Journal of Philosophical Investigations
Print ISSN: 2251-7960 Online ISSN: 2423-4419
Homepage: https://philosophy.tabrizu.ac.ir

Pragmatism, the Ethics of Artificial Intelligence and Sociotechnical Systems

Mohammad-Ali Ashouri-Kisomi[1] | **Maryam Parvizi**[2] | **Seyyed-Mohammad-Hassan Momeni**[3]

1. Corresponding Author, Ph.D. in Philosophy, Allameh Tabataba'i University, Tehran, Iran. E-mail: m_ashori@atu.ac.ir
2. Ph.D. in Comparative Philosophy and Research Expert, Allameh Tabataba'i University, Tehran, Iran. E-mail: maryam_parvizi@atu.ac.ir
3. Assistant Professor of Theology and Islamic Studies Department, Allameh Tabataba'i University, Tehran, Iran. E-mail: dr.smhmomeni@atu.ac.ir

Article Info

Article type:
Research Article

Article history:
Received 11 April 2025
Received in revised from 18 May 2025
Accepted 10 July 2025
Published online 00 November 2025

Keywords:
Artificial Intelligence, Sociotechnical Systems, Pragmatism, Interdisciplinary Study.

ABSTRACT

This paper investigates the ethical evaluation of artificial intelligence (AI) through the sociotechnical systems approach, with a particular emphasis on the insights of pragmatist philosophers. In recent years, the sociotechnical perspective has gained attention among AI ethics researchers. However, this approach has been interpreted and discussed in various ways, which can be broadly categorized into three main narratives. This paper introduces and critically examines all three, with a primary focus on the pragmatist narrative. Given the increasing relevance and relative novelty of this approach, a thorough evaluation is necessary to understand its potential and limitations. To this end, the study adopts an analytical-critical methodology to assess the framework and propose constructive solutions. Our analysis indicates that the pragmatist narrative of the sociotechnical systems approach emphasizes the real-world functioning of AI and its societal impacts. It seeks to identify the diverse social, technical, and institutional factors that contribute to ethical and societal outcomes. One of the central challenges of this approach lies in the inherent complexity of sociotechnical systems and the multitude of interacting elements they encompass. In response to this challenge, the paper draws on the work of pragmatist philosophers to propose interdisciplinary inquiry as an appropriate and effective method for studying sociotechnical systems. The findings suggest that, despite the complexity involved, this approach offers a viable path for comprehensively analyzing and ethically assessing AI technologies within their broader social contexts.

Cite this article: Ashouri-Kisomi, M. A.; Parvizi. M. & Momeni, S. M. H. (2025). Pragmatism, the Ethics of Artificial Intelligence and Sociotechnical Systems. *Journal of Philosophical Investigations*, 19 (52), 281-300. https://doi.org/10.22034/jpiut.2025.66643.4060

© The Author(s).
Publisher: University of Tabriz.

Extended Abstract

Introduction

The rapid development of AI has attracted increasing attention from humanities scholars. While some discussions focus on speculative risks such as superintelligent AI and existential threats (Bostrom, 2014; Russell, 2019), a substantial body of research (Prem, 2023; Levesque, 2017; Van de Poel, 2023) is concerned with guiding the responsible development of AI and reducing its potential harms. One approach that has recently gained traction is the sociotechnical systems perspective (Watson et al., 2024; Young & Coeckelbergh, 2024; Pfeiffer et al., 2023), which emphasizes that AI systems are part of larger sociotechnical systems (Van de Poel, 2020; Kudina & Van de Poel, 2024). While this approach offers a more nuanced framework for evaluating AI, its complexity poses challenges for ethical analysis. This paper explores various interpretations of the sociotechnical approach, critically assesses its limitations, and proposes a method for ethical evaluation grounded in this framework.

Discussion

Taking sociotechnical systems into account has specific advantages, providing us with a more precise understanding. However, this idea can be interpreted in multiple ways, leading to three general narratives. Dogmatic narrative focuses on the economic and social benefits of technology while often ignoring potential negative impacts (Watson et al., 2024). It assumes that technological advancements lead to social and economic progress, and thus, efforts to slow down or prevent technological progress is unethical. It is undeniable that technological advancement brought tremendous benefits to society. For instance, social media has played a significant role in increasing public awareness of climate change. However, we must not overlook the physical infrastructures that support these platforms. They can contribute to greenhouse gas emissions due to the large amounts of data they process, potentially becoming a cause of climate change. That is, in the meantime, the utopian future in mind doesn't resolve existing concerns and may even contribute to more harm in the future.

Contrary to the dogmatic narrative, the skeptical narrative places emphasis on the negative consequences of AI technology (Zuboff, 2023; Susser et al., 2019; Manheim & Kaplan, 2019). However, some skeptics view technology as the cause of social issues and reject the idea of improving technology to address these problems (Hoffmann, 2019). The downside of skepticism is that it can be overly simplistic and counterproductive as it advocates for restriction while not offering any constructive component. For instance, once again we can refer to the previous example and argue that social media technologies have contributed to improving public awareness of climate change, which skepticism tends to overlook. All in all, though the former narrative is reductive, the latter is limiting (Watson et al. 2024), and they present an incomplete picture of sociotechnical systems.

The pragmatic narrative offers a more balanced perspective, recognizing the limitations of both the dogmatic and skeptical narratives. It emphasizes progress, human agency, and the iterative nature of technology, encouraging collaboration across disciplines to address complex issues. This

narrative acknowledges uncertainty in the effects of AI on society, but focuses on adapting technology to meet evolving needs. Pragmatism supports the idea that technology and values co-evolve (Van de Poel & Kudina, 2022). In contrast to the deterministic view of the dogmatic narrative, the pragmatic approach recognizes human limitations but highlights the role of human agency in shaping technological advancement. This narrative emphasizes progress without aiming for perfection, acknowledging uncertainty and focusing on solutions that are adaptable to real-world problems.

The need for interdisciplinary study

Sociotechnical systems are made up of many interwoven elements. Paying close attention to these layers is essential, but it also complicates how we address ethical questions. The pragmatist approach encourages us to engage with this complexity head-on, yet it often leaves us without clear guidance for navigating the moral tensions that arise. We're asked to face the system, but not always shown how to respond to its ethical depth.

To move forward, and given the involvement of various factors in sociotechnical systems, it is clear that evaluating AI technology requires a problem-driven, interdisciplinary approach that engages with several factors. One critical point is that no single discipline is sufficient to address these issues (Decker, 2004). The complexity of AI systems calls for interdisciplinary research to understand their ethical implications, as focusing on one discipline alone may lead to unintended consequences (Schmidt, 2011; Alvargonzález, 2011; Maki, 2016). Thus, a pragmatic approach to interdisciplinary study is necessary to navigate the challenges posed by AI and ensure that solutions are aligned with societal norms and values. This approach focuses on continuous collaboration across disciplines and the need for adaptive solutions to meet the challenges posed by AI in a dynamic, evolving social landscape. In this regard, we proposed a framework for a pragmatist interdisciplinary study.

Conclusion

The results of this paper indicate that, given the various factors influencing the design and use of artificial intelligence technology, a pragmatic interdisciplinary study could be considered as a solution for a more thorough examination of AI systems. Such a study could identify the various impacts and outcomes related to AI ethics, while simultaneously preventing the complexity of the system under study from becoming overwhelming. Furthermore, an interdisciplinary study allows for the selection of a technology control method based on objective and empirical conditions. Using this method, influential factors can be identified based on context and application, and control and evaluation strategies can be assessed not only proactively but also, to some extent, reactively. Another advantage of using a pragmatic interdisciplinary study in sociotechnical systems is that it enables the combination of control and evaluation methods according to the specific application context. Relying on such an approach, the theories and control methods used can evolve over time based on the selected context, providing the opportunity to benefit from technology while avoiding its harms.

پژوهش‌های فلسفی
شاپا چاپی: ۷۹۶۰-۲۲۵۱ شاپا الکترونیکی: ۴۴۱۹-۲۴۲۳
Homepage: https://philosophy.tabrizu.ac.ir

پراگماتیسم، اخلاق هوش مصنوعی و سیستم‌های اجتماعی–فنی

محمدعلی عاشوری کیسمی[۱] | مریم پرویزی[۲] | سیدمحمدحسن مومنی[۳]

۱ نویسنده مسئول، دانش آموخته دکتری فلسفه، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه:m_ashori@atu.ac.ir
۲. دانش آموخته دکتری فلسفه تطبیقی و کارشناس پژوهش، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه:maryam_parvizi@atu.ac.ir
۳. استادیار گروه فقه و مبانی حقوق اسلامی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: dr.smhmomeni@atu.ac.ir

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۴/۰۱/۲۲
تاریخ بازنگری: ۱۴۰۴/۰۲/۲۸
تاریخ پذیرش: ۱۴۰۴/۰۴/۱۹
تاریخ انتشار: ۱۴۰۴/۰۹/۰۱

کلیدواژه‌ها:
هوش مصنوعی، اخلاق هوش مصنوعی، سیستم‌های اجتماعی–فنی، پراگماتیسم، مطالعه میان‌رشته‌ای.

چکیده

هدف از مقاله حاضر بررسی رویکرد سیستم‌های اجتماعی–فنی در بررسی اخلاقی هوش مصنوعی با تکیه‌بر نظرات فیلسوفان پراگماتیست است. در سالیان اخیر برخی پژوهشگران به استفاده از رویکرد سیستم‌های اجتماعی–فنی در حوزه هوش مصنوعی روی آورده‌اند. با این حال انواع روایت‌های مختلف از این رویکرد معرفی شده است که می‌توان آنها را به سه دسته اصلی تقسیم‌بندی کرد. در مقاله حاضر، هر سه دسته‌بندی معرفی و موردمطالعه قرار خواهند گرفت؛ البته تمرکز اصلی بر روایت پراگماتیستی خواهد بود. با توجه به اهمیتی که این رویکرد پیدا کرده است و با در نظر گرفتن تازگی آن، نیاز است پژوهشگران مختلف این رویکرد را به‌دقت مورد ارزیابی و مطالعه قرار داده و چالش‌های آن را شناسایی کنند. در راستای رسیدن به هدف اصلی این مقاله، از روش تحلیلی–انتقادی برای بررسی و ارائه راهکار استفاده شده است. بررسی‌های ما نشان می‌دهد که یک روایت پراگماتیستی از سیستم‌های اجتماعی–فنی بر عملکرد و اثرات هوش مصنوعی در جامعه تمرکز می‌کند و عوامل مختلف مؤثر که به آثار و نتایج اخلاقی می‌انجامند را شناسایی می‌کند. با این حال یک چالش اصلی این رویکرد پیچیدگی سیستم و وجود عوامل مختلف در آن است. در این راستا، با تکیه‌بر نظرات فیلسوفان پراگماتیست، مطالعه میان‌رشته‌ای به‌عنوان شیوه بررسی و مطالعه در سیستم‌های اجتماعی–فنی ارزیابی و معرفی شد. نتایج مقاله حاضر نشان می‌دهد، این راه‌کار با وجود عوامل مختلف و پیچیدگی سیستم؛ مطالعه و بررسی آن را امکان‌پذیر می‌کند.

