



## Virtue Ethics and the Moral Status of Artifacts

Zargar, Zahra 

Assistant Professor in Institute for Science and Technology Studies, Shahid Beheshti University.

(z\_zargar@sbu.ac.ir)

---

### ABSTRACT

In recent decades philosophers of technology have paid more attention to the doctrines of virtue ethics. Many virtue theorists claim that in dealing with certain issues, this view has significant advantages over utilitarian and deontological views. Therefore, appealing to virtue ethics' framework, they suggest prescriptions for designing and using technologies. In this paper, we focus on moral problems concerning human-artifact interaction, and specially on those cases in which artefacts appear as quasi-other for humans and enter in quasi-social relations with them. Our question is about the moral status of artefacts in these situations. In other words, we want to see whether there are moral obligations at work in dealing with artefacts in these cases or not. In this paper, after reviewing virtue philosophers' insights about the problem, we will explain and examine two main critiques that are leveled at virtue ethical claims about technology ethics: the question from originality of virtue ethic's views and "the paradox of simultaneous anthropomorphism and dehumanization."

**Keywords:** Virtue Ethics, Technology Ethics, Moral Status of Artefacts, Social Artefacts, Human-Technology Interaction.

---

Received: 2025/03/29 - Received in revised form: 2025/06/15 - Accepted: 2025/09/06 - Published online: 2025/09/25

---

□ Zargar, Z. (2025). Virtue Ethics and the Moral Status of Artifacts, *Journal of Philosophical Sophia Perennis*, 46(2), 165-189.

<https://doi.org/10.22034/iw.2025.514372.1836>

---

This article is distributed under the terms of the Creative Commons Attribution 4.0 <http://creativecommons.org/licenses/by/4.0>



## Extended Abstract

### 1- Introduction

Virtue Ethics is one of the three main theories of ethics. The roots of virtue ethics could be traced back to the ancient era in Aristotle's works and religious doctrines. By the beginning of the modern age, virtue ethics had been eclipsed by Utilitarian and Deontological theories of ethics. However, in recent decades, due to the emergence of non-human world ethical issues (i.e., animal rights, environmental ethics, and ethics of technology), Virtue ethics has again returned to the scene of ethical debates. Advocates of Virtue ethics find novel and unique capacities in its doctrines for articulating a moral framework that involves non-human entities.

Utilitarian theory assesses actions based on the amount of pleasure and pain they create. Actions toward entities that are not sentient neither cause pleasure nor pain, hence, can't be morally evaluated on this consequentialist basis. On the other hand, Deontological Ethics presupposes a system of rights and duties, which in turn determines the moral value of actions. Again, we can't attribute rights and duties to non-human entities as we do for human beings. But in Virtue ethics, actions are evaluated based on their impact on the character and flourishing of the agent. It doesn't matter whether one is acting toward a human being or a non-human. The only important thing is the effect of the action on the character of the actor. This feature of Virtue ethics provides it with the capacity for developing ethical prescriptions for human-artifact interaction.

In this paper, we explore the implications of Virtue ethics for human interactions with artifacts when the artifact appears to us as a social being or quasi-other. First, we consider arguments of virtue theorists in favor of the moral patiency of social artifacts. Then, we examine these views against two challenges: the originality of Virtue ethics' implications, and the Anthropomorphizing while Dehumanizing Robots Paradox.

### 2- Artefacts as Moral Patient

Are our actions toward social robots, virtual avatars, or Chatbots morally evaluable? And if yes, how can it be? There are two major approaches for answering these questions: the standard approach and the relational approach. According to the standard approach, the moral status of each entity depends on the existence of some properties in it. Hence, the question about the moral status of an entity amounts to the question of its properties. On the other hand, from the relational point of view, entities are defined based on their relations to other social and natural entities and how they appear for them (Coeckelbergh, 2014, p. 62-6).

Due to some difficulties and deficiencies of the Standard view, philosophers like Coeckelbergh and Cappuccio et al. admitted the relational approach for analyzing the moral status of artifacts. Also, they located this approach in the framework of Virtue ethics (Coeckelbergh, 2014; Cappuccio et al. 2019).

According to Coeckelbergh, the moral status of an entity is constructed through its interaction with other entities. These interactions occur in material and social relations which include various patterns of behavior, thinking, interpretation, and evaluation. Our moral experience of another entity, therefore, is both personal and dependent on the culture that gives meaning to our interactions (Coeckelbergh, 2014, p. 69)

Based on different ways of experiencing an entity, it can have a different moral status. Despite the standard model, in the relational view, the moral status of an entity isn't objective. Questioning about the moral status of an artefact is not either a scientific or a philosophical question; it is a practical question about how to

confront an entity and behave with it (Coeckelbergh, 2014, p. 65-66). Thus, our tacit knowledge acquired through interacting with an entity enables us to answer the question (Coeckelbergh, 2014, p. 70-71)

Cappuccio et al. also admit the relational approach toward moral status and emphasize on the consonance of this view with doctrines of Virtue ethics. According to Virtue ethics, for moral evaluation of an action, we don't need to examine the properties of the action's receiver to calculate the amounts of pleasure and pain, or consider its rights. What matters is the impact of the action on the agent, which in turn depends on the material and social background in which the interactions take place (Cappuccio et al., 2019, p. 13-14).

According to virtue theorists, if our relation with an artefact resembles our relations with human beings sufficiently to make similar impacts on us, then interacting with it affects our personality. Based on this, one can divide various artefacts according to their degree of anthropomorphism. Coeckelbergh thinks that by increasing the entanglement of human and technological worlds, we will need more accurate moral categories for describing the moral status of artefacts (Coeckelbergh, 2010, p. 239)

For example, artefacts which are designed in order to be social and emotional would be perceived more like humans, therefore, have a moral status closer to humankind (Cappuccio et al., 2019, p. 18). As an instance, violent computer games are designed to simulate a world as indiscernible from the real world as possible, to immerse the player. The player perceives the other avatars in the game much like real social agents, while the game encourages her and increases her score for doing violent actions against those quasi-real avatars. In this situation, the player's behavior with the avatar has impact on her character and fosters the vice of lacking empathy. Hence, we can evaluate violent action against the game's avatar as immoral (Coeckelbergh, 2007, p. 24-25).

### 3- The Originality of Virtue Ethics' Doctrines

The doctrines of Virtue ethics have been criticized from various aspects. For example, some philosophers believe that Virtue ethics doesn't provide us with straight and simple moral prescriptions, or isn't useful for making decisions in moral dilemmas, or that the list of virtues and vices is arbitrary, vague, and not universal. Putting all these critiques aside, we want to concentrate on another challenge for advocates of Virtue ethics: are the doctrines of Virtue ethics original, or we can derive them from Utilitarian views?

Utilitarians evaluate an action based on its consequences. It seems that in the case of Virtue ethics, we consider the consequence or impact of an action on the agent's character and her flourishing and eudemonia. So, one can reasonably ask what is the difference between the Utilitarian assessment of consequences and the Virtue ethics' one?

To answer this question, it is useful to consider the difference between versions of Virtue ethics. There are at least three distinct versions of Virtue ethics: the eudemonist version in which happiness and flourishing are the basic concepts, the agent-based version in which the moral exemplar is the basic concept and actions are examined according to the similarity with actions of moral exemplars in similar situations, and the target-centered version in which virtues and vices are the basic concepts and the morality of actions are measured based on their role in fostering virtues or vices (Hursthouse, 2022, pp. 4-7).

The critique for lack of originality just threatens the eudemonist version, since



the definition of eudemonia is very close to the basic concept of happiness in Utilitarian views, and sometimes equals it. But this is not the case for the other two versions. While in the agent-based version of Virtue ethics, the impact of the action on the agent is considered, this impact is far from the categorical properties obtained in the world's state of affairs, which are measurable. In this version, the action makes differences in the person's dispositional properties, which in the absence of stimulation may never manifest and change the world's state of affairs. Therefore, while the agent-based version of Virtue ethics examines the consequences of action, these consequences are metaphysically distinct from the consequences calculated in Utilitarian theories. The same is true about the target-based version of Virtue ethics. Hence, the critique from originality only would be levelled at the eudemonist version of Virtue ethics.

#### 4- Anthropomorphizing while Dehumanizing Robots Paradox

Another problem with Virtue theorists' views about the moral status of artefacts refers to a phenomenon called by Cappuccio et al. as "Anthropomorphizing while Dehumanizing Robots Paradox" (Cappuccio et al., 2019, p. 26).

This paradox refers to situations when an entity is both perceived as human and non-human. According to Cappuccio et al., by interacting with social artefacts, we experience them very similarly to our human fellows; they provoke emotions and behaviors in us just as those we have in our human-human interactions. But on the other hand, we know that they aren't human. We see them as artefacts that are in our service. Thus, we sell and buy them, turn them on and off, and if needed, we can throw them away. These two aspects are in contradiction.

More than just an emotional confusion, this contradiction has moral importance. Historically, some people, like slaves, and sometimes women and children, were treated like dehumanized humans. This immoral treatment has fostered the masterhood personality in people, which includes the vices of lack of empathy, selfishness, and irresponsibility. The revival of this manner of behavior (this time toward human-like artefacts) can cultivate those vices in the users of these social artefacts.

Coeckelbergh thinks that this contradictory experience results from the standard views, since the contradiction is constituted from the opposition of our propositional knowledge (artefacts aren't human) and our experiential knowledge (we perceive them like humans) (Coeckelbergh, 2014, p. 64). If we rely on our experiential knowledge, we don't need to consider the propositional knowledge anymore, and the confusion disappears.

Unfortunately, this solution isn't promising. Since the dehumanizing side of the paradox is fueled by our experience, too. It is not the case that we just theoretically know that artefacts aren't human; moreover, we experience them as our properties, which are aimed to be at our service and could be customized as we wish. Therefore, turning from the standard view to the relational one doesn't seem helpful for overcoming the problem.

In fact, from the Virtue ethics point of view, one can find reasons for elevating the anthropomorphism side of the paradox: if you conceive something like a human, the way you treat it affects your character, so you should treat it morally as a real human. Also, on the same ground, we can elevate the dehumanizing side: it is far from practical wisdom to treat a non-human as a human. Thus, Virtue ethics apparently doesn't suggest a clear and definite solution for dissolving the paradox.



## 5- Conclusion

Relational understanding is a basic concept in Virtue ethics. This implies that if an entity appears to us like a human moral patient, we have to adhere moral obligations in our treatment of it. Since our interactions can have an impact on our moral character and foster vices or virtues, which ultimately affect our eudemonia. This analysis makes it possible to confer the status of moral patient to some artefacts in some relations.

