640(8059), 577–578. https://doi.org/10.1038/d41586-025-01123-x

Sparrow, R. (2007). Killer Robots. *Journal of Applied Philosophy*, 24(1), 62–77. https://doi.org/10.1111/j.1468-5930.2007.00346.x

Strain, M. R. (2024, Summer). *The Case for AI Optimism*. National Affairs, 60. https://nationalaffairs.com/publications/detail/the-case-for-ai-optimism

Thaiduong, N. (2025). IT Professionals Versus the Public: Who's More Optimistic About AI's Future Impacts? *SAGE Open*, 15(2). https://doi.org/10.1177/21582440251348802

Vallor, S., & Vierkant, T. (2024). Find the Gap: AI, Responsible Agency and Vulnerability. *Minds and Machines*, 34(3), 20. https://doi.org/10.1007/s11023-024-09674-0

Wada, K., & Shibata, T. (2007). Living with seal robots—Its sociopsychological and physiological influences on the elderly at a care house. *IEEE Transactions on Robotics*, 23(5), 972–980. https://doi.org/10.1109/TRO.2007.906261

Wang, H., & Blok, V. (2025). Why putting artificial intelligence ethics into practice is not enough: Towards a multi-level framework. *Big Data & Society*, 12(2), https://doi.org/10.1177/20539517251340620

JISE
مطالعات میان‌رشته‌ای
اخلاق
Journal of Interdisciplinary Studies in Ethics

تحلیل مقایسه‌ای دو رهیافت اخلاقی خوش‌بینانه و بدبینانه به توسعه هوش مصنوعی با تمرکز بر معضل شکاف مسئولیت

مسعود طوسی سعیدی

پژوهشکده مطالعات بنیادین علم و فناوری، دانشگاه شهید بهشتی، تهران، ایران. m_tousi@sbu.ac

چکیده

این مقاله به‌طور مقایسه‌ای دو رهیافت خوش‌بینانه و بدبینانه نسبت به توسعه هوش مصنوعی را با تمرکز بر معضل «شکاف مسئولیت» بررسی می‌کند. نخست مبانی هر یک از این دو دیدگاه تشریح می‌شود: در رهیافت خوش‌بینانه تلاش می‌شود بر مزایای چشمگیر هوش مصنوعی برای جامعه، فقدان نگرانی در میان متخصصان و عدم قطعیت‌های فلسفی تأکید شود؛ در مقابل، رهیافت بدبینانه با استناد به سوگیری‌های الگوریتمی، افول توانایی تصمیم‌گیری انسانی و پیامدهای پیش‌بینی‌ناپذیر اجتماعی- اقتصادی، توسعه بدون محدودیت هوش مصنوعی را از نظر اخلاقی مردود می‌داند. پس از آن شکاف مسئولیت به‌عنوان معضلی بررسی می‌شود که امکان انتساب کامل پیامدهای منفی فنّاوری به فرد یا نهادی مشخص را از بین می‌برد و نمونه‌هایی از مواجهه با آن در نظام سلامت و سلاح‌های خودران ارائه می‌شود. در ادامه سه معیار محوری- مبنای خیر اخلاقی، دامنه قابلیت‌های متافیزیکی و عملی هوش مصنوعی و آسیب‌پذیری انسان در رواج هوش مصنوعی به‌عنوان چارچوبی برای تحلیل معقولیت دو رهیافت مطرح می‌شوند، اما تحلیل انجام شده نشان می‌دهد هیچ‌یک از استدلال‌های ارائه‌شده ذیل رهیافت‌های خوش‌بینانه و بدبینانه به‌طور مطلق پاسخ قطعی به این معیارها نمی‌دهند؛ بنابراین استدلالی مستقل درباره میزان معقولیت رهیافت خوش‌بینانه و بدبینانه با تأکید بر معضل شکاف مسئولیت ارائه خواهد شد که براساس آن، رهیافت خوش‌بینانه نسبت به رهیافت بدبینانه معقول‌تر دانسته می‌شود. در پایان تأکید خواهد شد که این استدلال و معقول‌تر بودن رهیافت خوش‌بینانه نسبت به رهیافت بدبینانه به معنای بی‌احتیاطی در توسعه هوش مصنوعی و نادیده‌انگاشتن معضل شکاف مسئولیت نیست.

کلیدواژه‌ها: معضل شکاف مسئولیت، اخلاق هوش مصنوعی، رهیافت خوش‌بینانه، رهیافت بدبینانه.

استناد به این مقاله: طوسی سعیدی، مسعود (۱۴۰۴). تحلیل مقایسه‌ای دو رهیافت اخلاقی خوش‌بینانه و بدبینانه به توسعه هوش مصنوعی با تمرکز بر معضل شکاف مسئولیت. *مطالعات میان‌رشته‌ای اخلاق*، ۱(۲)، ص۷-۲۴. https://doi.org/10.48308/jiethics.2026.241786.1030

تاریخ دریافت: ۱۴۰۴/۰۲/۱۲ ؛ **تاریخ اصلاح:** ۱۴۰۴/۰۳/۱۶ ؛ **تاریخ پذیرش:** ۱۴۰۴/۰۴/۱۹ ؛ **تاریخ انتشار:** ۱۴۰۴/۰۷/۱۰

نوع مقاله: پژوهشی jiethics.sbu.ac.ir

مقدمه

هوش مصنوعی (AI) گستره وسیعی از پرسش‌های فلسفی و اخلاقی را پیش روی جوامع انسانی قرار داده است. رشد سریع الگوریتم‌ها و سامانه‌های مبتنی بر هوش مصنوعی، افزون بر تأثیر بر محصولات فنّاورانه، زمینه‌ساز مکاتب فکری و رهیافت‌های اخلاقی فراوانی شده است. به نظر می‌رسد که دیدگاه‌ها و تحلیل‌ها درباره ابعاد اخلاقی توسعه هوش مصنوعی را می‌توان ذیل دو رهیافت اصلی و متضاد دسته‌بندی کرد: خوش‌بینانه و بدبینانه. هر دو رهیافت تلاش می‌کنند تا ابعاد اخلاقی و انسان‌محور توسعه این فنّاوری را تحلیل و ارزیابی کنند، اما تفاوت‌های عمیقی در پیش‌فرض‌ها، اهداف و نتایج آنها قابل مشاهده است. وجه اصلی این تقسیم، توصیه نهایی این رهیافت‌هاست.