استناد: عاشوری کیسمی، محمدعلی.؛ پرویزی، مریم و مومنی، سیدمحمدحسن. (۱۴۰۴). پراگماتیسم، اخلاق هوش مصنوعی و سیستم‌های اجتماعی–فنی، *پژوهش‌های فلسفی*، ۱۹ (۵۲)، ۳۰۰-۲۸۱. https://doi. org/10.22034/jpiut.2025.66643.4060

ناشر: دانشگاه تبریز.

© نویسندگان.

مقدمه

توسعه هوش مصنوعی[1]در سالیان اخیر به توجه متخصصان در حوزه‌های گوناگون علوم انسانی به این فناوری منجر شده است. چراکه این فناوری در بسیاری از جنبه‌های زندگی انسان حضور داشته و نمی‌توان از اهمیتِ مسائل اخلاقی و اجتماعی آن چشم پوشید. اکنون شاید بتوان گفت دوران هوش مصنوعی خوب قدیمی[2]سپری شده و شاهد فناوری‌های بسیار پیچیده‌تری نسبت به گذشته هستیم. همین امر سبب شده است که توجه به مسائل و چالش‌های این فناوری، به امری ضروری و با اولویت بالا از نظر پژوهشگران فلسفه و اخلاق بدل شود.

مباحث گوناگونی در حوزه اخلاق هوش مصنوعی مطرح شده و پژوهش‌های فراوانی در این حوزه صورت گرفته است. درحالی‌که برخی از مباحث حول موضوعاتی همچون هوش مصنوعی فوق هوشمند[3]یا خطر وجودی[4]این سیستم‌ها برای انسان‌ها است (بوستروم ۲۰۱۴، راسل ۲۰۱۹، فلوریدی ۲۰۱۶، مولر و کانن ۲۰۲۲)؛ بخش عمده‌ای از پژوهشگران چه در سطح بین‌المللی (پرم، ۲۰۲۳؛ لووک، ۲۰۱۷؛ وان دِ پُل، ۲۰۲۳) و چه در میان پژوهشگران فارسی‌زبان (عاشوری کیسمی، ۱۴۰۳ب؛ رمضانی و فیضی درخشی، ۱۳۹۲) به تعیین مسیر مطلوب پیشرفت این فناوری، هم‌سوسازی[5]با ارزش‌های انسانی و کاهش آسیب‌های آن پرداخته‌اند.[6] حال شاید این پرسش را مطرح کنیم که چرا هنوز هم نیاز به تکرار این مباحث از سمت پژوهشگران احساس می‌شود؟ آیا متخصصین طراح و توسعه‌دهنده سیستم‌های هوش مصنوعی اهمیتی به چالش‌های اخلاقی و اجتماعی آن نمی‌دهند؟

یکی از رویکردهایی که در سالیان اخیر برای بررسی مسائل مربوط به حوزه هوش مصنوعی، موردتوجه پژوهشگران قرار گرفته است، رویکرد سیستم‌های اجتماعی-فنی[7]است (واتسون و همکاران، ۲۰۲۴؛ یانگ و کوکل‌برگ، ۲۰۲۴). بر اساس این رویکرد، سیستم‌های هوش مصنوعی، جزئی از سیستم‌های بزرگتر اجتماعی-فنی در نظر گرفته می‌شوند و وجه اجتماعی مؤثر در فناوری نیز موردتوجه قرار می‌گیرد (وان دِ پُل، ۲۰۲۰). آن‌گونه که پژوهشگران طرفدار این رویکرد معتقدند، به این روش دریچه‌های جدیدی برای پژوهش در حوزه اخلاق و فلسفه هوش مصنوعی باز خواهد شد. به‌عبارتی‌دیگر، زمانی که سیستم هوش مصنوعی را به‌عنوان بخشی از سیستم اجتماعی-فنی موردتوجه قرار می‌دهیم، درک کامل‌تری از سیستم هوش مصنوعی، نحوه عملکرد، تأثیرات اجتماعی، اخلاقی، فرهنگی و حاکمیتی آن به‌دست می‌آوریم (کودینا و وان دِ پُل، ۲۰۲۴). یک ارزش‌افزوده این روش آن است که روشن می‌کند عوامل مختلفی درون یک سیستم نقش‌آفرینی می‌کنند و لذا سخن گفتن از ارزش‌های اخلاقی یا اجتماعی تنها منحصر به تصمیمات

---

[1] Artificial Intelligence (AI)

[2] Good old fashioned artificial intelligence (GOFAI): اصطلاحی است که برای اشاره به سیستم‌های هوش مصنوعی کلاسیک و سمبلیک استفاده می‌شود عموماً مقصود سیستم‌های هوش مصنوعی نسل‌های قدیمی (تا اواخر دهه ۱۹۸۰ میلادی) است که قابلیت‌های محدودی داشتند (Muller, 2016) و لذا مباحث اخلاقی و اجتماعی آن‌ها بسیار محدودتر از هوش مصنوعی دوره حاضر بود.

[3] Artificial superintelligence (ASI)

[4] Existential risk

[5] Value Alignment

[6] برخی از پرتکرارترین این مباحث شامل حریم خصوصی و نظارت (Privacy & Surveillance)، دست‌کاری در رفتار (Manipulation of Behaviour)، شفافیت (Opacity)، سوگیری (Bias)، تعامل انسان و ربات (Human-Robot Interaction)، اتوماسیون و کار انسانی (Automation and Employment)، سیستم‌های خودمختار (Autonomous Systems) و اخلاق ماشینی (Machine Ethics) هستند که در چندین فرامطالعه مورد بررسی قرار گرفته‌اند (Floridi, 2023; Muller, 2021; Corrêa et al , 2023 Hagendorff, 2020;).

[7] Sociotechnical System (STS): در ادبیات پژوهشی گاهی به‌صورت sociotechnical و گاهی به‌صورت socio-technical نوشته می‌شود.

طراحان سیستم هوش مصنوعی نیست. با وجود مزایای این رویکرد، هنوز نیازمند بررسی‌ها دقیق‌تر آن در حوزه هوش مصنوعی هستیم و بسیاری از جنبه‌های این رویکرد باید مورد پژوهش و بررسی قرار بگیرد. یک چالش اساسی می‌تواند این امر باشد که چنین رویکردی، با تأکید بر تعدد عوامل تأثیرگذار، بررسی مسائل اخلاقی را دشوار و یا غیرممکن سازد. لذا، هدف از مقاله حاضر بررسی رویکرد سیستم‌های اجتماعی–فنی در حوزه اخلاقی هوش مصنوعی و بررسی برخی از چالش‌های آن است. در این راستا، روایات مختلف رویکرد سیستم اجتماعی–فنی مورد بررسی قرار خواهند گرفت و سپس یک شیوه پیشنهادی برای بررسی اخلاقی بر اساس این رویکرد ارائه خواهد شد.

جهت دستیابی به هدف اصلی، این مقاله به سه بخش تقسیم شده است. در بخش اول ابتدا ضرورت توجه به رویکرد سیستم‌های اجتماعی–فنی، عناصر و ویژگی‌های این سیستم‌ها بیان خواهد شد. در بخش دوم روایات[۱]اصلی این رویکرد معرفی و مورد ارزیابی قرار خواهند گرفت. در بخش سوم، یک شیوه پیشنهادی برای بررسی‌های اخلاقی در رویکرد سیستم‌های اجتماعی–فنی ارائه خواهد شد.

۱. اخلاق هوش مصنوعی و رویکرد سیستم‌های اجتماعی–فنی

آنچه مشخص است، نگرانی‌ها در مورد فناوری هوش مصنوعی بسیار گسترده است؛ اما نباید فراموش کرد که شیوه پژوهش و بررسی نیز حائز اهمیت است. یک رویکرد می‌تواند این باشد که مباحث اخلاقی یا اجتماعی هوش مصنوعی را به شکلی عام موردبررسی قرار دهیم. به‌عنوان‌مثال، می‌توانیم خودمختاری[۲]و یا شفافیت را در ارتباط با «هوش مصنوعی» بررسی کنیم. عمده ایراد چنین پژوهشی، این خواهد بود که ممکن است انتزاعی باشد. به عبارت دقیق‌تر، هوش مصنوعی نامی عام است که به انواع سیستم‌ها با قابلیت‌ها و زمینه‌های کاربردی متفاوت اشاره دارد. ویژگی‌های مختلفی به هرکدام از این سیستم‌ها مترتب بوده و علاوه بر این در زمینه‌های کاربردی مشخص، ممکن است برخی از مباحث اخلاقی بسیار مهم بوده و برخی دیگر دارای اهمیت کمتری باشند. به‌عنوان‌مثال چندان محل مناقشه نیست که در یک سیستم هوش مصنوعی که در زمینه آموزش کودکان به کار می‌رود، ممکن است نگرانی در خصوص حریم خصوصی کم‌تر از یک سیستم هوش مصنوعی باشد که در حوزه اطلاعات مالی و یا پزشکی افراد به‌کار گرفته می‌شود.

رویکرد دیگر می‌تواند این باشد که مباحث اخلاقی و یا اجتماعی را با تمرکز بر یک تکنیک یا الگوریتم[۳]مشخص یا زمینه کاربردی خاص موردنظر قرار دهیم.[۴] به‌عنوان‌مثال، می‌توانیم چگونگی محافظت از خودمختاری انسان در سیستم‌هایِ تصمیم‌گیریِ خودکارِ یادگیریِ ماشین را در کاربرد آموزشی بررسی کنیم. درحالی‌که چنین تحقیقاتی مهم و روشنگر هستند، یک ایراد اصلی می‌توان در این رویکرد مشاهده کرد. به کلامی دقیق‌تر، چنین پژوهش‌هایی فاقد بررسی «عملکرد» و «اثرات» سیستم‌های هوش مصنوعی در «جامعه» خواهند بود و محدود به زمینه کاربردی و فنی می‌مانند. در راستای مثال پیشین، چنین رویکردی قادر به پاسخ‌گویی به سؤالاتی همچون اینکه چنین سیستمی چه تأثیراتی بر نظام آموزشی می‌گذارد و یا کارکنان سیستم آموزشی چگونه از این سیستم

---

[۱] Narratives

[۲] Autonomy

[۳] Algorithm

[۴] به‌عنوان‌مثال در پژوهش‌های بین‌المللی (Zhang and Zhang, 2023; Zhou and Nabus, 2023; Nguyen, 2023)، و در پژوهش‌های فارسی زبان (عباسی و تیموری، ۱۴۰۲؛ عاشوری کیسمی، ۱۴۰۳؛ سیفی و رزمخواه، ۱۴۰۲) می‌توان نمونه‌هایی از این قبیل پژوهش‌ها مشاهده کرد.

استفاده می‌کنند و یا نظام آموزشی چه جایگاهی برای این سیستم قائل است، نخواهد بود. درحالی‌که زمانی که سیستم هوش مصنوعی در جامعه به کار گرفته می‌شود، نمی‌توان این‌گونه پرسش‌ها را بی‌پاسخ گذاشت.

آنچه از بررسی اجمالی دو رویکرد بالا نمایان می‌شود، کمبودهای است که در رویکردی جایگزین باید بتوان آن‌ها را مرتفع ساخت. ازجمله این رویکرد باید بتواند عملکرد و اثرات فناوری در جامعه، توجه به زمینه مورداستفاده و وجوه فنی و تکنیکی فناوری را پوشش دهد.