But the Virtue ethics' doctrines are facing some theoretical challenges. One important challenge questions the originality of this view and the possibility of accommodating its prescriptions to Utilitarian ones. We saw that this challenge only threatens the eudemonist version of Virtue ethics and is not serious for agent-based and target-centered versions. The other challenge refers to the morally problematic status of artefacts which are simultaneously perceived as human and non-human. The anthropomorphism implies that we should treat them morally, while the dehumanizing side implies that they don't have moral status. It seems that current theorists of Virtue ethics haven't suggested a convincing answer to this problem.

# اخلاق فضیلت‌گرا و وضعیت اخلاقی مصنوعات

زهرا زرگر 

علمی - پژوهشی



استادیار پژوهشکده مطالعات علم و فناوری دانشگاه شهید بهشتی. (z\_zargar@sbu.ac.ir)

## چکیده

اخلاق فضیلت‌گرا در دهه‌های اخیر مورد توجه فیلسوفان فناوری قرار گرفته است. حامیان فضیلت‌گرایی معتقدند این دیدگاه در مواجهه با برخی مسائل نسبت به رویکردهای فایده‌گرا و وظیفه‌گرا مزایای قابل توجهی دارد و بر این اساس، تجویزهایی را درباره توسعه و کاربری فناوری ارائه می‌کنند. در این مقاله، از میان انواع مسائل حوزه اخلاق فناوری بر مسائل مربوط به تعامل انسان و فناوری تمرکز می‌کنیم و از میان این مسائل به مسئله وضعیت اخلاقی مصنوعات می‌پردازیم که برای کاربر جایگاه شبه‌دیگری داشته و با او وارد روابطی شبیه به روابط اجتماعی می‌شوند. پرسش از وضعیت اخلاقی مصنوعات در چنین مواردی، ناظر به لزوم یا عدم لزوم رعایت ملاحظات اخلاقی در تعامل با این نوع مصنوعات است. پس از شرح تحلیل و تجویز فضیلت‌گرایان درباره این موارد، دو چالش اصلی درباره رویکرد فضیلت‌گرایانه مورد بحث قرار می‌گیرند. چالش اول اصالت رویکرد فضیلت‌گرایی در مقایسه با فایده‌گرایی را زیر سؤال می‌برد؛ و چالش دوم به «پارادوکس انسان‌انگاری و انسانیت‌زدایی همزمان» می‌پردازد. با شرح و بررسی این دو چالش ظرفیت فضیلت‌گرایی برای پاسخگویی به آنها ارزیابی می‌شود.

**کلیدواژه‌ها:** اخلاق فضیلت‌گرا، اخلاق فناوری، وضعیت اخلاقی مصنوعات، مصنوعات اجتماعی، تعامل انسان-فناوری.

---

تاریخ دریافت: ۱۴۰۴/۰۱/۰۹ - تاریخ اصلاح: ۱۴۰۴/۰۳/۲۵ - تاریخ پذیرش: ۱۴۰۴/۰۶/۱۵ - تاریخ انتشار آنلاین: ۱۴۰۴/۰۷/۰۳

---

□ زرگر، زهرا؛ (۱۴۰۳) اخلاق فضیلت‌گرا و وضعیت اخلاقی مصنوعات، *جاویدان خرد*، ۴۶ (۲)، ۱۶۵ - ۱۸۹.

<https://doi.org/10.22034/iw.2025.514372.1836>

---

This is an open access article under the terms of the Creative Commons attribution 4.0  
<http://creativecommons.org/licenses/by/4.0>



## مقدمه

اخلاق فضیلت‌گرا در کنار فایده‌گرایی و وظیفه‌گرایی، سه نظریه اصلی فلسفی اخلاق را تشکیل می‌دهند. فایده‌گرایی در اخلاق، درست یا غلط بودن یک عمل را بر اساس وضعِ اموری که در پی انجام عمل حاصل می‌شود تعیین می‌کند. وظیفه‌گرایی، ارزیابی اخلاقی اعمال را بر اساس اصول و قواعدی دربارهٔ وظایف و حقوق افراد، انجام می‌دهد. اما اخلاق فضیلت‌گرا ارزش اخلاقی اعمال را بر اساس تأثیری که هر عمل بر شخصیت عامل می‌گذارد، معین می‌کند. در میان این سه رویکرد، اخلاق فضیلت‌گرا ریشه‌هایی باستانی دارد.

نخستین سرچشمه‌های فضیلت‌گرایی به آراء ارسطو، کنفوسیوس، و بودا بازمی‌گردد. همچنین مایه‌هایی از فضیلت‌گرایی را می‌توان در اخلاق اسلامی و مسیحی دید. در دوران روشنگری نظریه اخلاق وظیفه‌گرایانه کانت، و پس از آن فایده‌گرایی<sup>۱</sup> جان استوارت میل و جرمی بنتام به عنوان نظریه‌هایی مدرن در باب اخلاق مطرح و فراگیر شدند و مبنای تفکر اخلاقی در عصر مدرن قرار گرفتند (Mill et al., 2015). اما فضیلت‌گرایی به عنوان نظریه‌ای کمتر انتزاعی و فاقد ساختار قاعده - محور، خوشایند ذائقه فکری اندیشمندان مدرن نبود و به حاشیه رفت.

با این حال، در نیمه قرن ۲۰ اخلاق فضیلت‌گرا به صحنه نظریه‌های فلسفی بازگشت. طلوع این رجوع، مقاله آنسکام با عنوان «فلسفه اخلاق مدرن» بود که در آن بر اهمیت روانشناسی و توجه به شخصیت در مباحث اخلاقی تأکید شده، و محوریت «فضیلت» در تجویزهای اخلاقی مورد توجه قرار گرفته بود (Anscombe, 1958, p. 5). پس از این مقاله، فیلسوفانی چون مک‌این‌تایر و نوسباوم نیز در آثار خود به توسعه و احیای اخلاق فضیلت‌گرا پرداختند (Nussbaum, 1990; McIntyre, 1985). ظهور نظریه‌های اخلاق مراقبتی<sup>۲</sup> در اواخر قرن ۲۰ و اوایل قرن ۲۱ نیز به جهت وجوه اشتراک مهم این نظریه‌ها با رویکرد فضیلت‌گرا و انتقادی بودنشان نسبت به رویکردهای رقیب وظیفه‌گرا و فایده‌گرا، بر تحکیم موقعیت رویکرد فضیلت‌گرا در مباحث اخلاقی دهه‌های اخیر تأثیر گذاشت.<sup>۳</sup>

رویکرد فضیلت‌گرا به طور خاص مورد استقبال نظریه‌پردازان اخلاق کاربردی قرار گرفت. بسیاری از مباحث داغ در اخلاق معاصر، حول موضوعاتی چون حقوق حیوانات، اخلاق محیط زیست، و اخلاق فناوری شکل گرفته است؛ موضوعاتی که در آنها مسائل اخلاقی در ارتباط با اموری نا-انسانی مطرح می‌شوند. در کانون توجه قرار گرفتن وضعیت اخلاقی جهان نا-انسانی و مسائل پیرامون آن، نظریه‌های اخلاقی را با چالش‌های متعددی مواجه کرد و مدافعین رویکردهای مختلف در پی تحلیل و پاسخگویی به این مسائل برآمدند. در این میان، مدافعین

1. Utilitarianism

2. Care Ethics

۳. برای مثال (Held, 2005; Tronto 1993; Tronto, 2013).

رویکرد فضیلت‌گرا دیدگاه خود را واجد مزیتی نسبی در مقایسه با دو دیدگاه وظیفه‌گرا و فایده‌گرا یافتند. وظیفه‌گرایی برای تحلیل عمل اخلاقی، متکی بر نظامی از حقوق و وظایف است. در حالی که به نظر می‌رسد نمی‌توان برای جهان‌نا-انسانی حقوق و وظایفی شبیه به آنچه برای انسان‌ها قائل هستیم، در نظر بگیریم. بر این اساس، وضعیت اخلاقی جهان‌نا-انسانی (چه به عنوان عامل اخلاقی و چه مخاطب عمل اخلاقی) زیر سؤال می‌رود. از سوی دیگر، فایده‌گرایی هر عمل را بر اساس سود یا لذتی که می‌آفریند و زیان یا رنجی که می‌زداید ارزیابی می‌کند. در حالی که جهان‌نا-انسانی (و خصوصاً بی‌جان) ظرفیت بهره‌مندی، خسران، و لذت و رنجی مشابه انسان را ندارد. اما از آنجا که اخلاق فضیلت‌محور بر تأثیر عمل بر شخصیت عامل متمرکز می‌شود، می‌تواند با اجتناب از دشواری‌های فوق درباره اخلاقیات در مواجهات انسان با نا-انسان تجویزهایی داشته باشد. کتاب شخصیت و محیط زیست از رونالد سندلر (Sandler, 2009) درباره اخلاق محیط زیست، کتاب «فناوری و فضایل» از شنون والر (Vallor, 2016) درباره اخلاق فناوری، و کتاب «علم، فناوری و فضیلت» (Ratti and Stapleford, 2022) از جمله آثار متأخری هستند که مسائل اخلاقی فوق را از منظری فضیلت‌گرایانه تحلیل می‌کنند.

هدف مقاله حاضر بررسی دلالت‌های اخلاق فضیلت‌گرا درباره موارد خاصی از تعامل انسان و فناوری است. به طور کلی تعامل انسان و فناوری را می‌توان ذیل انواعی از روابط تحلیل کرد. برای مثال پیتر فریبک نقش واسطه‌گر فناوری را در هفت نوع رابطه میان انسان و جهان بر می‌شمرد. این روابط شامل روابط بدنمندی<sup>۱</sup>، هرمنوتیک<sup>۲</sup>، پس‌زمینه‌ای<sup>۳</sup>، غیریت<sup>۴</sup>، سایبرگی<sup>۵</sup>، استغراق<sup>۶</sup>، و افزونگی<sup>۷</sup> می‌شوند (Verbeek, 2015). در این مقاله ما بر مواردی تمرکز می‌کنیم که در آنها فناوری با انسان رابطه شبه‌دیگری پیدا می‌کند (رابطه غیریت در تقسیم‌بندی فریبک). منظور از شبه‌دیگری، حالتی است که در آن فناوری با انسان در نوعی رابطه انسانی قرار گرفته و جایگاهی را تصاحب می‌کند که سابقاً به انسانی دیگر تعلق داشته است. برای مثال اگر فناوری در جایگاه متصدی بانک، منشی، دستیار سخن‌گو یا پزشک قرار بگیرد، در نقش شبه‌دیگری ظاهر شده است. در حالی که فناوری‌هایی چون ماهواره، شبکه اینترنتی، یا پوشش‌های نانو کربن چنین خصلتی ندارند.

