

Critical Analysis of the “No Relevant Difference” Argument in Defense of the Rights of Artificial Intelligences

Ali Reza Mazarian*

Received: 31/10/2018 | Accepted: 28/02/2019

Abstract

There are many new philosophical queries about the moral status and rights of artificial intelligences; questions such as whether such entities can be considered as morally responsible entities and as having special rights. Recently, the contemporary philosophy of mind philosopher, Eric Schwitzgebel, has tried to defend the possibility of equal rights of AIs and human beings (in an imaginary future), by designing a new argument (2015). In this paper, after an introduction, the author reviews and analyzes the main argument and then deals with four criticisms which Schwitzgebel himself has presented and rejected and ultimately, the author presents and explains two new critiques against the argument. According to the writer, though the argument regarding his new criticisms and some other probable critiques is not convincing, or at least needs to be corrected, however, it holds appeal and raises questions and may open a new horizon for future questions and researches.**



Keywords

Artificial Intelligence, rights, ethics, philosophy of mind, Schwitzgebel.

* Ph.D. Graduate of the Institute for Cognitive Science Studies, Tehran | alirezamazarian@gmail.com

** I consider it my duty to thank Tyler Burge (UCLA), Eric Schwitzgebel (University of California, Riverside), Sam Cumming (UCLA) and David Chalmers (New York University) for their beneficial conversations and emails regarding Schwitzgebel's argument and the two final critiques. Also I am very grateful to the reviewers for the Journal of Philosophical Theological Research for their precise comments.

▣ Mazarian, A. (2019). Critical Analysis of the “No Relevant Difference” Argument in Defense of the Rights of Artificial Intelligences, *Journal of Philosophical Theological Research*, 21(79), 165 -190. <https://doi.org/10.22091/jptr-pfk.2019.3925.2023>.



Introduction

Along with the remarkable progress of science, varied types of super computers, robots and artificial intelligences have emerged. The science of artificial intelligence is still at the initial stages of its awesome path. But at its very beginning, it has inspired new philosophical, ethical and legal questions.

A group of these philosophical questions are moral ones. In such questions, the main focus is on moral considerations about robots and other similar beings. In principle, what moral status may AIs have? Do we have moral responsibility to robots? What is our responsibility towards robots? Can robots have rights similar to animals and humans? Can robots be considered as moral or amoral agents as well? While we are designing and programming robots, should we assume that they have a moral responsibility for their operations and their consequences?

Recently, some philosophers of mind (such as Schwitzgebel: 2015) and Mara Garza have tried to provide an argument (with the title: “No Relevant Difference”) in defense of AI rights. The author seeks to first describe and review the argument and then to criticize it. The “No Relevant Difference” argument has two versions and the main controversy is about the second one.

The second version: The similarity of the moral status of human and AIs

Premise 1. If Entity A deserves some particular degree of moral consideration and Entity B does not deserve that same degree of moral consideration, there must be some *relevant difference* between the two entities that creates the foundation of this difference in their moral status.

Premise 2. There are possible AIs who do not differ in any such relevant or involved respects in the discussion, from human beings and their rights are similar to those of human beings

Conclusion

Therefore, there are possible AIs who deserve a degree of moral consideration similar to that of human beings.

The most important critique against the argument is the supposed “possibility” of some particular beings in the second challenging premise of the argument. Many matters which can be considered epistemologically possible (in terms of their conceivability) are in fact impossible metaphysically and in the external world. There is an important distinction between genuine possibility and conceivability. If we really want to discover whether the property or state we are conceiving as possible to be realized in the world, is really a possible one or not, we need to reasons which mostly depend on our scientific and empirical knowledge. Although conceivability can be a method for discovering metaphysical possibles and impossibles, nevertheless, it is fallible and can show possibles as impossible and vice versa.

I think that Schwitzgebel, and Garza have mixed up the two kinds of possibility regarding the possibility of the emergence of AIs who have the same psychological states as that of humans, without having a human body. Based on epistemic possibility, it can be conceived or imagined that it is possible that some robots, without having any human body, can have psychological states such as phenomenal consciousness, free will, feeling, introspection and so on. But considering possible metaphysical worlds which are more limited than conceivable possible worlds, we find that such beings are not metaphysically possible. My claim is that it is not metaphysically possible that a being can have human psychological properties but is not human.

References

- Bostrom, N. (2003). Are We Living in a Computer Simulation? *Philosophical Quarterly*, 53, 243-55. doi: 10.1111/1467-9213.00309.
- Chalmers, D. (2002). Does Conceivability Entail Possibility? in T. S. Gendler & J. Hawthorne (Eds.) *Conceivability and Possibility*, (pp. 145-200). Oxford: Oxford University Press.
- Egan, G. (1994). *Permutation City*. London: Millennium.
- Egan, G. (1997). *Diaspora*. London: Millennium.
- Fine, K. (2002). The varieties of necessity. T. S. Gendler & J. Hawthorne (Eds.). *Conceivability and possibility*. (pp. 253-281). Oxford: Oxford University Press.
- Hume, D. (1777/1975). *Enquiries Concerning Human Understanding and Concerning the Principle of Morals*. eds. L. A. Selby-Bigge & P. H. Nidditch, Oxford: Oxford University Press.
- Kripke, S. (1980). *Naming and necessity*. Cambridge: Harvard University Press.
- Putnam, H. (1960). Minds and Machines. *Journal of Symbolic Logic*, 57-80. doi: 10.2307/2271581.
- Schwitzgebel, E. & Garza, M. (2015). A defense of the Rights of Artificial Intelligences. *Midwest Studies in Philosophy*, 39, 98-119. doi: 10.1111/misp.12032.
- Soames, S. (2011). Kripke on epistemic and metaphysical possibility: Two routes to the necessary a posteriori. In A. Berger (Ed.), *Saul Kripke*. (pp.167-188). Cambridge: Cambridge University Press.

تحلیل انتقادی استدلال «عدم تفاوت مربوط» در دفاع از حقوق هوش مصنوعی

علیرضا مازاریان *

تاریخ دریافت: ۱۳۹۷/۰۸/۰۹ | تاریخ پذیرش: ۱۳۹۷/۱۲/۰۹

چکیده

پرسش‌های فلسفی نوظهور درباره جایگاه اخلاقی و حقوق هویت دارای هوش مصنوعی فراوان است؛ پرسش‌هایی نظیر این که آیا می‌توان چنین هویت‌هایی را دارای مسئولیت اخلاقی و همچنین حقوق ویژه دانست. اخیراً، اریک شویتسگیل، فیلسوف ذهن معاصر، با طرح استدلالی جدید کوشیده است از امکان حقوق یکسان هوش مصنوعی و انسان در یک آینده فرضی دفاع کند. من در این نوشتار، پس از توضیحاتی مقدماتی، اصل استدلال را تشریح و تحلیل می‌کنم. سپس به چهار نقد مختلف در برابر این استدلال که شویته سگیل طرح و رد کرده، می‌پردازم و در نهایت، دو نقد جدید را در برابر این استدلال طرح و ارائه خواهم کرد. به نظر من اگرچه استدلال مزبور، با توجه به دو نقد من و نقدهای احتمالی دیگران قانع‌کننده نیست، یا دست‌کم نیاز به اصلاحیه‌هایی دارد، اما از جذابیت و پرسش‌برانگیزی خوبی برخوردار است و می‌تواند دریچه‌ای جدید به روی پرسش‌ها و پژوهش‌های بعدی بگشاید.

کلیدواژه‌ها

هوش مصنوعی، حقوق، اخلاق، فلسفه ذهن، استدلال، شویتسگیل.

مقدمه

با پیشرفت چشم‌گیر دانش بشر، انواعی رنگارنگ از آبرایانه‌ها، ربات‌ها و مصادیقی از هوش

* دکتری فلسفه ذهن، پژوهشکده علوم شناختی، تهران، ایران (نویسنده مسئول) | alirezamazarian@gmail.com

مازاریان، علیرضا. (۱۳۹۸). تحلیل انتقادی استدلال «عدم تفاوت مربوط» در دفاع از حقوق هوش مصنوعی. فصلنامه پژوهش‌های فلسفی-کلامی. ۲۱ (۷۹) ۱۶۵-۱۹۰. <https://doi.org/10.22091/jptr-pfk.2019.3925.2023>

1. super computer



مصنوعی^۱ امکان ظهور یافته‌اند. هنوز دانش هوش مصنوعی و رباتیک در آغاز مسیر حیرت‌برانگیز خود است، اما در همین آغاز، به پرسش‌های فلسفی، اخلاقی و حقوقی نوآمده دامن زده است. دسته‌ای از این پرسش‌ها وجودشناسانه است. پرسش‌هایی از این قبیل که واقعاً طبیعت^۲ ربات‌های فوق‌پیشرفته از چه نوع و جنسی است؟ هوش مصنوعی در کدام مقوله^۳ از اشیا می‌گنجد؟ آیا ربات‌های فوق‌پیشرفته که توانایی‌های محاسباتی^۴، عملکردی^۵ و واکنشی^۶ بسیار بالاتری از حد معمول و متعارف بشری دارند را می‌توان در کنار مقولات مادی عادی همچون سنگ و پلاستیک قرار داد؟ به ربات‌های فوق‌پیشرفته، انعطاف‌پذیر و پردازشگری فکر کنید که پیچیده‌ترین فعالیت‌های پزشکی و جراحی، مهندسی و فضاوردی را در کم‌ترین زمان ممکن انجام می‌دهند. آیا نمی‌توان این گونه موجودات بسیار توانمند، خلاق و تحلیل‌گری را شبه‌انسان و فراتر از اشیا مادی بسیط و در مرز میان حیوان و انسان قرار داد؟ آیا می‌توان ربات‌های فوق‌پیشرفته را شخص^۷ و دارای ویژگی‌های شخصیت^۸، عاملیت^۹ و اختیار یا اراده^{۱۰} آزاد دانست؟