۱-۱. رویکرد سیستم‌های اجتماعی-فنی

با در نظر گرفتن محدودیت‌های دو رویکرد پیشین، یک رویکرد جایگزین، رویکرد سیستم‌های اجتماعی-فنی است. این رویکرد، فناوری هوش مصنوعی را به‌عنوان بخشی از سیستم اجتماعی-فنی در نظر می‌گیرد؛ و نه به‌عنوان یک مصنوع مادی طراحی شده توسط انسان که در خلأ و خارج از زمان قرار دارد. به سخنی دیگر، «فناوری تنها طراحی و ظاهر مادی نیست، بلکه اجتماعی-فنی نیز هست»، یعنی در یک فرآیند پیچیده قرار دارد که عوامل مختلف اجتماعی، سیاسی، اقتصادی، فرهنگی و فناوری در آن نقش داشته و عملکردشان مؤثر است[۱](کمیسیون اروپا، ۲۰۲۳). بر اساس این رویکرد، اهمیت زمان، زمینه اجتماعی و همچنین عامل انسانی پررنگ شده و در مرکز توجه قرار می‌گیرد. این ایده اولیه می‌تواند بخشی از کمبودهای رویکردهای پیشین را جبران کند. وجوه اجتماعی و توجه به تأثیر و عملکرد سایر عوامل در اجتماع پررنگ می‌شود و لذا عملکرد و اثرات فناوری در جامعه از نظر دور نمی‌ماند. بااین‌حال چنین رویکردی فناوری را در جامعه بی‌اثر ندانسته و به عامل بودن فناوری و محدودیت‌های تکنیکی و فنی توجه دارد.

ایده در نظر گرفتن زمینه‌های اجتماعی-فنی سیستم‌های هوش مصنوعی در سالیان اخیر توجه پژوهشگران زیادی را به خود جلب کرده است (واتسون و همکاران، ۲۰۲۴؛ یانگ و کوکل‌برگ، ۲۰۲۴). چنین رویکردی می‌تواند مزایای خاص خود را داشته باشد و به ما کمک کند تا بتوانیم یک بررسی و راه‌کار دقیق‌تری در دسترس داشته باشیم. اما پرواضح است که سخن گفتن از اینکه سیستم هوش مصنوعی بخشی از یک سیستم اجتماعی-فنی است چندان روشن نیست. بلکه ابتدا نیاز است این رویکرد کمی دقیق‌تر بررسی و تبیین شود.

۱-۱-۱. عناصر سیستم‌های اجتماعی-فنی و هوش مصنوعی

در ادبیات پژوهشی، اشاره شده است که سیستم اجتماعی-فنی از سه عنصر اساسی یعنی فناوری، انسان‌ها و قواعد اجتماعی[۲]تشکیل می‌شوند (کروز و همکاران، ۲۰۰۶). بااین‌حال، سیستم‌های هوش مصنوعی دارای ویژگی‌هایی هستند که آن‌ها را به یک فناوری خاص تبدیل می‌کند. برای مشخص شدن این امر، توجه به ویژگی‌های هوش مصنوعی می‌تواند به‌وضوح بحث بیفزاید.

---

[۱] در اینجا باید توجه داشت که مطابق با این تعریف، نقش و عملکرد تنها به یکی از عوامل محدود نمی‌شود

[۲] درحالی‌که مفاهیم فناوری و انسان به‌عنوان دو عنصر اصلی در سیستم گویا است، عنصر سوم یعنی قواعد اجتماعی (Social rules) نیاز به توضیح بیشتری دارد قواعد اجتماعی، تعاملات میان عوامل انسانی در جامعه را امکان‌پذیر یا محدود می‌کنند (Crawford and Ostrom, 1995) آن‌ها ممکن است رسمی باشند، مانند قوانین و مقررات و یا غیررسمی باشند مانند عادات و فرهنگ جامعه این قواعد معمولاً از تعاملات گذشته در جامعه ناشی می‌شوند، اما در ارتباط با فناوری، ممکن است قواعد جدیدی نیز به‌صورت عمدی طراحی شوند به‌عنوان نمونه، برخی از فناوری‌ها ممکن است نیاز به قواعد جدیدی مانند دستورالعمل استفاده کاربران داشته باشند (Kudina and Van de Poel, 2024)

سیستم‌های هوش مصنوعی سه ویژگی استقلال[۱] در عمل، تعامل[۲] با محیط و تطبیق‌پذیری[۳] را دارند که پیشتر در سایر فناوری‌ها شاهدشان نبودیم. به عبارتی دقیق‌تر، سیستم‌های هوش مصنوعی بدون دخالت انسان تا حدی قادر به انطباق‌پذیری و عملکرد مستقل در پاسخ به ورودی‌هایی هستند که از محیط خود[۴] دریافت می‌کنند. این امر را می‌توان نوعی عاملیت غیرانسانی و غیر التفاتی سیستم هوش مصنوعی در نظر گرفت.[۵] علاوه بر این، در تعامل هوش مصنوعی و محیط، سیستم‌های هوش مصنوعی تطبیق یافته و برخی هنجارها[۶] شکل می‌گیرند که شبیه به قواعد اجتماعی عمل می‌کنند. به عبارت دیگر، در تعامل میان هوش مصنوعی و عامل انسانی در سیستم اجتماعی–فنی قواعد یا هنجارهایی فنی پدید می‌آیند که پیشتر در جامعه وجود نداشتند. بنابراین، سیستم‌های هوش مصنوعی دو عنصر دیگر را به سیستم‌های اجتماعی–فنی اضافه می‌کنند که در ادبیات پژوهشی با عنوان «عوامل مصنوعی[۷]» و «هنجارهای فنی[۸]» (وان دِ پُل، ۲۰۲۰؛ وان دِ پُل، ۲۰۲۳) شناخته می‌شوند.

۲. روایت‌های سیستم‌های اجتماعی–فنی

همان‌طور که در ابتدا هم اشاره شد، ایده سیستم‌های اجتماعی–فنی و در نظر گرفتن فناوری در زمینه اجتماعی آن می‌تواند دامنه پژوهش را گسترش داده و عوامل مؤثر در توسعه و کاربست تکنولوژی را در بر بگیرد؛ اما باید توجه داشت که این ایده به روش‌های مختلفی قابل تفسیر است، به شکلی که سه روایت کلی را می‌توان در این خصوص مشاهده کرد.[۹] برای روشن‌تر شدن بحث، ابتدا برخی روایات اصلی را به‌اجمال بررسی خواهیم کرد.

۲–۱. روایت جزم‌گرایانه سیستم‌های اجتماعی–فنی

روایت جزم‌گرایانه[۱۰] سیستم‌های اجتماعی–فنی بر مزایای اقتصادی و اجتماعی فناوری تمرکز دارد و اغلب تأثیرات منفی بالقوه آن را نادیده می‌گیرد (واتسون و همکاران، ۲۰۲۴). این روایت بر این اساس استوار است که پیشرفت‌های فناوری درنهایت موجب پیشرفت‌های اقتصادی و اجتماعی می‌شود؛ بنابراین، تلاش برای کند کردن روند یا ممانعت از پیشرفت فناوری تحت عنوان اخلاق یا مصالح دیگر اشتباه بوده و لذا مضر و غیراخلاقی است. در همدلی با این روایت می‌توان گفت که این امر غیرقابل‌انکار است که پیشرفت فناوری مزایای شگرفی به همراه دارد که تنها محدود به اقتصاد نمی‌شود. به‌عنوان‌مثال، سیستم‌های هوش مصنوعی به ما

---

[۱] Autonomy

[۲] Interactivity

[۳] Adaptivity

[۴] در اینجا محیط منظور به راه‌های ارتباطی و شیوه‌هایی است که سیستم هوش مصنوعی با خارج از اجزای خود در ارتباط بود و داده‌ها و یا اطلاعات دریافت می‌کند

[۵] در اینجا دو نکته حائز اهمیت است اول آنکه بدون وجود این ویژگی‌ها سیستم هوش مصنوعی چیزی بیش از یک برنامه کامپیوتری نخواهد بود به‌عبارتی‌دیگر این ویژگی‌ها بخشی از دلایلی است که به خاطر آن سیستم‌های هوش مصنوعی را طراحی می‌کنیم (Van de Poel, 2023) نکته دوم، ما ویژگی‌هایی مانند عاملیت را با عاملیت انسانی یکسان در نظر نخواهیم گرفت و صرفِ شباهت را به معنای این‌همانی نمی‌دانیم تا در دام انسان‌انگاری (anthropomorphism) هوش مصنوعی نیافتیم (عاشوری کیسمی و پرویزی، ۱۴۰۳)

[۶] Norms

[۷] Artificial agents

[۸] Technical norms

[۹] در اینجا ذکر دو نکته حائز اهمیت است اول آنکه دسته‌بندی این روایات پیشتر توسط پژوهشگران صورت گرفته است و ما نیز همین دسته‌بندی را موردنظر قرار می‌دهیم دوم آنکه هرکدام از این روایات را به‌عنوان یک اصطلاح چتری (Umbrella Term) باید در نظر گرفته شود که به طیف وسیعی از نظرات و پژوهش‌ها اشاره دارد (Watson et al , 2024) البته یکی از نوآوری‌های مقاله حاضر بررسی نظام‌مند این روایت‌ها و ارائه یک تصویر شفاف‌تر و دقیق‌تر از پژوهش‌های پیشین است

[۱۰] Dogmatism

کمک کرده‌اند تا شناخت بهتری از تغییرات اقلیمی به دست آوریم. بااین‌حال، حتی در همین زمینه کاربردی نیز، انرژی مصرفی حاصل از پردازش مقادیر عظیم داده‌های مورداستفاده در این سیستم‌ها، همراه با انتشار گازهای گلخانه‌ای است و لذا خود این سیستم‌ها می‌توانند به افزایش تغییرات اقلیمی منجر شوند (کاولز و همکاران، ۲۰۲۳). به‌عبارت‌دیگر، آینده آرمان‌شهری که این روایت آن را وعده می‌دهد، نمی‌تواند نگرانی‌های موجود را برطرف کند و حتی ممکن است به آسیب‌های بیشتر منجر شود.

مطابق با این روایت، اگر از ابتدا نتیجه توسعه فناوری را تعیین شده در نظر بگیریم — خیر نهایی و حتمی- و بررسی اخلاقی که بتواند به پاسخ‌های هنجاری بیانجامد را مضر و غیراخلاقی بدانیم؛ آنگاه در بهترین حالت بررسی و مطالعه ما چیزی جز توصیف جزمی مزایای توسعه فناوری نخواهد بود. لذا این روایت، اگرچه به وجود عناصر مختلف در سیستم اجتماعی-فنی توجه دارد، در مباحث اخلاقی می‌تواند بسیار تقلیل دهنده عمل کند.