حالت‌هایی که در آنها فناوری همچون شبه‌دیگری ظاهر می‌شود نیز، از لحاظ وضعیت اخلاقی فناوری به دو دسته دیگر قابل تقسیم هستند: حالتی که در آن فناوری مخاطب عمل اخلاقی<sup>۸</sup> است، و حالتی که در

1. Embodiment
2. Hermeneutic
3. Background
4. Alterity
5. Cyborg
6. Immersion
7. Augmentation
8. Moral Patient



آن فناوری به عنوان عامل اخلاقی<sup>۱</sup> شناخته می‌شود. در حالت اول، وضعیت و خصوصیات مخاطب یک عمل (و نه عامل آن) در ارزش اخلاقی عمل تعیین‌کننده است. برای مثال لگد زدن به یک تگه کلوخ از نظر اخلاقی خنثا است، اما لگد زدن به یک کودک چنین نیست. این تفاوت در ارزش‌گذاری اخلاقی، به تفاوت وضعیت کودک با تگه کلوخ به عنوان مخاطب عمل بازمی‌گردد. از جهت دیگر، در حالت‌هایی که فناوری مسبب و علت انجام عملی است و این عمل به لحاظ ارزشی خنثا نیست، عاملیت اخلاقی فناوری موضوعیت پیدا می‌کند. برای مثال اگر خودرویی خود-ران به شهروند پیاده‌ای برخورد کرده و منجر به فوت او شود، فناوری در آنچه رخ داده است دارای (دست‌کم سطحی از) عاملیت است. وضعیت اخلاقی فناوری‌ها چه به عنوان مخاطب عمل اخلاقی و چه به عنوان عامل اخلاقی با پیشرفته‌تر شدن فناوری‌ها و وارد شدن آنها به رابطه شبه‌دیگری با انسان‌ها موضوعیت پیدا کرده و بحث‌های فراوانی را برانگیخته است.

در این مقاله بر پرسش از وضعیت اخلاقی فناوری‌ها در روابط شبه‌دیگری متمرکز می‌شویم. به طور خاص به این پرسش خواهیم پرداخت که طبق دیدگاه فضیلت‌گرا آیا فناوری نیز می‌تواند در جایگاه مخاطب عمل اخلاقی بوده و رفتار با آنها مشمول ملاحظات اخلاقی شود؟ پس از بررسی رویکرد فضیلت‌گرایان نسبت به این مسئله، به دو چالش اصلی که دیدگاه فضیلت‌گرایی با آن مواجه است خواهیم پرداخت و قابلیت این دیدگاه برای فائق آمدن بر این دو چالش را ارزیابی خواهیم کرد. به این ترتیب، در قسمت دوم مقاله به دلالت‌های اخلاق فضیلت‌گرا درباره جایگاه فناوری به مثابه مخاطب عمل اخلاقی می‌پردازیم؛ استدلال‌های فیلسوفان فضیلت‌گرا در دفاع از جایگاه اخلاقی فناوری را مرور کرده و پیامدهای این دیدگاه را شرح می‌دهیم. در قسمت سوم، چالش اصالت فضیلت‌گرایی را شرح داده و پاسخ‌های ممکن به آن را بررسی می‌کنیم. و در قسمت چهارم چالش «پارادوکس انسان‌نگاری و انسان‌زدایی همزمان» را شرح داده و قابلیت اخلاق فضیلت‌گرا برای پاسخگویی به آن را محک می‌زنیم.

## ۱. فناوری به عنوان مخاطب عمل اخلاقی

آیا کتک زدن یک ربات از نظر اخلاقی اشکال دارد؟ آیا استفاده از ناسزا در گفت‌وگو با چت‌بات کاری غلط است؟ به خاک و خون کشیدن آواتارها در بازی‌های رایانه‌ای چگونه؟ واضح است که چنین رفتارهایی در قبال انسان‌ها و در شرایط عادی اخلاقاً غلط است. اما برای ارزیابی اخلاقی این اعمال در قبال امور فناورانه، باید به این پرسش پاسخ دهیم که آیا تعامل ما با ربات‌ها، چت‌بات‌ها و آواتارها و به طور کلی مصنوعات مجسم و

1. Moral Agent

مجازی، در دایره ارزش‌گذاری اخلاقی قرار می‌گیرد؟ آیا می‌توان برای مصنوعات نوعی شأنیت اخلاقی در نظر گرفت؟ و اگر بله، بر چه اساسی؟

در پاسخ به این پرسش دو رویکرد اساسی در پیش گرفته شده است؛ رویکرد استاندارد (یا عینی) و رویکرد رابطه‌ای<sup>۱</sup>. مطابق رویکرد استاندارد، منزلت اخلاقی هر چیز وابسته به وجود برخی ویژگی‌های خاص در آن چیز است. برای مثال، برخی قابلیت رنج کشیدن و لذت بردن را ویژگی‌ای معرفی می‌کنند که وجود آن در هر چیز، آن را دارای منزلت اخلاقی می‌کند. بر این اساس، پرسش از منزلت اخلاقی یک چیز، پرسشی وجودشناختی است که ما را به تحقیق و تأمل درباره ویژگی‌های آن رهنمون می‌کند.

در نقطه مقابل، طبق رویکرد رابطه‌ای، هر هویت بر اساس مجموعه روابط طبیعی و اجتماعی که با سایر هویت‌ها دارد تعریف می‌شود و وضعیت اخلاقی‌اش نیز به این باز می‌گردد که آن هویت با دیگر هویت‌ها چه ارتباطی دارد و برای آنها چطور پدیدار می‌شود (Cockelbergh, 2014, p. 62-63).

فیلسوفانی همچون کاکلبرگ، کاپوچو و دیگران در ارتباط با وضعیت اخلاقی مصنوعات، رویکرد رابطه‌ای را برگزیده‌اند.<sup>۲</sup> آنها برای این منظور اولاً بر نقایص و دشواری‌های رویکرد استاندارد تکیه کرده و ثانیاً رویکرد رابطه‌ای را فاقد چنان معضلاتی و در عین حال سازگار با رویکرد فضیلت‌گرا دانسته‌اند (Cockelbergh, 2014; Cappuccio et al. 2019).

کاکلبرگ ادعا می‌کند مدافعین رویکرد استاندارد با دو چالش جدی مواجه‌اند. چالش اول شناسایی ویژگی‌هایی است که مرتبط با اخلاقی بودن یک هویت هستند. و چالش دوم، تشخیص وجود این ویژگی‌ها در هویت‌ها در صورتی است که مرتباً با اخلاقی بودن یک هویت هستند. و چالش دوم، تشخیص وجود این ویژگی‌ها و بررسی وجود یا عدم آنها، نابسند هستند (Cockelbergh, 2014, p. 64). انتخاب از بین ویژگی‌هایی همچون آگاهی، داشتن عواطف، سوژه حیات بودن، و ... که گزینه‌هایی برای ویژگی‌های مرتبط با وضعیت اخلاقی هستند، و همچنین تحلیل این ویژگی‌ها مسائل مناقشه‌برانگیزی را پیش می‌کشد که به راحتی نمی‌توان به آنها پاسخ گفت. فیلسوفان هم در مورد این که کدام یک از این ویژگی‌ها بنیان وضعیت اخلاقی هستند دیدگاه‌های متفاوتی دارند، و هم در مورد امکان تحقق این ویژگی‌ها در مصنوعات، اختلاف نظر دارند.<sup>۳</sup> در وهله بعد، حتی اگر درباره ویژگی‌ها و ماهیت آنها به توافق نظر برسیم، به دلیل عدم دسترسی مستقیم به منظر هویت دیگر، تشخیص وجود این ویژگی‌ها در این هویت با مانعی معرفت‌شناختی همراه است.

#### 1. relational

۲. کاکلبرگ در کتاب «پرورش روابط اخلاقی» انواع دیدگاه‌های رابطه‌ای را تحلیل و نقد کرده و از میان آنها دیدگاه پدیدارشناختی-استعلایی را برای تعریف جایگاه اخلاقی برمی‌گزیند (Cockelbergh, 2012).

۳. برای نمونه، برای مشاهده دیدگاه‌های مختلف فیلسوفان درباره رابطه عواطف و ظرفیت اخلاقی مصنوعات این منبع را ببینید: (زرگر، ۱۴۰۴).



علاوه بر این، اگر رویکرد استاندارد را بپذیریم، چنین نتیجه می‌شود که رفتار ما با هویت مختلف باید مبتنی بر معرفت ما نسبت به وضعیت اخلاقی آن هویت باشد. مثلاً اگر می‌دانیم که هویتی فاقد آگاهی است، باید با آن به مثابه امری غیر شخصیت‌مند برخورد کنیم (Cockelbergh, 2014, p. 64).

اما مطالعات تعاملات انسان-ربات<sup>۱</sup> شیوع شخص‌پنداری ربات‌ها در میان کاربران انسانی را نشان می‌دهد. در یک نمونه یک سرهنگ نظامی با یک ربات مین‌یاب کار می‌کرد. این ربات پاهای متعددی داشت، تا با از بین رفتن پایش در اثر انفجار مین بتواند به کار خود ادامه دهد. در حین کار، ربات پاهایش را یکی یکی از دست داد و نهایتاً خود را با یک پا روی زمین می‌کشید تا مین دیگری را پیدا کند. سرهنگی که با این ربات کار می‌کرد برآشفته شد، کار را متوقف کرد و این آزمایش را غیرانسانی خواند. در نمونه‌ای دیگر، این مطالعات نشان‌دهنده شخص‌پنداری در مورد جاروبرقی خودکار رومبا<sup>۲</sup> است. در حالی که رومبا نه در طراحی و نه رفتار جنبه اجتماعی ندارد، صاحبانش به خاطر این که منزل را نظافت می‌کند، نسبت به آن حس قدردانی دارند. هرچند رومبا میان کاربر و صندلی و میز تفاوتی نمی‌گذارد، اما صاحبانش تمایل دارند برای رومبا کاری انجام دهند، مثلاً بخشی از نظافت را خودشان انجام دهند تا رومبا استراحت کند. بعضاً رومبا را به والدینشان معرفی می‌کنند یا او را با خود به سفر می‌برند و او را عزیز خود می‌خوانند. حتی خود پژوهشگران رباتیک نیز گزارش می‌کنند هنگامی که در آزمایشگاه ربات انسان‌نما تصادفاً روشن باقی می‌ماند و به صحبت حضار توجه نشان می‌دهد، از حضور او معذب می‌شوند؛ یا بعضاً میان خود و ربات پرده‌ای می‌کشند تا از احساس نگاه او تمرکزشان به هم نخورد (Scheutz, 2009, p. 6-9).

