Philosophy & Technology (2022) 35: 4
https://doi.org/10.1007/s13347-022-00498-3

RESEARCH ARTICLE

Images of Artificial Intelligence: a Blind Spot in AI Ethics

Alberto Romele[1]

Received: 21 June 2021 / Accepted: 11 January 2022 / Published online: 29 January 2022

Abstract

This paper argues that the AI ethics has generally neglected the issues related to the science communication of AI. In particular, the article focuses on visual communication about AI and, more specifically, on the use of certain stock images in science communication about AI — in particular, those characterized by an excessive use of blue color and recurrent subjects, such as androgyne faces, half-flesh and half-circuit brains, and variations on Michelangelo's *The Creation of Adam*. In the first section, the author refers to a "referentialist" ethics of science communication for an ethical assessment of these images. From this perspective, these images are unethical. While the ethics of science communication generally promotes virtues like modesty and humility, similar images are arrogant and overconfident. In the second section, the author uses French philosopher Jacques Rancière's concepts of "distribution of the sensible," "disagreement," and "pensive image." Rancière's thought paves the way to a deeper critique of these images of AI. The problem with similar images is not their lack of reference to the "things themselves." It rather lies in the way they stifle any possible forms of disagreement about AI. However, the author argues that stock images and other popular images of AI are not a problem per se, and they can also be a resource. This depends on the real possibility for these images to support forms of pensiveness. In the conclusion, the question is asked whether the kind of ethics or politics of AI images proposed in this article can be applied to AI ethics tout court.

Keywords AI images · Stock images · AI ethics · Ethics of scientific communication · Jacques Rancière

This article is part of the Topical Collection on *Philosophy of Technology and the French Thought*

✉ Alberto Romele
alberto.romele@uni-tuebingen.de

[1] IZEW, Universität Tübingen, Tübingen, Germany

agreements with stock imagery agencies, with the result that these institution's and company's communication services routinely use stock imagery to represent emerging technologies such as AI.[4]

The images we want to deal with have been for long time dismissed, both by "hard" and "soft" sciences, as mere fantasies. However, they are invasive in the current imagery of AI, and for this reason, they deserve to be questioned.

The article is structured in two sections. In the first section, we present two cases of stock images of AI in science communication, and we apply to them a standard ethics of science communication. The output is a foregone conclusion: from the perspective of this ethics, which is oriented by a form of "referentialism," these images of AI are unethical. Similar stock images of AI do not "humbly" represent the "things themselves"; they let more than what they are supposed to show be seen (certainly more than what is concretely done in technological innovation in AI). At this point, two choices seem possible. The first one consists of criticizing the use of similar images and inviting scholars, and, more generally, all stakeholders involved in science communication about technological innovation in AI, to be more cautious with how they visually represent AI. The second one, which we believe to be more interesting, consists in accepting that AI is difficult to visually represent, but is represented nevertheless. Similar images cannot simply be dismissed, because they are produced continuously and so occupy an important part of AI's present imagery. In the second section, we engage with the reflections on aesthetics, politics, and images developed by the French philosopher Jacques Rancière, in particular the notions of "distribution of the sensible" (Rancière, 2004), "disagreement" (Rancière, 1999), and "pensive image" (Rancière, 2009a). We contend that Rancière's perspective offers the possibility of a different critique of stock images and other popular images of AI. From his perspective, similar images of AI are problematic not because they are "unethical" but rather because they are "unpolitical." The problem with them does not lie in their lack of reference to the "things themselves." It rather lies in the way they mark a gap between experts and non-experts, insiders and outsiders; it also lies in their incapacity to promote forms of disagreement among concerned groups beyond a simplistic logic of oppositions—goodness/badness, risk/opportunity, humans/nonhumans,[5] etc. We contend as well that Rancière's perspective offers the possibility to think of these images beyond their criticisms, that is, not only as a danger but also as potential resources. This depends on the concrete possibility for these images to support, rather than stifle, forms of pensiveness. In the conclusion,

[4] See *Research*eu*, the monthly magazine of CORDIS, European Commission's primary source of results from the projects funded by the EU's framework programs for research and innovation. The magazine, as well as CORDIS' website, makes abundant use of images of science and technology retrieved from Shutterstock. https://cordis.europa.eu/research-eu/en. Accessed December 1st, 2021.

[5] In an article devoted to the visual representations of data centers, Taylor (2019) has theorized the notion of "technological wilderness." According to him, what characterizes these images is the absence of human beings. This corresponds to a representational strategy related to "to emic and etic fantasies and futures of human-free security, automation and data objectivity" (Taylor 2018, 3). Interestingly enough, popular imagery of AI is usually characterized by the presence of both humans and machines, most often represented in terms of transition from one to the other.

we briefly ask ourselves if the kind of ethics (or politics) of AI images we propose in this article can be applied to AI ethics tout court.