۲-۲. روایت شک‌گرایانه سیستم‌های اجتماعی-فنی

برخلاف روایت جزم‌گرایانه، در روایت شک‌گرایانه[۱]سیستم‌های اجتماعی-فنی، تمرکز بر مضرات فناوری است (زوبوف، ۲۰۲۳؛ سوسر و همکاران، ۲۰۱۹؛ منهیم و کاپلان، ۲۰۱۹). با توجه به مسائل و چالش‌های اخلاقی و اجتماعیِ سیستم‌های هوش مصنوعی چندان دور از ذهن نیست که باید آگاهانه و با چشم باز در پژوهش به این فناوری بنگریم و از این منظر می‌توان با این روایت همدل بود. بااین‌حال، برخی شک‌گرایان، فناوری را علت مشکلات اجتماعی می‌دانند و با نادیده گرفتن اهمیت ارائه راه‌حل، امکان پیشرفت در طراحی بهتر فناوری که برخی از چالش‌های موجود را مرتفع سازد را رد می‌کنند (هوفمان، ۲۰۱۹). جنبه منفی شک‌گرایی این است که می‌تواند ساده‌انگارانه باشد، زیرا از محدودیت بر فناوری حمایت می‌کند ولی هیچ پیشنهاد سازنده‌ای برای استفاده از آن ارائه نمی‌دهد. درمجموع می‌توان گفت، روایت اول تقلیل‌دهنده و روایت دوم محدودکننده است (واتسون و همکاران، ۲۰۲۴) و هر دو تصویری سودمند اما ناقص از سیستم‌های اجتماعی-فنی ارائه می‌دهند.

۲-۳. روایت پراگماتیستیِ سیستم‌های اجتماعی-فنی

مطابق با آنچه به‌اختصار در خصوص هر دو روایت پیشین مطرح شد، به نظر می‌رسد کاستی‌هایی در هر دو وجود دارد و نیازمند رویکرد دیگری خواهیم بود که بتواند آن‌ها را مرتفع سازد. به‌عبارتی دیگر، اگرچه مطابق این روایت‌ها وجود عوامل مختلف در سیستم موردنظر قرار گرفته‌اند، اما عملکرد و اثرات اجزای سیستم به‌صورت یک‌طرفه و محدود موردتوجه قرار می‌گیرند. برای روشن شدن محدودیت‌های این روایات و دستیابی به روایتی دقیق‌تر، توجه به تعاملات درون سیستم‌های اجتماعی-فنی روشنگر خواهد بود. لذا بررسی روایت پراگماتیستی را به پس از روشن شدن این محدودیت‌ها واگذار خواهیم کرد.

۲-۳-۱. تعاملات درون سیستم اجتماعی-فنی

[۱] Skepticism

اگرچه برخی از دیدگاه ابزارگرایانه[۱]در مورد فناوری طرفداری می‌کنند و معتقدند که فناوری غیرارزش‌بار[۲]است، زیرا که تنها یک ابزار در دست انسان است (پیت، ۲۰۱۳)، اجماع نظر بسیاری از محققان، امروزه متفاوت با این دیدگاه است (وان دِن هوون، ۲۰۱۲؛ وان دِ پُل و کروز، ۲۰۱۴؛ پارتو و کوکل‌برگ، ۲۰۲۴). یک ایراد دیدگاه ابزارگرایانه این است که چنین دیدگاهی این امر که هر فناوری می‌تواند اعمال و ارزش‌های خاصی را محدود یا تقویت کند، نادیده می‌گیرد. به‌عنوان‌مثال یک فناوری شبکه اجتماعی را در نظر بگیرید که در آن بیان نظرات دارای نوعی محدودیت است، چنین سیستمی امکان تقویت نظرات و ارزش‌های همسو را فراهم کرده و بیان نظرات مخالف را می‌تواند محدود کند. به صورت مشابه نیز می‌توان گفت اگر این سیستم به شیوه‌ای طراحی شود که هیچ گونه محدودیتی در بیان نظرات نداشته باشد، این امر خود تقویت ارزش‌های دیگری است که بر پایه عدم وجود محدودیت در بیان نظرات است.

دیدگاه دیگر ذات‌گرایی[۳]است که معتقد است اگرچه فناوری ارزش‌بار[۴]است، اما انسان‌ها بر آن تأثیر نمی‌گذارند. محدودیت ذات‌گرایی این است که تأثیر انتخاب‌ها و عاملیت انسان‌ها در طراحی و به‌کارگیری فناوری را نادیده می‌گیرد (سریواتسا و همکاران، ۲۰۱۷) و معتقد است که ذات فناوری ارزش‌بار است. در ایراد به این رویکرد می‌توان مجدداً بر ایده سیستم‌های اجتماعی-فنی تأکید کرد که فناوری در خلأ، خارج از زمان و بدون عامل انسانی طراحی و استفاده نمی‌شود. به‌عبارتی‌دیگر، می‌توان یک فناوری را به شیوه‌های مختلفی طراحی و از آن استفاده‌های متفاوتی کرد. چندان محل مناقشه نخواهد بود اگر اینگونه ایراد کنیم که هرکدام از انتخاب‌های ما طراحی فناوری و یا استفاده‌های گوناگون از یک فناوری می‌توانند ارزش‌های مشخصی را تقویت و یا تضعیف کنند.

سومین دیدگاه تعامل‌گرایی[۵]است که هم به وجه ابزاری فناوری توجه دارد و هم به عاملیت انسانی در طراحی و استفاده از آن. بر اساس این دیدگاه، در تعامل میان عامل انسانی با فناوری، چه در استفاده و چه در طراحی، ارزش‌ها بدنمند[۶]شده و فناوری ارزش‌بار می‌شود (سریواتسا و همکاران، ۲۰۱۷). این دیدگاه برخلاف دو دیدگاه پیشین تبیین دقیق‌تری ارائه داده و وجه ابزاری، نقش و تأثیر عاملیت انسانی و ارزش‌بار بودن فناوری را پوشش می‌دهد. بااین‌حال زمانی که دیدگاه تعامل‌گرایانه اتخاذ می‌کنیم، این پرسش مطرح می‌شود که در اثر این تعاملات، آیا ارزش‌ها ثابت خواهند بود و یا اینکه این تعاملات بر خود ارزش‌ها نیز اثر دارند؟ چنین پرسشی، ما را به بررسی تغییر یا ثبات ارزش‌ها درون سیستم‌های اجتماعی-فنی وارد می‌کند.

۲-۳-۲. تغییر ارزش‌ها درون سیستم اجتماعی-فنی

مطابق با دیدگاه تعامل‌گرا در یک سیستم اجتماعی-فنی، ارزش‌ها در تعامل عناصر سیستم، نقش مهمی را ایفا می‌کنند. بااین‌حال، این نکته قابل‌تأمل است که ارزش‌ها ممکن است در اثر تعاملات تغییر کنند. برای روشن شدن این موضوع، مثال ظهور ارزش پایداری[۷]را در نظر بگیرید. پایداری نمونه‌ای از ظهور ارزش جدید در جامعه است که تنها در چند دهه گذشته موردتوجه قرار گرفته است. این تغییر در ارزش، بر اولویت‌هایِ اجتماعیِ جوامع تأثیر گذاشته و در حوزه‌های مختلفی ازجمله در توسعه فناوری نیز ورود پیدا

---

[۱] Instrumentalism
[۲] Value-free
[۳] Substantivism
[۴] Value-laden
[۵] Interactionism
[۶] Value embodiment
[۷] Sustainability

کرده است (وان دِ پُل و کودینا، ۲۰۲۲). البته باید در نظر داشت که این تعامل دوسویه است و پیشرفت‌های فناوری نیز در ارزش‌ها تغییرات[۱] اساسی ایجاد کرده‌اند (هوپستر، ۲۰۲۱). در حقیقت، تغییرات فناوری گاهی فراتر از تأثیرگذاری محدود بر جامعه است و به قلمرو «ارزش‌ها» و «مفاهیم[۲]» انسانی نیز ورود پیدا می‌کنند. به‌عنوان‌مثال، طراحی و ساخت دستگاه تنفس مصنوعی[۳]مفهوم مرگ و ارزش‌های اخلاقی در خصوص پایان زندگی را پیچیده و چالش‌برانگیز کرد. اکنون می‌توان حتی پس از توقف تنفس طبیعی انسان، با استفاده از این دستگاه فرد را زنده نگاه داشته و مفهوم مرگ تنها محدود به توقف عمل فیزیکی و درونی بدن فرد نمی‌شود[۴](وان دِ پُل و همکاران، ۲۰۲۳) و لذا به‌تبع آن ارزش‌های اخلاقی که در خصوص تصمیم‌گیری درباره پایان زندگی نیز پیچیده‌تر خواهند شد. به سخنی دیگر، ظهور این فناوری مفهوم مرگ فیزیکی را تغییر داد. اکنون اگر تلاش برای نجات جان بیمار را بر اساس مفهوم قدیمی مرگ در نظر بگیریم، احتمالاً تلاش برای زنده‌نگاه داشتن بیمار و سایر تصمیمات اخلاقی پیرو آن نیز باید پس از توقف تنفس طبیعی پایان بیابد. در این راستا، پژوهشگران مثال‌های متعددی از اختلال فناوری در ارزش‌ها و مفاهیم ارائه کرده‌اند (هوپستر، ۲۰۲۱) و لذا به نظر می‌رسد تعاملات فناوری و جامعه همراه با تغییر در برخی از ارزش‌ها است.

آن‌طور که به نظر می‌رسد، حداقل به‌نوعی ارزش‌ها می‌توانند از طریق تعامل عناصر در سیستم‌های اجتماعی-فنی تغییر کنند. موضوع «تغییر در ارزش[۵]» و انواع مختلف آن بارها در ادبیات پژوهشی موردبحث قرار گرفته است[۶](بوئنینک و کودینا، ۲۰۲۰؛ دِ ویلدت و شوایزر، ۲۰۲۲؛ وان دِر ویج و همکاران، ۲۰۲۳). البته باید در نظر داشت زمانی که سخن از تعاملات درون سیستم اجتماعی-فنی به میان می‌آید، این تعاملات یک‌سویه و تنها از سمت تکنولوژی بر جامعه در نظر گرفته نمی‌شود؛ بلکه منبع تغییر ارزش می‌تواند توسعه اجتماعی، تغییر در قواعد اجتماعی یا تغییرات ناشی از فناوری باشد (وان دِ پُل، ۲۰۲۱ب). درواقع، بر اساس این دیدگاه، فناوری و ارزش‌ها در جامعه همزیستی می‌کنند و از طریق تعامل با یکدیگر تکامل[۷]می‌یابند (وان دپل و کودینا، ۲۰۲۲) یا به‌عبارت‌دیگر اجزای سیستم در یک «فرایند[۸]» و در روند همیشگی «شدن[۹]» هستند (کوکلبرگ، ۲۰۲۲ب).

