

Artificial intelligence and the challenge to the epistemology of creativity

Simona Chiodo*

ABSTRACT

I propose to reflect upon the relationship between artificial intelligence and creativity against the background of a possible epistemology of the latter. First, I introduce the notion of creativity by drawing at least on its essential cornerstones. Second, I introduce the notions of analysis, abstraction and idealisation as the foundation of a possible epistemology of creativity. Third, based on it, I question the relationship between creativity and artificial intelligence. Fourth, I propose to answer that one of the reasons why artificial intelligence cannot be creative is that it lacks the kind of epistemology of creativity described above as analysis, abstraction and, especially, idealisation.

KEYWORDS

Creativity, Epistemology of creativity, Artificial intelligence, Generative adversarial networks, *Edmond de Belamy*

1. *The essential cornerstones of the notion of creativity*

The necessarily limited space of a short article leads me to try to identify at least the cornerstones of the notion of creativity that are the most essential for my purpose, which is reflecting upon the relationship between artificial intelligence and creativity against the background of a possible epistemology of the latter.

Thus, I focus on what not only philosophers (in recent years, see at least Carruthers 2006, Stokes 2008, Garroni 2010 and Gaut 2010) but also psychologists (in recent years, see at least Amabile 1996, Csíkszentmihályi 1996, Mayer 1999 and Runco 2023) share most when it comes to defining creativity: there is creativity if there is the capability to produce something that is not only novel but also valuable. Indeed, according to Gaut's literature review, "[t] here is a broad consensus that creativity is the capacity to produce things that are original and valuable" (Gaut 2010, p. 1039). Being "original", i.e. novelty, means that we are creative if we can

* Politecnico di Milano, simona.chiodo@polimi.it

produce, for instance, a glass that is different from existing ones. Being “valuable” means that we are creative if we can produce, for instance, a glass that is better than existing ones – more precisely, we are creative if we can produce a glass that is novel precisely to be valuable, i.e. better (which is not the case of a glass whose novel, and even eccentric, shape makes our experience of it worse, from holding it worse with our hands to drinking from it worse with our mouths).

It goes without saying that there is a sense in which the *hendiads* of novelty and value dates back at least to Kant, who correlates the notion of genius with both the absence of “imitation” (Kant 1790, 5, 308) when it comes to novelty and the presence of “models, i.e. exemplary” (Kant 1790, 5, 308) artefacts, when it comes to value – which ultimately means that, again, there is creativity if there is the capability to produce something whose novelty is far from making it “original nonsense” (Kant 1790, 5, 308).

Interestingly enough, Garroni draws on Kant to define creativity as the capability to cross a kind of *no man’s land* extending from the universal rule, which is given, to its particular application, which is not given (see Garroni 2010). For instance, in the case of the glass, the universal rule may be the following: any glass is, first, an artefact for drinking liquids, second, suitable for human hands holding it and, third, suitable for human mouths drinking from it. And its particular application may be a glass with a round shape to fit both human hands and mouths – and we can go further and say that its (more) creative particular application may be a glass with a slightly oval shape to fit (better) both human hands and mouths.

Most of the authors, from psychologists to philosophers, also share the idea that, if it is true that creativity means novelty and value in the ways described above, it is also true that it requires producing agents capable of understanding, imagination and judgment, i.e. typically human capabilities.¹ Yet, in recent years, the rise of artificial intelligence has started making authors questioning the coincidence of the former with the latter. In other words, are we sure that creativity can be exclusively human? More precisely, are we sure that creativity cannot be artificial? For instance, according to Boden, “[t]he ultimate vindication of AI creativity would be a programme that generated novel ideas that initially perplexed or even repelled us, but which was able to persuade us that they were

¹ I cannot develop the relationship between creativity, on the one hand, and understanding, imagination and judgment, on the other hand, in the necessarily limited space of a short article. But I worked on it in Chiodo 2024a and forthcoming.

indeed valuable” (Boden 1998, p. 355. See also Boden 1991 and 2009 and Boden and Edmonds 2009). In 1998, she adds that “[w]e are a very long way from that” (Boden 1998, p. 355). Today, our question is the following: are we sure that artificial intelligence is “a very long way from that”, i.e. creativity?

2. *The foundation of a possible epistemology of creativity*

A possible answer to the question posed above may arise from reflecting upon the relationship between artificial intelligence and creativity against the background of a possible epistemology of the latter.