دسته‌ای خاص از این قبیل پرسش‌های وجودشناسانه در فلسفه ذهن معاصر بسیار پررنگ‌اند (پاتنم، ۱۹۶۰، ص ۱۹۶۷، ۱۹۶۴). پرسش‌هایی نظیر این که آیا ربات‌ها و کامپیوترها هم می‌توانند فکر کنند؟ آیا ربات‌ها دارای ویژگی آگاهی پدیداری^{۱۱} هستند؟ آیا حالات و رویدادهای ذهنی^{۱۲} نظیر احساسات و درد را می‌توان به هوش مصنوعی نیز اسناد^{۱۳} داد؟ و این که مرز میان موجودات ذهن‌دار و موجودات فاقد ذهن چیست؟ چنین پرسش‌هایی با پیش‌فرض‌ها و مبانی فلسفی گوناگون، پاسخ‌های متفاوتی خواهند داشت. وابسته به این قبیل مباحث قبلی که جهان را تشکیل شده از کدام انواع از هویت^{۱۴} تلقی کنیم، نفس یا ذهن را امری مادی بدانیم یا غیرمادی و مجرد، شخصیت و

1. artificial intelligence
2. nature
3. category
4. computational
5. operational
6. reactive
7. person
8. personality
9. agency
10. free will
11. phenomenal consciousness
12. mental states and events
13. attribute
14. entities

عاملیت را بر اساس چه عواملی تعریف کنیم ویژگی‌های کیفی^۱ را قائم بر چه مؤلفه‌هایی بدانیم و دسته‌ای دیگر از امور، پاسخ‌های متفاوتی را در برابر خواهیم داشت.

دسته‌ای دیگر از این پرسش‌های فلسفی، پرسش‌های اخلاقی^۲ اند. در پرسش‌های اخلاقی، محل بحث ملاحظاتی اخلاقی پیرامون ربات‌ها و دیگر موجودات مشابه است. اساساً هوش مصنوعی چه جایگاه اخلاقی‌ای می‌تواند داشته باشد؟ آیا ما در برابر ربات‌ها مسئولیت اخلاقی داریم؟ مسئولیت ما نسبت به ربات‌ها چیست؟ آیا ربات‌ها همانند حیوانات، طبیعت و انسان‌ها می‌توانند از حقوق برخوردار باشند؟ آیا ربات‌ها نیز می‌توانند یک عامل اخلاقی یا ضد اخلاقی تلقی شوند؟ آیا در هنگام طراحی و برنامه‌ریزی ربات‌ها باید توجه داشته باشیم که آنها در برابر عملکردشان و پی‌آمدهای آن مسئولیت اخلاقی دارند؟ همان‌طور که گفته شد، یکی از مباحث فلسفی - اخلاقی مربوط به هوش مصنوعی، بحث «حقوق هوش مصنوعی»^۳ است. ما صرف‌نظر از مباحث وجودشناسانه محض در اینجا می‌کوشیم به بحث اخلاقی حقوق هوش مصنوعی بپردازیم.

اخیراً برخی فیلسوفان ذهن و علوم شناختی نظیر مارا گارزا و اریک شویتسگبل (۲۰۱۵) کوشیده‌اند استدلالی با عنوان «عدم تفاوت مربوط»^۴ در دفاع از حقوق هوش مصنوعی ارائه کنند. در این نوشتار نخست استدلال مزبور را نقل و تحلیل کرده و سپس نقد می‌کنم.

استدلال عدم تفاوت مربوط

این استدلال دو تقریر^۵ دارد و عمده بحث بر سر تقریر دوم است (Schwitzgebel & Garza, 2015, p. 99)

تقریر اول: مشابهت جایگاه اخلاقی هوش مصنوعی با حیوان

مقدمه^۱: اگر موجود A شایسته میزان خاصی از مراعات اخلاقی^۶ است و موجود B شایسته آن میزان خاص از مراعات اخلاقی نیست، باید تفاوت مربوطی بین این دو موجود باشد که مبنای تفاوت در جایگاه اخلاقی آنان باشد.

مقدمه^۲: هوش مصنوعی‌های ممکن وجود دارند که از جنبه‌های مربوط و دخیل در بحث، هیچ

-
1. qualitative properties
 2. moral
 3. artificial intelligence rights
 4. no-relevant-difference argument
 5. version
 6. moral consideration

تفاوتی با حیوان‌ها ندارند.

نتیجه: بنا براین، هوش مصنوعی‌های ممکن وجود دارند که شایسته مراعات اخلاقی مشابه با حیوان‌ها هستند و حقوق آنها مانند حقوق حیوان‌هاست.

تقریر دوم: مشابهت جایگاه اخلاقی هوش مصنوعی با انسان

مقدمه ۱: اگر موجود A شایسته میزان خاصی از مراعات اخلاقی است و موجود B شایسته آن میزان خاص از مراعات اخلاقی نیست، باید تفاوت مربوطی بین این دو موجود باشد که مبنای تفاوت در جایگاه اخلاقی آنان باشد.

مقدمه ۲: هوش مصنوعی‌های ممکن وجود دارند که از جنبه‌های مربوط و دخیل در بحث، هیچ تفاوتی با انسان‌ها ندارند و حقوق آنها مانند حقوق انسان‌هاست.

نتیجه: بنا براین، هوش مصنوعی‌های ممکن وجود دارند که شایسته مراعات اخلاقی مشابه با انسان‌ها هستند.

می‌دانیم که انسان‌ها مستحق یا سزاوار توجه و مراعات اخلاقی بیشتری نسبت به حیوان‌ها هستند. پس همان‌طور که مشخص است تقریر دوم اصطلاحاً نه به معنای متعارف که ناظر به قوت دلیل یا شاهد است - قوی‌تر خواهد بود، یعنی مدعای بزرگ‌تری دارد و اگر اثبات شود، به طریق اولی، تقریر اول که ضعیف‌تر است و مدعای کوچک‌تری دارد، قابل اثبات است. پس بحث ما در کل بر تقریر دوم متمرکز است.

همان‌طور که روشن است، استدلال به لحاظ صوری معتبر است؛ یعنی نتیجه از مقدمات حاصل می‌شود. اگر دو مقدمه را نیز بپذیریم، استدلال درست^۱ خواهد بود. همچنین در کل، محتوای استدلال انتزاعی^۲ است. برای مثال، لزومی در این نمی‌بیند که به ذکر مصادیق «مربوط و دخیل» در «تفاوت مربوط و دخیل» پردازد. تنها به صورت کلی بیان می‌کند که وقتی دو موجود تفاوت مربوط و دخیلی در بحث با یکدیگر ندارند، به لحاظ استحقاق و شایستگی مراعات اخلاقی نیز یکسانند. فرض کنید استدلال را درباره استحقاق مراعات اخلاقی یک مرد و یک زن، برای مثال، مهدی و مهدیه در نظر می‌گیریم. روشن است که مهدی و مهدیه در جنسیت با هم تفاوت دارند، اما این تفاوت علی‌الاصول در بحث مراعات اخلاقی این دو تفاوت مربوط و دخیلی ایجاد نمی‌کند. پس اگر چه در امر جنسیت با هم متفاوتند، اما استحقاق مراعات اخلاقی یکسانی خواهند داشت. ما می‌توانیم درباره این که دقیقاً چه

1. sound

2. abstract

اموری به بحث مربوط و دخیل‌اند، دیدگاه‌های مختلفی داشته باشیم. اما استدلال از آنجا که به لحاظ محتوایی انتزاعی است، به چنین اموری نمی‌پردازد و نسبت به آنها خنثی^۱ است. در واقع، این استدلال ما را به تأملات بعدی در این امر فرامی‌خواند که چه اموری به تفاوت وضعیت اخلاقی موجودات مربوط است و چه اموری بی‌ارتباط است. البته، در ادامه، گزینه‌های فراگیر و چندسویه‌ای که محدود به دیدگاهی خاص نیست، معرفی خواهد شد.

نکته دیگر این که سخن از امکان تحقق پیدا کردن چنین هوش مصنوعی‌هایی است، نه از تحقق بالفعل آنها و بدین لحاظ نیز استدلال کم‌ادعا و اصطلاحاً ضعیف است. برخی معتقدند که فراتر از بحث امکان، در واقع، اکنون نیز انسان‌ها هیچ تفاوت مربوط و دخیل در بحثی با موجودات کامپیوتری ندارند (Egan, 2003; Bostrom, 1997)، اما این استدلال به طرح چنین ادعای قوی و دشواری نمی‌پردازد.

در باره مقدمه^۱ نمی‌توان شک و تردید قابل توجهی داشت. با توجه به این پیش‌فرض شهودی^۲ که علی‌الاصول، اخلاق معیارمند و غیرخودسرانه^۳ است، نمی‌توان در این تردید کرد که اگر دو موجود، تفاوت دخیل یا مربوطی نداشته باشند، به لحاظ استحقاق اخلاقی یکسانند. برای مثال، نمی‌توان در این تردید کرد که اگر دو حیوان تفاوت دخیل یا مربوطی نداشته باشند، به لحاظ استحقاق اخلاقی و حقوق نیز یکسانند. همین قضیه درباره انسان‌ها نیز صادق است.