درمجموع، مبحث تعاملات و تغییرات می‌تواند برخی از کاستی‌های روایت‌های شک‌گرایی و جزم‌گرایی را آشکار سازد. به‌عنوان‌مثال، برخلاف هر دو روایت، پیشرفت فناوری ممکن است به نتایج متفاوتی منجر شود. به‌عبارتی دیگر، اگر تعاملات و تغییرات را در نظر بگیریم، شیوه‌های طراحی و استفاده از فناوری دارای اهمیت بوده و در سیستم نقش و عملکرد بسزایی دارند. با توجه به محدودیت‌های شک‌گرایی و جزم‌گرایی، سومین روایت یعنی روایت پراگماتیستیِ سیستم‌های اجتماعی-فنی، در سال‌های اخیر

---

[۱] Change

[۲] Conceptions

[۳] Ventilator

[۴] به‌عنوان نمونه‌ای دیگر، هوش مصنوعی شبکه‌های عصبی خلاق (Creative Adversarial Networks - CAN)، مفهوم خلاقیت (Creativity) و تولید اثر هنری را تغییر داد است و اکنون سیستم‌های مختلف قادر به تولید این نوع آثار هستند (عاشوری کیسمی، ۱۴۰۳ الف) به عبارتی دقیق‌تر، اکنون نیازمند آن هستیم که برای مفاهیمی که از پیش داشتیم (در این نمونه خلاقیت)، تعریفی جدید ارائه یا تعاریف پیشین را مورد بازبینی قرار دهیم

[۵] Value change

[۶] به‌عنوان‌مثال، ظهور ارزش‌های جدید، تغییر در ارتباط ارزش‌های موجود با فناوری‌های خاص، تغییر در اولویت یا اهمیت نسبی ارزش‌ها، تغییر در مفاهیم ارزش‌ها (Van de Poel 2021b) و غیره

[۷] Coevolve

[۸] Process

[۹] Becoming

موردتوجه پژوهشگران قرار گرفته است (واتسون و همکاران، ۲۰۲۴؛ کودینا و وان دِ پُل، ۲۰۲۴؛ واتسون و موکاندر، ۲۰۲۳؛ گامبو و تاوِتر، ۲۰۲۱) که در تلاش است خود را از محدودیت‌های روایات پیشین رها کرده و چهارچوبی کامل‌تر ارائه دهد.

۳-۳-۲. روایت پراگماتیستی از سیستم‌های اجتماعی-فنی

از منظر تاریخی، پراگماتیسم ریشه در آثار چارلز سندرز پرس و ویلیام جیمز دارد. بااین‌حال، همانطور که هیلاری پاتنم (۱۹۹۵) اشاره می‌کند، پراگماتیسم «پرسشی گشوده» است. به‌عبارت‌دیگر، پراگماتیسم یک رویکرد فلسفی با تعابیر گوناگون و متنوع است. البته قابل‌توجه است که روایت پراگماتیستی از سیستم‌های اجتماعی-فنی را نباید به‌عنوان شاخه‌ای از پراگماتیسم در نظر گرفت. بلکه مقصود یک رویکرد اجتماعی-فنی است که از برخی ایده‌های فیلسوفان پراگماتیست و روش‌های آن‌ها متأثر است (وان دِ پُل و کودینا، ۲۰۲۲). برای روشن شدن این موضوع، در ادامه برخی از این ایده‌های اصلی را موردنظر قرار خواهیم داد.

بنا به تعبیر عام دانشنامه فلسفی استفنورد، پراگماتیسم را می‌توان یک رویکرد فلسفی دانست که بر اساس آن شناخت جهان از عاملیت درون آن جدایی‌ناپذیر است (لگ و هوک‌وی، ۲۰۰۸). برخلاف روایت جزم‌گرایانه که تصویری جبری و متعین از آینده آرمانی فناوری ارائه می‌دهد، روایت پراگماتیستی اگرچه محدودیت‌های انسانی را در نظر می‌گیرد (دوئی، ۱۹۱۷) بر نقش کنش و عاملیت انسانی نیز تأکید می‌کند. به‌عبارت‌دیگر، بر اساس این روایت، اگرچه محدودیت‌های فنی در آینده پیشرفت فناوری نقش مهمی‌دارند، اما این موضوع تا حدی هم متأثر از عوامل انسانی و در نتیجه غیرجبری و نامعین است (واتسون و همکاران، ۲۰۲۴). با در نظر گرفتن این موضوع، درحالی‌که این روایت به محدودیت‌های فنی در انتخاب عامل انسانی توجه دارد و آن را نادیده نمی‌انگارد، به نقش عوامل انسانی در میان انتخاب‌های موجود، بررسی نوآوری جدید به‌جای فناوری کنونی یا محدود کردن برخی از فناوری‌ها نیز توجه دارد. در این راستا، روایت پراگماتیستی از سیستم‌های اجتماعی بر عناصر انسانی و اجتماعی مانند نقش حاکمیت، نهادها و فرهنگ در عملکرد سیستم توجه کرده (کودینا و وان دِ پُل، ۲۰۲۴) و فناوری را تنها عامل مؤثر در نظر نمی‌گیرد. این امر به‌صورت ضمنی تأکید می‌کند که اگر امکان سودمندی فناوری برای جامعه انسانی وجود داشته باشد، رسیدن به چنین امری خودبه‌خودی نبوده و متأثر از تعاملات عناصر مختلف درون سیستم است.

برخلاف روایت شک‌گرایانه که عموماً تنها بر محدود کردن فناوری تأکید می‌کند و بر تلاش برای پیشرفت چشم می‌پوشد، روایت پراگماتیستی بر ایده پیشرفت[۱](دوئی، ۲۰۰۴) و خطاپذیری[۲](پیرس، ۱۹۵۵) استوار بوده و درعین‌حال تأکید می‌کند که اعتبار نظریه‌ها به سودمندی آن‌ها در انطباق با واقعیت (جیمز، ۱۹۰۷) بستگی دارد. به‌این‌ترتیب، در این روایت، اولین انتخاب در برابر مسائل اخلاقی و اجتماعی که هوش مصنوعی پدید می‌آورد، تأکید بر محدودیت یا کنار گذاشتن این فناوری نیست؛ بلکه بر پیشرفت تأکید دارد اما با تکیه بر خطاپذیری کمال را هدف خود قرار نمی‌دهد. این روایت، عدم قطعیت را اصل قرار می‌دهد و تأکید می‌کند که ارزیابی هوش مصنوعی باید در حوزه کاربردی آن صورت گرفته و بازتاب هنجارهای جامعه و منطبق با آن باشد (واتسون و همکاران، ۲۰۲۴). به این شیوه هم بر زمینه کاربردی تأکید می‌شود و هم فضایی برای در نظر گرفتن امکان تغییر ارزش‌ها در جامعه.

در مجموع ورود فناوری به جامعه با عدم قطعیت در مورد اثرات اجتماعی آن همراه است (کالینگریج، ۱۹۸۰). بااین‌حال، چندین عامل می‌توانند این عدم قطعیت را شدت بخشند. به‌عنوان‌مثال، فقدان آینده‌نگری و پیش‌بینی تأثیرات فناوری بر جامعه می‌تواند منجر

---

[۱] Progress

[۲] Fallibilism

به پیامدهای ناخواسته اجتماعی شود. به‌عنوان‌مثال، در طراحی یک سیستم هوش مصنوعی ممکن است گاهی برخی تأثیرات آن بر جامعه را از روی عدم بررسی دقیق و کافی در نظر نگرفته باشیم. لذا درنتیجه این عدم توجه، زمانی که هوش مصنوعی در جامعه مورد استفاده قرار می‌گیرد ممکن است به پیامدهای ناخواسته‌ای بیانجامد. اما علاوه بر این موضوع، باید در نظر داشت که ظرفیت معرفتی انسان محدود است و نه‌تنها فقدان آینده‌نگری و عدم پیش‌بینی، بلکه ناآگاهی معرفتی نیز می‌تواند به اثرات ناخواسته فناوری منجر شود. این امر را می‌توان به معنای توجه به محدودیت‌های معرفت علمی در نظر گرفت. اگرچه فقدان آینده‌نگری ممکن است از روی عدم التفات پدید آید، اما محدودیت ظرفیت معرفتی و دانش علمی شکلی متفاوت از پیامدهای ناخواسته را منجر می‌شود که بر اثر عدم التفات پدید نیامده است. شایان توجه است که عدم قطعیت در خصوص ورود فناوری به جامعه تنها به این موارد محدود نمی‌شود. در حقیقت چندین عامل مختلف در یک جامعه وجود دارند. همانطور که اشاره شد، عوامل فرهنگی، اقتصادی و غیره در یک جامعه نقش‌آفرینی می‌کنند که تعامل آن‌ها با فناوری جدید می‌تواند به تأثیرات اجتماعی مختلفی بیانجامد. لذا با در نظر گرفتن عواملی که در زمان طراحی و استفاده از یک سیستم هوش مصنوعی خارج از کنترل ما است، همواره با میزانی عدم قطعیت روبرو خواهیم بود (وان دِ پُل، ۲۰۲۳). به‌عبارتی دیگر، ورود فناوری به جامعه را می‌توان به‌عنوان یک «آزمایش اجتماعی[۱]» تصور کرد. تنها زمانی که فناوری در اجتماع به‌کار گرفته می‌شود و به‌صورت «تدریجی» برخی از اثرات اجتماعی فناوری قابل‌کشف است (وان دِ پُل، ۲۰۱۶). به عبارتی دیگر، بخشی از دانش ما درباره اثرات اجتماعی فناوری، امری پیسنی[۲]است و نمی‌توان به‌صورت پیشینی[۳]به آن‌ها دست یافت.

۳. پیچیدگی و سیستم‌های اجتماعی–فنی

از فراز آنچه موردنظر قرار گرفت، روایت پراگماتیستی از سیستم‌های اجتماعی–فنی می‌تواند تبیین کامل‌تری نسبت به دو روایت دیگر در اختیار ما قرار دهد. بااین‌حال این امر منجر نمی‌شود تا چشم بر محدودیت‌هایی که ممکن است این روایت با خود به همراه داشته باشد ببندیم. لذا لازم خواهد بود کمی دقیق‌تر به چالش‌های روایت پراگماتیستی توجه داشته باشیم.

در این راستا، پیش از هر چیز و با توجه به وجود عوامل مختلف در سیستم‌های اجتماعی–فنی دانستیم که ارزیابی فناوری هوش مصنوعی را باید یک تلاش مسئله‌محور[۴]و در حوزه کاربردی آن در نظر داشت که با عوامل سیاسی، اقتصادی، اجتماعی و غیره سر و کار دارد. با اینحال یادآور شدیم که قصد داریم به چالش‌های روایت پراگماتیستی نیز بپردازیم. برای روشن شدن چالش روایت پراگماتیستی، مثالی که کودینا و ون‌دِپول (۲۰۲۴) ارائه می‌دهند را بهتر است کمی بسط دهیم. فرض کنید قصد داریم یک سیستم هوش مصنوعی برای یک سازمان دولتی در کشوری خیالی و در حوزه رفاه اجتماعی طراحی کرده و استفاده کنیم. یکی از موضوعاتی که در طراحی و استفاده باید موردتوجه قرار بگیرد، «انصاف» است. انصاف بدون شک تا حد زیادی به الگوریتمِ هوش مصنوعی ما بستگی دارد. پس ما نیاز به توجه به مباحث اخلاق و همچنین محدودیت‌های فنی طراحی خواهیم داشت. اما می‌دانیم معیارهای

---

[۱] Social experiment

[۲] a posteriori

[۳] a priori

[۴] در اینجا مقصود آن است که ارزیابی، بر اساس مسائل در زمینه‌های کاربردی مشخص صورت می‌گیرد و نه آنکه مسائل اخلاقی و اجتماعی را به شیوه‌ای انتزاعی یا خارج از زمینه آن موردنظر قرار دهیم.