If it is true that creativity can be defined as the capability to cross a kind of no man’s land extending from the universal rule, which is given, to its particular application, which is not given, it is also true that it can be defined as the last step in the epistemological process from analysis to abstraction to idealisation (and which has been typical of Western culture for millennia. See at least Brzeziński *et al.* 1990a and 1990b, Coniglione, Poli and Rollinger 2004 and Cartwright and Jones 2005).² More precisely, creativity can be thought of, first, as the move from abstraction to idealisation (concerning the universal rule quoted above) and, second, as the move from idealisation to reality (concerning the particular application quoted above).

But let us start from the first step in the epistemological process of creativity. Analysis is usually defined as the capability to identify what varies between analogous things. If we consider ten analogous, but not identical, glasses, we analyse whenever we can identify their differences in shapes, materials, colours, uses and so forth.

The second step is abstraction, which is usually defined as the capability to identify what does not vary between analogous things. If we consider the ten glasses described above, we abstract whenever we can identify their identicalities. In other words, we abstract whenever we can answer the following question: after discarding the differences between the ten glasses described above, what is left? Our answer can be that, again, what is left is their being, first, artefacts for drinking liquids, second, suitable for human hands holding them and, third, suitable for human mouths drinking from them – what is left is their universal rule. As Locke specifies, what is left is what we can

² I worked on it especially in Chiodo 2018.

identify as their “universals” (Locke 1690, II, 11, 9). But Locke also specifies that abstracting does not mean creating something novel, which is one of the two necessary conditions of creativity, as we have seen (see Locke 1690, III, 3, 7). It is no coincidence that, according to contemporary philosophers working on abstraction, it means “omi[ssion] of properties” (Nowak 1990, p. 193) that vary and “omission of truth” (Jones 2005, p. 175), but not the creation of something novel.

Indeed, the creation of something novel starts emerging from the last step in the epistemological process of creativity: the move from abstraction to idealisation, which is usually defined as the capability to change something to improve it. If we consider the ten glasses described above, we idealise whenever we can change their universal rule to improve it and, consequently, its particular application – and, most interestingly, speaking of changing something to improve it means speaking of both the two necessary conditions of creativity: first, speaking of changing means speaking precisely of novelty and, second, speaking of improving means speaking precisely of value. It is no coincidence that, according to contemporary philosophers working on idealisation, it means “misrepresentation” (Jones 2005, p. 185) and “assertion of a falsehood” (Jones 2005, p. 175): we “misrepresent” and “assert [...] a falsehood” precisely because we change universal rules to improve them. For instance, we change the universal rule described above by adding the following subrule, which the ten glasses described above do not share: glasses are (also and necessarily) suitable for the kinds of drinks they contain. And, after changing their universal rule to improve it, we can improve its particular application by designing a glass that is novel (being different from our ten glasses) and valuable (being better than our ten glasses) – which means speaking precisely of creativity.

Thus, there is creativity when we move from abstraction to idealisation and when we move from idealisation to reality, i.e. the particular application of the universal rule we change. For instance, we are creative if we, first, add that glasses are (also and necessarily) suitable for the kinds of drinks they contain and, second, design a glass whose novel shape is better (than any other glass) for containing liqueur wines – which ultimately means that we are creative if we make our experience of drinking liqueur wines better.

Again, are we sure that artificial intelligence is “a very long way from that”, i.e. creativity as we can define it from an epistemological perspective?

3. Can (generative) artificial intelligence be creative?

Let us address a case in point: *Edmond de Belamy*, which is the portrait produced by the kind of generative artificial intelligence, specifically generative adversarial networks (GANs), that its inventor defines as follows:

The basic idea of GANs is to set up a game between two players. One of them is called the generator. [...] The other player is the discriminator. [...] The generator is trained to fool the discriminator. We can think of the generator as being like a counterfeiter, trying to make fake money, and the discriminator as being like police, trying to allow legitimate money and catch counterfeit money. To succeed in this game, the counterfeiter must learn to make money that is indistinguishable from genuine money, and the generator network must learn to create samples that are drawn from the same distribution as the training data (Goodfellow 2016, pp. 16-17. See also McCormack and d’Inverno 2012, Veale and Cardoso 2019, Anscomb 2021 and Bown 2021).

Basically, the discriminator cannot distinguish *Edmond de Belamy* from the thousands of human-made portraits painted between the 14th and the 19th centuries on which it is trained.

Obvious, which is the group of artists who generated *Edmond de Belamy* in 2018, specifies that it is a novel image (which is one of the two conditions of creativity, as we have seen). Indeed, it is neither a copy of the thousands of human-made portraits nor a collage of them nor an average of them, because it is produced by the generator, which does not know any of them. Obvious even uses the word “creation” to describe the production of *Edmond de Belamy*:

Does it mean that an artist working with GAN algorithms makes no contribution to the process in terms of inventiveness? No. It simply means that the artist focuses her creativity on other variables of the process or uses a different kind of creativity, and that the visual creation becomes more and more delegated to the tool. Some of the aspects on which the artist focuses are the choice of the theme, research related to the decision to treat this theme, the search for inspiration (which translates here into the search and choice of database components used as input for the algorithm), the programming and fine-tuning of the algorithm and the whole process based on trial and error, and the choice of means of expression (Obvious 2020, p. 168, my translation).

Thus, on the one hand, “creativity” is attributed to humans (“the artist focuses her creativity on other variables of the process”) and, on the other hand, “creation”, specifically “visual creation”, is attributed to artificial intelligence (“the visual creation becomes more and more delegated to the tool”). In the first case, “creativity” is at stake in several circumstances (“the choice of the theme,

research related to the decision to treat this theme, the search for inspiration [...], the programming and fine-tuning of the algorithm and the whole process based on trial and error, and the choice of means of expression”). In the second case, “creation” is at stake in one circumstance (“the visual creation”) – but speaking of “visual creation” means speaking of something crucial when it comes to creativity as the capability to produce something that is not only novel but also valuable, and especially the capability to produce something that is novel precisely to be valuable, i.e. better.

Indeed, speaking of “visual creation” means speaking of nothing less than the process that translates novelty and value into the artefact itself, i.e. the shape itself. We can choose both “the theme” (for instance, a kind of power that petrifies and devours) and the “means of expression” (for instance, painting). But we can hardly live up to the capability of “visual creation” proved by Picasso in *La Celestina*, which is defined “as the symbol for a devouring female whose power could petrify its victim” (Holloway 2006: p. 118). In other words, what most distinguishes Picasso from us is his exceptional capability to shape, precisely through his “visual creation”, “the theme” that both he and we can choose. Thus, the attribution of “visual creation” means the attribution of nothing less than the exceptional capability to shape a kind of power that petrifies and devours, to the point of even shaping “the symbol [itself] for a devouring female whose power could petrify its victim” – the attribution of “visual creation” to artificial intelligence means thinking of *Edmond de Belamy* as capable as *La Celestina* of shaping exceptionally complex human experiences. Does it make sense? For instance, can we think of *Edmond de Belamy* as “the symbol for” melancholy (because of its dark colours and soft edges), specifically as a novel and better “symbol for” melancholy?

4. *Why (generative) artificial intelligence cannot be creative*

In what follows, I will answer the question posed above against the background of the epistemology of creativity as I have argued it in the second section.

Let us compare *Edmond de Belamy* and *La Celestina* through the following logical steps. First, in the case of *La Celestina*, we can identify the kind of epistemological process I have defined as the move from abstraction to idealisation. Indeed, we can reasonably argue that Picasso, no matter to what extent consciously and to

what extent unconsciously, first, abstracts the “universals” of petrifying and devouring after experiencing (both directly and indirectly) a variety of their cases and, second, idealises by changing their universal rules to improve them and, consequently, their particular application, i.e. *La Celestina* – which means, again, shaping something whose novelty makes it a better “symbol for” what petrifying and devouring mean. Thus, we can ultimately argue that Picasso is creative. Specifically, we can say that he can add novelty and value by ingeniously correlating the idea of petrifying and devouring with the idea of partial blindness: the healthy eye can exercise power over the present and, in addition, the blind eye can exercise power over the future (according to the classical meaning of blindness in Western art).

Second, in the case of *Edmond de Belamy*, we cannot identify the kind of epistemological process I have defined as the move from abstraction to idealisation. Indeed, we cannot reasonably argue that generative artificial intelligence, first, abstracts the “universals” of melancholy after experiencing (both directly and indirectly) a variety of its cases and, second, idealises by changing its universal rule to improve it and, consequently, their particular application, i.e. *Edmond de Belamy*. Conversely, we can identify the following kind of process. If it is true that it is the generator (which does not know any of the thousands of human-made portraits), and not the discriminator (which knows them), that far more directly produces *Edmond de Belamy*, it is also true that its shape does not result from abstraction – *Edmond de Belamy* results neither from abstraction nor from the move from it to idealisation, but from nothing more than trial and error, i.e. a kind of randomness, which means precisely that the generator tries “to fool the discriminator” by producing “samples that are drawn from the same distribution as the training data”, as we have seen. In other words, *Edmond de Belamy* is far from a human being who, first, experiences melancholy directly in various ways, second, experiences it indirectly in various ways, third, reflects upon it, no matter to what extent consciously and to what extent unconsciously, and, fourth, shapes a novel and better “symbol for” it – *Edmond de Belamy* is far from a human being who makes sense of being a human being by shaping its meaning.³

But let us add the following thought experiment: let us imagine that the thousands of human-made portraits on which the discriminator is trained are chosen by Obvious as right representations of

³ I cannot develop the argument of sensemaking (as opposed to randomness) in the necessarily limited space of a short article. But I worked on it in Chiodo 2024a and 2024b.

melancholy. Can we think of *Edmond de Belamy* as resulting from abstraction and idealisation? In other words, can we think of generative artificial intelligence as creative?

I propose to answer no for the following reason. Again, if it is true that it is the generator, and not the discriminator, that far more directly produces *Edmond de Belamy*, it is also true that its shape results from nothing more than trial and error, i.e. the kind of randomness described above. More precisely, even though there is a sense in which the discriminator, which is trained on right representations of melancholy, may lead the generator to produce *Edmond de Belamy* as something that may make us reflect upon melancholy, it is also true that it would happen randomly. Again, it would happen through trial and error when it comes “to fool[ing] the discriminator” by producing “samples that are drawn from the same distribution as the training data” – again, it would not happen through abstraction and idealisation as the foundation of the epistemological process of a human being who makes sense of being a human being by experiencing it directly, experiencing it indirectly, reflecting upon it and shaping its meaning.

Thus, *Edmond de Belamy* can be novel, but it cannot be valuable. If we happen to reflect upon melancholy when we see it, it is as random as when we happen to reflect upon the passage of time when we see traffic lights. Traffic lights have nothing important (and reliable) to teach us about the meaning of the passage of time. Analogously, *Edmond de Belamy* has nothing important (and reliable) to teach us about the meaning of melancholy. Conversely, *La Celestina* has something exceptionally important (and reliable) to teach us about the meaning of petrifying and devouring – more precisely, it is a novel shape to better teach us what petrifying and devouring mean, which is the reason why Picasso is exceptionally creative.

If it makes sense, does it mean that artificial intelligence is totally useless when it comes to creativity? I propose to answer no for the following reason: it can offer us novel opportunities to create value whenever we try to shape our making sense of ourselves as humans. For instance, *Edmond de Belamy* can (randomly) offer us insights into the correlation between dark colours and soft edges, on the one hand, and melancholy, on the other hand. Analogously, traffic lights can (randomly) offer us insights into the correlation between sudden changes, on the one hand, and the passage of time, on the other hand.

And we, as humans, can randomly shape something (as we do on a daily basis). For instance, if we scribble curved lines and Picasso, after seeing them and exclaiming “I got it!”, draws one of

his doves, we are not creative and he is creative, because the curved lines we scribble do not result from his (exceptional) capability to draw a novel shape to better teach us what peace means, but from nothing more than the (trivial) capability to scribble (random) curved lines.

Thus, it may make more sense to stop comparing what artificial intelligence does with what we do rather than continuing to compete in a kind of race that is simply not playable by artificial intelligence: making sense of who we are – which is not only one of the best reasons to exercise our creativity but also one of the best things it can do.