برخی افراد با اشتیاق وعده‌های هوش مصنوعی را در پرتوی رشد سریع و توسعه آن می‌پذیرند؛ درحالی‌که گروهی دیگر با دیده تردید و نگرانی به آن می‌نگرند و معتقدند که AI موجب اختلال در بازار کار می‌شود و انسان‌ها را بیکار می‌سازد. پرسش‌هایی از این دست، از دغدغه‌های اخلاقی درباره هوش مصنوعی به شمار می‌روند: «آیا هوش مصنوعی واقعاً مشاغل را نابود می‌کند و سبب بیکاری و بحران‌های اجتماعی می‌شود؟» و «آیا این فنّاوری در روزی از انسان پیشی خواهد گرفت و بر او غلبه خواهد کرد؟».

در رهیافت خوش‌بینانه تلاش می‌شود که نشان دهد دغدغه‌های اخلاقی درباره توسعه هوش مصنوعی، آن‌قدر اساسی و نگران‌کننده نیستند که لازم باشد روند توسعه این فنّاوری محدود گردد. در تقابل با این نتیجه‌گیری، دیدگاه‌های ذیل رهیافت بدبینانه می‌کوشند نشان دهند که توسعه بدون محدودیت هوش مصنوعی از دیدگاه اخلاقی قطعاً ناموجه است و از این جهت باید محدودیت‌هایی در روندهای مرتبط با توسعه این فنّاوری اعمال شود.

در این میان، یک معضل مهم که به طور خاص نقطه تمرکز این مقاله است «شکاف مسئولیت»[1] نام دارد. شکاف مسئولیت معضلی مربوط به تشخیص دادن مسئولیت فرد یا نهادی مشخص درباره کارکردهای مخرب هوش مصنوعی است. شناخت تحلیلی این شکاف، نقشی کلیدی در ارزیابی کارآمدی هر رهیافت اخلاقی به توسعه AI دارد و می‌تواند مبنایی برای سنجش مقایسه‌ای معقولیت آنها باشد.

در ادامه، نخست مبانی و دیدگاه‌های اصلی رهیافت خوش‌بینانه بررسی شده، سپس به رهیافت بدبینانه و ایده‌های کلیدی آن پرداخته می‌شود. پس از آن، تحلیل تطبیقی دو رهیافت از دیدگاه شکاف مسئولیت ارائه خواهد شد. سؤال اصلی مدنظر مقاله پیش‌رو این است که کدام‌یک از دو رهیافت با تأکید بر معضل شکاف مسئولیت، معقول‌تر است؟

---

1. Responsibility Gap

۱. رهیافت خوش‌بینانه به توسعه هوش مصنوعی

یک مقدمه اصلی در رهیافت خوش‌بینانه به توسعه هوش مصنوعی این مطلب است که فنّاوری‌های پیشرفته از جمله AI، ظرفیت ایجاد مزایای چشمگیر برای بشر را دارند و با طراحی و مدیریت درست می‌توانند به فرصت‌هایی برای ارتقای کیفیت زندگی انسانی بدل شوند. مقدمه دیگر اینکه استدلال‌ها درباره معضلات و مشکلاتی که ممکن است درنتیجهٔ توسعه هوش مصنوعی پدید آیند، قطعیت ندارند، کمابیش احتمالی و حتی گاهی مبتنی بر پیش‌فرض‌های فلسفی‌اند؛[1] از این‌رو، برخی جلوگیری از توسعه هوش مصنوعی یا محدود کردن آن براساس دغدغه‌های اخلاقی را موجه نمی‌دانند.

یادداشت «دلایل خوش‌بینی به هوش مصنوعی» که میشل استرین[2] آن را پس از ایراد سخنرانی خود در مجلس سنای ایالات متحده منتشر کرد، نمونه‌ای از این قسم دیدگاه‌هاست. در این یادداشت تلاش شده است با استناد به آمارهای تاریخی اقتصاد تبیین شود که فنّاوری‌ها در طول تاریخ، با وجود نگرانی‌های اولیه درنهایت و همواره به منافع انجامیده‌اند (Strain, 2024). او می‌کوشد با استدلالش، سناریوهای بدبینانه و مبتنی بر حدسیات درباره آینده کار و مشاغل را با اشاره به شواهد تجربی حاصل از توسعه فنّاوری‌ها در گذشته به چالش کشد.

نمونه دیگر مقاله «هوش مصنوعی به عنوان یک فنّاوری معمولی» است. نویسندگان این مقاله توضیح می‌دهند که تعبیر معمولی برای هوش مصنوعی در تضاد با دیدگاه‌های آرمان‌شهری و ویران‌شهری از آینده هوش مصنوعی است که تمایل مشترکی به برخورد با آن به عنوان گونه‌ای جداگانه، یک موجود بسیار خودمختار و ابرهوشمند دارند (Narayanan & Kapoor, 2025b). آنها مدعی‌اند بدین شیوه، تصویری واقع‌گرایانه از آینده هوش مصنوعی ـ مبتنی بر داده‌های قطعی فعلی ـ ارائه می‌دهند و از سناریویی دفاع می‌کنند که در آن AI پیشرفته است، اما همچنان تحت کنترل انسان‌ها و سازمان‌ها باقی می‌ماند و نقش انسان‌ها بیشتر به هدایت و نظارت بر AI تبدیل می‌شود. در این چارچوب، هوش مصنوعی به‌عنوان یک فنّاوری «عادی» برداشت می‌شود، نه پدیده‌ای خارق‌العاده یا فراهوشمند. روشن است که این قسم دیدگاه ها به تحلیل متفاوتی از مخاطرات و ملاحظات اخلاقی می‌انجامد. نتیجه این‌گونه تحلیل‌ها، این است که به‌جای اقدامات قاطع می‌بایست بر کاهش عدم‌قطعیت و افزایش تاب‌آوری تمرکز کرد.

---

۱. منظور از احتمالی بودن این است که مثلاً تحلیل‌ها درباره بحران‌های بیکاری ناشی از هوش مصنوعی و پیامدهای اجتماعی آن تخمینی است و از این‌رو نباید به سبب یک تخمین از مزایای قطعی توسعه هوش مصنوعی صرف‌نظر کرد. منظور از پیش‌فرض‌های فلسفی نیز این است که مثلاً رسیدن هوش مصنوعی به قابلیت‌های ذهنی انسان یا عبور از آن، مبتنی بر ایده‌هایی فلسفی درباره ذهن و آگاهی می‌باشد که مشابه با سایر آرا و ایده‌های فلسفی، از سوی متفکران دیگر، نقدهای گوناگونی به آنها وارد دانسته شده است.