2 The Unethics of AI Images

Type "Artificial Intelligence" in a web browser and look for images: Among the results, you will see unreal holographic interfaces, half-flesh half-circuit brains, lines of code waving in space, robots tapping on smart touchscreens, and at least one of the hundred variations of Michelangelo's *The Creation of Adam* in a human–robot version. Most of these images are stock images. What usually characterizes stock images is their clichéd way of representing aspects of reality. Stock images have been mocked for this, for instance where women are pictured laughing alone eating salad[6] or seem unable to drink water from a bottle or glass.[7] In the case of AI and other emerging technologies, stock images have the tendency to be overly "unrealistic" and "hyperbolic." A limited group of scholars in media studies has undertaken analyses of stock images and their social consequences, without, however, focusing on the images of science and technology or specifically on AI—see, in particular, Frosh (2003, 2020) and Turlow, Aiello, and Portmann (2019). It has been observed that while stock images are generally dismissed as the "wallpaper" of consumer culture, they are also "central to the ambient image environment that defines our visual world" (Aiello, 2016, np).

Stock images of AI have not only invaded the popular Web. They are widely used, both online and offline, to communicate about events, publications, courses, etc., on AI proposed and organized by scientific institutions that are often considered to be leading in the field of AI research (be it in engineering or in social sciences and humanities). In this regard, without any claim of exhaustiveness, we started to collect stock images of AI used in science communication and marketing through an Instagram profile called "ugly.ai."[8] We collected from the profile over eight months (May–December 2021), collecting more than a hundred images. From these images, we choose in this article to focus on two images, both relating to the field of AI ethics. It is indeed interesting to note that AI ethics itself sometimes shows little attention to the ethical implications of visual representations of AI. It is important to stress that it is not our intention to offer a detailed analysis of these images. In fact, this article has a theoretical intent. Also, it is important to highlight that we do not want to criticize the use of images when communicating about AI or AI ethics in general, nor the use of stock images as such. Rather, we want to problematize the use, however abundant, of a certain type of stock image, which is characterized by some common traits, for example: colors (mainly blue), subjects (robots,

[6] https://www.thehairpin.com/2011/01/women-laughing-alone-with-salad/. Accessed December 1st, 2021. The example is retrieved from Aiello and Woodhouse (2016).

[7] https://www.boredpanda.com/women-dont-know-how-to-drink-water-stock-photos/?utm_source=google&utm_medium=organic&utm_campaign=organic. Accessed December 1st, 2021.

[8] https://www.instagram.com/ugly.ai/. Accessed December 1st, 2021.

The fact is that in both cases, the use of narrative has the sole function of paving the way to a dynamic that is entirely internal to science and technology. The same holds for the second consideration, which is about the appropriate level of accuracy (strict or relaxed) to maintain in the use of narrative in science communication. Once again, there is no recognition of narrative per se: its ethical value is always measured on the basis of its "accuracy," that is, its capacity to properly refer to the things themselves. Finally, the last consideration is about the possibility of not using narrative at all, which means that narrative in science communication is somehow reduced to a sometimes necessary, but always unpleasant, stratagem to realize the scopes of science. With theological terminology, we could say that the logic of the use of narrative in science communication is a logic of *kenosis*, which in Ancient Greek means "self-emptying."

Images such as images 1 and 2 follow an opposite logic. They are not humble, honest, sincere, or transparent. Rather, they are arrogant, and overconfident. In sum, they are not "accurate." They indicate more scientific progress than they should, certainly more than actually exists in current science and technology. No human head/brain/mind has ever been turned into, and probably will never be, "digital particles"; the robotic hand depicted in Image 2 is a fantasy: whoever has visited a prosthetic center, or even a scientific laboratory working on upper-limb prostheses, knows that the status of research and innovation in the field is very different. Not to mention the transparent touch screen, which is very different from the screens we deal with in our everyday lives. According to the "referentialist" framework proposed by Dahlstrom and Ho, these images are simply unethical.

