STUDIA UBB. PHILOSOPHIA, Vol. 69 (2024), Sp. Iss., pp. 49-60
(RECOMMENDED CITATION)
DOI:10.24193/subbphil.2024.sp.iss.03

# Creating with AI: On recent debates about authorship revisiting the influence of Barthes and Foucault

**Iulia-Dana PUȘCAȘU***

**ABSTRACT.** Artificial Intelligence (AI) models are now capable of producing artifacts that mimic human creation, such as visual art, text or music. The remarkable sophistication of these results reignited the debate on authorship, calling into question issues such as intent, originality, autonomy or aesthetic engagement. I will present and explain the main positions on authorship that have emerged from this questioning, drawing on Emanuele Arielli's recent account in *AI-aesthetics and the artificial author* (2023). Furthermore, I will show how Roland Barthes' *The Death of the Author* (1967) and Michel Foucault's *What is an Author?* (1969) are central to understanding the philosophical implications of the debate and how conceptualisations pursued in these works inform current perspectives on authorship when AI is involved.

***Keywords:*** *artificial intelligence, authorship, intention, the death of the author, author-function.*

## 1. Introduction. The issue with authorship and AI

The field of artificial intelligence has seen notable progress in recent years, prompting conversations regarding the impact of AI across various domains. AI systems designed to produce media such as images, text, or music have shown

* Phd student, Doctoral School of Philosophy, Babeș-Bolyai University, Cluj-Napoca, Romania. Email: iulia.puscasu@ubbcluj.ro.

remarkable sophistication. Their outputs are often indistinguishable from human-authored works and sometimes even preferred according to certain criteria[1].

While the objective to synthesise new artifacts using computers is not new, the specific methods involved in achieving this have undergone a significant shift, from step-by-step or rule-based programming to a sub-symbolic approach. This novel approach does not require explicit rules and is instead based on machine learning and artificial neural networks[2], making it possible for an AI system itself to extract deep structure from a set of artifacts and synthesise new ones as a result[3]. The innovative architecture of these networks and the computing power now available have had a major impact on the progress of this field. However, their success heavily relies on the input data used for their training. The large amount of data that is necessary for this process—commonly called the 'corpus' in literature—is mined and collected from various sources. It is then analysed and processed by the AI, constituting a foundational base for its function and outputs.

Given this situation, there has been significant debate surrounding questions of authorship and their implications—philosophical, socio-cultural, ethical, or legal. Several key questions that arise in almost every discussion and are of concern to this paper include: Can AI be considered an author in the traditional sense? Is the claim to authorship from the humans involved legitimate? What does proper attribution of authorship mean given the amount and variety of data necessary for an AI system to perform? Not last, will AI force us to reconsider or reshape our understanding of authorship and how it should be granted?

The claims in addressing this question, both within academia and the public sphere, are difficult to reconcile. Some critics argue that AI systems are mere tools or instruments, citing the absence of intentionality and creative agency to support their position, while others, applying the criterion of autonomy, argue that AI systems engender a new form of expression, specific and proper to the system

[1] For example, Sunspring, a 2016 science fiction film written entirely by an AI, which was placed top ten in Sci-Fi London's annual film festival. In Carys J. Craig, Ian R. Kerr, "The Death of the AI Author" (March 25, 2019), in *Osgoode Legal Studies Research Paper* (March 25, 2019). Available at SSRN: https://ssrn.com/abstract= or http://dx.doi.org/10.2139/ssrn.3374951, p. 36.

[2] Hannes Bajohr, "Writing at a Distance: Notes on Authorship and Artificial Intelligence" (march, 2023), in *German Studies Review*, 47(2).2024, The German Studies Association, p. 321.

[3] Lev Manovich, "Defining AI Arts: Three Proposals", in *AI and Dialog of Cultures*, Hermitage Museum, St Petersburg, 2019, p. 5.

itself, given that human control may be minimal[4]. An alternative view posits that AI's outputs are in a sense authorless by default, given that they depend on viewer interpretation and are thus inherently meaningless[5], or that the very idea of AI as an author is conceptually inconsistent and inherently flawed[6].

I do not intend to present a conclusive answer or validate any specific standpoint over another. Instead, I will outline the primary positions on authorship that emerge from this questioning of the issue of authorship, explaining their underlying assumptions, aims, and implications. I will do this in the next section, primarily building on Emanuele Arielli's recent account on authorship in *AI-aesthetics and the artificial author* (2023). Furthermore, I will show how Roland Barthes' *The Death of the Author* (1967) and Michel Foucault's *What is an Author?* (1969) are central to understanding the implications of the debate and how the conceptualisations pursued in these works inform current perspectives on authorship when AI is involved.