کاکلبرگ می‌گوید طبق رویکرد استاندارد، در مواردی - شبیه به موارد فوق - که با ربات‌ها همچون انسان رفتار می‌شود، میان معرفت و عمل ما یک شکاف وجود دارد. در بسیاری از موارد ما می‌دانیم که مصنوعی که با آن در تعامل هستیم فاقد ویژگی‌های مورد نظر است، اما با آن طوری رفتار می‌کنیم که گویی می‌تواند مخاطب عمل اخلاقی باشد. در چنین مواردی ناگزیر باید عمل را تخطئه کنیم و تنها یک شیوه رفتار با مصنوع را صحیح بدانیم، شیوه‌ای از رفتار که با خصوصیات وجودشناختی یک هویت همخوانی دارد. اما به عقیده کاکلبرگ محدود کردن شیوه صحیح رفتار به یک شیوه خاص قابل قبول نیست.

در نقطه مقابل، در رویکرد رابطه‌ای هر هویت در زمینه‌ای از روابط تعریف می‌شود. اگر بخواهیم بدانیم که یک هویت خاص جایگاه اخلاقی دارد یا خیر، باید ببینیم که در چه زمینه مادی، فرهنگی، و اجتماعی جای

۱. برای دیدن نمونه‌های بیشتر مطالعات تعامل انسان و ربات‌های اجتماعی این منبع را ببینید: (Korn, 2019)

گرفته و قوام یافته است، دیگر هویت با او چگونه ارتباط برقرار می‌کنند، و برای آنها چطور پدیدار می‌شود (Coeckelbergh, 2014, p. 64). با این حال، مفهوم رابطه در اینجا بر خلاف مفهوم ویژگی‌ها در رویکرد استاندارد، مفهومی انتزاعی نیست. کاکلبرگ تأکید می‌کند که برای شناسایی این که طبق رویکرد رابطه‌ای یک هویت منزلت اخلاقی دارد یا خیر، نباید مجدداً به روش‌های فلسفی و علمی برای مطالعه و تحلیل روابط بین هویت متوسل شویم. بلکه باید ببینیم ابژه بر سوژه چطور پدیدار می‌شود. به عبارت دیگر از جمله ارتباطات تعیین‌کننده در وضعیت اخلاقی یک هویت، خود رابطه میان سوژه شناساگر و ابژه‌ای است که وضعیت اخلاقی‌اش مورد سؤال است.

وضعیت اخلاقی یک هویت، چیزی نیست که مستقل از چگونگی تجربه شدن آن برای سوژه وجود داشته باشد. وضعیت اخلاقی هر هویت در تجربه تعاملی سوژه با آن هویت برساخته می‌شود؛ و البته چنین تعاملی در بستری از روابط مادی و اجتماعی شکل می‌گیرد که خارج از کنترل سوژه است. این بستر، در واقع یک شکل زندگی است و شامل الگوهای مختلفی از رفتار، تفکر، تفسیر، و ارزش‌گذاری می‌شود که در یک فرهنگ ظهور پیدا کرده‌اند. تجربه اخلاقی یک سوژه از یک هویت، از یک سو شخصی بوده و از سوی دیگر به واسطه مستقر بودن در بستری مادی و اجتماعی، غیر شخصی است. به عبارت دیگر، در وضعیت اخلاقی یک مصنوع برای یک انسان، علاوه بر این که آن مصنوع چطور برای آن انسان پدیدار می‌شود، این که فرهنگی که آن فرد در آن زیست می‌کند ماشین‌ها را چطور تفسیر کرده و به آنها معنا می‌بخشد نیز تعیین‌کننده است (Coeckelbergh, 2014, p. 69).

بر این اساس، شیوه‌های مختلف تجربه شدن یک هویت برای سوژه‌های مختلف می‌تواند دلالت‌های متفاوتی در ارتباط با وضعیت اخلاقی آن هویت داشته باشد. نمی‌توان از یک حقیقت عینی و مستقل از سوژه در ارتباط با وضعیت اخلاقی یک هویت سخن گفت. با توجه به این موضوع، کاکلبرگ پرسش از وضعیت اخلاقی را بر اساس رویکرد رابطه‌ای نه پرسشی فلسفی، بلکه پرسشی عملی درباره چطور مواجه شدن با هویت و رفتار با آنها می‌داند (Coeckelbergh, 2014, p. 65-66).

به این ترتیب در رویکرد رابطه‌ای بر خلاف رویکرد استاندارد، خود پرسش از وضعیت اخلاقی مصنوع در خلال یک رابطه مطرح می‌شود و اهمیت می‌یابد، و نیز در خلال رابطه به آن پاسخ داده می‌شود. معرفت سوژه نسبت به وضعیت اخلاقی مصنوع طرف تعامل، دانشی گزاره‌ای از سنخ «می‌دانم که آن ربات فاقد آگاهی است» که محصول تأملات فلسفی و علمی است، نیست. بلکه معرفتی ضمنی<sup>۱</sup> است که سرچشمه‌هایی غیرعلمی دارد و در بردارنده شناختی است که از تجربه مرتبط بودن با آن مصنوع خاص حاصل شده است. در نتیجه، از معضل معرفت‌شناختی که در مورد رویکرد استاندارد وجود داشت (عدم دسترسی معرفتی به

---

1. Tacit.



ویژگی‌های هویت) نیز اجتناب می‌شود. در رویکرد رابطه‌ای، معرفتی که لازم داریم؛ همان معرفت ضمنی است که انسان به واسطه رابطه‌اش با مصنوع به دست می‌آورد (Coeckelbergh, 2014, p. 70-71).

کاپوچو و همکارانش ضمن همراهی با کاکلبرگ در برتر دانستن رویکرد رابطه‌ای، این رویکرد را در چارچوب اخلاق فضیلت‌گرا قرار می‌دهند. می‌توان دید که رویکرد استاندارد با رویکردهای فایده‌گرا و وظیفه‌گرا در اخلاق به خوبی سازگار است: در رویکرد استاندارد ما به دنبال کشف ویژگی‌هایی هستیم که می‌توانند مبنای ادراک لذت و رنج، یا داشتن حقوق برای یک هویت قرار بگیرند. اما برای رویکرد فضیلت‌گرا، پیامدهای یک عمل برای طرف مقابل یا جایگاه طرف مقابل در نظامی از حقوق و وظایف، اصالت ندارد. آنچه مهم است این است که یک عمل در زمینه تعاملات اجتماعی فرد، چه تأثیری بر شخصیت عامل آن می‌گذارد. در واقع تفکر رابطه‌ای یکی از مؤلفه‌های رویکرد فضیلت‌گرا است (Cappuccio et al., 2019, p. 13-14).

به پرسش آغازین بازگردیم. از رویکرد رابطه‌ای درباره وضعیت اخلاقی مصنوعات مختلف و ملاحظات مربوط به تعامل با آنها چه نتیجه‌ای می‌توان گرفت؟ همانطور که دیدیم، از دید فیلسوفان فضیلت‌گرا این مسئله پاسخی زمینه‌مند دارد. باید دید که رابطه فرد با یک مصنوع تا چه حد با روابط انسانی که دایره متعارف اخلاق را شکل می‌دهند شباهت دارد. نمی‌توانیم از پیش مصنوعات را در دسته‌های مشخصی جای دهیم و برخی را دارای وضعیت اخلاقی و برخی را فاقد آن بدانیم. اما آیا این به این معناست که ما در نهایت برای تشخیص وضعیت اخلاقی هویت ملاکی نداریم؟ کاپوچو و همکارانش به این اشکال پاسخ می‌دهند. از منظر اخلاق فضیلت‌گرا شاخصه ما برای رعایت ملاحظات اخلاقی، تأثیر یک عمل بر شخصیت ماست. پس هر قدر رابطه ما با یک مصنوع شباهت بیشتری با رابطه ما با انسان‌ها داشته باشد، چگونگی تعامل ما با آن بر شخصیت اجتماعی ما تأثیر می‌گذارد. بر این اساس، ما می‌توانیم با تکیه بر میزان انسان‌انگاری<sup>۱</sup> تفکیکی میان انواع مصنوعات داشته باشیم. کاکلبرگ نیز به این نکته اشاره می‌کند که با گذشت زمان و درهم تنیدگی بیشتر جهان انسانی و فناورانه، ما برای توصیف وضعیت اخلاقی ربات‌ها نیز مقولات اخلاقی بیشتر و دقیق‌تری خواهیم داشت. همان طور که مثلاً در جهان حیوانات، در برخورد با حیوان خانگی و حشره‌ای مودی ملاحظات اخلاقی متفاوتی را در نظر می‌گیریم، در دایره مصنوعات نیز می‌توانیم به دسته‌بندی‌های ظریف‌تری برسیم (Coeckellbergh, 2010, p. 239).

برخی مصنوعات به طور خاص برای شبیه‌سازی رابطه انسان - انسان طراحی شده‌اند. هدف طراحان ربات‌های اجتماعی این است که یک رابطه معاشرتی را با جزئیات بدن‌مند، عاطفی و روانشناختی‌اش تقلید کنند. به گونه‌ای واکنش نشان دهند که انسان‌ها آن را واکنشی اجتماعی برداشت کنند، و آنها نیز نسبت به

کنش‌های طرف انسانی، واکنش‌های انسان‌وار بروز دهند. در صورتی که این هدف به میزان قابل قبولی برآورده شود و ربات‌ها بتوانند الگوهای مشابهی از عواطف و اعمال را در انسان‌ها برانگیزند، می‌توان گفت که آنها مشابه انسان‌ها درک می‌شوند (Cappuccio et al., 2019, p. 18). پس چگونگی رفتار ما با آنها می‌تواند بر چگونگی رفتار ما با دیگر انسان‌ها نیز اثر بگذارد.

کاکلبرگ، کاپوچو و همکارانش، علاوه بر دفاع از جایگاه اخلاقی فناوری‌ها، ملاحظات اخلاقی مربوط به تعامل با برخی فناوری‌ها را نیز از منظر فضیلت‌گرایی بررسی می‌کنند. کاپوچو و همکارانش به تأثیر تعامل با ربات‌ها و دستیارهای دیجیتال بر فضیلت مسئولیت‌پذیری می‌پردازند. افراد در بستر تعاملات اجتماعی یکدیگر را به رسمیت می‌شناسند، انتظارات خود از دیگران را بروز می‌دهند و در معرض انتظارات دیگران قرار می‌گیرند. از یک سو این که هر فرد خود را در قبال چه وظایفی نسبت به دیگران پاسخگو می‌بیند، به او تصویری درباره جایگاه و مسئولیت‌های خودش ارائه می‌دهد. از سوی دیگر هر فرد به اندازه‌ای که توسط دیگران به رسمیت شناخته شود و انتظارات ایشان را برآورده کند، اعتبار کسب می‌کند. این دو فرآیند، سازوکارهای اصلی شکل‌گیری شخصیت فرد هستند.

هنگامی که تعامل با ربات‌ها یا دستیاران دیجیتالی بخشی از تعاملات اجتماعی انسان‌ها شود، انتظارات متقابل انسان‌ها و ربات‌ها در این تعامل نیز در فرآیند شکل‌دهی به شخصیت دخیل می‌شود. کاپوچو و همکارانش بر وضعیت نامتقارن انسان و ربات در این تعاملات تأکید می‌کنند. در رابطه بین انسان و ربات، ربات‌ها انسان را به عنوان شخص به رسمیت می‌شناسند، اما نه به عنوان شخصی برابر با خود. در این تعاملات همواره انسان‌ها دست‌پرتر را دارند. آنها دستور می‌دهند و ربات‌ها اطاعت می‌کنند، درست شبیه به رابطه ارباب و برده در قرون سابق. (Cappuccio et al., 2019, p.20) در این رابطه کاربر فناوری بدون آن که صلاحیت خاصی داشته باشد در موقعیت فرماندهی قرار می‌گیرد، بی‌آن که نسبت به فرمانبر خود توجه یا پاسخگویی داشته باشد. در این تجربه، انسان‌ها توسط ربات‌ها قضاوت نمی‌شوند، مورد بازخواست واقع نمی‌شوند و توقعی از ایشان نمی‌رود. در حالی که بر خلاف بسیاری از مصنوعات دیگر که در آنها نیز مصنوع در نقش ابزار به کار می‌رود، در اینجا فرآیند تشخیص اجتماعی رخ می‌دهد. انسان‌ها در تعامل با ربات‌ها احساس می‌کنند به لحاظ اجتماعی به رسمیت شناخته شده‌اند و ربات نیز به‌گونه‌ای واکنش نشان می‌دهد که گویی انسان او را به عنوان یک معاصر اجتماعی به رسمیت شمرده است. بر این اساس، تعاملات انسان- ربات/دستیاران دیجیتال بر شکل‌گیری شخصیت افراد اثر می‌گذارد و این تأثیر به‌نحوی است که شخص به داشتن نقش اجتماعی مسئولیت‌زدایی شده خو می‌کند.

کاکلبرگ به تأثیر بازی‌های رایانه‌ای خشن بر فضیلت همدلی می‌پردازد (Coeckelbergh, 2007, p. 219-220). در



بازی‌های رایانه‌ای خشن، بازیکن خود را جای یک شخصیت می‌گذارد و در جهان ساخته شده در فضای بازی، به مبارزه با شخصیت‌های دیگر می‌پردازد. این فعالیت چند ویژگی دارد که تضعیف‌کننده همدلی است. اولین ویژگی، خشونت موجود در این بازی‌ها است. بازیکن برای برنده شدن باید به کشتن افراد و تخریب اشیاء دست بزند. تکرار چنین اعمالی، آن‌ها را برای بازیکن عادی می‌کند. ممکن است گفته شود وجود عنصر خشن در این بازی‌ها نمی‌تواند باعث خشونت‌ورزی افراد در تعامل با انسان‌های دیگر شود. چون کاربران می‌توانند تفاوت فضای مجازی و حقیقی را درک کنند. ضمن این که فضاها و رفتارهای خشن فقط در بازی‌های رایانه‌ای وجود ندارند، بلکه مثلاً در رمان‌ها و فیلم‌ها نیز شاهد آن هستیم. پس چرا درباره آنها نمی‌توان همین حکم را داد؟ برای پاسخ به این دو ایراد، کاکلبرگ به ویژگی دوم این بازی‌ها اشاره می‌کند: شبیه‌سازی تعامل در جهان واقعی. هرچند ما در تماشای فیلم یا خواندن رمان نیز خود را جای شخصیت‌ها قرار می‌دهیم، اما در بازی‌های رایانه‌ای، به عنوان عاملی فعال در یک فضا حاضر می‌شویم و با کنش‌هایمان یک شخصیت را می‌سازیم. به این ترتیب بازیکن نه فقط به عنوان تماشاگری منفعل، بلکه چون کنشگری فعال در بازی نقش آفرینی می‌کند. سعی طراحان این بازی‌ها بر این است که تا حد ممکن بازیکن واسطه‌ای میان خود و جهان مجازی بازی درک نکند، بازیکن در فضای بازی غرق شود و جهان بازی را مشابه جهان واقعی تجربه کند. فاصله‌برداری و غرق‌شدگی، خصوصاً در بازی‌هایی که ما بدن آوارتارمان را نمی‌بینیم، بلکه فقط زاویه دید او را داریم و از منظر او به جهان بازی می‌نگریم، بیشتر نیز می‌شود (Coeckelbergh, 2007, p. 24-25).

ویژگی دیگر این بازی‌ها این است که در آن بازیکن در ازای رفتارهای خشن پاداش می‌گیرد و زشتی و پیامدهای منفی خشونت در آن دیده نمی‌شود. در حالی که خشونت بخشی از جهان واقعی نیز هست و می‌توان گفت این بازی‌ها تصویری از خشونت‌های موجود در جهان انسانی را بازنمایی می‌کنند، اما در روایتی که این بازی‌ها ارائه می‌دهند، خشونت‌ورزی به پایانی خوش منتهی می‌شود و زشتی و پیامدهای دردناک آن نشان داده نمی‌شود (Coeckelbergh, 2007, p. 27).

با در نظر داشتن ویژگی‌های فوق، کاکلبرگ بر این نکته تأکید می‌کند که رشد شخصیتی افراد یک فرآیند پیوسته است. نمی‌توان برای انجام بازی‌های رایانه‌ای، شخصیت را در محفظه‌ای نفوذناپذیر قرار داد و پس از بازی آن را بیرون آورد و به زندگی واقعی بازگرداند. حتی اگر بازیکنان بتوانند تفاوت فضای مجازی بازی و فضای حقیقی را تشخیص دهند، اما کنش‌هایی که بازیکنان در این بازی‌ها انجام می‌دهند، عاداتی را در آن‌ها شکل می‌دهد. عادت به انجام خشونت و لذت بردن از آن، توانایی درک رنج و درد دیگران را در بازیکنان کم می‌کند و این زمینه‌ساز رفتارهای پرخاشگرانه و تهاجمی در ایشان می‌شود.

## ۲. آیا فضیلت‌گرایی رویکردی اصیل است؟

آراء فیلسوفان فضیلت‌گرا دربارهٔ تعامل انسان و فناوری با انتقادات مختلفی مواجه شده است. برخی از این انتقادات اشکالاتی کلی را در ارتباط با رویکرد فضیلت‌گرا در اخلاق مطرح می‌کنند؛ برای مثال آموزه‌های فضیلت‌گرایی را قابل تبدیل به دستورالعمل‌های صریح و ساده نمی‌دانند؛ آنها را برای تصمیم‌گیری در تنگناهای اخلاقی ناکارآمد می‌دانند؛ یا ادعا می‌کنند فهرست فضایل و شخصیت‌های الگو نامتعین، دلبخواهی، و غیرجهان‌شمول<sup>۱</sup> است. برخی دیگر از این انتقادات بر آراء خاص فضیلت‌گرایان در باب فناوری متمرکز شده است. برای مثال کروزر دیدگاه والر دربارهٔ فضایل فناورانه-اخلاقی و تأثیر فناوری بر فضایل (Vallor, 2016) را نقد می‌کند. او در اهمیت فضایی که والر به عنوان فضایل فناورانه-اخلاقی معرفی می‌کند تردید کرده و ادعا می‌کند بر خلاف دیدگاه والر، فناوری‌های جدید زندگی فضیلت‌مندان را دشوارتر نکرده‌اند؛ تنها شکل چالش‌های موجود برای زندگی فناورانه را تغییر داده‌اند (Cruzer, 2017, p. 288).

گذشته از این دست انتقادات، یک انتقاد کلی نسبت به فضیلت‌گرایی دربارهٔ فناوری، پرسش از اصیل بودن این دیدگاه است. به عبارت دیگر، می‌توان پرسید آیا دیدگاه فضیلت‌گرایانه دربارهٔ تعامل انسان و فناوری، قابل فروکاستن به دیدگاه‌های دیگری همچون وظیفه‌گرایی یا فایده‌گرایی نیست؟

در دیدگاه‌های فیلسوفان فضیلت‌گرا - چه در ارتباط با زندگی در بستر فناورانه و چه در مورد تعامل با فناوری به مثابه شبه‌دیگری - یک خط استدلالی مشترک برجسته است: استفاده ما از فناوری‌ها و تعامل ما با آنها در صورتی اخلاقی است که به تقویت عاداتی در ما بینجامد که از ما شخصیت‌هایی فضیلت‌مند می‌سازند. فیلسوفان فضیلت‌گرا برای تفصیل این خط اجمالی، تأثیر عادات فناورانه بر ویژگی‌های شخصیتی را بررسی می‌کنند. مثلاً برای ارزیابی بازی‌های خشن رایانه‌ای، باید ببینیم این بازی‌ها چه عاداتی را در بازیکن تشدید می‌کنند، این عادات چه ویژگی‌های شخصیتی را در او ایجاد می‌کنند، و در نهایت این ویژگی‌ها با تأثیری که روی رابطه فرد با دیگر انسان‌ها می‌گذارند، چطور میزان بهره‌مندی فرد از «زندگی خوب» را متأثر می‌کنند.

در این روند ارزیابی، مایه‌هایی از فایده‌گرایی قابل مشاهده است. در وهله اول برای ارزیابی یک عمل فناورانه، پیامد آن در مورد عادات فرد در نظر گرفته می‌شود. در وهله بعد، پیامد این عادات بر شخصیت فرد مورد توجه قرار می‌گیرد. در نهایت نسبت خصایل شخصیتی کسب شده با زندگی خوب محک زده می‌شود. در حالی که مفهوم زندگی خوب در نهایت با مفاهیمی چون «شادمانی<sup>۲</sup>» و «شکوفایی<sup>۳</sup>» تعریف می‌شود و با توجه به ماهیت رابطه‌ای انسان در دیدگاه فضیلت‌گرایانه، شادمانی و شکوفایی در نهایت در بستری جمعی معنا پیدا می‌کند.

۱. برای مرور انتقادات مختلف به اخلاق فضیلت‌گرایی و پاسخ‌های وارد شده به آنها این منبع را ببینید: (Hursthouse, 2000).

2. happiness

3. eudemonia



به این ترتیب، ظاهراً رویکرد فضیلت‌گرا از آنجا که پیامدهای اعمال را لحاظ می‌کند و این پیامدها را بر اساس نسبتی که با شادمانی و شکوفایی دارند ارزش‌گذاری می‌کند، به فایده‌گرایی بسیار نزدیک می‌شود. بنابراین می‌توان به درستی پرسید آیا می‌توان از اصالت و پایه‌ای بودن مفهوم فضیلت در تحلیل‌های فوق دفاع کرد یا خیر؟ آیا در نهایت ملاک ما برای اخلاقی یا غیر اخلاقی شمردن یک عمل، میزان فایده یا ضرر آن برای جامعه انسانی نیست؟ و اگر این طور باشد، تفاوت یا مزیت رویکرد فضیلت‌گرا در مقایسه با فایده‌گرایی چه خواهد بود؟ اگر در نهایت تمام تجویزهای فضیلت‌گرایان قابل ترجمه به تجویزهای فایده‌گرایی و متناظر با آنها باشند، به نظر می‌رسد نمی‌توان فضیلت‌گرایی را رویکردی اصیل دانست.

در پاسخ به این پرسش توجه به دسته‌بندی‌های درون رویکردهای فضیلت‌گرا اهمیت دارد. در میان رویکردهای فضیلت‌گرا می‌توان سه رویکرد سعادت - محور<sup>۱</sup>، عامل - محور<sup>۲</sup>، و هدف - محور<sup>۳</sup>، را از هم متمایز ساخت<sup>۴</sup>. در رویکرد سعادت - محور، مفهوم سعادت، شادمانی، یا شکوفایی، مفهومی پایه‌ای است و هر عمل در نسبت با آن ارزش‌گذاری می‌شود. در رویکرد عامل - محور، اشخاصی به عنوان «الگوهای اخلاقی»<sup>۵</sup> مینا قرار می‌گیرند. هر فرد شخصیت‌هایی را به عنوان الگوی اخلاقی شناسایی می‌کند و در هر موقعیت میزان انطباق مقاصد و گرایش‌های خود را با مقاصد و گرایش‌های الگوهای خود در موقعیت مشابه مقایسه می‌کند، و بر این اساس درباره اخلاقی بودن یا نبودن یک عمل داوری می‌کند. در رویکرد هدف - محور، فهرستی از فضایل (مانند سخاوت، صداقت، شفقت و...) مفاهیم پایه هستند و انگیزه‌ها و اعمال هر فرد به میزان انطباقشان با این فضایل، اخلاقی محسوب می‌شوند (Hursthouse, 2022).

ایراد تحویل‌پذیری فضیلت‌گرایی به فایده‌گرایی، بیش از همه در مورد فضیلت‌گرایی سعادت-محور جدی است. اگر نقش یک عمل در پیشبرد یا تهدید سعادت فردی و جمعی، ملاک ارزیابی آن عمل قرار بگیرد، به دشواری می‌توان این دیدگاه را از تحویل‌پذیری به فایده‌گرایی حفظ کرد. خصوصاً که رویکردهای فایده‌گرا مفهوم فایده را با گشاده‌دستی تعریف می‌کنند، چنان‌که در بسیاری از قرائت‌های آن سعادت نه تنها بر اساس لذت‌های مادی، بلکه بر اساس لذت‌های معنوی (از جمله شکوفایی) تعریف می‌شود. به این ترتیب می‌توان تمام آنچه

1. Eudaemonist

2. Agent-Based

3. Target-Centered

۴. برای دیدن نمونه‌ای از رویکرد سعادت-محور این آثار را ببینید: (Bloomfield, 2014)(Badhwar, 2014). به‌عنوان نمونه‌ای از رویکرد عامل-محور این اثر را ببینید: (Zagzebski, 2017).

5. Moral exemplars

به عنوان ممیزات رویکرد فضیلت‌گرا مطرح است، مانند اهمیت پرورش شخصیت اخلاقی، نقش عادات در شکل‌گیری شخصیت و... را به مثابه زنجیره‌هایی در نظر گرفت که بر اساس تأثیری که در وضع امور نهایی می‌گذارند، و میزان شادکامی یا رنجی که می‌آفرینند محک زده می‌شوند.

اما آیا این ایراد به دو نسخه دیگر فضیلت‌گرایی نیز وارد است؟ رابرت اسپارو در مقاله‌ای که در آن ابعاد اخلاقی طراحی ربات‌های جنسی تجاوزپذیر را بررسی می‌کند، به تمایز سه رویکرد فضیلت‌گرا، فایده‌گرا و وظیفه‌گرا، در نسبت با این مسئله می‌پردازد. اسپارو می‌گوید اگر ما طراحی این ربات‌ها را غیر اخلاقی بدانیم به این دلیل که استفاده از این ربات‌ها موجب ترویج رفتار تجاوز جنسی در جامعه (یا بالا رفتن احتمال ارتکاب این عمل در زندگی اجتماعی کاربر) می‌شود، رویکردی فایده‌گرایانه را در پیش گرفته‌ایم. اگر دلیل ما برای محکوم کردن این فناوری این باشد که ساختن رباتی به شکل یک زن به عنوان ابژه تجاوز جنسی، حرمت زنان را نقض می‌کند، در این صورت ما با رویکردی وظیفه‌گرایانه این فناوری را نقد کرده‌ایم. اما اگر این ربات‌ها را به این دلیل غیر اخلاقی بدانیم که استفاده از آنها گرایش‌ها و امیال منفی کاربر را افشا می‌کند و انعکاسی منفی در فرد پیدا می‌کند، رویکرد فضیلت‌گرا را در پیش گرفته‌ایم. از منظر اخلاق فضیلت‌گرا آنچه تعرض به یک ربات را غیر اخلاقی می‌کند این است که فرد از آن لذت می‌برد و مدت زیادی به آن می‌پردازد. ایجاد یک فانتزی منفی در فرد، ردیلتی را در او آشکار ساخته و تقویت می‌کند، حتی اگر هرگز منجر به این نشود که او چنین رفتاری را با یک موجود زنده داشته باشد (Sparrow, 2017, p. 472).

اسپارو ادعا می‌کند رویکرد فضیلت‌گرا در محکوم کردن چنین ربات‌هایی موفق‌تر از رویکرد فایده‌گرا است. برای ارائه استدلالی فایده‌گرایانه در محکومیت ربات‌های تجاوزپذیر، باید بتوانیم شواهدی قاطع ارائه کنیم که حاکی از تأثیر استفاده از این ربات‌ها بر افزایش سوءرفتارهای اجتماعی باشد. اما چنین شواهد قاطعی وجود ندارد، بلکه برخی شواهد نیز خلاف این را نشان می‌دهند. مناقشه بر سر این که آیا لذت بردن فرد از بازنمایی یک رفتار (در رمان، فیلم، بازی، و...) احتمال ارتکاب آن رفتار را بالا می‌برد یا نه، بر اساس شواهد موجود فیصله نیافته است. حتی برخی در این باره مدافع نظریه تطهیر هستند و معتقدند قرار گرفتن در معرض رفتارهای خشن یا انجام آنها در قبال امور غیر انسانی می‌تواند هیجانانگیز منفی فرد را تخلیه کند و این احتمال که در رابطه با افراد واقعی مرتکب چنین اعمالی شود را کاهش دهد (Sparrow, 2017, p. 469).

دیگر نظریه‌پردازان فضیلت‌گرا نیز نسبت به نابسندگی شواهد درباره پیامدهای منفی فناوری‌های مختلف توجه داشته‌اند. برای مثال کاکلبرگ در نقد بازی‌های خشن رایانه‌ای، بحث خود را با این باور رایج آغاز می‌کند که رفتارهای پرخاشگرانه و تهاجمی در میان کسانی که بازی‌های رایانه‌ای خشن می‌کنند رایج‌تر است. اما بعد



ادعا می‌کند نتایج آزمایش‌های تجربی در این مورد همگرا نیستند و تأیید واضحی نسبت به تأثیر بازی‌های رایانه‌ای خشن بر خشونت‌ورزی افراد ندارند. علاوه بر این که، حتی اگر پژوهش‌ها وجود همبستگی میان وقت‌گذرانی با بازی‌های خشن و بروز رفتارهای خشن را نشان دهند، همچنان برای آن که رابطه‌ای علی (فراتر از همبستگی) میان انجام بازی‌های خشن و خشونت‌ورزی ثابت کنیم، نیازمند کار بیشتری هستیم. به همین دلیل، رویکرد فایده‌گرا را مبنای مناسبی برای غیراخلاقی دانستن بازی‌های رایانه‌ای خشن نمی‌داند (Coeckelbergh, 2007, p. 219-220).

رویکردی که اسپارو اتخاذ کرده است، در دسته فضیلت‌گرایی هدف-محور قرار می‌گیرد. اسپارو تأثیر استفاده از ربات‌های تجاوزپذیر را بر گرایش‌های عاطفی، احساسی، و شناختی فرد در نظر می‌گیرد. ویژگی‌های گرایش‌ی<sup>۱</sup>، ویژگی‌هایی هستند که در صورت وجود محرک بروز پیدا می‌کنند و در غیاب محرک ممکن است هرگز بروز نیابند (Marmadoro and Mayr, 2019, p. 64). برای مثال یک جام بلور دارای گرایش شکنندگی است، حتی اگر هرگز در معرض ضربه قرار نگرفته و نشکند. به طور مشابه، ویژگی‌های شخصیتی نیز به‌عنوان ویژگی‌های گرایش‌ی ممکن است هرگز بالفعل نشوند. ممکن است فردی خصلت شخصیتی شرارت را پیدا کند، اما هرگز فرصتی برای بروز شرارت پیدا نکند. از منظر اخلاق فضیلت‌گرا، ملاک، پیدایش این گرایش منفی است و نه تغییری که ممکن است در وضع امور جهان بیرونی ایجاد کند. در اینجا تفاوت رویکرد فضیلت‌گرا و فایده‌گرا مشهود می‌شود.

هرچند فضیلت‌گرایی پیامد یک عمل را در ارزیابی آن مرتبط می‌داند، اما نقش پیامد در فضیلت‌گرایی با نقش آن در فایده‌گرایی فرق می‌کند. فایده‌گرایی تغییراتی را که یک عمل در وضع امور جهان رقم می‌زند لحاظ می‌کند و با محاسبه تغییراتی که در لذت و رنج جمعی حاصل شده است، آن عمل را ارزش‌گذاری می‌کند. در حالی که فضیلت‌گرایی هدف-محور، تأثیر تداوم یک عمل بر گرایش‌های شخصیتی فرد را بررسی می‌کند. اولاً از آنجا که تحول شخصیتی ملاک است، برای بررسی تأثیر هر عمل، پیامدهای تکرار آن عمل در طول زمان در نظر گرفته می‌شود و نه انجام یک باره آن. ثانیاً، از آنجا که تحول در گرایش‌های شخصیتی فرد عامل اهمیت دارد، تغییرات در ویژگی‌های مقولاتی<sup>۲</sup> جهان (یعنی تغییراتی که در وضع امور جهان رخ داده است) در نظر گرفته نمی‌شود. در نهایت، بر خلاف رویکرد فایده‌گرایانه برای تصمیم‌گیری راجع به یک عمل، رویکردی محاسباتی مبتنی بر دانش گزاره‌ای و تجربی در پیش گرفته نمی‌شود، بلکه نوعی تشخیص مبتنی بر معرفت ضمنی در ارزش‌گذاری نهایی تعیین‌کننده است.

1. Dispositional

2. Categorical

همین وضعیت با اندکی تفاوت دربارهٔ رویکرد عامل-محور نیز برقرار است. در رویکرد عامل-محور برای ارزش‌گذاری یک عمل باید بررسییم اگر یک الگوی اخلاقی در این موقعیت بود چطور عمل می‌کرد؟ پیامدی که در اینجا به ارزش‌گذاری مرتبط است، میزان شبیه شدن (یا نشدن) عامل به الگوی اخلاقی‌اش است. در این حالت نیز به این دلیل که این شباهت یا تفاوت مربوط به شخصیت فرد است که شامل ویژگی‌های گرایشی او می‌شود، این پیامد از سنخ پیامدهای مورد نظر فایده‌گرایان نیست و شیوه بررسی آن نیز محاسبه‌گرایانه نیست. به این ترتیب می‌توان گفت طبق دو نسخه عامل-محور و هدف-محور اخلاق فضیلت‌گرا، در عین حال که پیامد عمل به طور کلی می‌تواند امری مرتبط با ارزیابی اخلاقی باشد، اما پیامدهای مرتبط و شیوه تأثیرگذاری آنها در ارزیابی عمل با رویکرد فایده‌گرایانه متفاوت است. همین ویژگی باعث می‌شود اخلاق فضیلت‌گرا برای ارزیابی اعمال فناورانه نیازی نداشته باشد خود را متکی و محدود به شواهد تجربی مرتبط بکند. بلکه می‌تواند از مناقشات پیرامون ادعاهای تجربی اجتناب کرده و با تکیه بر دانشی ضمنی دربارهٔ تأثیر عادات بر گرایش‌های شخصیتی، دربارهٔ اعمال فناورانه داوری کند.

### ۳. پارادوکس انسان‌انگاری و انسانیت‌زدایی

فیلسوفان فضیلت‌گرا در تعیین جایگاه اخلاقی فناوری‌های تعاملی، رویکردی رابطه‌ای اختیار می‌کنند. همان‌طور که دیدیم طبق این رویکرد، وضعیت اخلاقی هر هویت برای یک فرد بر اساس نقشی که آن هویت در شبکه روابط اجتماعی فرد دارد، و شیوه پدیدار شدنش برای فرد تعیین می‌شود. بر همین اساس، این فیلسوفان رعایت ملاحظات اخلاقی را در ارتباط با ربات‌های اجتماعی لازم می‌دانند. می‌توان این قاعده را برای هر فناوری دیگری غیر از ربات‌ها که جایگاه شبه‌دیگری پیدا می‌کنند (مثل آواتارها، چت‌بات‌ها و...) به کار بست.

اما این رویکرد، معضلی را به وجود می‌آورد. اساس این که یک هویت جایگاه شبه‌دیگری پیدا کند، مبتنی بر انسان‌انگاری آن هویت است. یعنی این که آن هویت در شبکه تعاملات اجتماعی الگوهایی از عواطف و گرایش‌ها و افکار را برانگیزد که در روابط انسان-انسان برانگیخته می‌شود و این موجب شود آن هویت انسان‌وار درک شود. این در حالی است که این فناوری‌ها بر خلاف انسان‌ها، به دست انسان‌ها طراحی می‌شوند، خرید و فروش می‌شوند، و همواره در خدمت برآورده‌سازی نیازهای انسان‌ها هستند. به عبارت دیگر، در بخشی از شبکه تعاملات اجتماعی، از این فناوری‌ها انسانیت‌زدایی<sup>۱</sup> می‌شود.

انسان‌انگاری و انسانیت‌زدایی توأمان این فناوری‌ها وضعیت تناقض‌آمیزی را رقم می‌زند که کاپوچو و همکارانش آن را «پارادوکس انسان‌انگاری و انسانیت‌زدایی همزمان»<sup>۲</sup> می‌خوانند. استانداردهای فرهنگی رابطه انسان-ربات

1. Dehumanization

2. Anthropomorphizing while Dehumanizing Robots Paradox- ADP



ایجاب می‌کند در این رابطه کیفیت انسان‌واری ربات‌ها تأیید و همزمان انکار شود (Cappuccio et al., 2019, p. 26).

این معضل دو مشکل را با خود به همراه دارد: تناقضی عاطفی و فکری، و فساد اخلاقی.

پارادوکس انسان‌انگاری و انسانیت‌زدایی، وضعیتی است که در وهله اول می‌توان آن را در قالب یک تناقض منطقی صورت‌بندی کرد. اما فراتر از شکل منطقی آن، به صورت یک تعارض و سردرگمی عاطفی تجربه می‌شود. افراد در عین حال که می‌دانند طرف رابطه آنها انسان نیست، اما احساساتی مشابه احساساتشان در رابطه با انسان‌ها را تجربه می‌کنند و این آنها را دچار حس تعارض و سردرگمی می‌کند. کاکلبرگ به چنین وضعیتی توجه داشته و آن را مولود رویکرد استاندارد به مسئله تعامل انسان و ربات می‌داند. طبق رویکرد استاندارد، شاخص ما برای تعیین وضعیت اخلاقی یک هویت، ویژگی‌های عینی آن مانند آگاهی و ظرفیت احساسی است. ما در قالب دانش گزاره‌ای می‌دانیم که فناوری‌ها فاقد چنین ویژگی‌هایی هستند ولی هنگامی که با آنها وارد تعامل می‌شویم، آنها برای ما طوری پدیدار می‌شوند که گویی مانند انسان‌ها هستند. کاکلبرگ می‌گفت رویکرد استاندارد برای فائق آمدن بر این تناقض ناگزیر است درک پدیداری ما از ربات‌ها و شیوه رفتاری ما با آنها را تصحیح کرده و شیوه‌ای صحیح برای رفتار با ربات‌ها را توصیه کند، که از نظر او تجویز قابل قبولی نیست.

کاکلبرگ این را مزیت رویکرد رابطه‌ای می‌داند که معرفت‌مبنایی ما برای تشخیص وضعیت اخلاقی را نه معرفت گزاره‌ای درباره ویژگی‌ها، بلکه معرفت‌ضمنی ما درباره چگونگی تجربه شدن ربات‌ها در نظر می‌گیرد. پس طبق آن، لازم نیست رفتار ما بر اساس دانش گزاره‌ای تصحیح شود. وضعیت اخلاقی ربات‌ها بستگی به این دارد که در شبکه تعاملات اجتماعی چگونه تجربه می‌شوند. هر قدر که ربات‌ها را انسان‌وار تجربه می‌کنیم، باید با آنها اخلاقی رفتار کنیم. اما آیا این رویکرد می‌تواند از تناقض فوق اجتناب کند؟ به نظر نمی‌رسد این طور باشد.

همان‌طور که در صورت‌بندی پارادوکس اشاره شد، حقایقی که مقوم جنبه انسانیت‌زدایی این پارادوکس هستند و باعث می‌شوند ربات‌ها فاقد جایگاه انسانی باشند نیز، بخشی از تجربه تعاملی ما با ربات‌ها هستند. ما ربات‌ها را خریدوفروش می‌کنیم، به آنها به چشم فرمانبردار و خادم خود می‌نگریم، از آنها توقع اطاعت بی‌چون و چرای داریم، با کمک طراحان آنها را طبق خواسته و ذائقه خود شخصی‌سازی می‌کنیم، و ملزم به پاسخگویی به آنها نیستیم. این تجربیات نیز همچون تجربیاتی که کاکلبرگ در نظر دارد، بخشی از ادراک پدیداری فناوری‌ها را شکل می‌دهند. پس چرخش از معرفت گزاره‌ای به معرفت‌ضمنی، و از دیدگاه استاندارد به دیدگاه رابطه‌ای مورد نظر فضیلت‌گرایان، مسئله تناقض را حل نمی‌کند.

یک پاسخ ممکن به این مسئله، راه حلی است که ما را از تناقض بیرون برده اما با مشکل دوم این معضل یعنی فساد اخلاقی روبه‌رو می‌کند. طبق این راه حل، ما می‌توانیم به مقولات خود دسته‌دیگری بیفزاییم:

انسان‌وار انسانیت‌زدایی شده. در قرون و دهه‌های گذشته هویتی همچون بردگان، و گاه زنان و کودکان در این مقوله جای می‌گرفتند؛ کسانی که برای انسان کامل بودن چیزی کم داشتند. جنبه‌هایی از رابطه فرد با این هویت بر اساس استانداردهای ارتباط با انسان‌های همچون-خود پیش می‌رفت، اما جنبه‌هایی دیگر از رابطه مبتنی بر شیء‌انگاری و ابزارانگاری این هویت بود. اگر در کنار انسان‌های همچون-خود و امور غیر انسانی، وجود این دسته سوم را به رسمیت بشناسیم، هنجارهای مستقلی را برای تعامل با آن تعریف می‌کنیم که از هر دو دسته دیگر متفاوتند. بنابراین دیگر تناقض و سردرگمی از جهت همزمانی همچون انسان پدیدار شدن و همچون انسان پدیدار نشدن ایجاد نمی‌شود. چون آنها از جنبه‌های متفاوتی انسان‌وار و غیرانسان‌وار هستند و این الگوی جدید آنها را در مقوله‌ای تعریف شده جای می‌دهد.

اما درستی تعریف این مقوله به لحاظ اخلاقی مورد تردید است. به رسمیت شناختن این مقوله در طول تاریخ در خدمت باور به برتری ذاتی یک گروه بر گروه‌های دیگر و زمینه‌ساز سلطه و استبداد بوده است. اعضای این مقوله در عین حال که شباهت بسیاری به اربابان خود دارند، اما هرگز جایگاه اربابان را ندارند، قصدمندی و غایت‌مندیشان مهم شمرده نمی‌شود و همواره همچون ابزار در خدمت خواسته‌های اربابان هستند. اربابان خود را در آینه این گروه می‌شناسند و هرچه بردگان در تعاملات اجتماعی بیشتر به اربابان شباهت پیدا کنند، ارزش سلطه اربابان بر آنها بیشتر می‌شود. به نظر می‌رسد در پیش گرفتن راه حل فوق برای اجتناب از وضعیت تناقض آمیز، این مقوله مسئله‌دار را از نو احیا می‌کند.