It must be acknowledged that representing AI to a non-expert audience is a real challenge. Especially when it comes to showing not only the technology itself, but also its social and cultural implications. In this regard, one could distinguish three levels of visual representation: (1) The first is the one that wants to be closer to the "thing itself," that is, the algorithm. Think of the representation of a decision tree learning, of a network of artificial neurons, or the way the algorithm is encoded in a computer program. Yet, not only might one wonder if such representations really show the "thing itself." One could say that such a representation does not take into account AI as a social and cultural phenomenon; (2) The second is the one that represents AI as being embedded in different technologies (drones, smartphones, mechanical arms, etc.) and specific contexts (agriculture, medicine, military actions, etc.). In this case, however, AI is clearly black-boxed into another technology. Moreover, such images are often already third-level images, for example, when they choose specific technical objects (in particular, humanoid robots), or when they "augment" existing technologies (e.g., by adding elements that come out three-dimensionally from the screen of a smartphone), or even when they place simple objects against backgrounds (e.g., sunsets or particularly clear skies) that instill feelings of hope or fear; (3) Finally, the third level is that of the images we consider in this article. From a referentialist point of view, these are definitely the worst ones. The fact is that they do not so much refer to the "thing themselves" as they do to expectations and imaginaries, whether those of engineers, organizations, and companies, or potential spectators. Each of these levels, we believe, is legitimate in its own way, but under certain conditions. Therefore, it is important to emphasize that

between yes and no, positioned at their point of intersection. It marks its distance both from the paradigm of crisis and that of critique" (Esposito 2018, p. 16). In either case, he says, one ends up in an impasse of thought. The Italian Thought — Esposito thinks of a tradition in political philosophy going from Machiavelli to Agamben — would instead be able to avoid this impasse, being an "affirmative thought": "it can be argued that, by and large, the main effort of Italian philosophers has been to think not in a reactive but in an active, productive, affirmative way" (Esposito 2018 p. 17). Following this distinction (admittedly simplistic in many ways, but nonetheless useful), we assert that Rancière's thought is properly a representative of the French *Thought* because, while embracing a certain critical, and even neutral, attitude of French theorists (evident especially in the notion of "distribution of the sensible"), he also embraces the affirmative attitude of Italian thinkers, as emerges especially from the concepts of "disagreement" and "pensive image." For instance, as will be shown in the conclusion, disagreement does not coincide with mere chaos, but rather with the concrete possibility of thinking and building a new form of (technological) democracy.

Our first hypothesis is that AI images like Images 1 and 2 are "unpolitical" because they contribute to the framing of a specific "distribution of the sensible" in the technological innovation in AI. For Rancière (2004, p. 12), the expression indicates "the system of self-evident facts of sense-perception that simultaneously disclose the existence of something in common and the delimitations that define the respective parts and positions within it." In other words, the distribution of the sensible regards the constitution of a shared time, space, and horizon of understanding, and the distribution of access and roles (that is, recognition, legitimacy, and ultimately power) within such a delimited space, time, and horizon of understanding. The distribution of the sensible, and the consequent distribution of access and roles, imply exclusions, sometimes from specific access and roles, sometimes from the whole space, time, and horizon of understanding. The distribution of the sensible is for Rancière a political practice, because "politics revolves around what is seen and what can be said about it, around who can see and the talent to speak, around the properties of space and the possibilities of time" (Rancière, 2004, p. 13). Politics and aesthetics are strongly connected, where "aesthetics" is to be understood both in the sense of the Greek *aisthesis*, which means "perception," and in the sense of art and cultural productions in general. On the one hand, politics is a matter of distribution (or exclusion from) roles and access to perception—seeing/being seen, listening/being listened to, etc. On the other hand, art and cultural productions can either contribute to the reproductions of the dominant regimes of perception or contribute to their suspension and eventual transformation.[16]

[16] In what concerns the link between art and politics, see Rancière (2009b). For him, there is a continuity between authentic art and politics insofar as authentic art, as well as politics, represent the possibility to suspend the ordinary forms of the sensible experience, that is, the ordinary distribution of the sensible. In Rancière's words (2009b, p. 25–26. Translation modified. Italics is our), "art and politics do not constitute two permanent separate realities […]. They are two *suspended* forms of distribution of the sensible […]." The English translation does not include the adjective "suspended," hence radically distorting the meaning of the sentence here. Indeed, for Rancière, what characterizes both authentic art and politics is the possibility to suspend, that is, to offer alternatives to, the dominant regime of the sensible exercised by the police and of which what is called and legitimized as art is mostly an expression.

We contend that the dominant imagery of AI implies a specific distribution of the sensible whose ultimate effect is to mark a gap between experts and non-experts, insiders and outsiders. It has been argued that the use of images in science popularization has an introductory function. For instance, Gigante (2018) coined the term "portal images." However, we contend that stock images of AI in science communication are "screen images," where "screen" refers to its etymology, meaning "to cut, divide, cover, shelter, and separate." The fact is that one can watch thousands of similar images of AI without having to develop any critical reasoning about AI. These images instead have an "anesthetic" effect, which means that the reiterated contact with them makes non-experts and outsiders less and less sensitive to the most urgent issues related to AI and increases their feelings of resignation about AI.