مقدمه^۱ جدا از این که به نظر صادق می‌رسد، بصیرتی مهم را نیز در بر دارد. این بصیرت در قالب «آزمون تفاوت» قابل ارائه است. آزمون تفاوت به ما می‌گوید که اگر می‌خواهیم بفهمیم موجودی با موجود دیگر به لحاظ استحقاق مراعات اخلاقی هم‌سان یا ناهم‌سان است، باید به این توجه کنیم که آیا این دو تفاوت مربوط و دخیلی با یکدیگر دارند، یا نه. در واقع، این آزمون تلویحاً مستلزم این نکته است که بدون وجود تفاوت‌های مقتضی و قابل توجه، نمی‌توانیم به تفاوت جایگاه اخلاقی دو موجود قائل باشیم؛ چه در مقایسه انسان‌های مختلف با یکدیگر و چه در مقایسه انسان‌ها یا انواع دیگر موجودات، نظیر حیوانات، ربات‌ها و... باید حتماً تفاوت مربوط و دخیلی وجود داشته باشد، تا بتوانیم آنها را به صورتی موجه و معقول، به لحاظ استحقاق مراعات اخلاقی و همچنین جایگاه اخلاقی متفاوت بدانیم. این آموزه را می‌توان آموزه‌ای مبتنی بر عدالت اخلاقی و مبانی جهان‌شمول^۴ اخلاق دانست.

مبانی یا بنیان‌های جهان‌شمول اخلاق با تکیه بر چنین آزمون‌هایی، برداشت‌های شخصی و گروهی ساده‌اندیشانه که بدون پشتوانه‌های عقلی مقتضی، به تبعیض اخلاقی دامن می‌زنند، را به چالش می‌کشد.

1. neutral
2. intuitional
3. non-arbitrary
4. universal

تمایزاتی نظیر رنگ پوست، سن، جنسیت و... نمی‌توانند تفاوت دخیل و مربوطی در بحث وضعیت اخلاقی دانسته شوند. پس نمی‌توان مبتنی بر چنین تمایزهایی، تفاوت در استحقاق اخلاقی را نتیجه گرفت. اما دربارهٔ مقدمهٔ ۲ احساس می‌شود که می‌توان شک و تردیدهایی داشت. که در ادامه، به نقدهای گوناگون دربارهٔ آن خواهیم پرداخت.

محدودسازی مقدمهٔ ۱: ارتباط وضعیت اجتماعی - روان‌شناختی با جایگاه اخلاقی موجودات

شویتسگیل و گارزا دربارهٔ قید «تفاوت مربوط» در مقدمهٔ ۱ ادعا می‌کنند که این که یک موجود چه نوع بدن یا اندام‌هایی دارد، ارتباطی به وضعیت استحقاق اخلاقی او ندارد، مگر این که چگونگی نوع بدن او در ویژگی‌های روان‌شناختی - اجتماعی او تأثیر بگذارد (Schwitzgebel & Garza, 2015, p. 100). مبنای این ادعا نیز یک استقراء دربارهٔ گزینه‌های مختلف احتمالی است. سعی شده است مدل پیه‌شنهادی نسبت به این گزینه‌ها جامع با شد. یکی از گزینه‌ها محصول یک نگرش کانتی است که بر اساس آن مربوط‌ترین تفاوت در این بحث، توانایی‌های عقلانی است. به این معنا که دو نوع از موجودات، برای مثال، ربات و انسان که به لحاظ توانایی‌های عقلانی متفاوت‌اند، به لحاظ حقوق و جایگاه اخلاقی نیز متمایز خواهند بود و بالعکس. گزینه‌ای دیگر نگرشی سودانگاران^۱ است که بر مبنای آن مهم‌ترین تفاوت‌های مربوط به لحاظ اخلاقی، تفاوت به لحاظ احساساتی نظیر لذت و درد است. به این معنا که دو نوع از موجودات، برای مثال، ربات و انسان که به لحاظ احساساتی نظیر لذت و درد متفاوت‌ند، به لحاظ جایگاه اخلاقی نیز متمایز خواهند بود. بر اساس این گزینه، انسان و ربات نمی‌توانند به لحاظ جایگاه و حقوق اخلاقی یکسان باشند، مگر این که به لحاظ احساساتی نظیر لذت و درد یکسان باشند. بر اساس گزینه‌ای دیگر روابط و موقعیت اجتماعی، تفاوت دخیل یا مربوط از حیث اخلاقی است.

همچنین ادعا می‌کنند که این که بدن یک موجود چه طراحی و چیدمان زیرینی دارد، نیز به وضعیت و جایگاه اخلاقی او ارتباطی ندارد، مگر این که آن طراحی و چیدمان زیرین در ویژگی‌های روان‌شناختی و اجتماعی او تأثیر بگذارد. پس طبق نظر آنان تنها ویژگی‌های روان‌شناختی و اجتماعی موجودات در تفاوت وضعیت اخلاقی، مربوط و دخیل خواهند بود. منظورشان از ویژگی‌های روان‌شناختی نیز هم

ویژگی‌های شناختی و کارکردی^۱ مثل توان استدلال ریاضیاتی و هم ویژگی‌های پدیداری^۲ و آگاهانه^۳ مثل استعداد تجربه درد هنگام آسیب دیدن است. آنها به این بحث نمی‌پردازند که آیا ویژگی‌های پدیداری و آگاهانه، به ویژگی‌های شناختی و کارکردی، تحویل‌پذیر هست یا نه، بلکه تنها مدعی‌اند هر دو نوع ویژگی‌ها را در توضیح «تفاوت مربوط» به عنوان ویژگی‌های روان‌شناختی مد نظر دارند. همچنین منظورشان از ویژگی‌های اجتماعی ویژگی‌های مربوط به روابط اجتماعی نظیر جایگاه خانوادگی، شهروندی، عضویت در یک جامعه خاص و... است.

روشن است که این دیدگاه تنها نسبت به وضعیت روان‌شناختی موجودات توجه ندارد، پس یک منظر روان‌شناختی محض نیست. یک منظر روان‌شناختی محض و وضعیت اخلاقی موجودات را تنها منوط به ویژگی‌های روان‌شناختی آن موجودات می‌داند. در این صورت، بر مبنای یک منظر روان‌شناختی محض که وضعیت روان‌شناختی را تنها عامل مهم در جایگاه اخلاقی می‌داند، اگر موجودی جایگاه اجتماعی داشته باشد، اما جایگاه روان‌شناختی نداشته باشد، یا ویژگی‌های روان‌شناختی خاصی نداشته باشد، جایگاه و موقعیت اخلاقی او دستخوش تغییر خواهد شد، اما دیدگاه این دو فیلسوف چنین نیست، بلکه افزون بر ویژگی‌های روان‌شناختی، روابط اجتماعی را نیز دخیل در بحث می‌دانند. پس طبق نظر این دو وضعیت روان‌شناختی شرط لازم است نه کافی. طبق این دیدگاه ویژگی‌های روان‌شناختی و همچنین اجتماعی در موقعیت اخلاقی بالفعل و خلاف واقع^۴ و همچنین گذشته و آینده موجودات دخیل است.

در اینجا این پرسش پیدا می‌شود که چرا باید چنین دیدگاهی را پذیرفت؟ به بیان دیگر، بر اساس چه دلیلی، می‌توانیم این دیدگاه را بپذیریم که تنها بر اساس تفاوت یا عدم تفاوت ویژگی‌های روان‌شناختی و اجتماعی، موقعیت اخلاقی موجودات می‌تواند متفاوت یا یکسان باشد. دو فیلسوف مزبور دو دلیل برای این دیدگاه ارائه می‌کنند:

دلیل اول

طبق گزارش این دو فیلسوف، بر اساس همه تبیین‌های عرفی فلسفه جدید وضعیت و جایگاه اخلاقی موجودات و افراد مختلف به ویژگی‌های روان‌شناختی و اجتماعی آنها، مثل توانایی تفکر عقلانی،

1. functional
2. phenomenal
3. conscious
4. counterfactual

لذت، درد و روابط اجتماعی منوط است. بدین ترتیب، بر اساس این تلقی، موجوداتی که ویژگی‌های روان‌شناختی و اجتماعی یکسانی دارند، به لحاظ موقعیت اخلاقی و استحقاق مراعات اخلاقی نیز یکسانند و بالتبع حقوق یکسانی نیز باید داشته باشند.

این دلیل از منظری دینی و همچنین ارسطویی قابل نقد است: بر اساس یک منظر دینی، دو موجود که به لحاظ روان‌شناختی و اجتماعی یکسان باشند، ممکن است با توجه به روابط بنیادینی که در نحوه خلقت یا تکوین خود با خدا دارند، به لحاظ اخلاقی متفاوت باشند؛ برای مثال، یک ربّات جراح فوق پیشرفته و یک آدم جراح که به لحاظ اجتماعی و روان‌شناختی، فرضاً جایگاه و کارکردهای یکسانی دارند، می‌توانند با توجه به تفاوت تکوین در رابطه با خدا، جایگاه اخلاقی متفاوتی داشته باشند. خداوند یکی از آنان را، یعنی آدم جراح را طبق این دیدگاه، دارای روح آفریده است و دیگری را انسان‌ها ساخته‌اند.

شویتسگبل و گارزا این نقد را وارد نمی‌دانند؛ چون معتقدند حتی بنا بر این فرض که نحوه رابطه خدا با یک موجود در جایگاه اخلاقی او دخیل باشد، این رابطه باید در ویژگی‌های روان‌شناختی و اجتماعی او تأثیری بگذارد. در غیر این صورت، پذیرش چنین امری معقول نیست. در واقع، آنان به این رأی ملتزمند که خداوند نمی‌تواند برای دو موجود که ویژگی‌های روان‌شناختی و اجتماعی یکسانی دارند، جایگاه اخلاقی متفاوتی در نظر بگیرد.