2. Michael R. Strain

نمونه دیگری از رهیافت خوش‌بینانه در آثاری دیده می‌شود که می‌کوشند از تفاوت میان دریافت‌های متخصصان و برداشت عموم از مخاطرات هوش مصنوعی و این مقدمه که اصولاً متخصصان فهم عمیق‌تر و دقیق‌تری دارند به سود رهیافت خوش‌بینانه استدلال کنند. دومقاله «متخصصان فنّاوری اطلاعات در مقابل عموم مردم: چه کسی در مورد تأثیرات آینده هوش مصنوعی خوش‌بین‌تر است؟» (Thaiduong, 2025) و «آیا هوش مصنوعی زندگی شما را بهبود خواهد بخشید؟ این چیزی است که ۴ هزار محقق فکر می‌کنند» (Schwaller, 2025)، نمونه‌ای از این موارد‌ند. شوالر شواهد یک نظرسنجی را ارائه می‌دهد که «دانشمندانی که روی هوش مصنوعی کار می‌کنند، نسبت به عموم مردم اطمینان بیشتری دارند که این فنّاوری به نفع مردم خواهد بود» و از این نتیجه می‌گیرد که درک نزدیک از آنچه در فرایندهای واقعی توسعه هوش مصنوعی می‌گذرد، به جای نگرانی‌های حدسی به ارزیابی‌های خوش‌بینانه‌تر و مبتنی بر شواهد می‌انجامد. تایدونگ نشان می‌دهد که متخصصان فنّاوری اطلاعات هم مزایا و هم محدودیت‌های بالقوه را می‌پذیرند و در عین حال دیدگاه متعادل‌تری نسبت به هوش مصنوعی در مقایسه با افکار عمومی دارند که به طور فزاینده‌ای بدبین هستند.

جان‌داناهر در مقاله پر ارجاع خود (۲۰۱۹) استدلالی ارائه می‌دهد که به سود رهیافت خوش‌بینانه است و مطلبی افزون بر مطالب دیدگاه‌های یادشده دربردارد. او یادآوری می‌کند بحث متافیزیکی درباره اینکه ربات‌های هوشمند به معنای قوی نمی‌توانند از ویژگی‌های متقابل[1] برخوردار باشند، هنوز جمع‌بندی قطعی ندارد. منظور از ویژگی‌های متقابل، وجود ویژگی‌های هم‌سنخی همچون برخوردار بودن از عواطف، آگاهی، التفات و عقلانیت در طرفین روابط انسان ـ انسان می‌باشد، اما در روابط از نوع انسان ـ ربات، وجود این ویژگی‌ها در طرف ربات مورد تردید است. نقد داناهر، نقدی اساسی‌تر به آن دسته از استدلال‌های بدبینانه می‌باشد که در مقدمات خود بر این مضمون تأکید دارند که ربات‌ها به طور کلی و از حیث درونی هم‌تراز انسان نیستند. داناهر تصریح می‌کند که در حال حاضر ممکن است از نظر فنی[2] محدودیت‌هایی وجود داشته باشد، اما از نظر متافیزیکی تکلیف مشخص نیست و جمع‌بندی قاطعی وجود ندارد (Danaher, 2019a: 12–23).

به این ترتیب در رهیافت خوش‌بینانه دست‌کم یکی از سه مضمون زیرهسته اصلی استدلال‌ها را تشکیل می‌دهد:

ـ فواید توسعه هوش مصنوعی بیش از زیان‌های آن است و از این‌رو بدبینی نسبت به توسعه آن بدین سبب که جوامع را از فواید آن محروم می‌سازد، از نظر اخلاقی موجه نیست.

ـ بدبینی نسبت به توسعه هوش مصنوعی، ریشه در انگاره‌های غیرفنی و غیرتخصصی درباره این

1. Mutual
2. Technical

فنّاوری دارد؛ درحالی‌که از نظر تخصصی این نگرانی‌ها برطرف می‌شود. با توجه به آنکه رهیافت بدبینانه، ریشه در این انگاره‌ها دارد موجه نیست.

ـ استدلال‌های بدبینانه مبتنی بر پیش‌فرض‌های متافیزیکی خاصی هستند که هیچ نوع جمع‌بندی قطعی و نهایی درباره آنها وجود ندارد؛ از این‌رو ترجیح آنها بر استدلال‌های خوش‌بینانه موجه نیست.

۲. رهیافت بدبینانه به توسعه هوش مصنوعی

رهیافت بدبینانه مبتنی بر تردیدهای عمیق نسبت به ماهیت یا توانایی پیش‌بینی و کنترل پیامدهای اخلاقی توسعه AI است. در این رهیافت گاهی بر ماهیت برخی پیامدهای AI و گاهی بر عوارض پیش‌بینی‌نشده و مهارنشدنی این پیامدها تأکید می‌شود. بخش مشترک نتیجه استدلال‌های ذیل رهیافت بدبینانه این است که از نظر ملاحظات مهم و اصیل اخلاقی، توسعه بدون محدودیت هوش مصنوعی موجه نیست.

نویسنده یادداشت «خوش‌بین به فنّاوری یا بدبین به هوش مصنوعی: پیامدگرایی و اخلاق هوش مصنوعی» ضمن مرور تجارب در دسترس تأکید می‌کند شواهد قابل تأملی وجود دارد که احتمالاً هوش مصنوعی قدرتمند، خلّاق و مایل به دروغ گفتن و نقشه‌های سوء کشیدن برای رسیدن به اهداف خود است؛ درحالی‌که عملکردهای این فنّاوری، شواهد قابل اندازه‌گیری، تأییدشده و مشخصی در مورد امکانش برای انجام اقدامات خوب به ما نمی‌دهد تا بتوانیم قضاوتی واقعاً آگاهانه در مورد اخلاقی بودن ادامه توسعه هوش مصنوعی با سرعت سرسام‌آور فعلی داشته باشیم (Prentice, 2025). او نتیجه می‌گیرد که احتیاط در توسعه هوش مصنوعی لازمه این شرایط است و ساده‌گیری و خوش‌بینی، با توجه به شواهد موجود و دغدغه‌های مربوط به آنها موجه نیست.

یکی از مهم‌ترین دغدغه‌های رهیافت بدبینانه، سوگیری‌های الگوریتمی است که موجب به‌وجود آمدن نتایج مخرب می‌شود. مقاله «اخلاق و تبعیض در سامانه‌های مبتنی بر هوش مصنوعی» پدیده‌ای به نام *سوگیری در ورودی و خروجی* را معرفی کرده است؛ جایی که سوگیری‌های اجتماعی موجود در داده‌ها به تبعیض الگوریتمی می‌انجامد که نابرابری‌های تاریخی موجود در جوامع و فرهنگ‌ها را به آینده منتقل و حتی شاید آنها را تشدید کند. این پژوهش نشان می‌دهد که سامانه‌های هوش مصنوعی به‌جای حل نابرابری‌های اجتماعی، آنها را به‌طور سیستماتیک بازتولید و تشدید می‌کنند (Chen, 2023).