We propose to apply these considerations to our object of study. In particular, we introduce the notion of "an*ae*sthetics," a word referring to the fact that the distribution of the sensible related to similar images (aesthetics) has anesthetic effects on those who are "outside." The concept of an*ae*sthetics is also important for another reason. One might think that the loss in terms of both ethics and politics at the level of the single image of AI is somehow retrieved at the level of the context in which the image is used, and to which it finally belongs. Hence, a possible criticism of our discourse might consist in affirming that there is no ethics or politics of similar images per se, because similar images are always used in context, and the ethical or political assessment should be made not on the single image, but with regard to the whole context. To put it plainly, science communication on AI is full of ugly and bad images, yet these images can still be used ethically or politically whenever they are integrated into a rigorous discourse. However, such criticism not only forgets that in the media environment in which we live, images are most often detached from, and perceived outside from, their context. Think of how often we content ourselves with scrolling the home screen of our news feeds without actually reading the article or even the titles. This criticism also forgets that similar images can, through their "force," anesthetize the communicational context in which they are supposed to be embedded and on which they are supposed to depend.[17]

Our second hypothesis is that stock images of AI are also unpolitical because they impede or anesthetize any form of "disagreement." Above, we have argued that politics has to do with the distribution of the sensible. However, on other occasions, Rancière proposes distinguishing more carefully between politics and police. We might say that the distribution of the sensible as a form of domination is related to

[17] One of the anonymous reviewers stated that "according to certain theories of interpretation (see the intentionalism of Quentin Skinner), authorial intention is what matters." For this reason, space should be given to the author's reasons — in this case, either the author of the images or the one who chose them. To this objection, one can reply that there are many other theories of interpretation that argue that authorial intentions do not matter or that it would not be fair to take them into account. Once produced, a text, an image, or the articulation of the two is instead to be considered as autonomous both in its contents and in its effects. We are referring, in particular, to authors such as Gadamer and Ricoeur. The latter writes, for example: "Dialogue is an exchange of questions and answers; there is no exchange of this sort between the writer and the reader. Rather, the book divides the act of writing and reading into two sides, between which there is no communication. The reader is absent from the act of writing; the writer is absent from the act of reading" (Ricoeur 1991, p. 107).

of techniques and technologies—we do not know how the black boxes move. It rather refers to AI as an imaginary, which is, however, not anesthetized according to the easy opposition between fear and hope. *Black Box* inspires both fascination and uncanniness, attraction, and repulsion. The black boxes move, they behave and seem alive, and yet they cannot be understood. A second example is the *Anatomy of an AI System* by Crawford and Joler,[27] whose goal is to present Amazon Echo as "an anatomical map of human labor, data and planetary resources." We believe that this map can be approached from two different levels. The first one is the level of representativeness. For instance, one can download and *read* the map in its details to have a better understanding of AI not in isolation, but rather in its multiple human and environmental implications. The second other one consists of *perceiving* the map as a whole. In this second case, the spectator is taken by a kind of vertigo, given the complexity and the many dimensions that are suggested by the opening of the AI black box — like the opening of a human body and the arrangement of all its internal organs. The effect, after all, is not unlike that of the *Black Box*. Certainly, this latter work extremizes opacity, while the other one extremizes "monstration." Yet, in both cases, it is a matter of problematizing AI and our daily relationship with it.

We believe that the main challenge for the ethics of AI images would consist of going beyond the limits of the artistic (and hence most often elitist) production to import the pensiveness of works like *Black Box* and the *Anatomy of an AI System* in more popular contexts, in particular in the context of the production of stock images about AI, and science and technology in general.

4 Conclusion

In this paper, we have argued that there is a blind spot in the current debate about the ethics of AI. This blind spot consists of ignoring the ethical issues related to science communication about AI. In particular, we have focused on visual communication, and even more specifically on the use of certain stock images of AI. In the first section, we have referred to Dahlstrom and Ho (2012), who investigated the ethical implications of using narrative to communicate science, with a view to making an ethical assessment of the dominant imagery in science communication about AI. The result has been a foregone conclusion: similar images are unethical. While the ethics of science communication generally promotes the practice of virtues like modesty, humility, sincerity, transparency, openness, honesty, and generosity, stock images and other popular visual representations of AI are arrogant, pompous, and overconfident. In this section, we have also sketched the outlines of a general theory of visual representability of AI — which is today mostly identified with machine learning algorithms. We have distinguished between (1) the possibility of representing the algorithm itself; (2) the depiction of those technologies (drones, autonomous vehicles, etc.) in which AI is embedded; (3) the images, like those considered in this article, that focus on the expectations, fears, and hopes about AI. Our idea is that (1)

[27] https://anatomyof.ai/. Accessed December 1st, 2021.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH ("Springer Nature").

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users ("Users"), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use ("Terms"). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com