انصاف متعدد و حتی متناقضی وجود دارند. برای تصمیم‌گیری در مورد اینکه کدام معیارها باید موردتوجه قرار بگیرد، ابتدا نیاز است در خصوص زمینه کاربرد این هوش مصنوعی بیشتر بدانیم. بی‌شک سؤالات متفاوتی هم برای ما پیش خواهد آمد. به‌عنوان‌مثال باید بدانیم بی‌عدالتی اصلی که ممکن است در این زمینه کاربردی رخ دهد، چیست که می‌تواند منصفانه بودن سیستم را به خطر بیندازد؟ در کشور خیالی ما، کدام افراد بیشتر در معرض آسیب‌های ناشی از بی‌عدالتی هستند که لازم است توجه ویژه‌ای به آن‌ها داشته باشیم؟ و یا اینکه آیا باید از تمامی تصمیمات ناعادلانه جلوگیری شود؟ لذا علاوه بر اخلاق و مهندسی، باید به مباحث حوزه‌هایی مانند علوم اجتماعی هم وارد شویم. حال فرض کنیم تمامی این سؤالات را پاسخ داده‌ایم. به‌عنوان‌مثال معیارهای اخلاق، زمینه موردنظر، آسیب‌ها و بی‌عدالتی‌ها را شناخته‌ایم. اکنون شروع به طراحی سیستمی بر اساس پاسخ‌هایمان می‌کنیم. اما سیستم ما نمی‌تواند همواره نتایجی داشته باشد که در آن انصاف به‌صورت کامل رعایت شود. یک دلیل عمده این است که عوامل دیگری هم در سیستم دخیل هستند. به‌عنوان‌مثال، رفتار کارمندان این سازمان دولتی، فضای سیاسی در رابطه با انصاف در این کشور، اطلاع شهروندان از حقوقشان و معیارهای انصاف به‌کار رفته در سیستم و یا قوانین داخلی در سازمان دولتی و بسیاری عوامل دیگر همگی می‌توانند در نتایج منصفانه برای افرادی که رفاهشان به این سیستم وابسته است نقش داشته باشند. حال اگر نتیجه‌ای غیرمنصفانه‌ای رخ دهد چطور باید عمل کرد؟ فرض کنیم مشکل از طراحی سیستم نباشد. فرد یا افرادی که اطلاعات شهروندان را درون سیستم وارد کرده‌اند، مرتکب اشتباه در ورود داده‌ها شوند و این موضوع منجر به نتیجه‌ای غیرمنصفانه شود. باید بدانیم آیا در سازمان ما امکان عدم توجه به قضاوت‌های غیرمنصفانه وجود دارد؟ آیا به شهروندان امکان اعتراض به تصمیم یا ارائه مدارک برای توجیه تصمیمات سیستم ارائه می‌شود؟ یا ازآنجایی‌که تصمیمات سیستم است آن را بدون خطا می‌دانیم؟ لذا از یک منظر قوانین و نهادهای قانونی بر عملکرد و عادلانه بودن سیستم اجتماعی که الگوریتم هوش مصنوعی در آن تعبیه شده است نیز نباید دور از منظر بماند. حتی چندان دور از ذهن نیست که این موضوعات در طول زمان دچار تغییرات نیز بشوند. به عنوان مثال، قوانین و مقررات، شرایط اجتماعی و سایر عوامل در جامعه در طول زمان شاهد تغییرات هستند.

بر اساس آنچه مثال بالا به ما نشان می‌دهد در بررسی چالش‌های اجتماعی و اخلاقی هوش مصنوعی، اگر از این وجوه مختلف چشم‌پوشی کنیم، حتی ممکن است که عمل به پیشنهادات هنجاری ما نیز تأثیرات ناخواسته‌ای به دنبال داشته باشد. به‌عبارتی دیگر، همانطور که در انتهای بخش پیشین موردنظر قرار گرفت، عدم آینده‌نگری و پیش‌بینی چالش‌هایی که از طریق سایر زمینه‌های بررسی قابل شناسایی هستند، می‌تواند منجر به افزایش عدم قطعیت نامطوب ورود فناوری به جامعه شود. در طرف مقابل، توجه به این عوامل می‌تواند همراه باشد با دشواری در پاسخ به پرسش‌های اخلاقی. به‌عبارتی دیگر، یک چالش روایت پراگماتیستی سیستم‌های اجتماعی–فنی نحوه پاسخگویی به پیچیدگی سیستم است؛ چراکه ما را دعوت می‌کند با سیستمی پیچیده روبرو شویم اما به نظر می‌رسد راه‌حلی برای این امر ارائه نمی‌کند.

به‌عنوان یک راه حل برای برون‌رفت از پیچیدگی روایت پراگماتیستی سیستم‌های اجتماعی–فنی ما مطالعه میان‌رشته‌ای[1]را مورد ارزیابی قرار می‌دهیم. در ادبیات پژوهشی، بارها بر اهمیت پژوهش میان‌رشته‌ای در حوزه فلسفه علم و تکنولوژی تأکید شده است (اشمیت، ۲۰۱۱؛ آلوارگونثالس، ۲۰۱۱؛ ماکی، ۲۰۱۶) و حتی برخی از پژوهشگران معتقدند تنها با این شیوه مطالعه، تأملی اخلاقی که با دنیای واقعی، فناوری و جامعه ارتباط برقرار کند امکان‌پذیر می‌شود (دِکِر، ۲۰۰۴). باید توجه داشت که انواع نظریات مختلف در

[1] Interdisciplinary

خصوص پژوهش میان‌رشته‌ای وجود دارد. از منظر فلسفی، رویکردهای مختلفی مانند رویکردهای میان‌رشته‌ای ابزارگرایانه (روکو و بِین‌بریج، ۲۰۰۲)، انتقادی[۱](فْرودِمان و میچَم، ۲۰۲۷)، راهبردی[۲](واین‌گارت، ۲۰۰۰) و غیره را در این خصوص می‌توان مشاهده کرد. در این راستا، ما رویکردی پراگماتیستی[۳]از مطالعه میان‌رشته‌ای را به‌عنوان راه‌حل موردنظر قرار می‌دهیم.[۴] در این راستا، راه‌حل جایگزین باید بتواند دو امر را پوشش دهد. نخست آنکه این راه‌حل باید با روایت پراگماتیستی از سیستم‌های اجتماعی–فنی هم‌خوانی داشته باشد. و دوم، چالش‌ها و مشکلاتی که در بررسی مثال ابتدایی این بخش با آنها روبرو شدیم را مرتفع سازد.

با توجه به مباحثی که در بالا مطرح شد، اکنون زمان آن است که مشخص کنیم مقصودمان از مطالعه میان رشته‌ای با رویکرد پراگماتیستی چیست. همانطور که پیش‌تر ذکر شد، پراگماتیسم به‌جای ایده کمال و قطعیت بر ایده پیشرفت و خطاپذیری تأکید می‌کند. لذا همین امر بخشی از نگرانی ابتدایی ما از پیچیدگی سیستم و دشواری دستیابی به پاسخ را تا حدودی کاهش می‌دهد. به‌عبارتی دیگر، هدف از مطالعه میان‌رشته‌ای در سیستم‌های اجتماعی–فنی با رویکردی پراگماتیستی، دست‌یابی به قطعیت و کمال نخواهد بود. پراگماتیسم بر فرآیند همکاری مداوم و حرکت به سمت درکی جامع‌تر از مسائل پیچیده، ما را دعوت می‌کند که اگرچه سویه‌های شک[۵]یا عدم تعین حضور مؤثری دارند، اما این عدم تعین نقشی مولد دارد (ولچ، ج،۴، ۲۰۱۴). در حقیقت بررسی میان‌رشته‌ای به این شیوه، باید وجود عوامل مختلف را در نظر بگیرد اما همزمان بر خطاپذیری و روند پیشرفت توجه داشته باشد.

باید در نظر داشت که شاید بتوان با یک نگاه تقلیل‌گرایانه[۶]امکان دست‌یابی به قطعیت را به شکلی صوری افزایش داد، اما این قطعیت مطلوب نخواهد بود، چراکه محصول عدم بررسی عوامل مختلف است. به‌عنوان‌مثال می‌توانیم بررسی درون رشته‌ای یا مطالعه‌ای که در آن به زمینه‌های محدود توجه داشته باشیم را ملاک خود قرار دهیم. به‌عنوان‌مثال فرض کنید قصد داریم تا بررسی برای طراحی یک سیستم هوش مصنوعی را ترتیب دهیم و برای این منظور، به بررسی فلسفی محدود می‌مانیم. از منظر معرفتی، ما بر پایه دانش فلسفی قادر نخواهیم بود بسیاری از زمینه‌ها را به درستی بررسی کنیم. به‌عنوان‌مثال، مطالعه ما فاقد آینده‌نگری و پیش‌بینی تاثیرات جامعه‌شناختی یا روانشناختی سیستم هوش مصنوعی بر جامعه خواهد بود. چندان دور از ذهن نیست که این امر اگر با التفات صورت گرفته باشد، خود می‌تواند امری غیراخلاقی تلقی شود.

برخلاف نگاه تقلیل‌گرایانه، رویکرد پراگماتیستی با تکیه بر آراء دیویی (۱۹۱۷؛ ۱۹۹۸) در خصوص دینامیک میان «عدم تعین» و «ظرفیت عاملیت انسانی»، تأکید می‌کند که تنها با یک «گفتگوی» معرفت‌شناختی میان رشته‌های مختلف حرکت به سمت پیشرفت و امکان بررسی موضوعات در سیستم‌های پیچیده امکان‌پذیر است (ولچ، ج،۴، ۲۰۱۴). این امر از عدم قطعیت در معرفی فناوری جدید در جامعه را می‌تواند مرتفع سازد. به‌عبارت‌دیگر، اگرچه عدم توجه به سایر رشته‌های علمی، محصول محدودیت‌های معرفت علمی

---

[۱] Critical

[۲] Strategic

[۳] قابل‌توجه است که اگرچه این رویکرد پیش‌تر در خصوص مطالعه فلسفی و اخلاقی میان‌رشته‌ای بررسی شده است (Welch IV, 2011) اما در خصوص سیستم‌های اجتماعی‌فنی و هوش مصنوعی تاکنون موردتوجه پژوهشگران قرار نگرفته و این موضوع یکی دیگر از نوآوری‌های مقاله حاضر است. همچنین باید در نظر داشت که ایده ما از این مطالعه میان‌رشته‌ای بیش از آنکه بر اساس پژوهش‌های گذشته (ibid) باشد، با توجه به نیازهای روایت پراگماتیستی سیستم‌های اجتماعی‌فنی صورت‌بندی شده و از این منظر نیز در نوع خود بدیع است.

[۴] البته باید خاطرنشان کرد، شیوه پیشنهادی می‌تواند یکی از راه‌کارها باشد و شایسته است سایر پژوهشگران وجود راه‌کارهای دیگر را مورد ارزیابی قرار دهند.

[۵] باید توجه کرد که این شک، به‌قصد افتادن در دام شک‌گرایی نیست.

[۶] Reductionist

نیست؛ اما در یک مطالعه میان‌رشته‌ای بر اساس گفتگوی معرفت‌شناختی میان رشته‌های مختلف، می‌توان بخشی از محدودیت‌های معرفت علمی را برطرف ساخت. به سخنی دیگر، از پس گفتگو میان مجموعه‌های ناهمگون دانش، به دانشی دست خواهیم یافت که پیش‌تر و با تکیه‌بر یک زمینه علمی ناشناخته و یا دست‌نیافتنی بود.