همچنین بر اساس یک منظر ارسطویی، هدف یا غایت^۱ موجودات بر جایگاه اخلاقی آنان تأثیر می‌گذارد. بدین ترتیب، دو موجود، مثل همان آدم جراح ماهر و همان ربّات فوق پیشرفته جراح که ویژگی‌های روان‌شناختی و اجتماعی یکسانی دارند، با توجه به هدف و غایت متفاوت‌شان جایگاه اخلاقی ناهم‌سانی خواهند داشت.

دو فیلسوف مزبور این نقد را نیز وارد نمی‌دانند، چون معتقدند اولاً چنین تفسیری مبهم است و بعید است که ارسطوییان آشکارا این تفسیر را بپذیرند. ثانیاً، تفاوت در غایات نیز باید در تفاوت ویژگی‌های روان‌شناختی و اجتماعی منعکس شود و گرنه به خودی خود قابل قبول نیست (Schwitzgebel & Garza, 2015, p. 103).

دلیل دوم

به نظر این دو فیلسوف، با تمرکز بر برخی سناریوهای خلاف واقع^۲ هنری درمی‌یابیم که شهود ما مویذ دیدگاه

1. telos

2. counterfactual

روان‌شناختی-اجتماعی است. به فیلم‌های علمی-تخیلی مرتبط با ربات‌ها فکر کنید. در این قبیل فیلم‌ها، ما با توجه به افکار، احساسات، دردها و نقش اجتماعی ربات‌ها جایگاهی اخلاقی برای آنها در نظر می‌گیریم. وقتی چنین فیلم‌هایی را می‌بینیم، یا چنین داستان‌هایی را می‌خوانیم، برای ما مهم نیست که بدن این موجودات از چه ساخته شده است، آیا در رگ‌هایشان مثل ما انسان‌ها خون جاری است و یارگ‌هایشان سیلیکونی و پلاستیکی است، یا این‌که چهره‌شان دقیقاً چه شکلی است. ما تنها با توجه به ویژگی‌های روان‌شناختی و اجتماعی آنها، آنها را دارای جایگاه و موقعیت اخلاقی و مستحق مراعات اخلاقی می‌دانیم؛ از کشته شدن خیالی آنها در فیلم‌ها متأثر می‌شویم و پیروزی‌شان را در همان سناریوهای فرضی آرزو می‌کنیم. چنین قضاوت‌هایی حاکی از این نکته شهودی است که ویژگی‌های روان‌شناختی و اجتماعی موجودات در موقعیت اخلاقی آنها دخیل‌اند، نه نوع یا کیفیت بدن آنها و این‌که بدنی انسانی یا شبه‌انسانی داشته باشند یا نه. بدن‌مندی آنها و نوع بدن‌شان اگر مهم باشد، تنها تا آنجا اهمیت دارد که بر ویژگی‌های روان‌شناختی و اجتماعی آنها تأثیر می‌گذارد و فی‌نفسه، در این بحث، تعیین‌کننده نیست.

با توجه به همین بحث، بار دیگر باید متذکر شویم که اگر چه در واقع، فعلاً دانش رباتیک و هوش مصنوعی ربات‌هایی نساخته است که امکان ربات‌های دارای بدن مشابه با انسان را به وضوح عملی سازد، اما دور از انتظار نیست که چنین کند. ممکن است که در آینده نزدیک یا دور به دست بشر ربات‌های فوق پیشرفته‌ای ساخته شوند که بخش‌هایی یا تمامی بدن آنها نه سیمی یا سیلیکونی، بلکه طبیعی و مشابه با انسان باشند. چنین امکانی باز به وضوح نشان می‌دهد که ما برای کشف موقعیت اخلاقی یک موجود، به ساختار بدن او نمی‌نگریم، بلکه به ویژگی‌های روان‌شناختی و اجتماعی او توجه می‌کنیم. به این توجه می‌کنیم که چه جایگاه و نقشی در جامعه دارد و می‌تواند داشته باشد و چه حالات و احساسات ذهنی‌ای دارد و می‌تواند داشته باشد.

اگرچه ما با توجه به همین نکته می‌توان دریافت که وقتی از «هوش مصنوعی» سخن می‌گوییم، واژه «مصنوعی» به‌صورتی مبهم و احتمالاً اختلاف‌برانگیز به کار می‌رود. اگر در آینده ربات‌هایی داشته باشیم که با پیوند طبیعی با بدن انسان تکامل یابند و دارای رگ و خون و برای مثال، باشند، آیا باز هم آنها را «مصنوعی» می‌نامیم، یا «طبیعی».

همین امروز نیز در باره ویروس‌ها یا برخی حیوانات که دست‌کاری ژنتیک شده‌اند، چنین پرسش‌هایی مطرح است، اما دیدگاه این دو فیلسوف اساساً حاکی از این است که در بحث موقعیت اخلاقی، مهم نیست که چه چیزی را مصنوعی یا طبیعی می‌نامیم و مهم نیست که موجودات چگونه بدنی دارند، بلکه ویژگی‌های روان‌شناختی و اجتماعی مهم است.

از همین جا مسیر ساخت یک استدلال جدید آماده می‌شود. انسانی را در نظر بگیرید که به تدریج بخش‌هایی از مغز و بدنش با قطعات مصنوعی پیشرفته جایگزین می‌شود. یا جینی را در نظر بگیرید که

در موقعیت آزمایشگاهی و ابتدائاً از ترکیبات غیرطبیعی و به تدریج در فرایند اصلاح ژنتیک متولد می‌شود. به نظر می‌رسد چنین موجوداتی اگر چه بدنی متفاوت با بدن ما انسان‌ها دارند، اما جایگاه اخلاقی آنان مشابه ماست و تنها امر مهم در این میان ویژگی‌های روان‌شناختی و اجتماعی آنهاست.

استدلال جایگزینی بدن طبیعی با قطعات مصنوعی

مقدمه ۱: با جایگزین سازی یک قطعه مصنوعی کوچک درون بدن یک موجود که دارای حقوقی است، اگر در ویژگی‌های روان‌شناختی و اجتماعی آن موجود تغییری ایجاد نشود، در حقوق و موقعیت اخلاقی آن موجود تغییری ایجاد نمی‌شود.

مقدمه ۲: فرایندی که در مقدمه ۱ گفته شد، می‌تواند آن قدر تکرار شود که یک موجود طبیعی دارای حقوق به یک موجود مصنوعی دارای همان حقوق قبلی تبدیل شود. نتیجه: پس ممکن است موجود مصنوعی‌ای ساخته شود که همان حقوق موجود طبیعی و انسانی را داشته باشد.

با توجه به این‌که اکنون پیوند قلب مصنوعی به آسانی و فراوانی صورت می‌پذیرد، می‌توان تصور کرد که در روزگاری دور یا نزدیک، عمل جراحی پیوند مغز مصنوعی نیز تحقق یابد. در چنین شرایطی بعید نیست، اجزای بدن یک انسان، به تدریج و یکی پس از دیگری با قطعات مصنوعی جایگزین شود، تا آنجا که در نهایت، هیچ عضو طبیعی از بدن قبلی او باقی نماند. در چنین حالت فرضی و در عین حال ممکن، حقوق و جایگاه اخلاقی این انسان پیش و پس از عمل جراحی تغییری نمی‌کند، مگر آن‌که ویژگی‌های روان‌شناختی و اجتماعی او تأثیر پذیرفته و تغییر کند. پس دوباره مشخص می‌شود که جایگاه اخلاقی موجوداتی همچون انسان‌ها، حیوانات و ربات‌ها وابسته به ویژگی‌های روان‌شناختی و اجتماعی آنها نظیر افکار و احساسات و نقش اجتماعی آنهاست، نه نوع طبیعی یا مصنوعی بدن آنها.

نقدها

نقد اول: تفاوت‌های روان‌شناختی انسان‌ها و هوش مصنوعی

با توجه به این‌که ربات‌ها و دیگر مصادیق هوش مصنوعی تفاوت‌های بارز روان‌شناختی با انسان‌ها دارند، هم در شرایط فعلی و هم در آینده بسیار بعید است که بتوان این تفاوت‌ها را از میان برداشت. برای مثال، سرل (۱۹۸۰) ربات‌ها و دیگر موجودات مشابه را فاقد آگاهی می‌داند. لاولیس (۱۹۸۴) چنین موجوداتی را فاقد اختیار تلقی می‌کند. و پن‌روز (۱۹۹۹) معتقد است که هوش مصنوعی فاقد درون‌نگری است.

پاسخ

استدلال «عدم تفاوت مربوط» سخنی در باره میزان احتمال ساخت یا پیدایش چنین موجوداتی نمی‌گوید؛ تنها ادعا می‌کند که تحقق چنین شرایط فرضی‌ای ممکن است. استدلال، ادعا می‌کند که ممکن است که هوش مصنوعی‌های فوق‌پیشرفته‌ای در آینده وجود پیدا کنند که به لحاظ ویژگی‌های روان‌شناختی و اجتماعی همانند انسان‌ها باشند و تفاوت مربوطی به لحاظ استحقاق مراعات اخلاقی با ما نداشته باشند. منتقدان باید نشان دهند که چنین چیزی ممکن نیست. استدلال نشان می‌دهد با توجه به این‌که پیدایش چنین موجوداتی ممکن است، نمی‌توان موقعیت اخلاقی تمام انواع هوش مصنوعی را یکسان دانست^۱ (Schwitzgebel & Garza, 2015, p. 103).