## 2. Authorship perspectives, presuppositions and related concepts

In his recent publication, *AI-aesthetics and the artificial author* (2023), Emanuele Arielli provides a way of looking at and synthesising the problem concerning the identification and necessity or relevance of an author in creations that are made with the use of AI, that is "AI-generated works". While he seeks to avoid side-taking in the matter of pointing out where the authorship lies, his approach is motivated by an interest in how different views on the authorship of AI-generated works influence their aesthetic experience and vice versa.

Arielli begins by locating the concern with the legitimacy of applying the notion of authorship within developments of structuralism and post-structuralism, which have provided a context or framework that undermines the privileged position of an author as the sole or ultimate source of a work[7]. As he suggests by pointing out an alignment of this type of effort with the objectives of New Criticism

[4] This is explained by E. Arielli in Emanuele Arielli, "AI-aesthetics and the artificial author", in *Proceedings of the European Society for Aesthetics*, The European Society for Aesthetics, 15.2023, pp. 41-42.

[5] Ibidem.

[6] Position advanced in Carys J. Craig, Ian R. Kerr, Op. Cit.

[7] Arielli, Op. Cit., p. 43.

and Beardsley's intentional fallacy—which criticised the reliance on external evidence, such as the author's intention, in the assessment of a work[8]—the scepticism over authorial intent was something that became generally shared in literary studies and criticism. This scepticism is seen to culminate in Roland Barthes' *Death of the Author* and, although concerned beyond the problem of intention, in Michel Foucault's *What is an Author?*, which Arielli references with the merit of having discredited "the image of the unique individual artist or author"[9]. However, in his view, these philosophical achievements or theories have not yet truly manifested phenomenologically in our consideration of human-made cultural products, as we would continue to think in terms of authorial intent and the figure behind the work, whose motives and intentions we consider[10].

Departing from this last claim, Arielli continues by addressing the necessity for an author in aesthetic appreciation. In this, he links "agency and intentionality attribution" and "the need for a recognisable subject behind an artifact" to the perception of authorial depth and views the results of AI as a possible test to determine in which forms of cultural production is it necessary[11]. This is "the threshold of authorial relevance", which is concerned with when authorship is relevant or even crucial for aesthetic appreciation and what type of work requires an author in this sense. The other threshold that he identifies as relevant is "the threshold of instrumentality". That is, what separates human-made from machine-made? Recent art has given us examples of great collaborations between man and machine, where "the complexity of the mechanism […] does not shift the locus of artistic authorship"[12], meaning that AI remains a tool for artists to achieve their intent. On the other hand, new models are increasingly responsible for "creative decisions"[13], that is, "creating" and influencing the aesthetic qualities of a work, producing intricate outputs with minimal human input.

Based on these thresholds, which Arielli considers to be at the centre of authorship questioning, the main perspectives on authorship in the context of AI use emerge. He identifies and categorises them into five main groups. The first is the "human-centric view", which considers the author to be "the first designer", the

[8] Ibidem.
[9] Ibidem.
[10] Ibidem, pp. 43-44.
[11] Ibidem, pp. 45-46.
[12] Ibidem, p. 47.
[13] Ibidem.

initiator of the creation process for an "original idea or concept"[14], who is therefore linked to the work by an essential causal relationship that is established outside a later stage of execution. This perspective considers AI as a tool used to fulfil the author's intention, including the idea of "author as selector" or "author as prompt-engineer"[15]. The second perspective regards "AI as a full author" within a framework that anticipates the moment when the outputs of AI would be seen as the result of their own agency[16]. The future delineation and acceptance of this sense of agency are crucial because, according to Arielli, authorship in a work is to emerge and be recognised from the attribution of intentionality and autonomy rather than solely from the characteristics of the work itself[17]. The third perspective is "Remixed authoriality", which Arielli aligns with theories proclaiming the death of the author and the post-productive stance[18]. This emphasises that AI-generated works are the result of an interplay and blending of different sources, influences, and pre-existing materials, and therefore, a "reflection of collective human intentionality" that rather points towards a form of artificially transformed collective authoriality[19] than to a sole author figure. The fourth perspective draws from narrative theory and is concerned with "implied authoriality". The implied author is "the voice grounded in the text and expressed by its content and style", indicative of a construct that "emerges from the work, over and above the original source that produced it"[20]. This can occur even when the viewer is aware of the lack of intentionality behind a work, becoming an "actively imagined authoriality" that allows ascribing meanings to something inherently inanimate[21]. Rather than trying to assign 'real' authoriality, this perspective focuses on the implied author as it is constructed through the engagement with a work, either separate from the 'real persona' of the author or actively imagined in its absence. Finally, there is the option to disregard concerns about authorial intention and attribution, marking a potential change in how we engage with certain works. For Arielli, this could shift our focus, for example, towards formal and aesthetic qualities, regardless of the creator's identity or origins of a