البته باید در نظر داشت که این معرفت علمی بر یک گفتگوی دو یا چند سویه تمرکز دارد و لذا در چنین مطالعه‌ای، یک رشته علمی یا نظریه به‌عنوان معیار یا برتر از سایر نظریات و رشته‌های دیگر در نظر گرفته نمی‌شود. بلکه همانطور که پیرس (۱۹۰۵) بر توجه به زمینه معرفت‌شناختی تأکید دارد و جیمز (۱۹۰۷) از اعتبار نظریه با توجه به کاربرد آن سخن می‌گوید، زمینه‌های کاربردی است که معیار بررسی میان‌رشته‌ای قرار خواهد گرفت. بر این اساس، یک نگرانی در خصوص مطالعه میان‌رشته‌ای مرتفع خواهد شد. به‌عبارت‌دیگر، زمانی که از حضور عوامل مختلف سخن به میان می‌آوریم، این امر ممکن است خود دام چالش پیچیدگی بررسی را به همراه داشته باشد. بااین‌حال، زمانی که از زمینه معرفت‌شناختی و کاربردی سخن به میان می‌آید این نگران تا حدود زیادی برطرف خواهد شد. به سخنی دیگر، در مطالعه میان‌رشته‌ای سیستم اجتماعی–فنی، تأکید ما بر این نیست که مطالعه میان تمامی رشته‌های علمی صورت بگیرد. بلکه این مطالعه میان‌رشته‌ای، در حوزه‌های معرفتی صورت می‌گیرد که در زمینه کاربردی سیستم هوش مصنوعی، کاربرد دارند.

البته مطالعه میان‌رشته‌ای یک ویژگی دیگر نیز نیاز خواهد داشت و آن هم عدم محدودیت به زمان طراحی است. دیویی (۱۹۲۷) اهمیت تجربه در اخلاق را به ما نشان داد. همانطور که ذکر شد، برخی از آثار و نتایج معرفی فناوری جدید بر جامعه، پسینی بوده و به‌تدریج خود را نمایان می‌سازند. لذا مطالعه میان‌رشته‌ای باید علاوه بر زمان طراحی، در زمان توسعه و استفاده از فناوری نیز ادامه یابد تا بتوان در طول زمان این آثار را شناسایی و راهکارهای مناسب آن‌ها ارائه کرد. در این راستا، چندان دور از ذهن نیست که در جریان این روند و با نمایان شدن آثار فناوری، نظریات و رشته‌های علمی دخیل برای بررسی نیز دچار تغییر شوند. اگر به خطاپذیری توجه داشته باشیم و اهمیت تجربه را در نظر بیاوریم، برای اینکه شناخت کامل‌تری از سیستم هوش مصنوعی، نحوه عملکرد و تأثیرات آن در سیستم اجتماعی–فنی به‌دست بیاوریم، باید همانطور که اجزای سیستم در یک «فرایند» و در روند همیشگی «شدن» هستند، مطالعه میان‌رشته‌ای را نیز یک فرایند که در طول روند صورت می‌گیرد در نظر گرفت.

در این راستا، باید در نظر داشت که اگرچه به‌صورت پسینی بسیاری از آثار و نتایج فناوری خود را نمایان می‌سازند، اما این موضوع به معنای عدم اهمیت مطالعه میان‌رشته‌ای پیش از طراحی سیستم نیست. به‌عبارتی دیگر، این موضوع را نباید به‌اشتباه به این معنا دریافت که به‌صورت پیشینی نمی‌توان هیچ‌کدام از آثار و یا نحوه عملکرد سیستم را بررسی کرد. در مباحث پیشین، دانستیم که با تکیه بر رویکرد تعامل‌گرایانه، فناوری ارزش‌بار است؛ اما این ارزش‌ها ذاتی نبوده و فناوری در زمان طراحی و استفاده ارزش‌بار می‌شود. حال اگر مطالعه فناوری را به پس از معرفی آن در جامعه محول کنیم، حداقل از دو جهت این امر قابل ایراد است. نخست آنکه معیارهای اولیه در طراحی فناوری تنها به تصمیمات برنامه‌نویسان محول کرده‌ایم. مثال ابتدایی این بخش به ما نشان داد که معیارهایی که نیاز است برای سیستم در نظر گرفته شوند بسیار گسترده بوده و محدود کردن تصمیمات در انتخاب ارزش‌ها به برنامه‌نویسان توجیه‌پذیر نخواهد بود.

دوم، در حقیقت اگر به اهمیت موضوع تغییر در ارزش‌ها توجه داشته باشیم، این امر برایمان روشن خواهد شد که همواره با تغییرات روبرو هستیم. به‌عبارتی دیگر، هر بررسی پسینی، خود در معرض تغییرات در آینده خواهد بود. اگر مقصود بررسی ما از بررسی پسینی، موکول کردن به آینده نامشخص باشد، یا هیچ‌گاه رخ نخواهد داد و یا اگر رخ دهد نمی‌تواند تاثیر عملی داشته باشد. برای روشن شدن این موضوع، باید در نظر داشت که با توجه به امکان تغییرات دائمی، تنها زمانی تغییر جدیدی رخ نخواهد داد که دیگر از یک فناوری استفاده نکنیم. پر واضح است که این امر نمی‌تواند هیچ گونه تاثیر عملی در استفاده یا طراحی فناوری را با خود به همراه بیاورد. اما نظر به تجربه‌های پیشین، بسیاری از آثار فناوری را می‌توان پیش از طراحی آن به‌صورت مطالعه میان‌رشته‌ای مورد ارزیابی قرار داد. لذا در مجموع می‌توان اینگونه جمع‌بندی کرد که مطالعه میان‌رشته‌ای مورد نظر، باید از زمان طراحی سیستم آغاز شود و تا پایان زمان استفاده از آن همواره ادامه یابد.

نتیجه‌گیری

نتایج این مقاله نشان می‌دهد، رویکرد اجتماعی–فنی در بررسی موضوعات و مسائل اخلاقی و اجتماعی تصویری جامع از عوامل مؤثر بر فناوری و تأثیرات آن بر جامعه ارائه داده و ضرورت دارد پژوهشگران بیشتری به این رویکرد توجه و دقت نظر داشته باشند. همچنین با توجه به وجود عوامل مختلف تأثیرگذار در طراحی و استفاده از فناوری هوش مصنوعی که می‌توانند آثار و نتایج اخلاقی به همراه داشته باشند، یک مطالعه پراگماتیستی میان‌رشته‌ای می‌تواند برای بررسی دقیق‌تر به‌عنوان راه‌کار در نظر گرفته شود. چنین مطالعه میان‌رشته‌ای شکلی از تعامل و هم‌افزایی میان رشته‌ها خواهد بود که بر اساس آن یک سوال فراگیر از طریق چندین رشته مورد بررسی قرار می‌گیرد، که هدف آن ادغام مجموعه‌های ناهمگن دانش در یک نتیجه کلی از دیدگاه‌های متعدد است. چنین مطالعه‌ای می‌تواند آثار و نتایج گوناگونی را در زمینه اخلاق هوش مصنوعی شناسایی کرده و همزمان از افتادن در دشواری پیچیدگی سیستم موردبررسی جلوگیری کند. همچنین برخلاف شیوه‌هایی که تمرکز را بر قوانین و مقررات، سیاست‌گذاری، کدهای اخلاقی، طراحی، کنترل با عامل انسانی و سایر شیوه‌های دیگر می‌گذارند، مطالعه میان‌رشته‌ای فرصت این را فراهم می‌سازد که انتخاب شیوه کنترل فناوری را بر اساس شرایط عینی و تجربی انجام دهیم. با استفاده از چنین روشی، می‌توان عوامل مؤثر را بر اساس زمینه و کاربرد شناسایی کرده و راه‌کارهای کنترل و بررسی را نه به‌صورت پیشینی، بلکه تا حدودی به‌صورت پسینی نیز ارزیابی کنیم. از دیگر مزایای استفاده از مطالعه پراگماتیستی میان‌رشته‌ای در سیستم‌های اجتماعی–فنی نیز این خواهد بود که امکان ترکیب روش‌های کنترل و بررسی با توجه به زمینه کاربردی را فراهم می‌کند. با تکیه‌بر چنین رویکردی، نظریات مورداستفاده و شیوه‌های کنترل بر اساس زمینه انتخاب شده، در طول زمان قابل‌تغییر بوده و این فرصت فراهم می‌شود که هم از سودمندی فناوری بهره ببریم و هم از آسیب‌های آن جلوگیری کنیم.

منابع

رمضانی، مجید و فیضی درخشانی، محمدرضا. (۱۳۹۲). اخلاق ماشین: چالش‌ها و رویکردهای مسائل اخلاقی در هوش مصنوعی و ابرهوش. *فصلنامه اخلاق در علوم و فناوری*، ۸(۴)، ۴۳-۳۵. https://dorl.net/dor/20.1001.1.22517634.1392.8.4.4.7

سیفی، آناهیتا و رزمخواه، نجمه. (۱۴۰۲). نقدی بر پیش‌نویسِ توصیه‌های یونسکو در اخلاقِ هوش مصنوعی از منظر حق بر محیط زیست سالم. *پژوهش حقوق عمومی*، ۲۴(۷۸)، ۹-۴۷. https://doi.org/10.22054/qjpl.2022.63030.2659

عاشوری کیسمی، محمدعلی. (۱۴۰۳الف). طرح و بررسی برخی از مسائلِ اخلاقیِ هوش مصنوعی در هنر. *متافیزیک*، *۱۶*(۳۷)، ۹۳-۱۱۰. https://doi.org/10.22108/mph.2024.138105.1488

عاشوری کیسمی، محمدعلی. (۱۴۰۳ب). همگرایی حریم خصوصی و شفافیت، محدودیت‌های طراحی هوش مصنوعی. *حکمت و فلسفه*، ۲۰(۷۸)، ۷۳-۴۵. https://doi.org/10.22054/wph.2024.75680.2183

عاشوری کیسمی، محمدعلی و پرویزی، مریم. (۱۴۰۳). بررسی عاملیت اخلاقی هوش مصنوعی عام. پژوهش‌های *علم و دین*، *۱۵*(۱)، ۱۲۵-۱۵۱. https://doi.org/10.30465/srs.2024.49282.2162

عباسی، محمود و تیموری، مهرداد. (۱۴۰۲). مروری بر چالش‌های اخلاقی و حقوقی کاربرد هوش مصنوعی در نظام سلامت. *مجله اخلاق پزشکی*، *۱۷*(۴۸)، ۱-۱۱. https://doi.org/10.22037/mej.v17i48.44053

صراف‌زاده، شقایق و ابوطالب، احسان. (۱۴۰۲). اهمیت اخلاق در استفاده از هوش مصنوعی در آموزش پزشکی. *پژوهش در آموزش علوم پزشکی*. *۱۵*(۲)، ۱-۴. http://dx.doi.org/10.32592/rmegums.15.2.1

References

Abbasi, M. & Teymouri, M. (2024). A Review of the Ethical and Legal Challenges of Using Artificial Intelligence in the Health System. *Quarterly Journal of Medical Ethics*, *17*(48), 1–11. https://doi.org/10.22037/mej.v17i48.44053 (*in Persian*)

Alvargonzález, D. (2011). Multidisciplinarity, interdisciplinarity, transdisciplinarity, and the sciences. International studies in the philosophy of science, 25(4), 387-403. https://doi.org/10.1080/02698595.2011.623366

Ashouri Kisomi, M. A. & Parvizi, M. (2024). Investigation of the Ethical Agency of Artificial General Intelligence. *Science and Religion Studies*, *15*(1), 125-151. https://doi.org/10.30465/srs.2024.49282.2162 (*in Persian*)

Ashouri Kisomi, M. A. (2024a). Investigating some ethical issues of artificial intelligence in art. Metaphysics, 16(37), 93-110. https://doi.org/10.22108/mph.2024.138105.1488 (*in Persian*)

Ashouri Kisomi, M. A. (2024b). Convergence of privacy and transparency, limitations of artificial intelligence design. Wisdom And Philosophy, 20(78), 45-73. https://doi.org/10.22054/wph.2024.75680.2183 (*in Persian*)

Boenink, M., & Kudina, O. (2020). Values in responsible research and innovation: from entities to practices. Journal of Responsible Innovation, 7(3), 450-470. Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. MIT press. https://doi.org/10.1080/23299460.2020.1806451

Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford: Oxford University Press.