مسیر طراحی ربات‌های هوشمند به این سمت می‌رود که این ربات‌ها هرچه بیشتر خودمختار، قصدمند، و انسان‌وار به نظر برسند و کاربران را به سرمایه‌گذاری عاطفی روی خود سوق دهند. کاربران این ربات‌ها، خود را مالک موجوداتی -ظاهراً- قصدمند می‌بینند و در عین حال قصدمندی آنها را انکار کرده و از آنها صرفاً برای مقاصد خود استفاده می‌کنند. به این ترتیب رابطه کاربر و ربات اجتماعی هرچه بیشتر به رابطه ارباب و برده شبیه می‌شود. در حالی که در چنین رابطه‌ای کاربر حتی شاید کمتر از اربابان گذشته احساس مسئولیت و پاسخگویی داشته باشد. حتی اگر یک ربات به عنوان همدم و مونس یک فرد مورد استفاده قرار بگیرد، اگر کاربر به هر دلیلی از آن خسته شود می‌تواند به سادگی آن را خاموش کند یا دور بیندازد. رفتاری که در قبال بردگان نیز به این سادگی رخ نمی‌داد. به عبارت دیگر، اگر رویکرد فضیلت‌گرایان را بپذیریم و مطابق تفکر رابطه‌ای فناوری‌های اجتماعی را دارای شأنیت اخلاقی بدانیم، در این صورت در شرایطی که این فناوری‌ها شباهت بسیار زیادی به انسان‌ها داشته باشند، صرف کتک زدن، ناسزا گفتن، و آسیب نرساندن به این فناوری‌ها برای اخلاقی رفتار کردن با آنها کفایت نمی‌کند. بلکه همین که آنها را خرید و فروش می‌کنیم، به آنها دستور می‌دهیم و هرگاه بخواهیم آنها را خاموش می‌کنیم نیز مصداق بی‌اخلاقی به حساب می‌آید. این در حالی



است که این بخش از تجربیات ما در ارتباط با فناوری‌های اجتماعی (تملیک و تسلط) بخشی اساسی از تعاملات ما بوده و قابل انکار یا نادیده گرفتن نیست. بنابراین اگر مطابق آموزه‌های اخلاق فضیلت‌گرا، ایجاد خوی سلطه‌گری و بی‌تفاوتی را منفی ارزیابی کنیم، نمی‌توانیم به رسمیت شناختن مقوله سوم را به عنوان راه حل بپذیریم.

در نتیجه، وضعیت تناقض‌آمیز همچنان به جای خواهد ماند. کاپوچو و همکارانش می‌گویند برای گریز از این وضعیت می‌توان یکی از دو کفه انسان‌انگاری یا انسانیت‌زدایی را تعدیل کرد، هرچند خود این شیوه را مناسب نمی‌دانند (Cappuccio et al., 2019, p. 26). از آموزه‌های فضیلت‌گرایی، هم می‌توان تعدیل انسان‌انگاری را نتیجه گرفت و هم به نفع تعدیل انسانیت‌زدایی استدلال کرد. برای مثال می‌توان گفت حکمت عملی ایجاب می‌کند از انسان انگاشتن آنچه مانند یک انسان کامل درک نمی‌شود، اجتناب کنیم. یا می‌توان گفت فضایی چون همدلی، شفقت، و مراقبت ایجاب می‌کنند از آن چه شبیه به یک انسان درک می‌شود، انسانیت‌زدایی نکرده و با آن اخلاقی رفتار کنیم. بر این اساس، به نظر می‌رسد ابهام فضیلت‌گرایی درباره این مسئله موجب می‌شود به پاسخی صریح و واحد نرسیم و پدیده ربات‌های اجتماعی (و به طور عام‌تر فناوری‌های شبه اجتماعی مثل چت‌بات‌ها) و جایگاهشان در شبکه تعاملات اجتماعی انسانی، همچنان به عنوان چالشی دشوار برای فیلسوفان فضیلت‌گرا باقی می‌ماند.

### نتیجه‌گیری

اخلاق فضیلت‌گرا به واسطه ویژگی‌های برجسته و متمایزی که از رویکردهای فایده‌گرا و وظیفه‌گرا دارد به عنوان منبعی برای آموزه‌های اخلاقی مورد نیاز در عصر فناوری‌ها حاضر شناخته شده است. از جمله ویژگی‌های مهم این رویکرد زمینه‌گرایی، شخصیت-محوری، تأکید بر اهمیت معرفت‌ضمنی، و تفکر رابطه‌ای است. این ویژگی‌ها توجه فیلسوفان فناوری را به ظرفیت‌های رویکرد فضیلت‌گرا برای مسائل اخلاقی عصر حاضر جلب کرده است. نقطه تمایز و برتری رویکرد فضیلت‌گرا در این حیطة، از این جهت است که تجویزهای خود را بر وجود داده‌های تجربی قاطع درباره پیامدهای عادات فناوری‌ها مبتنی نمی‌کند. بلکه از آنجا که در فضیلت‌گرایی عامل-محور و هدف-محور ملاک ارزیابی اعمال تأثیر آنها بر گرایش‌های شخصیتی افراد است، صرف این که عمل و عادت بر گرایش‌های فرد اثر مثبت (یا منفی) بگذارد، دلیلی برای تجویز (یا تحذیر) نسبت به آن عمل به دست می‌دهد. یکی از موضوعاتی که فیلسوفان فضیلت‌گرا بر آن متمرکز شده‌اند، مسئله جایگاه فناوری‌ها به مثابه مخاطب عمل اخلاقی است. این فیلسوفان با الهام از فضیلت‌گرایی، رویکردی رابطه‌ای نسبت به مسئله وضعیت اخلاقی در پیش گرفته و پیشنهاد می‌کنند برای این که یک فناوری دارای جایگاه اخلاقی باشد، کافی است

که در تعاملات اجتماعی برای کاربران همچون انسان پدیدار شود. از دید این فیلسوفان، از آنجا که رفتار ما با فناوری‌ای که شبه‌دیگری است بر شخصیت ما اثر گذاشته و بر گرایش‌های در رفتار با انسان‌های دیگر منعکس می‌شود، رعایت ملاحظات اخلاقی در برخورد با این فناوری‌ها لازم است.

با این حال این رویکرد در مورد فناوری‌های اجتماعی با چالشی جدی مواجه است. چرا که از یک سو انسان‌انگاری این فناوری‌ها را مجاز می‌داند و از سوی دیگر نمی‌تواند مانع از انسانیت‌زدایی از آنها شود. طبق رویکرد استاندارد، می‌توانیم از این تناقض اجتناب کنیم. چرا که طبق این رویکرد این فناوری‌ها فاقد ویژگی‌هایی هستند که به آنها شأنیت اخلاقی ببخشد و بنابراین نباید آنها را انسان‌انگاری کنیم. اما رویکرد رابطه‌ای به ما اجازه می‌دهد به جنبه‌هایی از تجربه پدیداری خود درباره فناوری‌های اجتماعی اصالت دهیم و آنها را همچون انسان در نظر بگیریم. با این حال از آنجا که همواره در بخشی از تجربه پدیداری ما این فناوری‌ها همچون انسان درک نمی‌شوند (همچنان که طبق دانش گزاره‌ای ما نیز انسان نیستند) و به مثابه ابزار و مایملک بر ما پدیدار می‌شوند، ما با وضعیتی تناقض‌آمیز مواجهیم. اگر برای اجتناب از این تناقض بخواهیم مقوله جدیدی را برای اموری تعریف کنیم که با آنها رابطه عاطفی شبیه به انسان‌ها داریم اما مالک آنها هستیم و نسبت به آنها پاسخگو نیستیم، امکان رشد رذایل و فساد اخلاقی ایجاد می‌شود. بنابراین به نظر می‌رسد در مورد مصنوعات اجتماعی، رویکرد فضیلت‌گرایانه با چالشی جدی مواجه است.



## منابع

زرگر، زهرا، ۱۴۰۴. رابطه عواطف و ظرفیت اخلاقی در فناوری‌های هوش مصنوعی، مجله پژوهش‌های فلسفی، دوره ۱۹ شماره ۵۰، صفحات ۱۹-۴۰.

## Reference

- Anscombe, G.E.M. (1958). "Modern Moral Philosophy". *Philosophy*, 33 (124): 1-19.
- Badhwar, N. (2014). *Well-Being: Happiness in a Worthwhile Life*. Oxford University Press.
- Bloomfield, P. (2014). *The Virtue of Happiness: A Theory of Good Life*. Oxford University Press.
- Cappuccio, M., Peeters, A., McDonald, W. (2019). "Sympathy for Dolores: Moral Consideration for Robots Based on Virtue and Recognition". *Philosophy & Technology*, 33 (1): 9-31.
- Coeckelbergh, M. (2007). "Violent Computer Games, Empathy, and Cosmopolitanism". *Ethics and Information Technology*, 9 (3): 219-231.
- Coeckelbergh, M. (2010). "Moral Appearances: Emotions, Robots, and Human Morality", *Ethics of Information Technology*, 12: 235-241.
- Coeckelbergh, M. (2012). *Growing Moral Relations: Critique of Moral Status Ascription*. Palgrave Macmillan.
- Coeckelbergh, M. (2014). "The Moral Standing of Machines: towards a Relational and Non-Cartesian Moral Hermeneutics", *Philosophy of Technology*, 27 (1): 61-77.
- Held, V. (2005). *The Ethics of Care: Personal Political and Global*. Oxford University Press.
- Hursthouse, R. (2000). *On Virtue Ethics*. Oxford University Press.
- Hursthouse, R. (2022). "Virtue Ethics". *Stanford Encyclopedia of Philosophy*. Online Available on: <https://plato.stanford.edu/entries/ethics-virtue/#FormVirtEthi>
- Korn, O. (2019). *Social Robots, Technological, Societal and Ethical Aspects of Human-Robot Interaction*. Springer.
- MacIntyre, A. (1985). *After Virtue*. Duckworth, 2<sup>nd</sup> Edition.
- Marmodoro, A., Mayr E. (2019). *Metaphysics: An Introduction to Contemporary Debates and Their History*. Oxford University Press.
- Ratti, E., Stapleford, T. (2022). *Science, Technology, and Virtue: Contemporary Perspectives*. Oxford University Press.
- Sandler, R. (2009). *Character and Environment*. Columbia University Press.
- Scheutz, M. (2009). "The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots". Manuscript.
- Sparrow, R. (2017). "Robots, Rape, and Representation", *International Journal of Social Robotics*, 9 (4): 465-477.
- Tronto, J. (1993). *Moral Boundaries: A Political Argument for an Ethics of Care*. Routledge.
- Tronto, J. (2013). *Caring Democracy: Markets, Equality and Justice*. New York University Press.
- Vallor, Sh. (2016). *Technology and the Virtues: A Guide to a Future Worth Wanting*. Oxford University Press.
- Verbeek, P. (2015). "Beyond Interaction: A Short Introduction to Mediation Theory". *Interactions*. 22 (3): 26-31.
- Zagzebski, L. (2017). *Exemplarist Moral Theory*. Oxford University Press.