مشابه با همین رهیافت، امیلیو فرارا[1] در مقاله‌ای با عنوان «عدالت و سوگیری در هوش مصنوعی»، سوگیری را خطای سیستماتیک در فرایندهای تصمیم‌گیری تعریف می‌کند که به نتایج ناعادلانه می‌انجامد. این مقاله مستندات گسترده‌ای از نحوه شکل‌گیری سوگیری در سامانه‌های هوش مصنوعی ارائه می‌دهد که سبب به‌وجود آمدن الگوهای تبعیض سیستماتیک در حوزه‌های حیاتی مانند عدالت

1. Emilio Ferrara

کیفری، تشخیص‌های پزشکی و استخدام می‌شود (Ferrara, 2024).

معضل دیگری که موجب بدبینی برخی اندیشمندان نسبت به توسعه هوش مصنوعی از نظر اخلاقی شده است، عوارض کاربرد این فنّاوری بر افول اراده انسانی و توان تصمیم‌گیری اوست. احمد و همکارانش در مقاله‌ای با عنوان «تأثیر هوش مصنوعی بر کاهش توان تصمیم‌گیری انسانی» شواهد تجربی از آثار منفی هوش مصنوعی بر توانایی‌های شناختی انسان ارائه می‌دهند. این پژوهش نشان می‌دهد که وابستگی فزاینده به هوش مصنوعی به موارد زیر می‌انجامد (Ahmad et al., 2023):

ـ از دست رفتن توان تفکر انتقادی و حل مسئله خلّاقانه؛

ـ افت مهارت‌های حرفه‌ای در مواقع نبود کمک هوش مصنوعی؛

ـ جایگزینی سیستماتیک تصمیم‌گیری خودمختار انسانی با انتخاب‌های الگوریتمی.

این تحقیق نمونه‌ای از دیدگاه‌های بدبینانه درباره توسعه هوش مصنوعی است که بر تغییرات اساسی در کاهش رشد شناختی انسان و افول زیستی ـ شناختی ظرفیت‌های تصمیم‌گیری او تأکید می‌کند.

برخی دیگر از استدلال‌های بدبینانه به پیامدهای بلندمدت پیش‌بینی‌ناپذیر استناد می‌کنند. مقاله «چرا به‌کارگیری اخلاق هوش مصنوعی کافی نیست» به روابط غیرخطی و پیامدهای پیش‌بینی‌ناپذیر ــ و در مواردی برخلاف اهداف اولیه ـ در پیاده‌سازی سامانه‌های هوش مصنوعی اشاره می‌کند. نویسندگان این مقاله درصددند نشان دهند محدوده متعارف بررسی‌های اخلاق هوش مصنوعی که محدود به آثار این فنّاوری بر اشخاص است (مانند موارد یادشده در بندهای پیشین)، برای توسعه مسئولانه آن کفایت نمی‌کند و پیامدهای نامطلوب بلندمدت و پیش‌بینی‌ناپذیر این فنّاوری مقیاس‌های کلان اقتصادی ـ اجتماعی را نیز متأثر می‌سازد و وضع کلی جوامع انسانی و نظم جهانی در شئون گوناگون ـ از روابط اجتماعی گرفته تا شرایط زیست‌محیطی ـ را نامطلوب خواهد ساخت (Wang & Blok, 2025).

بنابراین ذیل رهیافت بدبینانه مضامین زیر هسته اصلی استدلال‌ها را تشکیل می‌دهند:

ـ برخی تجربه‌های در دسترس از هوش مصنوعی نشان می‌دهند که این فنّاوری (احتمالاً) دارای قابلیت‌های رذیلانه و ضداخلاقی همچون دروغ گفتن و اراده سوء داشتن است؛ درحالی‌که استدلال قاطعی برای اخلاقی شدن آن نیست و این وضعیت توسعه آن را از نظر اخلاقی اصولاً ناموجه می‌سازد.

ـ توسعه هوش مصنوعی عوارض نامطلوبی مانند نهادینه شدن نابرابری‌ها در جوامع و افول اراده و تفکر در انسان‌ها دارد که از نظر اخلاقی به هیچ‌وجه نمی‌تواند موجه باشد و هیچ چیزی آنها را جبران نمی‌کند.

ـ محدوده ملاحظات اخلاقی درباره توسعه هوش مصنوعی، محدود به شئون هنجاری زندگی انسان نیست، بلکه توسعه این فنّاوری به سبب آسیب‌هایی می‌باشد که احتمالاً برای طبیعت و زیست‌بوم حیات دارد و اصل زندگی انسان را تهدید می‌کند؛ از این‌رو توسعه بی‌ملاحظه آن از نظر اخلاقی ناموجه است.

۳. شکاف مسئولیت و مواجهه‌های خوش‌بینانه و بدبینانه با آن

در مقاله «شکاف مسئولیت: نسبت دادن مسئولیت به کنش‌های ماشین‌های یادگیرنده» که از سوی آندریاس ماتیاس در سال ۲۰۰۴ منتشر شد، مفهوم «شکاف مسئولیت» به عنوان یکی از چالش‌های اساسی در اخلاق هوش مصنوعی مطرح شده است. ماتیاس استدلال می‌کند که با افزایش توانایی سیستم‌های هوشمند در یادگیری و تصمیم‌گیری مستقل، امکان پیش‌بینی کامل رفتار آنها از سوی طراحان یا کاربران کاهش می‌یابد. این امر به وضعیتی می‌انجامد که در آن هیچ‌کس ــ نه برنامه‌نویس، نه کاربر و نه خود ماشین ـ نمی‌تواند به طور مشخص، مسئول پیامدهای تصمیمات سیستم شناخته شود. از نظر ماتیاس این شکاف مسئولیت، پرسش‌های عمیقی درباره پاسخگویی اخلاقی و حقوقی در عصر فنّاوری‌های خودمختار ایجاد می‌کند و نیازمند بازنگری در چارچوب‌های سنتی مسئولیت‌پذیری است (Matthias, 2004).

مقاله ماتیاس، یک متن کلاسیک برای معضل شکاف مسئولیت به شمار می‌رود. از زمان انتشار این مقاله تاکنون، آثار گوناگونی در راستای پاسخ به آن ارائه شده، و هر دو رهیافت خوش‌بینانه و بدبینانه در آنها نمایان است. شاید دیدگاه ماتیاس مصداق رهیافت بدبینانه قلمداد شود؛ زیرا دیدگاه او، درنهایت این جمع‌بندی را بازگو می‌کند که یا باید قابلیت انتساب مسئولیت را حفظ کرد و از ورود سیستم‌های یادگیرنده به جامعه جلوگیری کرد و یا اجازه داد این سیستم‌ها وارد زندگی فردی و اجتماعی انسان‌ها شوند و پذیرفت که در موارد بسیار (برخلاف گذشته) نمی‌توان مسئول اخلاقی یک اتفاق نادرست را معین کرد (Matthias, 2004: 182–183).

مقاله «ربات‌های قاتل» اسپارو، نمونه صریح‌تری از رهیافت بدبینانه نسبت به معضل شکاف مسئولیت است. اسپارو در این مقاله، یکی از پروژه‌های ارتش ایالات متحده برای توسعه ربات‌های هوشمند نظامی را مورد توجه قرار می‌دهد؛ پروژه «سامانه‌های رزمی آینده» ارتش ایالات متحده که هدف آن ساخت یک «ارتش رباتیک» است. این مقاله به بررسی اخلاقی تصمیم اعزام ربات‌های هوشمند به میدان جنگ می‌پردازد و این پرسش را مطرح می‌کند که در صورت ارتکاب جنایت جنگی ـ به معنای شناخته‌شده حقوقی ـ از سوی یک سامانه تسلیحاتی خودمختار، چه کسی باید مسئول شناخته شود. او استدلال می‌کند که نه طراحان و مهندسان، نه فرماندهان و نه حتی خود سامانه نمی‌تواند مسئول ارتکاب آن جنایت باشد؛ درحالی‌که یکی از شرایط ضروری مشروعیت یک جنگ این است که بتوان فردی را به‌طور عادلانه مسئول مرگ‌هایی دانست که در آن جنگ رخ می‌دهد. او نتیجه می‌گیرد از آنجا که این شرط در مورد مرگ‌هایی قابل تحقق نیست که از سوی سامانه‌های تسلیحاتی خودمختار رخ می‌دهد، توسعه و استفاده از چنین سامانه‌هایی در جنگ از نظر اخلاقی ناپذیرفتنی خواهد بود (Sparrow, 2007).

برخی محققان با رهیافتی خوش‌بینانه به معضل مسئولیت پرداخته‌اند. نمونه اخیر آن پیشنهاد متفاوت

داناهر درباره معضل مسئولیت است. داناهر که پیش‌تر به دیدگاه خوش‌بینانه‌اش در سال ۲۰۱۹ اشاره شد، معضل شکاف مسئولیت را یک فرصت معرفی می‌کند که به جای نگرانی، باید به استقبال آن رفت و آن را یکی از مزایای ماشین‌های خودران دانست که به ما امکان می‌دهند برخی انواع خاص از شکاف‌های مسئولیت را بپذیریم. این دیدگاه بر پایه چنین ایده‌ای استوار است که «اخلاق انسانی اغلب تراژیک است». ما با موقعیت‌هایی روبه‌رو می‌شویم که در آن ملاحظات اخلاقی متضاد در جهت‌های مختلف کشیده می‌شوند و رسیدن به تعادلی کامل میان آنها غیرممکن است و این امر بار مسئولیت تصمیم‌گیری را سنگین‌تر می‌کند. ما با تراژدی انتخاب اخلاقی به شیوه‌های مختلفی کنار می‌آییم: گاهی خود را فریب می‌دهیم و باور می‌کنیم که انتخاب‌مان تراژیک نبوده است (توهم‌گرایی)، گاهی تصمیم دشوار را به دیگری واگذار می‌کنیم (واگذاری)، و گاهی خودمان تصمیم می‌گیریم و پیامدهای روانی آن را می‌پذیریم (مسئولیت‌سازی). هر یک از این راهبردها مزایا و البته هزینه‌های خاص خود را دارند و تأکید داناهر بر این هزینه‌هاست. یکی از مزایای بالقوه ماشین‌های خودران این است که امکان واگذاری با هزینه روانی کمتر را فراهم می‌کنند (Danaher, 2022).

نمونه دیگر از رهیافت خوش‌بینانه نسبت به معضل شکاف مسئولیت که برای حل آن بر تغییر چارچوب اخلاقی سنت غربی تأکید می‌کند، مقاله «شکاف‌های مسئولیت ناشی از ورود هوش مصنوعی به سلامت: رهیافتی مبتنی بر اوبونتو[1]» است (Ferlito et al., 2024). نویسندگان استدلال می‌کنند که چگونه بهره‌گیری از فلسفه آفریقایی اوبونتو می‌تواند به حل معضل شکاف مسئولیت در حوزه سلامت کمک کند.

آنها توضیح می‌دهند در چارچوب‌های فلسفه اخلاقی غربی، مسئولیت معمولاً فردی است و این دیدگاه در برابر تصمیم‌گیری‌های پیچیده و غیرشفاف AI ناکارآمد است؛ درحالی‌که در فلسفه اوبونتو که بر همبستگی، وابستگی متقابل و مسئولیت جمعی تأکید دارد، دیدگاهی متفاوت ارائه می‌شود و به جای تمرکز بر مقصر، بر مراقبت مشترک و پاسخگویی جمعی تأکید دارد. آنها اوبونتو را به عنوان یک ایده فرهنگی و رفتاری، بلکه به مثابه یک چارچوب اخلاق عملی معرفی می‌کنند که مسئولیت را به‌صورت آینده‌نگر و رابطه‌محور تعریف می‌کند (Ferlito et al., 2024: 34).

نویسندگان بر این اساس تأکید دارند در نظام سلامت که تصمیم‌ها اغلب حاصل تعامل چندین عامل هستند، این دیدگاه بسیار کارآمد است و به معنای طراحی سیستم‌هایی می‌باشد که شفاف، مشارکتی و منطبق با ارزش‌های اجتماعی هستند. آنها معتقدند اوبونتو چارچوبی اخلاقی را ارائه می‌دهد که براساس آن مسئولیت میان توسعه‌دهندگان، پزشکان، بیماران و نهادها تقسیم می‌شود (Ferlito et al., 2024: 35). در پایان مقاله تأکید می‌شود که اوبونتو می‌تواند به ایجاد «زیست‌بوم اخلاقی» کمک کند؛ محیطی که در

1. Ubuntu

آن مسئولیت از طریق روابط انسانی و مراقبت متقابل پرورش می‌یابد. در این پیشنهاد که مبتنی بر فلسفه اخلاق اوبونتو ارائه می‌شود، «اخلاقی بودن» وصف زیست‌بومی است که در آن کنشگران انسانی، توسعه‌دهندگان ابزارها، سرمایه‌گذارها و مدیران و نهادها به صورت توأمان فعالیت دارند و نه وصف افراد منفرد انسانی. این چارچوب اخلاقی نه‌تنها شکاف مسئولیت را کاهش می‌دهد، بلکه اعتماد و همبستگی را در نظام سلامت تقویت می‌کند.

استفاده از برخی شواهد علوم شناختی مبنی بر اینکه انسان‌ها نیز با چالش‌های مشابهی در آگاهی از آنچه انجام می‌دهند و توانمندی کنترل آن روبه‌رو هستند، زمینه مناسبی برای ارائه نمونه دیگری از رهیافت خوش‌بینانه به معضل شکاف مسئولیت را فراهم کرده است. نویسندگان مقاله «یافتن شکاف: هوش مصنوعی، عاملیت مسئولانه و آسیب‌پذیری» ابهام معرفتی و کنترل ضعیف بر رفتار را منحصر به هوش مصنوعی نمی‌دانند (اگرچه با هوش مصنوعی تشدید شوند). آنها تلاش می‌کنند نشان دهند که می‌توان برای توسعه هوش مصنوعی از نحوه بازنگری اخیر فیلسوفان در مفهوم سنتی مسئولیت اخلاقی در قالب دیدگاه‌های ابزاری مسئولیت درس‌های مهمی گرفت که بر نقش آینده‌نگر و انعطاف‌پذیر پرورش عاملیت تأکید دارند (Vallor & Vierkant, 2024).

براساس این مرور، استدلال‌های زیر مضامین اصلی بدبینی نسبت به معضل شکاف مسئولیت را بازگو می‌کنند:

ـ معضل شکاف مسئولیت پاسخی ندارد؛ درحالی‌که اقتضای اخلاق، قابلیت سرزنش و انتساب مسئولیت اخلاقی است. بنابراین با توجه به آنکه هوش مصنوعی ناگزیر به بروز شکاف‌های مسئولیت در حوزه‌های گوناگون می‌انجامد، توسعه آن موجه نیست.

ـ شکاف مسئولیت در برخی روندهای تحقیقاتی و توسعه‌ای در حوزه هوش مصنوعی همچون توسعه سامانه‌های تسلیحاتی خودکار، با اصول اخلاقی مسلم در تعارض است؛ از این‌رو پشتیبانی از این روندها از نظر اخلاقی ناموجه است.

همچنین این مرور نشان می‌دهد برخی محققان براساس استدلال‌های زیر، موضعی خوش‌بینانه نسبت به توسعه هوش مصنوعی از نظر شکاف مسئولیت دارند:

ـ برخلاف برداشت سنتی درباره اخلاق و ضرورت قابلیت سرزنش و قبول مسئولیت اخلاقی، شکاف مسئولیت یک امکان اخلاقی برای کاهش برخی هزینه‌ها و فشارهای روانی است و از این دیدگاه می‌تواند از نظر اخلاقی موجه باشد؛ پس بدین‌گونه نیست که هوش مصنوعی، تنها به این سبب که معضل شکاف مسئولیت را تشدید می‌کند از نظر اخلاقی ناموجه باشد.

ـ ابهام در آگاهی و معرفت به آنچه انجام می‌دهیم و ناتوانی در کنترل افعال، انحصار به هوش مصنوعی ندارد و درباره انسان نیز صادق است؛ از این‌رو یکی از مبانی اساسی استدلال‌های اخلاقی

بدبینانه نسبت به توسعه هوش مصنوعی براساس تشدید شکاف مسئولیت، مخدوش است و به این ترتیب بدبینی موجه نیست.

در ادامه تلاش می‌شود در سطحی بالاتر از این استدلال‌ها، تحلیلی از معقولیت خوش‌بینی یا بدبینی نسبت به توسعه هوش مصنوعی براساس معضل شکاف مسئولیت ارائه شود.

٤. معضل شکاف مسئولیت و تحلیل معقولیت رهیافت‌های خوش‌بینانه و بدبینانه

در بخش‌های اول و دوم این بررسی، برخی استدلال‌های خوش‌بینانه و بدبینانه نسبت به توسعه هوش مصنوعی مرور و روشن شد که با تنوعی از مضامین در این استدلال‌ها مواجهیم. در بخش سوم، این تنوع به طور مشخص درباره معضل شکاف معناداری نشان داده شد. با توجه به تنوع مضامین و تفاوت‌ها در مبانی استدلال‌ها، به نظر می‌رسد که تحلیلی سطح بالاتر درباره میزان معقولیت رهیافت‌های خوش‌بینانه و بدبینانه به توسعه هوش مصنوعی بتواند به همگرایی تلاش‌های محققان کمک کند.

از منظری کلان‌تر، ریشه اختلاف نظر میان استدلال‌های رهیافت‌های خوش‌بینانه و بدبینانه به تفاوت دیدگاه‌ها درباره سه مقوله «مبنای خیر اخلاقی»، «آسیب‌پذیری انسان» و «قابلیت‌های هوش مصنوعی» بازمی‌گردد؛ به طور مشخص:

ـ اینکه در حوزه اخلاق، چه مبنایی مدنظر است و خیر اخلاقی چیست: آیا قابلیت سرزنش و قبول مسئولیت برای اخلاق ضروری است (Matthias, 2004) یا در مواردی مثل مشروعیت جنگ ضروری است (Sparrow, 2007)؟ یا لزوماً این‌گونه نیست (Danaher, 2022) و مثلاً خیر اخلاقی ریشه در مشارکت و مراقبت دارد (Ferlito et al., 2024)؟

ـ اینکه هوش مصنوعی از دیدگاه متافیزیکی چه قابلیت‌هایی دارد: آیا قابلیت‌هایی متمایزکننده نسبت به سایر فنّاوری‌ها دارد که از نظر اخلاقی موجه نیست (Chen, 2023; Ferrara, 2024; Prentice, 2025; Wang & Blok, 2025)؟ یا یک فنّاوری متعارف است که در مجموع زندگی را بهبود می‌بخشد (Strain, 2024) و نمی‌توان قاطعانه مدعی شد تمایزی اساسی با سایر فنّاوری‌ها دارد (Danaher, 2019b; Narayanan & Kapoor, 2025a; Schwaller, 2025; Thaiduong, 2025)؟

ـ اینکه انسان چقدر در برابر هوش مصنوعی آسیب‌پذیر است: آیا قابلیت‌های انسان (مثلاً قابلیت‌های شناختی) در نتیجه رواج استفاده از هوش مصنوعی افول می‌کند (Ahmad et al., 2023)؟ یا آنکه این قابلیت‌ها به صورت بالقوه در طبیعت انسانی موجود است و تغییری اساسی در آنها پیش نمی‌آید و حتی برعکس هوش مصنوعی سبب افزایش قابلیت‌های انسان و بهینه‌تر شدن زندگی او می‌شود (Wada & Shibata, 2007)؟

به نظر می‌رسد که در ادبیات تخصصی اخلاق هوش مصنوعی، در حال حاضر درباره دوگانه‌های

یادشده جمع‌بندی روشنی وجود ندارد. درباره مبانی اخلاقی اختلاف نظر در مواضع به طور کامل مستقل از قابلیت‌های هوش مصنوعی است. اینکه هوش مصنوعی از دیدگاه متافیزیکی چه قابلیت‌هایی دارد (یا خواهد داشت) از مسائل باز فلسفه هوش مصنوعی است؛ به بیان دیگر به نظر می‌رسد که درباره این مسئله مشابه مسائلی مانند آگاهی، حیث التفاتی در فلسفه ذهن، و حیات در فلسفه زیست‌شناسی، همیشه اختلاف نظر وجود داشته باشد. دوگانه سوم نیز همواره پاسخ بینابینی خواهد داشت، به این معنا که هوش مصنوعی سبب تقویت برخی قابلیت‌های انسان و تضعیف برخی دیگر می‌شود.

هنگامی که هیچ‌یک از این سه محور پاسخ قطعی ندارند، تمامی استدلال‌های کنونی صرفاً بخشی از دغدغه‌ها و مسائل را نمایندگی می‌کنند؛ از این‌رو تعمیم آنها و اتخاذ یک موضع خوش‌بینانه یا بدبینانه درباره توسعه هوش مصنوعی، تعمیمی نابجا است. پس باید این پرسش را به صورت مستقل پاسخ داد: از نظر معضل شکاف مسئولیت، رهیافت خوش‌بینانه به هوش مصنوعی معقول‌تر است یا رهیافت بدبینانه؟

استدلال مستقلی که در اینجا ارائه می‌شود، مبتنی بر تحلیل ماهیت مسئولیت‌پذیری و قابلیت سرزنش و تفاوت قضاوت‌پذیری اخلاقی و اخلاقی بودن با تأکید بر قابلیت‌های هوش مصنوعی است:

۱) قابلیت سرزنش و قبول مسئولیت لازمه قضاوت اخلاقی است و نه اخلاقی بودن؛ به بیان دقیق‌تر، فرض اینکه انسان‌ها یا هر موجود دیگری برخوردار از این ویژگی‌ها باشد و درباره آنها شکاف مسئولیت وجود نداشته باشد (که درباره هر دو مورد می‌توان تردید داشت)، صرفاً قضاوت درباره مسئولیت اخلاقی را مشروع می‌کند. در واقع می‌توان وضعیتی اخلاقی را تصور کرد که عوامل آن قابلیت سرزنش نداشته باشند.

۲) در همین راستا و افزون بر این به نظر نمی‌رسد که کنترل عوامل هوشمند از کنترل عوامل انسانی دشوارتر باشد؛ به بیان دقیق‌تر، انسان برخوردار از اراده آزاد، باز هم از نظر اخلاقی کنترل‌ناپذیر است و قضاوت‌پذیر بودن او سبب اخلاقی بودن او نمی‌شود. پس اضافه شدن عوامل هوشمند به جهان و جامعه از نظر اخلاقی، با فرض آنکه این عوامل نسبت به آنچه انجام می‌دهند معرفت ندارند ـ که فرض مورد نیاز برای اشاره به شکاف مسئولیت است ـ تفاوتی با به دنیا آمدن فرزندان و افزوده شدن انسان‌های جدید به جامعه و جهان ندارد.

۳) بر این اساس رهیافت بدبینانه نسبت به توسعه هوش مصنوعی دچار یک تناقض درونی است؛ توضیح بیشتر آنکه:

ـ براساس مقدمه اول، بدبینی نسبت به توسعه هوش مصنوعی با تأکید بر معضل شکاف معناداری، وضعیت اخلاقی را تضمین یا حتی محتمل‌تر نمی‌کند.

ـ به این ترتیب بدبینی اخلاقی نسبت به توسعه هوش مصنوعی می‌بایست معطوف به اخلاقی بودن عوامل هوشمند باشد، نه قضاوت‌پذیری آنها.

ـ اما اگر بدبینی اخلاقی به توسعه هوش مصنوعی بر چنین مبنایی قرار داده شود ـ صرف امکان گسترش معضلات اخلاقی ـ آنگاه براساس مقدمه دوم می‌بایست نسبت به دنیا آمدن انسان‌ها نیز بدبین بود و آن را از نظر اخلاقی موجه ندانست.

ـ به این ترتیب یک دلالت رهیافت بدبینانه نسبت به توسعه هوش مصنوعی برای صیانت از امر هنجاری اخلاق و شأنی از شئون زندگی انسان اتخاذ شد، نوعی بدبینی نسبت به اصل زندگی انسان را در پی دارد؛ به بیان دقیق‌تر، رهیافت بدبینانه دچار چنین تناقضی است که برای حفظ شأنی هنجاری از زندگی انسان به نام اخلاق، اصل آن را محدود می‌سازد.

۴) از این‌رو رهیافت بدبینانه نسبت به توسعه هوش مصنوعی با تأکید بر معضل شکاف مسئولیت به سبب دچار بودن به تناقض درونی و مستقل از دوگانه‌های حل‌ناشدنی یادشده نامعقول است.

بر این اساس رهیافت‌های خنثی و خوش‌بینانه نسبت به توسعه هوش مصنوعی با تأکید بر معضل شکاف مسئولیت نسبت به رهیافت بدبینانه ترجیح می‌یابند، ولی با توجه به نتیجه این استدلال (ترجیح ترکیب عطفی رهیافت‌های خنثی و خوش‌بینانه)، با وجود نامعقول دانستن رهیافت بدبینانه، احتیاط نسبت به عوارض سوء مرتبط با شکاف مسئولیت درباره برخی پروژه‌های خاص کنار گذاشته نمی‌شود.

نتیجه‌گیری

مقایسه منسجم دو رهیافت خوش‌بینانه و بدبینانه به توسعه هوش مصنوعی و بررسی معضل شکاف مسئولیت نشان داد که هر یک از استدلال‌ها تنها بخشی از واقعیت را نمایندگی می‌کنند و قادر به ارائه پاسخ قطعی به معیارهای اساسی ـ مبنای خیر اخلاقی، قابلیت‌های متافیزیکی و عملی هوش مصنوعی و آسیب‌پذیری انسان ـ نیستند.