Coeckelbergh, M. (2022b). *Digital technologies, temporality, and the politics of co-existence*. Palgrave Macmillan.

Collingridge, D. (1980). The social control of technology. London: Frances Pinter.

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., ... & de Oliveira, N. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. Patterns, 4(10). https://doi.org/10.1016/j.patter.2023.100857

Crawford, S. E. S., & Ostrom, E. (1995). A Grammar of Institutions. *American Political Science Review*, 89(3), 582–600. https://doi.org/10.2307/2082975

de Wildt, T. E., & Schweizer, V. J. (2022). Exploring value change. *Prometheus*, 38(1), 25-44. http://doi.org/10.13169/prometheus.38.1.0025

Decker, M. (2004). The role of ethics in interdisciplinary technology assessment. *Poiesis & Praxis*, 2, 139-156. https://doi.org/10.1007/s10202-003-0047-0

Dewey, J. (1917). The need for a recovery of philosophy (1917). In L.A. Hickman & T.M. Alexander (Eds.), The essential Dewey (pp. 46-70). Bloom- ington and Indianapolis: Indiana University Press.
Dewey, J. (1927). *The public and its problems*. New York: Holt.
Dewey, J. (2004). *Reconstruction in Philosophy*. Courier Corporation.
European Commission. (2023). EU-U.S. Terminology and Taxonomy for Artificial Intelligence. Retrieved 22-11-2024 from https://digital-strategy.ec.europa.eu/en/library/eu-us-terminology-and-taxonomy-artificial-intelligence
Floridi, L. (2016). Should We Be Afraid of AI? Machines Seem to Be Getting Smarter and Smarter and Much Better at Human Jobs, yet True AI Is Utterly Implausible. Why? Aeon, May 9. Retrieved 12-1-2025 from https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible
Floridi, L. (2023). The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities. United Kingdom: OUP Oxford.
Frodeman, R., Mitcham, C. (eds.) (2004). New Directions in the Philosophy of Science. Toward a Philosophy of Science Policy, Special Issue. Philosophy Today 48(5).
Gambo, I. P., & Taveter, K. (2021, July). A Pragmatic View on Resolving Conflicts in Goal-oriented Requirements Engineering for Socio-technical Systems. In *ICSOFT* (pp. 333-341). http://doi.org/10.5220/0010605703330341
Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and machines*, 30(1), 99-120. https://doi.org/10.1007/s11023-020-09517-8
Hoffmann, A. L. (2019). Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society*, 22(7), 900-915. https://doi.org/10.1080/1369118x.2019.1573912
Hopster, J. (2021). What are socially disruptive technologies? *Technology in Society*, 67, 101750. https://doi.org/10.1016/j.techsoc.2021.101750
James, W. (1907). Pragmatism: A New Name for some Old Ways of Thinking. Cambridge: Harvard University Press.
Kroes, P., Franssen, M., Poel, I. V. D., & Ottens, M. (2006). Treating socio technical systems as engineering systems: some conceptual problems. *Systems Research and Behavioral Science: The Official Journal of the International Federation for Systems Research*, 23(6), 803-814. https://doi.org/10.1002/sres.703
Kudina, O., & Van de Poel, I. (2024). A sociotechnical system perspective on AI. *Minds and Machines*, *34*(3), 21. https://doi.org/10.1007/s11023-024-09680-2
Legg, C., & Hookway, C. (2008). Pragmatism. The Stanford Encyclopedia of Philosophy. Retrieved 14-1-2025 from https://plato.stanford.edu/entries/pragmatism
Levesque, H. J. (2017). *Common sense, the Turing test, and the quest for real AI*. MIT Press.
Mäki, U. (2016). Philosophy of interdisciplinarity. what? why? how?. *European Journal for Philosophy of Science*, 6, 327-342. https://doi.org/10.1007/s13194-016-0162-0
Manheim, K., & Kaplan, L. (2019). Artificial intelligence: Risks to privacy and democracy. *Yale JL & Tech.*, 21, 106.
Muller, V. C., & Cannon, M. (2022). Existential risk from AI and orthogonality: Can we have it both ways? *Ratio*, 35(1), 25-36. https://doi.org/10.1111/rati.12320
Muller, V.C. (2016). New Developments in the Philosophy of AI. In: Müller, V.C. (eds) Fundamental Issues of Artificial Intelligence. Synthese Library, vol 376. Springer, Cham. https://doi.org/10.1007/978-3-319-26485-1_1
Pareto, J., & Coeckelbergh, M. (2024). Social assistive robotics: an ethical and political inquiry through the lens of freedom. *International journal of social robotics*, 16(8), 1797-1808. https://doi.org/10.1007/s12369-024-01161-x
Peirce, C. S. (1955). The scientific attitude and fallibilism. *Philosophical writings of Peirce*, 42-59.

Peirce, C.S. (1905). What pragmatism is. In E.C. Moore (Ed.), *Charles sanders Peirce: The essential writings* (pp. 262-280). New York: Harper & Row.

Pfeiffer, J., Gutschow, J., Haas, C., Möslein, F., Maspfuhl, O., Borgers, F., & Alpsancar, S. (2023). Algorithmic fairness in AI: an interdisciplinary view. *Business & Information Systems Engineering*, 65(2), 209-222. https://doi.org/10.1007/s12599-023-00787-x

Pitt, J. C. (2013). "Guns don't kill, people kill"; values in and/or around technologies. In *The moral status of technical artefacts* (pp. 89-101). Dordrecht: Springer Netherlands.

Prem, E. (2023). From ethical AI frameworks to tools: a review of approaches. *AI and Ethics*, *3*(3), 699-716. https://doi.org/10.1007/s43681-023-00258-9

Putnam, H. (1995). Pragmatism. In *Proceedings of the Aristotelian Society*. 95, 291–306. http://www.jstor.org/stable/4545222

Ramezani, M. & Feizi-Derakhshani, M. R. (2014). Machine Ethics: Ethical Challenges and Strategies in Artificial Intelligent and Superintelligence. *Ethics in Science and Technology*. 8 (4), 35-43. https://dorl.net/dor/20.1001.1.22517634.1392.8.4.4.7 (*in Persian*)

Roco, M. C., Bainbridge, W. S. (eds.) (2002). Converging Technologies for Improving Human Performance. Nanotechnology, Biotechnology, Information Technology and Cognitive Science. Arlington, VA: National Science Foundation.

Russell, S. (2019). *Human Compatible*. New York: Viking Press.

Sarrafzadeh, S. & Aboutaleb, E. (2023). The Importance of Ethics in Using of Artificial Intelligence in Medical Education. *Research in Medical Education*, *15*(2), 1-4. http://dx.doi.org/10.32592/rmegums.15.2.1 (*in Persian*)

Schmidt, J. C. (2011). What is a problem? On problem-oriented interdisciplinarity. Poiesis & Praxis, 7, 249-274. https://doi.org/10.1007/s10202-011-0091-0

Seifi, A. & Razmkhah, N. (2023). An In-depth Analysis of the Initial Draft of the Recommendation on the Ethics of Artificial Intelligence: Exploring the Right to Environment. *Public Law Researsh*, *24*(78), 9-47. https://doi.org/10.22054/qjpl.2022.63030.2659 (*in Persian*)

Srivatsa, N., Kaliarnta, S., & Kormelink, J. G. (2017). Responsible innovation: From MOOC to book.

Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. Internet policy review, *8*(2). https://heinonline.org/HOL/P?h=hein.journals/gtltr4&i=13

UNICEF. (2021). Policy guidance on AI for children. United Nations Children's Fund.

Van de Poel, I. (2016). An ethical framework for evaluating experimental technology. *Science and engineering ethics*, *22*(3), 667-686. https://doi.org/10.1007/s11948-015-9724-3

Van de Poel, I. (2021). Design for value change. Ethics and Information Technology, *23*(1), 27-31. https://doi.org/10.1007/s10676-018-9461-9

Van de Poel, I. (2023). AI, Control and Unintended Consequences: The Need for Meta-Values. In Rethinking Technology and Engineering: Dialogues Across Disciplines and Geographies (pp. 117-129). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-25233-4_9

Van de Poel, I., & Kudina, O. (2022). Understanding technology-induced value change: A pragmatist proposal. *Philosophy & Technology*, *35*(2), 40. https://doi.org/10.1007/s13347-022-00520-8

Van de Poel, I., Hermann, J., Hopster, J., Lenzi, D., Nyholm, S., Taebi, B., & Ziliotti, E. (2023). *Ethics of socially disruptive technologies: An introduction*. Open Book Publishers. https://doi.org/10.11647/OBP.0366

Van de Poel, I., Kroes, P. (2014). Can Technology Embody Values? In: Kroes, P., Verbeek, PP. (eds) The Moral Status of Technical Artefacts. Philosophy of Engineering and Technology, vol 17. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-7914-3_7

Van den Hoven, J. (2012). Neutrality and technology: Ortega Y Gasset on the good life. In *The Good Life in a Technological Age* (pp. 327-338). Routledge.

Van Der Weij, F., Steinert, S., Van De Poel, I., Alleblas, J., Melnyk, A., & De Wildt, T. (2023). *Value Change and Technological Design. IEEE Technology and Society Magazine*, 42(3), 25-32.

Watson, D. S., Mökander, J., & Floridi, L. (2024). Competing narratives in AI ethics: a defense of sociotechnical pragmatism. *AI & SOCIETY*, 1-23. https://doi.org/10.1007/s00146-024-02128-2

Watson, D., & Mökander, J. (2023). In defense of sociotechnical pragmatism. In *The 2022 Yearbook of the Digital Governance Research Group* (pp. 131-164). Cham: Springer Nature Switzerland.

Weingart, P. (2000). Interdisciplinarity; The Paradoxical Discourse. In: Weingart, P., Stehr, N. (eds.) (2000). Practising Interdisciplinarity. Toronto: University of Toronto Press, pp. 25–41.

Welch IV, J. (2011). The Emergence of Interdisciplinarity from Epistemological Thought. *Issues in Integrative Studies*, 29, 1-39.

Werthner, H., Prem, E., Lee, E. A., & Ghezzi, C. (2022). *Perspectives on digital humanism* (p. 342). Springer Nature.

Young, M. T., & Coeckelbergh, M. (2024). Keeping Things Going: Maintenance and Philosophy of Technology. In *Maintenance and Philosophy of Technology* (pp. 1-36).

Zuboff, S. (2023). The age of surveillance capitalism. In *Social theory re-wired* (pp. 203-213). Routledge.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی