XXVII Generative Art Conference - GA2024

Generative Authorship

Ian O'Loughlin, PhD.
Department of Philosophy, Pacific University, Forest Grove, USA
www.ianoloughlin.com
e-mail: ian.m.oloughlin@gmail.edu

Abstract

Artist attribution in cases of generative art is complicated and controversial. For one very particular class of cases, one hitherto overlooked possibility has merit. If a generative novel has been produced in the style of a human author, and the resultant work is indiscernible from other works with the same author even by competent critics, then there is reason to ascribe authorship of the generative piece to the human author whose work it was styled after. Although this has some counterintuitive and surprising consequences, these do not constitute insurmountable obstacles. This view also has a number of interesting implications, and offers some ways forward with regard to other arguments about artist attribution in generative pieces.

1. Introduction

Questions regarding whom to credit as the *artist* in cases of generative art remain complicated and fraught. Scholars are divided not only on how to answer this question, but also on how to approach it [1, 2, 3]. In the following I will examine one very special case of generative art—fiction in the style of a human author—and, by way of introducing a Turing Test-style condition, make a relatively strong, albeit conditional, claim about authorship attribution. My thesis, strictly speaking, only regards generative works of fiction in the style of a particular human author, but I will address possible and broader connections to other areas of generative art in due course. Additionally, of course, an argument for there being at least some cases of generative art for which artist attribution claims can be rightly made, it is hoped, already constitutes one small bit of progress in these larger pursuits.

To these ends, I will first briefly introduce some of the ways authorship attributions have been discussed and characterized by scholars interested in prose and fiction generated by artificial intelligence, and also some general considerations about authorship of prose works. After laying out some of the possible moves here, I will introduce and motivate a test in the style of the Turing Test, but that targets critics' perceptions of authorship. I will more or less directly lay out an argument that authorship can be rightly attributed in cases where a generative work passes this test. I will then consider a series of objections, worries, and implications.

2. The life of the author

Generative methods already produce significant prose works. Undoubtedly these are still in their infancy, and very likely in only a few years' time the prose works so produced will be far greater in both number and substance. Traditionally, the vast majority of prose works are straightforwardly attributable to an author or authors, and this is especially the case for long, complicated, self-contained and internally coherent works such as novels. While there are a few interesting cases of questionable authorship of novels (lost authors, collaborative works, complicated fan fiction, forgeries, etc.), almost all novels—more so than poetry or nonfiction—are clearly attributable to a definite and small collection of authors, usually to just one human being. This relative paucity of problematic cases is perhaps more the case for novels than it is for other art forms, and may make novels a good starting point for questions of artificial intelligence authorship.

This being said, questions of human authorship are not entirely simple. Roland Barthes famously argued that the role of the author *qua* author had been previously overemphasized in culture and literary criticism, and successfully reshaped the way that authorship questions are bound up with criticism of works [4]. Even more relevant to questions about the *nature* of authorship, Sophus Helle has argued that prose authorship exists in the "middle ranges" of agency, splitting the difference between complete originality and mere mediation, in part because of the heavy debt any author owes to previous works and elements of human culture [5].

In the case of generative authorship the apparent options are laid out nicely in a recent article by Jean-Marc Deltorn and Franck Macrez [6]. Deltorn and Macrez enumerate three broad options: to assign authorship to no one, to assign it to the programmers, developers, or deployers of the generative system, or to assign it to the generative system itself. On the face of it, this is a sensible and satisfying array of possibilities, and it tracks intuitions and discussions of artist attributions in cases other than authorship as well. It is not hard to imagine or find scholars in various disciplines, in different particular cases, taking up each of these causes in turn. For example, Jane Ginsburg and Luke Budiarjo dismiss the possibility of machine attribution, and argue that

authorship questions only remain between the various humans that were involved programming or prompting the system in any given instance of generative authorship, with the distinct possibility that some works of art are authorless [7]. In other areas of art, some scholars argue for the possibility of machine authorship, but only under certain conditions [1].

While I will not in the present paper take sides on whether artificial intelligence systems can rightly be afforded author status in some cases, I will argue that in the case of a particular class of generative fiction, one possibility has been overlooked. In the following section, I will propose a test, and make the claim that any work that passes this test is neither authorless, nor the work of the programmers, prompters, or machine.

3. A criticism test and the implications of indiscernibility

The particular class of generative novels with which we are presently concerned is those novels written "in the style" of an existing human author. The tools for generative novels are already extant, and the tools for generating prose in a style particular to a single human author are also so, even though as of 2024 this has not yet culminated in a library of full and convincing novels in the style of dead novelists. Nevertheless, it is hard to imagine how these will not be readily available in the near future. For the sake of authorship questions, I propose a test to which any such work can be put.

This test is in the spirit of the Turing Test, but instead of competent judges of intelligent conversation, we will turn to competent judges of authorship. For the sake of this thought experiment, we must imagine a critic who is eminently familiar with a given author's style, and can confidently and rightly distinguish a novel by that author from other novels. Also for the sake of the thought experiment, we must imagine a critic who is not already familiar with this given author's full *oeuvre*. Admittedly, this requirement may strain credibility in many cases in the real world—it is perhaps plausible for Balzac or Murdoch, but not for Kafka or Proust. At present let us imagine a world where such critics are to be found as needed, by careful sequestration or localized amnesia if need be, and we will return to some of the real-world considerations eventually.

In such a world, any given generative fiction work in the style of a particular human author will lend itself to a simple test: the competent literary critic will either be able to discern the generative novel from the traditionally-produced novels of the author, or not. For example, let us imagine a generative novel, *Cyril Ambrose*, written in the style of George Eliot. We might give this novel, along with one of hers written in the traditional style by her in her lifetime, to discerning critics who were unfamiliar with either of these particular works (again, pardoning what may be the real-world implausibility of this for the sake of the present argument), and ask the critics to attempt to identify which is the

generatively constructed novel. If the critics do no better than chance at this task, then *Cyril Ambrose* has passed the test.

Who or what is the author of such a work? I want to suggest that there is an author, and this author is none of the usual suspects (programmers, prompters, system). The author, if *Cyril Ambrose* has passed the above criticism test, is George Eliot. Characterizations of the nature of authorship and artistry lend support here. What an author is, according to Helle, is someone who mediates, who weaves together elements of culture into a new mode to be transmitted forward [5]. What an artist is, according to Alva Noë, is someone who provides a new way for us to understand our experience, offering a method for organizing the world as we find it [8]. Indeed, in other areas of art, we find historical cases of attribution that can serve as models: art historians sometimes attribute works to "Rafael" regardless of which of his assistants' hands may have been those in physical contact with the work [9]. It is treated as clear in these cases that Rafael is the person *responsible* for the work. In our own case, Eliot has done just this, offered a particular and new perspective, showcased in her novels. The generative system, and humans deploying it, have acted as her assistants, albeit separated from her in time. As is evidenced by even the most discerning readers being unable to tell the difference, this authorial contribution particular to Eliot is as present in the new generative work *Cyril Ambrose* as it is in any of her other works.

This somewhat expansive view of attribution is not entirely without precedent even in the arena of literature. Ascribing authorship of a text to Enheduanna or Homer even when the scribes who committed their texts to written form lived long after the authors' deaths has been relatively standard practice for centuries. Although it may at first seem like this represents an antiquated view, especially in light of Barthes' well-known *de*-emphasis of the author, since attributing generative posthumous works to an author might seem to emphasize or expand the role of the author, the reasoning in play can be interpreted as in keeping with Barthes. The text speaks for itself here—any details of the author's human life that are not already discernibly present in the text are not relevant. If the author of a collection of novels is, from a text-centered perspective, merely the zero-dimensional origin point, relationally defined as the organizing principle already immanent in the continuities among the perspectives offered by the texts, then this author "belongs" to a text only inasmuch as the author *comes through* in the text. Our criticism test is aimed at determining exactly this. These arguments, however, will be best made clear by considering a series of plausible objections.

4. Objections

In writing her novels, Eliot drew on a wealth of experience and had an array of

intentions and goals directing these works' creation. Aren't these a necessary component of authorship?

In considering attributing artist status to generative systems, Adam Linson makes a compelling case for the necessity of experiential connectedness to human experience and culture, arguing that only if machines have broad engagement with the world, enabling the development of something like conscience, can they rightly be considered the artist [1]. Although I am not here arguing in favor of *machine* authorship, I *am* arguing that our hypothetical *Cyril Ambrose* is an authored work, and Linson's caveats are relevant, since one natural worry is that the apparent context-free generation of the text lacks just these. However, one question is whether this broad experience is manifest in the texts. If it is not, then, in keeping with Barthes, it is nor our problem. If it is, however, then a discerning reader would be able to detect its absence (otherwise what could we *mean* by 'manifest in the text'?). So if *Cyril Ambrose* has passed our criticism test, then any necessary authorial antecedents have been satisfied.

Another way to characterize the acceptance of these (or similar) necessary conditions for authorship is to point out that the objection begs the question against the thesis: to say that Eliot's intentions and experiences are absent in the creation if *Cyril Ambrose* is just to already deny authorship. On the present view, her experiences and intentions are immanent in *Cyril Ambrose* in just the same ways that they are immanent in Middlemarch. The method of execution—whether penned on paper, or generated by clever artificial intelligence systems two centuries later—is a feature to which authorship questions are indifferent.

All of that being said, it is also worth noting that one could offer up related reasons for predicting that it will be very difficult for generative pieces to pass the criticism test above. This I readily accept. It may be quite difficult. The point in contention is whether it would be correct to attribute the work to Eliot *if* it were to pass.

On your view, someone can write a novel long after they have died. Whatever we mean by 'authoring a novel', doesn't it exclude this spooky posthumous action?

Admittedly, this implication applies pressure to our standard and preconceived notions of artistic creation, not to mention action more broadly. Ultimately, however, the point must be conceded. If *Cyril Ambrose*, for example, is written in 2030, and if the author is Eliot, then Eliot will have in fact authored a novel long after she died. The counterintuitive nature of this may be mitigated by considering nearby thought experiments, however. We can imagine Rafael in his workshop, giving explicit instructions to one of his assistants, and then leaving on a trip where he unexpectedly dies without the assistant knowing. The assistant starts and

finishes the work, just as many are thought to have done while Rafael was merely busy or absent, and—*voila!*—Rafael has painted something after he has died.

Notably, the question here is not whether Rafael can rightly be said to be the artist in this example—that is a complicated question in its own right—the relevant question is whether we would have *less* reason to call Rafael the artist if he had died on that trip, than we would if he had merely gone and spent some time in Florence, then returned after the work was complete. It seems absurd, in this case, to say that the artist attribution somehow depends on whether Rafael was still alive in Florence or not, even though that had no casual interaction with the creation of the work. The mere fact of death seems to not be as much on an obstacle to attribution as it may at first seem.

Your examples use dead authors, but by the logic of the argument, would the same claim not be made for a generative novel written in the style of a living author? In which case, are you suggesting that someone could write a novel without knowing it, even if they encountered and disavowed it as theirs?

I concede that dead authors make for clearer exemplars, and probably for more rhetorically convincing instances (perhaps because we are in the habit of not quite granting full personhood to the dead). However, this objection must be upheld; if the above arguments are accepted, then the textual indiscernibility speaks for itself, regardless of the state of the author's human body. It is true that this means that a living person could author a novel without knowing it, and it is true that this seems to us a strange state of affairs. It is worth noting, however, that there are nearby cases that are less counterintuitive: we certainly admit that living humans can be responsible for ideas, and deserve credit, even if these ideas were instantiated without their knowledge (more or less all of copyright law is predicated on just this). To say that some living person can author a novel without knowing it is not as dramatic a revision to our concepts of origination as it may be thought.

To the point of disavowal, it is helpful to remember that this does happen. An author may disavow one of their works. This does not make the work not authored by them, it makes it a work authored by them that was then disavowed. These are interesting cases (both the traditional and generative versions), but I think not particularly damaging to the argument at hand.

The criticism test you propose will probably never actually be able to be enacted. Why should we be interested?

It is a contingent feature of our world that authors happen to usually have a relatively small number of novels, and that anyone who has undergone the requisite training to be able to confidently and readily tell their work from any other is almost certainly familiar with the entire

oeuvre. In the end, this must be accepted, but only as contingent. The main thesis of this paper is that *if* a generative novel were the sort of thing that *would* pass this test, then the author is the human author whose works it was styled after. That is, any work that is indiscernible from an authors' other works, even by the most astute critics, deserves to be credited to that author. The fact that it would be quite difficult to uncontroversially determine which cases this applied to is neither here nor there with regard to the conceptual claim.

The criticism test you mention seems to depend on some kind of "authorial essence". If a human author's works were sufficiently different from one another so as to keep critics guessing, would that just mean any and every generative novel was theirs?

Technically, as stated, the arguments above are vulnerable to this objection. Unlike the above objection, which gets at a contingent feature of the way the test happens to be constrained by the real world, this objection gets at a conceptual feature of the idea of the test itself—the critics in question must be able to tell this authors' novels from other work. As it happens, I suspect that most novelists' work would lend itself to the needed continuity here, but for the sake of completeness we can be clear about the conditions of the test: the critics are people who must be able to tell the author's work from other work. If a particular author is not amenable to this in any way then they may, at most, successfully immunize themselves against authoring generative novels.

This is absurd. If I ask 'how many novels did Woolf write', the answer simply cannot be 'eight, for now, but we'll see what happens next year'. You must be talking about something other than real authorship.

One of the reasons that novels, in particular, make a good target for this investigation is that they constitute a form of art that is entirely comprised of text—which by its nature transcends its physical implementation—but also this form of art is oriented toward the creation of a coherent whole, an invented world that nevertheless takes from and gives to the world we all collectively inhabit, in a way that is recognizably the author's work, conferring a particularized and human perspective on the reader. If such a coherent whole has been created, and the most competent judges deem it to offer the same perspective as a person's other works, then that author deserves credit for inventing the world in question. Moreover, if no one can tell the difference between that generatively-constructed whole and the traditionally-written novels, then we are forced to ask what it is we want from authorship. It seems clear that *Cyrus Ambrose* could not have existed without George Eliot. It also seems clear that, by virtue of it passing our criticism test, it has successfully created a coherent and invented world that confers Eliot's authorial perspective on the reader. Is this not authorship? If a ninth Woolf

novel is released next year, so much the better.

In addition to objections, it may also serve our clarificatory aims to canvas a series of questions aimed at the implications of such a view.

5. Implications

You focus on novels. Could this be expanded to other art forms? Could we make analogous tests, and ascribe artist status to other generative art instances that are done in the style of a particular human?

These are worthy, if complex, avenues of investigation. There are a number of reasons why novels seem like a good starting point. Any art form that involves performance or too deeply involves its physical implementation invites another layer of complexity. Shorter form prose works are perhaps less plausible in the details of a hypothetical criticism test. While some visual art is perhaps not dependent on physical implementation (e.g. photography), these tend to simply involve techniques that are—as of right now—more difficult to imagine being executed indiscernibly by generative and artificial intelligence methods. In the end, these are all compelling as directions for future investigation, but there is some reason to begin with the simplest and least problematic cases. If we can show that there is a single class of art that rightly credits artists posthumously with its creation, we have made progress. This is so even if serious and widespread questions remain.

We sometimes speak metaphorically about the way that people who have died live on through their ideas. Does accepting this view of authorship imply that a human life, and particularly human agency, is more, and more literally, extended across time than is conventionally accepted?