[14] Ibidem, p. 48.
[15] Ibidem.
[16] Ibidem, p. 49.
[17] Ibidem, pp. 49-50.
[18] Ibidem, p. 50.
[19] Ibidem.
[20] Ibidem, p. 51.
[21] Ibidem.

work[22]. However, it remains questionable whether and in what way we would still be open to engaging with machine-produced works if we were to completely give up considerations related to their authorship[23].

In short, the first aspect of the problem involves questioning the legitimacy of attributing authorship to AI-generated works with respect to the concern of whether there can be a discernible author who confers meaning that is essential to the appreciation of a work as such. The second issue concerns the key thresholds identified by Arielli regarding authorship: that of author relevance and that of instrumentality. When is there a crucial need for an author and at what point does instrumentality end? In addition, the distinct perspectives on authorship that have been presented, together with their argumentative unfolding, point to several related presuppositions and concepts, such as the primacy of the original idea or concept, intention, autonomy, agency, or the mental construction of a person behind the work. These points can be better understood through the conceptualisations pursued in Roland Barthes' *Death of the Author* and Michel Foucault's *What is an Author?*.

## 3. Roland Barthes and the *Death of the Author* (1967)

Against the theoretical backdrop of the divide between structuralism and post-structuralism, Roland Barthes writes and publishes his seminal essay *Death of the Author* (1967), which questions the actual role of the author in its ties to the text. He does not ultimately undermine the legitimacy of the attribution of authorship *per se,* but criticises the search for the ultimate meaning or truth of a work in the figure of the author, and the construction of the identity of the author as such. With Barthes, the question of who is speaking remains unanswered as he argues that writing, in its multiplicity, erases any point of origin, thereby making way to proclaim the death of the author and, shifting the locus of a text's unity, the birth of the reader.

Piecing together a brief history of writing and authorship, Barthes held that the celebration of the author is—in contrast to what he calls 'ethnographic societies'[24]—a product of modern Western society and the prestige it has bestowed

[22] Ibidem.
[23] Ibidem.
[24] Barthes, Roland, "The death of the author", in *Image, music, text*, Fontana, London, 1990, p. 142.

on the individual. Drawing on the linguistics and semiotics of that time, he sees language as taking on its own life: "the whole of the enunciation is an empty process, functioning perfectly without there being any need for it to be filled with the person of the interlocutors […] the author is never more than the instance writing, just as the I is nothing other than the instance saying I"[25]. This 'truth' of language reveals that writing acts "outside of any function other than that of the very practice of the symbol itself"[26] and, in this, "has no other origin than language itself, language which ceaselessly calls into question all origins"[27].

Holding these views, and in this particular framework, the author, if such a thing can still be argued to exist, is an extremely fragile entity, subject to transformations like language itself, and becomes marginal in the appreciation of a work. This situation, though not necessarily proof of the complete disappearance of the author, has led to a general undermining of the idea that the author's identity and intentions permeate the text and of his authority over the final work. If the question of the legitimacy of attribution for AI-generated works is taken to be whether we can truly point towards an author just by engaging with a work, we are left in a difficult position to ponder.

Barthes contends that trying to find the lineage from the text to the author, which is to locate the truth, creates an artificial division into before and after, ultimately turning the author into a god-like figure through direct attribution of creation[28]. For him, this paradigm cannot hold as "a text is not a line of words releasing a single 'theological' meaning (the 'message of the author-god')"[29]. He opposed this view the idea of a 'modern scriptor', which is "born simultaneously with the text"[30] and does not claim originality and authority over the text, but rather emerges as a master of the narrative code: "Succeeding the Author, the scriptor no longer bears within him passions, humours, feelings, impressions, but rather this immense dictionary from which he draws a writing that can know no halt"[31].

[25] Ibidem, p. 145.
[26] Ibidem, p. 142.
[27] Ibidem, p. 146.
[28] Ibidem, p. 145.
[29] Ibidem, p. 146.
[30] Ibidem, p. 145.
[31] Ibidem, p. 147.