حتی اگر آگاهی، اختیار، درون‌نگری و این قبیل ویژگی‌ها مستلزم یک روح مجرد باشند، می‌توان ادعا کرد که همان‌طور که طبق یک ادعا وقتی بدن یا جسم به سطحی از تکامل و پیچیدگی رسید، نفس به عنوان امری جسمانی‌الحدوث در آن ظهور پیدا می‌کند، به همین ترتیب، آن هوش مصنوعی‌های فوق‌پیشرفته ممکن نیز خواهند توانست واجد این ویژگی‌های مجرد شوند.^۲

نقد دوم: تکثیرپذیری^۲ هوش مصنوعی و منحصر به فرد بودن انسان‌ها

بر اساس یک انتقاد که هانکینز مطرح کرده (Hankins, 2015)، می‌توان استدلال «عدم تفاوت مربوط» را چنین نقد کرد که حتی در آن شرایط فرضی ممکن که ربات‌های فوق‌پیشرفته‌ای وجود دارند، تفاوتی جدی بین انسان‌ها و آن موجودات وجود دارد. هر انسانی یک موجود منحصر به فرد^۳ و تکثیرناپذیر است، اما ربات‌ها حتی ربات‌های فوق‌پیشرفته قابلیت تکثیر و تولید انبوه دارند. به‌راحتی می‌توان دو ربات یکسان را ساخت و تکثیر کرد، اما هر انسانی برخوردار از ویژگی‌های بی‌همتا و منحصر به فردی است که اجازه نمی‌دهد شخصی از همه جهات مشابه او تولید شود. بدین ترتیب، در آن شرایط فرضی، ضربه‌زدن یا حتی از بین بردن آن‌گونه ربات‌های فوق‌پیشرفته تفاوتی بنیادین با ضربه‌زدن یا از بین بردن انسان‌ها دارد. پس نمی‌توان موقعیت اخلاقی و استحقاق مراعات اخلاقی را انسان‌ها و آنها را یکسان انگاشت.

پاسخ

۱. این‌که آیا ممکن است هوش مصنوعی، حالات ذهنی انسان‌ها که دارای ویژگی‌های پدیداری و بازنمودی-الثباتی‌اند، را داشته باشد، نیازمند بحث بیشتر و مستقل است.

۲. تورینگ (۱۹۵۰) پاسخ مشابهی ارائه کرده است.

3. duplicability

4. unique

به نظر می‌رسد که می‌توان به راحتی تصور کرد که ممکن است ربات‌های فوق پیشرفته‌ای ساخته شوند که منحصر به فرد باشند. برای مثال، هر کدام از ساختار و اجزای صنعتی، فیزیکی و شیمیایی ویژه و منحصر به فردی بهره‌مند باشند. در چنین شرایطی ممکن است باید پذیرفت که جایگاه اخلاقی انسان‌ها با چنین موجودات منحصر به فردی یکسان است (Schwitzgebel & Garza, 2015, p. 105).

نقد سوم: تفاوت و دیگربودگی آن انواع

با توجه به مبانی فلسفی هابز، (Hobbes, 1651/1996, p. 89) می‌توان ادعا کرد که بر اساس اشتراک ما انسان‌ها در نوع زیست‌شناختی و فراهم ساختن قراردادهای اجتماعی مبتنی بر همین نوع زیست‌شناختی ما به این درجه از رشد و حیات رسیده‌ایم. ما با توجه به اشتراکات زیست‌شناختی مان با دیگر انسان‌های هم‌نوع خود و ایجاد تفاهم و قراردادهای میان‌نوعی بر دیگر حیوانات غلبه پیدا کرده‌ایم. این واقعیت که اشتراکات نوعی زمینه ساز تمدن، اخلاق و فرهنگ ما است در واقعیت‌های روزمره نیز مشاهده می‌شود. انسان‌ها معمولاً به لحاظ اخلاقی نسبت به خانواده و دوستان، خویشان، همسایگان، همشهریان، هم‌وطنان خود به همین ترتیب اهمیت، توجه بیشتری دارند. به همین ترتیب، اگر موجوداتی فضایی بر اساس تزهایی بر زمین بیابند با توجه به منافع و مصالح شان خطا نیست که ما انسان‌ها را که از نوع دیگری هستیم، بکشند. پس در واقع، مسئله نوع زیست‌شناختی در ملاحظات اخلاقی دخیل خواهد بود. در نتیجه، در آن سناریوی فرضی، با توجه به تفاوت ما انسان‌ها و آن ربات‌ها در نوع نمی‌توانیم جایگاه اخلاقی یکسانی را برای هر دو نوع موجود فرض کنیم (Singer, 1975; 2002; 2009).

پاسخ

تقریرهای رادیکال این تزهایی از لحاظ اخلاقی، به هیچ روی قابل قبول نیست. نمی‌توان پذیرفت که تنها با توجه به اشتراکات نوعی، ملاحظات اخلاقی تحت تأثیر قرار گیرد و پذیرش چنین تزی مستلزم گونه‌ای «نوع‌گرایی افراطی» است.

همچنین می‌توان به آسانی فرض کرد که در آن شرایط ممکن است مرز میان نوع انسان‌ها و ربات‌ها بسیار پیچیده شود، به صورتی که بسیاری از ربات‌ها مغز انسان داشته باشند، بسیاری از انسان‌ها مغز مصنوعی داشته باشند. در چنین حالتی نیز روشن است که صرف تفاوت زیست‌شناختی و تفاوت در نوع، حتی اگر قابل احراز باشد، نمی‌تواند مبنایی برای تفاوت اخلاقی باشد، مگر این‌که تفاوت‌هایی روان‌شناختی یا

اجتماعی در میان باشد. در آن شرایط ممکن، یک ربات فوق پیشرفته و یک انسان که حالات ذهنی و موقعیت اجتماعی یکسانی دارند (همانند یک ربات فوق پیشرفته جراح و یک انسان جراح) وضعیت اخلاقی و استحقاق مراعات اخلاقی یکسانی نیز خواهند داشت. صرف دیگربودگی زیست‌شناختی بدون تأثیر گذاشتن بر تفاوت‌های روان‌شناختی و اجتماعی تأثیری بر ملاحظات اخلاقی نخواهد گذاشت (Schwitzgebel & Garza, 2015, p. 106).

نقد چهارم: وابستگی وجودی ربات‌ها به انسان‌ها

این انتقاد (DeGrazia, 2009) حاکی از این است که ربات‌ها از آنجا که مصنوعی هستند، مصنوع و دست‌ساز بشرند و وابستگی وجودی به انسان‌ها دارند. پس بنا بر این نقد، انسان‌ها آنها را به وجود آورده‌اند و هر گاه که بخواهند، می‌توانند آنها را از بین ببرند یا به آنها آسیب برسانند، بی‌آنکه مرتکب خطایی اخلاقی شده باشند. برای مثال، ربات فوق‌پیشرفته‌ای را در نظر بگیرید که انسانی صد سال بعد می‌تواند بخرد و از آن استفاده کند. یک روز مالک این ربات می‌بیند که گزینه اقتصادی بهتری پیش روی اوست. برای مثال، به جای این که ماهی صد دلار برای نگهداری و زنده ماندن این ربات بدهد، می‌تواند چند مجله بخرد. او تصمیم می‌گیرد این ربات فوق پیشرفته انسان‌نما را از بین ببرد. ربات گریه می‌کند و از او درخواست می‌کند که اجازه بدهد به حیاتش ادامه بدهد. مالک در مقابل می‌گوید «من تو را ساخته‌ام. تو وجودت را مدیون من هستی. من هرگاه تصمیم بگیرم می‌توانم تو را از بین ببرم» برای این اساس، احساس می‌شود از آنجا که حتی آن ربات‌های فوق پیشرفته نیز وجود و زندگی شان را مدیون ما هستند، تفاوتی بنیادین به لحاظ اخلاقی با انسان‌ها ندارند.

پاسخ

نقد فوق مستلزم این پیش‌فرض است که وقتی در ایجاد شدن یا زندگی یافتن موجودی نقش یا سهم داریم، در از بین بردن یا ضربه زدن به او نیز دست ما باز است و اخلاقاً مجازیم که هر کار که با او خواهیم بکنیم. چنین پیش‌فرضی نادرست است؛ زیرا اگر چنین پیش‌فرضی درست باشد والدینی که تصمیم گرفته‌اند بچه‌دار شوند، می‌توانند بعد از چند سال، برای مثال، کودک سه‌ساله‌شان را از بین ببرند و مدعی شوند که «از آنجا که آغاز حیات و وجود این کودک وابسته به خواست و عمل ما بوده است، مرگ او نیز در دست ما و در محدوده اختیار ماست، بی‌آنکه مرتکب خطای اخلاقی شده باشیم!» همان‌طور که مشاهده می‌شود، مسئولیت اخلاقی نسبت به یک موجود ارتباطی با این که آن موجود توسط چه کسی ساخته یا ایجاد شده است، ندارد؛ حتی کسی که در وجود و حیات موجودی برای مثال، یک ربات

یا کودک، نقشی اساسی یا زمینه‌چینانه دارد، اختیار مطلق نسبت به او ندارد و در دامنه همین مسئولیت اخلاقی قرار دارد (Schwitzgebel & Garza, 2015, p. 108).

افزون بر آن، همین نکته که ما نقشی در ایجاد و ساخت آن ربات‌های فوق پیشرفته داریم، خود دلیلی بر این است که ما مسئولیت اخلاقی بیشتری در قبال آنها داریم. ما آنها را می‌سازیم، پس باید شرایط زیست و رشد بهینه آنان را نیز فراهم کنیم. همان‌طور که این نکته دربارهٔ کودکان صادق است، دربارهٔ آن ربات‌های فوق پیشرفته ممکن نیز صادق خواهد بود. ما روابطی همچون «سازنده یا آفریننده با ساخته شده»، «والدین با فرزند» و یا «کارفرما با کارگر» با آنها خواهیم داشت و چنین روابطی مستلزم حقوق و ملاحظات اخلاقی دوسویه‌ای است.