References

- Amabile T.M., *Creativity in Context*, Wesview, Boulder 1996.
- Anscomb C., *Creative Agency as Executive Agency. Grounding the Artistic Significance of Automatic Images*, in “The Journal of Aesthetics and Art Criticism”, 79/4 (2021), pp. 415-427.
- Boden M.A. *The Creative Mind. Myths and Mechanisms*, Basic Books, New York 1991.
- Id., *Creativity and Artificial Intelligence*, in “Artificial Intelligence”, 103 (1998), pp. 347-356.
- Id., *Computer Models of Creativity*, “AI Magazine”, 30/3 (2009), pp. 23-34.
- Boden M.A., Edmonds E.A., *What Is Generative Art?*, in “Digital Creativity”, 20/1-2 (2009), pp. 21-46.
- Bown O., *Beyond the Creative Species. Making Machines That Make Art and Music*, The MIT Press, Cambridge 2021.
- Brzeziński J., Coniglione F., Kuipers T.A.F., Nowak L. (eds.), *Idealisation I. General Problems*, in “Poznań Studies in the Philosophy of the Sciences and the Humanities” 16 (1990a) (the entire issue).
- Brzeziński J., Coniglione F., Kuipers T.A.F., Nowak L. (eds.), *Idealisation II. Forms and Applications*, in “Poznań Studies in the Philosophy of the Sciences and the Humanities” 17 (1990b) (the entire issue).
- Carruthers, P., *The Architecture of the Mind. Massive Modularity and the Flexibility of Thought*, Clarendon Press, Oxford 2006.
- Cartwright N., Jones M.R. (eds.), *Idealisation XII. Correcting the Model. Idealisation and Abstraction in the Sciences*, in “Poznań Studies in the Philosophy of the Sciences and the Humanities” 86 (2005) (the entire issue).

- Chiodo S., *Come pensa un europeo. Epistemologia di un agire comune*, Carocci, Roma 2018.
- Ead., *Predicted Humans. Emerging Technologies and The Burden of Sensemaking*, Routledge, Abingdon-New York 2024a.
- Ead., *What AI 'Art' Can Teach Us About Art*, in "Journal of aesthetics & culture", 16/1 (2024b), pp. 1-13.
- Ead., *Artificial Creativity?* (forthcoming).
- Coniglione F., Poli R., Rollinger R. (eds.), *Idealisation XI. Historical Studies on Abstraction and Idealisation*, in "Poznań Studies in the Philosophy of the Sciences and the Humanities" 82 (2004) (the entire issue).
- Csíkorszenthalmási M., *Creativity. Flow and the Psychology of Discovery and Invention*, Harper Collins Publishers, New York 1996.
- Garroni E., *Creatività*, Quodlibet, Macerata 2010.
- Gaut B., *The Philosophy of Creativity*, in "Philosophy Compass", 12/5 (2010), pp. 1034-1046.
- Goodfellow I., *NIPS 2016 Tutorial. Generative Adversarial Networks*, in "arXiv" 1701/160 (2016), pp. 1-15.
- Holloway M., *Making Time. Picasso's Suite 347*, Peter Lang Publishing, Bern-New York 2006.
- Jones, M.R. (2005), *Idealisation and Abstraction. A Framework*, in Cartwright and Jones 2005, pp. 173-217.
- Kant I., *Kritik der Urteilskraft*, Edition Holzinger, Berlin und Libau 1790; transl. *Critique of the Power of Judgment*, ed. by Paul Guyer, Cambridge University Press, Cambridge-New York 2000.
- Locke J., *An Essay Concerning Human Understanding* [1690], ed. by P. Phemister, Oxford University Press, New York 2008.
- Mayer R.E., *Fifty Years of Creativity Research*, in R.J. Sternberg (ed.), *Handbook of Creativity*, Cambridge University Press, Cambridge 1999, pp. 449-460.
- McCormack J. d'Inverno, M. (eds.), *Computers and Creativity*, Springer, Berlin-Heidelberg 2012.
- Nowak L., *Abstracts Are Not Our Constructs. The Mental Constructs Are Abstracts*, in Brzeziński et al. 1990a, pp. 193-206.
- Obvious, *La 'Famille de Belamy' e i 'Sogni elettrici di Ukiyo'. Reinterpretazioni e accelerazioni*, in A. Barale (a cura di), *Arte e intelligenza artificiale. Be my GAN*, Jaca Book, Milano 2020, pp. 166-193.
- Runco M.A., *Updating the Standard Definition of Creativity to Account for the Artificial Creativity of AI*, in "Creativity Research Journal", 2023, pp. 1-5.

- Stokes D., *A Metaphysics of Creativity*, in K. Stock, K. Thomson-Jones (eds.), *New Waves in Aesthetics*, Palgrave Macmillan, Basingstoke 2008, pp. 105-124.
- Veale T., Cardoso F.A. (eds.), *Computational Creativity. The Philosophy and Engineering of Autonomously Creative Systems*, Springer, Cham 2019.