"The book is only a tissue of signs" suggests that there is nothing that is purely original as meaning is shaped through a collective negotiation of understanding and the text is seen as "a multi-dimensional space in which a variety of writings, none of them original, blend and clash"[32]. This way of looking at a work has inevitably been linked by multiple scholars to the process behind AI-generated works, which, by building on patterns extracted from training on a large corpus—data from diverse sources, encapsulating varying ideas in different forms—appears as a mastery of the code and annihilates the possibility of pointing out any precise origin. If we are to put anthropocentrism aside, considering Barthes' philosophical developments, this can be seen as reinforcing the idea that there is not always a real need for a unique individual author outside the work.

## 4. Michel Foucault's analysis in *What is an author?* (1969)

In *What is an author?* (1969), Michel Foucault is interested in the relationship between the author and a text, that is "the manner in which a text apparently points to this figure who is outside and precedes it"[33]. Building on similar theoretical principles as Barthes, Foucault asserts that a key attribute of the modern text is that "[it] has freed itself from the necessity of 'expression'; it only refers to itself"[34]. In the context of the broader linguistic and cultural shift that favoured signification over representation, he refers to the same idea that linguistic formations seem to exist on their own and, being detached from representation, the ceaseless transformation in the logic of signification leads to the situation that "the writing subject endlessly disappears"[35]. The work attained the right to "murder" its author, in the sense of effacement or cancellation of "the signs of his particular individuality"[36]. However, here is where Foucault, despite proclaiming the "death" or disappearance of the author in an apparently shared gesture, distances himself from the Barthesian approach and ventures into a much more complex process of dissection.

[32] Ibidem, p. 146.
[33] Michel Foucault, "What is an author?" (1969), in *Language, counter-memory, practice. Selected Essays and Interviews*, Cornell University Press, Ithaca, New York, 1977, p. 115.
[34] Ibidem, p. 116.
[35] Ibidem.
[36] Ibidem.

One key difference is the inquiry ground that Foucault opens up to discuss "the equally problematic nature of the word 'work'", because what constitutes a work, "if a work is not something written by a person called an 'author'?"[37]. He ponders what Sade's works that he wrote in prison before he was consecrated as an author were, or whether, by referring to Nietzsche, all that an author wrote should be considered part of their body of work[38]. Further analysing the use of an author's name compared to that of the proper name, he concludes that the former only accompanies certain types of text, characterising and serving as a function within only certain types of discourse[39]. In terms of attribution of authorship and its legitimacy, Foucault notes that modern criticism, in its strategies of defining authorship and displaying its function, employ devices derived from the Christian tradition of authenticating texts, which sought to prove their value or truthfulness by establishing the "holiness" of their author. He references here the four criteria in Saint Jerome's *De Viris Illustribus*, which show that the author ensures and defines a certain level of quality, conceptual or theoretical coherence, stylistic uniformity, and, nonetheless, constitutes a historical figure. Having noted that the search for authorial markers has to do with the way we handle texts and that particular signs in a text that seem to refer to an author do not actually refer to the writer, it is out of this "scission" that the author-function arises[40]. This function of discourse is explained by its four main features, conveniently summarised as follows:

> The 'author-function' is [1] tied to the legal and institutional systems that circumscribe, determine, and articulate the realm of discourses; [2] it does not operate in a uniform manner in all discourses, at all times, and in any given culture; [3] it is not defined by the spontaneous attribution of a text to its creator, but through a series of precise and complex procedures; [4] it does not refer, purely and simply, to an actual individual[41].

By replacing the conventional author figure with the author as a 'function of discourse', which, critics argue, "authorises the very idea of 'author'"[42], Foucault

[37] Ibidem, p. 117.
[38] Ibidem, pp. 118-119.
[39] Ibidem, p. 120.
[40] Ibidem, p. 130.
[41] Ibidem, p. 130.
[42] Adrian Wilson, "Foucault on the 'Question of the Author': a critical exegesis", in *The Modern Language Review*, 99 (2). pp. 339-363.

emphasises its constructed-ness, which can simultaneously give rise to multiple selves or subjects[43] and makes it impossible to think that we can simply and straightforwardly point to a real individual outside the text. The author figures within a specific moment of power/knowledge relation, making it a constrained entity rather than solely someone who creates something new.