This question involves a complex nexus of concepts, some of which cannot be disentangled here, but the short and simple answer is yes. The authorship in question is “real” authorship, which means that George Eliot will very likely continue producing novels over the next centuries. A living author who had this understanding of authorship might well, today, see their work as a beginning, rather than a completion; they are writing novels that will (or may) generate a coherent and fecund perspective that leads to a much larger supply of novels in future centuries. Although to some extent this may seem counterintuitive, in keeping with some of the responses to the above objections, it is worth remembering that we have long accepted that someone’s ideas may take new forms, that their projects may take on new lives, or that their actions may continue to develop, long after the deaths of their bodies.

If it is true that at least some sufficiently indiscernible works of art are rightly attributable to the human whose corpus

was used in the training of the system that created the work, does this have consequences for the legal ownership and attributability for generative art?

Although this present paper is concerned foremost with the authorship concept itself, this view almost certainly has implications for the legality of attribution, insasmuch as intellectual property law attempts to track actual authorship as at least one component of deciding under which conditions credit is due.

One relatively straightforward implication of the above arguments seems to be that any system that is trained and prompted *too* well, on just one artist's work, such that the results would fool a competent judge, may mean that the legal rights to the resultant work would be accidentally relegated to the original artist. While this was not the aim of the above argumentation about authorship, it seems an acceptable consequence. I make no claims, however, that this refined understanding of authorship offers very broad or powerful solutions to the complex collection of difficulties and questions artificial intelligence has brought to legal questions of intellectual property and origination.

6. Conclusions

Understanding how to rightly make artist attributions in cases of generative art is no simple task, and there are numerous and distinct features of this discussion each of which is convoluted and fraught in its own right. Although authorship questions in these cases have traditionally been limited to choosing from among the system as author, the programmers as authors, the prompter as author, or no one at all as author, this overlooks an intriguing possibility in cases where a generative piece is produced in the style of a particular human artist: that original artist as author. Whatever other conditions may suffice for authorship being credited thus, in a case where a generative work is indistinguishable from other works produced by the artist in question, this possibility gains real traction. At least in the case of novels, there is reason to think that the appropriate way to attribute authorship in these cases is to simply attribute authorship to the human whose work the system was tuned to and trained on. That is, if a generative novel is written in the style of a human author, and it is indistinguishable even by competent judges from other works by that author, then it is, itself, by that author. This is a way to respect the text speaking for itself, and to dismiss anything that is not relevant to the work. Although there are some counterintuitive consequences to accepting human authors' ability to "produce" novels long after their death, or without their knowledge, or in ways utterly separate from the traditional causal chain linking human and work, none of these succeed in undermining the reasons for believing that these are the work of the human author in question. This view of authorship has a number of implications for human agency, intellectual property law, and discussions of artist attribution in generative art, and merits further investigation.

10. References

[1] Linson, A. (2016). Machine Art or Machine Artists?: Dennett, Danto, and the Expressive Stance. *Fundamental Issues of Artificial Intelligence*, 443-458.

[2] Ghosh, A., & Fossas, G. (2022). Can there be art without an artist?. *arXiv preprint arXiv:2209.07667*.

[3] Browne, K. (2022). Who (or what) is an AI Artist?. *Leonardo*, *55*(2), 130-134.

[4] Barthes, R. (1967). The Death of the Author*, Aspen 5-6*.

[5] Helle, S. (2019). What Is an Author? Old Answers to a New Question. *Modern Language Quarterly*, 80(2), 113–139.

[6] Deltorn, J. M., & Macrez, F. (2019). Authorship in the Age of Machine learning and Artificial Intelligence, in *The Oxford Handbook of Music Law and Policy.* Oxford University Press.

[7] Ginsburg, J. C., & Budiardjo, L. A. (2019). Authors and machines. *Berkeley Tech. LJ*, *34*, 343.

[8] Noë, A. (2000). Experience and Experiment in Art. *Journal of Consciousness Studies*, *7*(8-9), 123-35.

[9] Jones, R. & Penny, M. (1983). *Rafael*. Yale University Press.