نقد پنجم: امکان در استدلال عدم تفاوت مربوط معرفتی است، نه متافیزیکی

چهار نقد گذشته، نقدهایی بود که طراحان استدلال با الهام از دیگران، ارائه کرده و کوشیده بودند پاسخ‌هایی به آن بدهند. در اینجا، من دو نقد جدید طرح می‌کنم که می‌تواند استدلال را نامعتبر یا دست‌کم مبهم و نیازمند تقریرهای دیگری سازد.

نقد اول این است که – همان‌طور که گذشت – در مقدمهٔ چالش‌برانگیز دوم استدلال «عدم تفاوت مربوط»، سخن از امکان موجوداتی خاص بود. مقدمهٔ دوم می‌گفت: هوش مصنوعی‌های ممکن وجود دارند که از جنبه‌های مربوط و دخیل در بحث، هیچ تفاوتی با انسان‌ها ندارند و حقوق آنها مانند حقوق انسان‌هاست.

پیش از پرداختن به نقد جدیدی که می‌خواهیم طرح کنیم، مناسب است مقدماتی را توضیح دهیم. می‌دانیم که ملهم از کریپکی^۱، تمایز مهمی در فلسفهٔ تحلیلی معاصر میان امکان متافیزیکی^۲ و امکان معرفتی^۳ برقرار و شناخته شده است. در یک بیان تقریبی، امکان معرفتی ناظر به طرق یا گونه‌هایی^۴ است که اشیاء می‌توانند به‌لحاظ تصور^۵ باشند، اما امکان متافیزیکی ناظر به طرق یا گونه‌هایی است که اشیاء واقعاً می‌توانند باشند یا بوده‌اند. با استمداد از مفهوم جهان ممکن^۶ یا جهان‌های ممکن می‌توان این تمایز

۱. کریپکی، وجهیت معرفتی (epistemic modality) را در مقابل وجهیت‌های ناظر به واقع اصیل (genuine de re modalities) قرار می‌دهد که منظور از آنها شامل وجهیت متافیزیکی می‌شود (Kripke, 1980, p. 142).

2. metaphysical
3. epistemic
4. ways
5. conceivably
6. possible world

را بهتر توضیح داد.

جهان‌های ممکن به لحاظ متافیزیکی، یا حالات جهان‌های ممکن به لحاظ متافیزیکی، تمام طرق یا گونه‌هایی است که همین جهان عینی واقعی می‌توانست و می‌تواند باشد، یا به تعبیری بیشترین ویژگی‌های است که همین جهان عینی واقعی می‌توانست و می‌تواند متمثل سازد. برخی ویژگی‌ها در جهان ممکن است که متمثل یا دارای مصداق شوند. این ویژگی‌ها و ویژگی‌های ممکن به لحاظ متافیزیکی اند، اما برخی ویژگی‌ها ممکن نیست که در جهان متمثل یا دارای مصداق شوند. این ویژگی‌ها و ویژگی‌های ناممکن به لحاظ متافیزیکی اند. برای مثال، یک بازی دوستانه را در نظر بگیرید. این بازی هم‌اینک بین دو نفر در حال انجام گرفتن است. این وضعیت را «وضعیت بالفعل» می‌نامیم. ممکن است که این بازی در جهان یا شرایط ممکن دیگری بین چهار نفر برقرار می‌بود. این وضعیت را «وضعیت ممکن متافیزیکی» می‌دانیم، اما ناممکن است که این بازی بین صفر نفر یا هیچ‌کس هیچ‌کس انجام می‌شد. این وضعیت را ناممکن یا محال متافیزیکی می‌دانیم، چون انجام یا تحقق بازی بدون بازیکن ممکن نیست.

اما جهان‌های ممکن به لحاظ معرفتی، یا حالات جهان‌های ممکن به لحاظ معرفتی، تمام طرق یا گونه‌هایی اند که می‌توان به صورتی سازگار تصور کرد که جهان باشد، یا به تعبیری تمام ویژگی‌هایی که می‌توان تصور کرد که جهان می‌تواند متمثل کند. شخص به صورت پیشینی نمی‌تواند بداند که چنین ویژگی‌هایی متمثل شده است یا نه.

با توجه دقیق نسبت به این تمایز، درمی‌یابیم که ممکن است که یک نفر ویژگی‌هایی که تحقق شان به لحاظ متافیزیکی ناممکن است را ممکن تصور کند. برای مثال، در روزگار باستان بسیاری عقاید غیرعلمی و نامدل وجود داشته که بنا بر آنها تصور می‌کرده‌اند که ممکن است جهان ویژگی‌های خاصی داشته باشد، اما به لحاظ متافیزیکی ناممکن است که جهان چنین باشد. همچنین بعید نیست و بسیار چنین رخ می‌دهد که فرد ویژگی‌های ممکن متافیزیکی را ناممکن فرض کند. برای مثال، روزگار پیش از اختراع هواپیما را در نظر بگیرید. بسیاری در آن روزگار به احتمال فراوان تصور می‌کرده‌اند که در هیچ جهان یا شرایط ممکن انسان‌ها نمی‌توانند بر فراز ابرها حرکت کنند، اما اگر چه به لحاظ معرفتی و از دید ناظری که قوه تصور دارد، چنین چیزی ناممکن بود، اما به لحاظ متافیزیکی ناممکن نبود و ما در جهانی ممکن تولد یافتیم که شاهد چنین تحقیقی بوده است.

بدین ترتیب درمی‌یابیم بسیاری از اموری که به لحاظ معرفتی (تصورپذیری) ممکن تلقی می‌شوند، به لحاظ متافیزیکی و در ظرف عالم خارج ناممکن اند. در واقع، تمایز مهمی میان امکان اصیل و

تصورپذیری برقرار است.^۱ حال برای این که در یابیم حقیقتاً ویژگی یا حالتی که برای تحقق در جهان، ممکن تصور می‌کنیم، ممکن است یا نه، یا آن چه ناممکن می‌دانیم، ناممکن است یا نه، نیاز به دلایلی داریم که در بسیاری موارد بر معرفت علمی و شواهد تجربی ما مبتنی است. درست است که تصورپذیری می‌تواند دریچه‌ای برای کشف ممکن‌ها و ناممکن‌های متافیزیکی باشد، اما در عین حال خطاپذیر است و می‌تواند اموری ممکن را ناممکن نشان دهد و یا بالعکس.

به نظر من، دو فیلسوف مزبور درباره امکان پیدایش هوش مصنوعی‌هایی که بدون این که بدنی انسانی داشته باشند، حالات روان‌شناختی یکسانی با انسان‌ها داشته باشند، دچار خلط یا درهم‌آمیزی این دو معنای امکان شده‌اند. مبتنی بر امکان معرفتی، می‌توان تصور کرد که ممکن است ربات‌هایی بدون این که بدن انسانی داشته باشند، حالات روان‌شناختی نظیر آگاهی‌پدیداری، اختیار، احساسات، درون‌نگری و... داشته باشند، اما با توجه به جهان‌های ممکن متافیزیکی که محدودتر از جهان‌های ممکن تصورپذیر هستند، درمی‌یابیم که چنین موجوداتی امکان متافیزیکی ندارند. به ادعای نگارنده، به لحاظ متافیزیکی ممکن نیست که موجودی انسان نباشد، اما ویژگی‌های روان‌شناختی انسانی داشته باشد.

در اینجا لازم است قدری بیشتر درباره نقد خود توضیح دهم. مقدمه دوم را دوباره مرور کنیم. این مقدمه می‌گفت که «هوش مصنوعی‌های ممکن وجود دارند که از جنبه‌های مربوط و دخیل در بحث، هیچ تفاوتی با انسان‌ها ندارند و حقوق آنها مانند حقوق انسان‌هاست» این مقدمه، به صراحت، نمی‌گوید که ما می‌توانیم تخیل^۲، فرض^۳ یا تصور^۴ کنیم که یک هوش مصنوعی فرضی یا خیالی، هیچ تفاوتی با انسان‌ها نداشته باشد و حقوقش مانند حقوق انسان‌ها باشد. اگر، به صراحت، چنین مطلبی را بیان کند که استدلال تنها ناظر به تخیل یا تصورپذیری است^۵، استدلال اگرچه درست است، اما از جذابیت خاصی برخوردار نخواهد بود. ما می‌توانیم تصور یا تخیل کنیم که ماه به قورباغه تبدیل شود، خورشید ریش درآورد و پی‌اچ‌دی بگیرد و یک خرس مریخ را به جای صبحانه بخورد، اما این که آیا واقعاً چنین چیزهای ممکن است یا ما به اشتباه، توهم امکان آنها را مساوی با امکان واقعی و متافیزیکی آنها فرض کرده‌ایم،

۱. فاین (Fine, 2002, p. 255) و نیز ویلسون (Wilson, 2013, p. 657) نیز ضرورت (و به همین قرار امکان) متافیزیکی را و نه ضرورت معرفتی (و بدین قرار امکان معرفتی) را گونه‌اصیل ضرورت می‌داند.