رهیافت خوش‌بینانه مزایای گسترده فنّاوری، فهم تخصصی توسعه‌دهندگان و عدم‌قطعیت‌های متافیزیکی را برجسته می‌کند؛ درحالی‌که رهیافت بدبینانه بر سوگیری‌های الگوریتمی، افول اراده و پیامدهای اجتماعی ـ اقتصادی پیش‌بینی‌ناپذیر تأکید دارد. این تنوع مقدمات و اهداف، نیاز به تحلیلی مستقل از دوگانه‌های حل‌ناشدنی را آشکار ساخت.

تحلیل مستقل معطوف به ماهیت مسئولیت‌پذیری، تفاوت قضاوت‌پذیری اخلاقی با اخلاقی بودن را نشان داد و اینکه شکاف مسئولیت در هوش مصنوعی از شکاف مسئولیت در خود انسان‌ها پیچیده‌تر نیست. کنترل‌ناپذیری و ابهام معرفتی ویژگی مشترک همه فاعل‌های اخلاقی است و حتی با فرض کنار گذاشتن این انتقادات، اضافه شدن موجودات هوشمند جدید از نظر اخلاقی بودن وضعیت، تفاوت بنیادینی با ورود انسان‌های جدید به جهان ندارد.

بر این اساس نشان داده شد که بدبینی اخلاقی که توسعه هوش مصنوعی را به سبب گسترش شکاف

مسئولیت ناموجه می‌داند، در درون خود از تناقض رنج می‌برد؛ زیرا اگر قابلیت سرزنش شرط اخلاقی بودن است، باید نسبت به تولد انسان‌ها نیز بدبین بود. پس ترکیب رویکردهای خنثی و خوش‌بینانه در مواجهه با معضل شکاف مسئولیت از نظر منطقی برتر و معقول‌تر از بدبینی مطلق است. این نتیجه‌گیری نامعقول بودن رهیافت بدبینانه را در کلیت آن و از این طریق، ترجیح رهیافت‌های خوش‌بینانه و خنثی را نشان می‌دهد، اما بی‌احتیاطی نسبت به آثار سوء معضل شکاف مسئولیت در برخی پروژه‌های خاص را تجویز نمی‌کند.

ملاحظات اخلاقی

انجام این پژوهش به صورت مستقل بوده و در اجرای آن از حمایت مالی مؤسسه خاصی استفاده نشده است؛ همچنین نویسنده تعارض منافعی نداشته‌است.

منابع

Ahmad, S. F.; Han, H.; Alam M. M.; Rehmat, M.; Irshad, M.; Arraño-Muñoz, M.; & Ariza-Montes, A. (2023). Impact of artificial intelligence on human loss in decision making, laziness and safety in education. *Humanities and Social Sciences Communications*, 10(1), 1–14. https://doi.org/10.1057/s41599-023-01787-8

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications*, 10(1), 1–12. https://doi.org/10.1057/s41599-023-02079-x

Danaher, J. (2019a). The Philosophical Case for Robot Friendship. *Journal of Posthuman Studies*, 3(1), 5–24. https://doi.org/10.5325/jpoststud.3.1.0005

Danaher, J. (2019b). The Philosophical Case for Robot Friendship. *Journal of Posthuman Studies*, 3(1), 5–24. https://doi.org/10.5325/jpoststud.3.1.0005

Danaher, J. (2022). Tragic choices and the virtue of techno-responsibility gaps. *Philosophy & Technology*, 35(2), 26. https://doi.org/10.1007/s13347-022-00519-1

Ferlito, B., Segers, S., De Proost, M., & Mertes, H. (2024). Responsibility Gap(s) Due to the Introduction of AI in Healthcare: An Ubuntu-Inspired Approach. *Science and Engineering Ethics*, 30(4), 34. https://doi.org/10.1007/s11948-024-00501-4

Ferrara, E. (2024). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci*, 6(1), 3. https://doi.org/10.3390/sci6010003

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 175–183. https://doi.org/10.1007/s10676-004-3422-1

Narayanan, A., & Kapoor, S. (2025a). *AI as normal technology*. Knight First Amend. Inst. https://thedocs.worldbank.org/en/doc/d6e33a074ac9269e4511e5d44db2f9ac-0050022025/original/AI-as-Normal-Technology-Narayanan-Kapoor-Final.pdf

Narayanan, A., & Kapoor, S. (2025b). *AI as normal technology: An alternative to the vision of AI as a potential superintelligence*. Knight First Amendment Institute, Columbia University, https://kfai-documents.s3.amazonaws.com/Documents/C3cac5a2a7/AI-as-Normal-Technology%E2%80%94Narayanan%E2%80%94Kapoor.Pdf

Prentice, R. (2025). *Techno-Optimist or AI Doomer? Consequentialism and the Ethics of AI*. Ethics Unwrapped. https://ethicsunwrapped.utexas.edu/techno-optimist-or-ai-doomer-consequentialism-and-the-ethics-of-ai

Schwaller, F. (2025). Will AI improve your life? Here's what 4,000 researchers think. *Nature*, 640(8059), 577–578. https://doi.org/10.1038/d41586-025-01123-x

Sparrow, R. (2007). Killer Robots. *Journal of Applied Philosophy*, 24(1), 62–77. https://doi.org/10.1111/j.1468-5930.2007.00346.x

Strain, M. R. (2024, Summer). *The Case for AI Optimism*. National Affairs, 60. https://nationalaffairs.com/publications/detail/the-case-for-ai-optimism

Thaiduong, N. (2025). IT Professionals Versus the Public: Who's More Optimistic About AI's

Future Impacts? *SAGE Open*, 15(2). https://doi.org/10.1177/21582440251348802

Vallor, S., & Vierkant, T. (2024). Find the Gap: AI, Responsible Agency and Vulnerability. *Minds and Machines*, 34(3), 20. https://doi.org/10.1007/s11023-024-09674-0

Wada, K., & Shibata, T. (2007). Living with seal robots—Its sociopsychological and physiological influences on the elderly at a care house. *IEEE Transactions on Robotics*, 23(5), 972–980. https://doi.org/10.1109/TRO.2007.906261

Wang, H., & Blok, V. (2025). Why putting artificial intelligence ethics into practice is not enough: Towards a multi-level framework. *Big Data & Society*, 12(2). https://doi.org/10.1177/20539517251340620