Concerning the discussion about the authorship of AI-generated works, we can see that Foucault's conceptualisations carefully address the need for an author figure to be attached to certain works. He emphasised that what constitutes a work in this sense is not a constant but is determined through a negotiation between different factors about the modes in which it exists and circulates within discourse. From here, the question follows: do AI-generated works require the attribution of authorship, and in what way would they be considered "works" without the attachment to such a figure? Foucault challenges the traditional view of authorship precisely by revealing that the "real authority" that governs a work is not an individual author but discourse itself[44]. Furthermore, he pointed out the division, in the case of text, between the author and the writer or producer, and that the author-function is one of the "subjects" that emerge in the distance between the two. By problematising the figure of the author in this way, Foucault signalled a potential "crisis" in the mechanisms for legitimate attribution of authorship. Having noted these, the framework in which Foucault treats the issue proves insufficient to attribute authorship over solely intentionality or execution in a creation process, whether we would try to argue in favour of a single human, machine, or collectively shared authorship.

Another subject that emerges in the distance between author and writer, distinct from the author-function and identified by Foucault through its link to the authorial markers present in a text as a 'second self', is what Arielli mentioned in relation to narrative theory as the implied author. The influence that Foucault exerted over this notion is that, while the concept was already present in the narrative theory of the 60s as an authorial construct, by positing the author-figure as a construct of the reader, he opened up the way for a series of later constructivist conceptions such as the 'postulated author', 'fictional author' or 'interpretative author'[45].

[43] Foucault, Op. Cit., p. 131.
[44] Wilson, Op. Cit.
[45] Ibidem, p. 343.

Nonetheless, Foucault looked forward to the possibility of moving away from the search for the author and instead of "arresting the possibility of genuine change" to explore the gap left by its disappearance. In this sense, he does not advocate a complete abandonment of the subject but proposes that we focus on the subject not in relation to its creative role, but as "a complex and variable function of discourse"[46]. However, at the end of his lecture and essay, Foucault opens the discussion on the possibility of a culture in which discourses circulate and unfold in complete anonymity, and in which new questions will thus arise amid "the murmur of indifference: 'What matter who's speaking?' "[47], prompting us to rethink how we engage with and inquire into works.

## 5. Conclusions

The field of artificial intelligence has made significant progress in recent years, leading to discussions about the influence of artificial intelligence in various fields. Artificial Intelligence systems focused on the creation of media content, such as images, text, or music, have shown remarkable complexity, which has sparked significant debate around questions of authorship from philosophical, socio-cultural, ethical, or legal perspectives. Approaching the question of authorship from a philosophical point of view has resulted in various perspectives that are difficult to reconcile, precisely because of the questions surrounding the foundations of the notion itself. These perspectives, which I have drawn from Arielli's *AI-aesthetics and the artificial author*, new and conflicting as they may be, are ultimately rooted in a philosophical questioning that can be said to have challenged thinkers at least since the 1960s. As the case was to discuss the works of Barthes and Foucault, their developments did not result in the emergence of a new conception of the author and method for its attribution. Instead, they demonstrated the underlying complexity of the matter and the careful consideration it necessitates, perhaps their accomplishment being a change in our presuppositions and attitudes when questioning the author's figure. With the increasing diversification and use of AI in the production of different works, it is imperative that we engage with the enduring discussions raised by Barthes and Foucault, as the conceptualisations pursued in their work inform our perspectives on authorship when AI is involved and shape our attitudes towards their implications.

[46] Foucault, Op. Cit., p. 138.
[47] Ibidem, p. 138.

## BIBLIOGRAPHY

1. Arielli, Emanuele, “AI-aesthetics and the artificial author”, in *Proceedings of the European Society for Aesthetics*, The European Society for Aesthetics, 15.2023, pp. 40-54.
2. Bajohr, Hannes, “Writing at a Distance: Notes on Authorship and Artificial Intelligence” (march, 2023), in *German Studies Review*, 47(2).2024, The German Studies Association, pp. 315-337.
3. Barthes, Roland, “The death of the author” (1967), in *Image, music, text*, Fontana London, 1977, pp. 142-148.
4. Craig, Carys J. and Kerr, Ian R., “The Death of the AI Author” (march, 2019), in *Osgoode Legal Studies Research Paper*, 2019, pp. 31-86, accessed at: http://dx.doi.org/10.2139/ssrn.3374951.
5. Foucault, Michel, “What is an author?” (1969), in *Language, counter-memory, practice. Selected Essays and Interviews*, Cornell University Press, Ithaca, New York, 1977, pp. 113-138.
6. Manovich, Lev, “Defining AI Arts: Three Proposals”, in *AI and Dialog of Cultures*, Hermitage Museum, St Petersburg, 2019, pp. 1-9.
7. Wilson, Adrian, “Foucault on the ‘Question of the Author’: a critical exegesis”, in *The Modern Language Review*, Modern Humanities Research Association, 99(2).2004. pp. 339-363.