2. imagine

3. suppose

4. conceive

۵. بسیاری از فیلسوفان - برای مثال، هیوم (Hume, 1777/1975) - بر این باوراند که تخیل کردن و تصور کردن یکسانند، اما برخی مانند چالمرز، ادعا می‌کنند که تخیل فقط آن نوع از تصوراتی است که همراه با یک تصویر ذهنی باشد (Chalmers, 2002) و البته، این تمایز در بحث ما دخیل نیست.

در گرو اقامه دلایل و توجه به واقعیات جهان است. درک امکان به معنای تصورپذیری وابسته به قدرت ذهن ماست. حال آن‌که امکان به معنای متافیزیکی وابسته به ساختار واقعی جهان است. این استدلال نیز به همین طریق، دقیقاً این نکته را که منظور از «ممکن» چیست، بیان نمی‌کند.^۱ اساساً نیازی نیست برای اثبات این‌که ما قدرت چنین تصور و تخیلی را داریم، استدلالی اقامه کرد. روشن است که ما می‌توانیم حتی تخیل‌ها و تصورات عجیب‌تری درباره آن ربات‌های فرضی داشته باشیم. هیوم می‌گفت که هیچ چیزی آزادتر از تخیل بشر نیست (Hume, David, 1777/1975, p. 47).^۱ اساساً تخیل انسانی در بند واقعیت نیست و یا دست‌کم دقیقاً همان گونه‌ای که ادراکات حسی و باورها در گرو محتوای باز نمودی واقعی خود هستند، در بند واقعیت نیستند. استدلال برای این‌که از جذابیت لازم برخوردار باشد، باید نشان دهد واقعاً (و نه تنها در عالم خیال و تصور) وجود چنین ربات‌هایی محال نیست و ممکن است. اگر استدلال مزبور، در مقدمه ۲، به قید «امکان معرفتی» یا «امکان به لحاظ تصورپذیری» صراحتاً اشاره می‌کرد، آن‌گاه ما از دو فیلسوف مدافع استدلال می‌پرسیدیم که چگونه می‌توانید توضیح دهید که چیزی که تصور یا تخیل کرده‌اید، محال نیست و واقعاً هم امکان به معنای متافیزیکی دارد؟ چگونه می‌توانید اثبات کنید که یک ناممکن را ممکن فرض نکرده‌اید؟ کدام واقعیت تجربی یا غیرتجربی در جهان می‌تواند نشان دهد که درک چنین امکانی مبتنی بر یک توهم یا خلط نیست؟

همان‌طور که پیش‌تر گفته شد، تصورپذیری و تصور جهان‌های ممکن بسیار خطا پذیر است. در عالم خیال و تصور، بسیاری از ممکنات واقعی، ناممکن کرده می‌شود و بسیاری از محال‌های واقعی، ممکن در نظر گرفته می‌شود. ما برای این‌که استدلال جذابی داشته باشیم، باید درباره یک سری ربات واقعاً ممکن نه تنها تصورپذیر استدلال اقامه کنیم و این دو فیلسوف با توجه به کاربرد مبهم یا دست‌کم فاقد قید «ممکن» دچار چنین مشکلی شده‌اند.

من از ادعای خود با توضیحی مربوط به هر کدام از دو رویکرد شایع در فلسفه ذهن، یعنی رویکرد فیزیکیالیستی - مادی‌انگاره و نیز رویکرد دوگانه‌انگاره دفاع می‌کنم. در صورتی که یک مادی‌انگار درباره ذهن (به عنوان یک جوهر^۲) یا حالات و رویدادهای ذهنی (به عنوان تمثیل و ویژگی‌ها^۳) باشیم، انسان را یک اندام‌واره یا بدن خاص با سطحی از پیچیدگی یا رشد زیست‌شناختی و فیزیکی خاص خواهیم دانست که هر وقت چنین پیچیدگی یا رشدی حاصل شد، تحقق حالات و رویدادهای ذهنی نیز ممکن و بالفعل خواهد شد. در صورتی که یک دوگانه‌انگار درباره ذهن یا حالات و رویدادهای ذهنی

۱. شویتسگیل در پاسخ به نقد من در یک ایمیل تلویحاً پذیرفته است که مقاله نیاز به اصلاحیه (amendment) دارد و باید به بحث امکان متافیزیکی پرداخته و دست‌کم تصریح شود.

2. substance

3. mental events and states as instantiations of properties

با شیم، انسان را مرکب از دو جوهر مادی (بدن) و مجرد (نفس یا ذهن) یا تنها جوهری مادی (جسمی خاص، یعنی بدن انسانی) که دارای ویژگی‌های ذهنی (و احتمالاً تحویل‌ناپذیر به ویژگی‌های فیزیکی) است، خواهیم دانست که هر وقت چنین بدنی همراه با آن جوهر یا ویژگی‌های مجرد یا غیرفیزیکی خاص محقق شد، ویژگی‌های روان‌شناختی نیز ممکن و بالفعل خواهد شد. هیچ یک از این تقریرهای گوناگون مادی‌انگاری و دوگانه‌انگاری که همگی در میان فیلسوفان طرفداران فراوانی دارند، نخواهند پذیرفت که بدون این که موجودی انسان باشد، صرف‌نظر از این که انسان همین بدن است یا چیزی فراتر از آن و غیرمادی نیز در آن دخیل است، بتواند ویژگی‌های روان‌شناختی انسانی را داشته باشد. طبق هر دو نگره مادی‌انگارانه و دوگانه‌انگارانه ذهن و ویژگی‌های ذهنی به بدن و ویژگی‌های بدنی انسان وابسته‌اند. آموزه وابستگی وجودشناختی^۱ مستلزم هیچ یک از دو نگره نیست و با هر دو سازگار است.

همچنین می‌توانم از نقد خود به گونه‌ای دیگر نیز دفاع کنم. همان‌طور که واژگان انواع طبیعی همانند واژه «آب» در هر جهان ممکن متافیزیکی به یک گروه خاص از موجودات یعنی مصادیق H_2O دلالت دارد، هیچ موجودی نمی‌تواند همه ویژگی‌های آب (خود آب نه شکل‌واژه یا مصوت «آب») را در واقع، داشته باشد، جز آن که آب باشد. به همین ترتیب، هیچ موجودی در هیچ جهان ممکن متافیزیکی نمی‌تواند همه ویژگی‌های انسانی روان‌شناختی انسان را داشته باشد، مگر این که انسان باشد.

پس بدین قرار، با توجه به دو تقریری که از نقد خود ارائه کردم، روشن است که ربات‌ها صرف‌نظر از تصورپذیری، به لحاظ متافیزیکی ممکن نیست که در تمام ویژگی‌های ذهنی انسانی همانند انسان باشند، مگر این که انسان باشند، یعنی بدن انسانی داشته باشند. گفتیم که طبق هر دو فرض گونه‌ای وابستگی میان ذهن یا ویژگی‌های ذهنی و بدن وجود دارد و موجودی نمی‌تواند با بدنی غیرانسانی انسان باشد. البته، اگر دو فیلسوف مزبور بخواهند از این سناریوی فرضی دفاع کنند که ربات‌ها به تدریج با تکامل دانش بشر، بدن‌هایی انسانی پیدا کنند، در این صورت، روشن است که عنوان یا نام «هوش مصنوعی» بر آنها صدق نخواهد کرد و شاهد چیزی شبیه حیوانات تولید شده در آزمایشگاه که در روزگار ما امکان تحقق یافته خواهیم بود. در آن صورت، آن ربات‌ها، انسان خواهند بود، چون بدن انسانی دارند و در ویژگی‌های روان‌شناختی با ما اشتراک دارند و هیچ مشکلی نیست که آنها را دارای حقوق و جایگاه اخلاقی مشابه با انسان‌ها فرض کنیم.

صرف‌نظر از این که به لحاظ علمی امکان دارد که چنین ربات‌هایی تولید شوند، به لحاظ متافیزیکی ممکن است که موجوداتی بدن انسانی داشته باشند و ویژگی‌های ذهنی انسانی نیز داشته باشند، حال چه ما با مسامحه آنها را «هوش مصنوعی» بنامیم و چه با دقت زبانی و مفهومی، آنها را انسان بدانیم.

1. ontological dependence

بدین ترتیب، به نظر من استدلال عدم تفاوت مربوط نامعتبر است، چون یا چیزی را که امکان متافیزیکی ندارد، ممکن به لحاظ متافیزیکی می‌داند، حال آن‌که می‌دانیم صرف تصورپذیری برای موفق شدن استدلال کافی نیست و یا این‌که واژه «هوش مصنوعی» را به صورت مبهم به کار می‌برد، به صورتی که می‌تواند شامل انسان‌ها و ربات‌هایی که کاملاً بدنی انسانی خواهند داشت، نیز بشود. هر دو مشکل خللی جدی در استدلال ایجاد می‌کند و مستلزم رد یا پالایش و تجدید نظر در استدلال است.

نقد ششم: لازمه استدلال، پذیرش تغییر جوهری است

به نظر من، یکی از لوازم این استدلال این است که بپذیریم ممکن است تغییر جوهری^۱ در هوش مصنوعی رخ بدهد؛ حال آن‌که چنین لازمه‌ای به دو دلیل قابل قبول نیست. نخست این‌که بر اساس بسیاری از مبانی و رویکردهای فلسفی، پذیرش تغییر جوهری بی‌دلیل است. دوم این‌که حتی اگر بپذیریم در برخی موارد چنین تغییری ممکن است، دلیلی نداریم که بپذیریم در مورد هوش مصنوعی چنین امری امکان دارد.

در توضیح بیشتر نقد باید بگویم که روشن است که استدلال ادعا نمی‌کند که اکنون به صورت بالفعل هوش مصنوعی‌های هم‌تراز انسان به لحاظ ویژگی‌های روان‌شناختی و اجتماعی وجود دارند. در واقع، در این تردیدی نیست که هوش مصنوعی‌های فعلی از جهات دخیل در بحث مثل انسان نیستند، بلکه تنها چنین ادعا می‌شود که ممکن است که در آینده چنین هوش مصنوعی‌های مشابه انسانی ساخته شوند. صرف نظر از این‌که امکان مزبور در این ادعا معرفتی است یا متافیزیکی که در نقد پنجم به این بحث پرداختم، گونه‌ای تبدل یا تغییر جوهری باید ممکن باشد.

ربات‌های فعلی به لحاظ ویژگی‌های روان‌شناختی مربوطه با انسان‌ها یکسان نیستند، اما برای این‌که در آینده با انسان‌ها یکسان شوند، باید در لحظه‌ای خاص تغییری جوهری در ذات یا ویژگی‌های ذاتی آنها رخ دهد. ما نه تنها برای پذیرش چنین تغییری جوهری دلیلی نداریم، بلکه فکر می‌کنیم ممکن نیست چنین تغییری جوهری در ربات‌ها رخ دهد. آن فیلدسوفانی که تغییر جوهری را در مرحله‌ای خاص از رشد انسانی (مثلاً در جهش و تغییر از وضعیت طبیعی غیرآگاهانه یک سلول به وضعیت آگاهانه یک شخص) ممکن می‌دانند، چنین تغییری را از آنجا می‌پذیرند که ویژگی‌های نوظهور تنها بر اساس ویژگی‌های پیشین طبیعی را تبیین‌پذیر نمی‌دانند.

در واقع، انگار احساس یا دست‌کم ادعا می‌شود در مواجهه با مجموعه‌ای از درشت‌مولکول‌ها (مثل

1. substantial change

پروتئین) که بعداً تبدیل به اسپرم و بعد انسان می‌شوند و هر کدام ویژگی‌های جداگانه و منحصر به فردی دارند، با چند ذات یا ویژگی ذاتی گوناگون رو به رو هستیم و نیاز به پذیرش فرض تغییر جوهری وجود دارد، اما دربارهٔ ربات‌های ممکن مد نظر این دو فیلسوف، چنین نیازی وجود نخواهد داشت.

از آنجا که ربات‌ها برنامه‌ریزی و کدگذاری می‌شوند و در کل توسط ما ساخته می‌شوند، تمام ویژگی‌های آنها بر اساس برنامه‌ها، کدها و طرح‌ها قابل محاسبه و تبیین است و نیازی به پذیرش این فرض نیست که در ربات‌ها از مرحله‌ای به بعد تغییری جوهری و تبیین‌ناپذیر بر اساس ویژگی‌های قبلی اتفاق خواهد افتاد. البته، اگر کسی چنین فرض کند که در مورد هوش مصنوعی، ممکن است بر خلاف طرح پیشین و ساختار مهندسی شدهٔ قبلی ویژگی‌های روان‌شناختی پیش‌بینی‌ناپذیری ظهور کنند، باید نشان دهد که چگونه چنین چیزی ممکن است. یعنی باید توضیح دهد که ربات‌هایی که بدن انسانی ندارند، چگونه ممکن است که در مرحله‌ای از ساخت خود، ناگهان دارای ویژگی‌های روان‌شناختی خاصی شوند که اساساً بر مبنای زیرساخت‌های فیزیکی آنها قابل تبیین نیست.

البته، پیش فرض نقد من این است که بپذیریم دربارهٔ انسان‌های ربات‌نما سخن نمی‌گوییم، بلکه دربارهٔ ربات‌ها و در کل هوش مصنوعی‌های واقعی سخن می‌گوییم. انسان‌های ربات‌نما و ربات‌های انسان‌نما هویات احتمالاً ممکن هستند که در آینده از ترکیب یا پیوند بخش‌هایی از بدن انسان و ربات ساخته می‌شوند. چنین هویاتی محل بحث ما نیست. اگر مرز میان هوش مصنوعی و هوش طبیعی و انسانی را مرزی واقعی و دقیق ندانیم، در واقع، در مقدمهٔ ۲ استدلال ادعا کرده‌ایم که «انسان‌های ممکن وجود دارند که ما (با ابهام و عدم دقت) آنها را «هوش مصنوعی» می‌نامیم و از جنبه‌های مربوط و دخیل در بحث (مثل ویژگی‌های روان‌شناختی)، هیچ تفاوتی با انسان‌ها ندارند». این مقدمه به بیان ساده‌تر یعنی «انسان‌های ممکن وجود دارند که ویژگی‌های انسانی دارند!» روشن است که چنین مقدمه‌ای بی‌مایه و شبه‌بدییه^۱ و نااطلاع‌بخش^۲ است و نتیجهٔ استدلال را نیز بی‌اثر و بی‌اهمیت خواهد ساخت. پس در واقع، بر اساس نقد ششم من، دو فیلسوف مزبور، یا مرز میان هوش مصنوعی و هوش طبیعی انسانی را مرزی واقعی می‌دانند که باید نشان دهند چرا در آینده چنین تبدل و تغییر جوهری رخ خواهد داد و یا این‌که چنین مرزی را نمی‌پذیرند که استدلال آنها نااطلاع‌بخش خواهد بود.

1. trivial

2. uninformative

نتیجه

من پس از بیان نکاتی مقدماتی به معرفی و تحلیل استدلال پرداختم. استدلال همراه با توضیحات دیگر در پی این است که مشابهت یا یکسانی جایگاه اخلاقی برخی هوش مصنوعی‌های ممکن با انسان را نشان دهد. سپس چهار نقد بیان شده در مقاله اصلی را مرور کردم. در پایان نیز خود دو نقد جدید طراحی و ارائه کردم. به نظر می‌رسد با توجه به دو نقد من، یعنی نقدهای ناظر به امکان معرفتی و تغییر جوهری، استدلال شویتسگبل قانع‌کننده نیست و در بهترین حالت، نیاز به اصلاح و ارائه تقریرهای جدید هست. بر اساس نقد امکان معرفتی، از آنجا که واژه ممکن به صورت مبهم در استدلال به کار رفته، نیاز به تقیید دارد. امکان معرفتی برای حصول نتیجه کافی نیست و امکان متافیزیکی نیز در چارچوب فعلی استدلال قابل دفاع نیست و نیاز به دلایل جداگانه دارد.

بر اساس نقد تغییر جوهری، مدافعان استدلال یا مرز میان هوش مصنوعی و هوش طبیعی انسانی را مرزی واقعی می‌دانند که باید نشان دهند چرا در آینده چنین تبدیل و تغییر جوهری میان هوش مصنوعی و هوش طبیعی رخ خواهد داد و نیاز به دلیل جداگانه دارد و یا این که چنین مرزی را نمی‌پذیرند که خروجی استدلال آنها نااطلاع‌بخش و شبه‌بدیهی خواهد بود.

References

- Bostrom, N. (2003). Are We Living in a Computer Simulation? *Philosophical Quarterly*, 53, 243-55. doi: 10.1111/1467-9213.00309.
- Chalmers, D. (2002). Does Conceivability Entail Possibility? in T. S. Gendler & J. Hawthorne (Eds.) *Conceivability and Possibility*, (pp. 145-200). Oxford: Oxford University Press.
- Egan, G. (1994). *Permutation City*. London: Millennium.
- Egan, G. (1997). *Diaspora*. London: Millennium.
- Fine, K. (2002). The varieties of necessity. T. S. Gendler & J. Hawthorne (Eds.). *Conceivability and possibility*. (pp. 253-281). Oxford: Oxford University Press.
- Hankins, P. (2015). Crimbots. Blogpost at Conscious Entities. <http://www.consciousentities.com/?p=1851>. February 2, 2015.
- Hobbes, T. (1651/1996). *Leviathan*. (ed.) R. Tuck. Cambridge: Cambridge University Press.
- Hume, D. (1777/1975). *Enquiries Concerning Human Understanding and Concerning the Principle of Morals*. eds. L. A. Selby-Bigge & P. H. Nidditch, Oxford: Oxford University Press.
- Kripke, S. (1980). Naming and necessity. Cambridge: Harvard University Press.
- Putnam, H. (1960). Minds and Machines. *Journal of Symbolic Logic*, 57-80. doi: 10.2307/2271581.
- Putnam, H. (1964). Robots: Machines or Artificially Created Life?, *Journal of Philosophy*, 61 (21), 668-691. doi: 10.2307/2023045.
- Putnam, H. (1967). The Mental Life of Some Machines. In H. Castaneda (ed.), *Intentionality, Minds and Perception*. (pp. 177-200). Wayne State University Press.
- Schwitzgebel, E. & Garza, M. (2015). A defense of the Rights of Artificial Intelligences. *Midwest Studies in Philosophy*, 39, 98-119. doi: 10.1111/misp.12032.
- Searle, J. (1980). Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3, 417-457. doi: 10.1016/B978-1-4832-1446-7.50007-8.
- Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press
- Singer, P. (1975/2002). *Animal Liberation*. New York: Ecco Press.
- Singer, P. (2009). Speciesism and Moral Status. *Metaphilosophy*, 40, 567-581. doi: 10.1111/j.1467-9973.2009.01608.x.
- Soames, S. (2011). Kripke on epistemic and metaphysical possibility: Two routes to the necessary a posteriori. In A. Berger (Ed.), *Saul Kripke*. (pp.167-188). Cambridge: Cambridge University Press.
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59, 433-460. doi: 10.1093/mind/LIX.236.433.
- Wilson, A. (2013). Schaffer on laws of nature. *Philosophical Studies*, 164, 653-67. doi: 10.1007/s11098-012-9878-7.