AI-Assisted Authorship

How to Assign Credit in Synthetic Scholarship

Authored by: Ryan Jenkins, PhD
Patrick Lin, PhD
California Polytechnic State University
Ethics + Emerging Sciences Group
San Luis Obispo, California

GPT-3.5
OpenAI
San Francisco, California

Significant revision: 30 January 2023

Version: 1.0

Suggested citation: Jenkins, Ryan and Patrick Lin. “AI-Assisted Authorship: How to Assign Credit in Synthetic Scholarship.” Report. Ethics + Emerging Sciences Group. 2023.

Abstract

This report proposes principles for determining when it is required to credit an artificial intelligence (AI) writer for its contributions to scholarly work. We begin by critiquing a policy recently published by the journal *Nature*, which forbids acknowledging AI writers as authors. We question the justification and breadth of this policy. We then suggest two fundamental considerations that we think are more relevant: continuity (how substantially are the contributions of AI writers carried through to the final product?), and creditworthiness (would this kind of product typically result in academic or professional credit for a human author?). We draw upon brief reflections on the nature and value of authorship to justify these considerations. This report provides a starting point for academics and the broader scholarly community in the emerging debate on determining when and how to credit AI writers' contributions.

01

Introduction

New technologies often provoke a renegotiation of the norms and expectations of a profession. This conversation is taking place now in academia as prompted by recent advances in generative artificial intelligence (AI) applications such as GPT-3, ChatGPT, and others. This includes not just challenging concepts of authorship and the value of human labor, but it's also prompting a reexamination of morally laden expectations about claiming credit for one's contributions. In this brief report, we consider when it is required to credit artificial intelligence for its contributions to scholarly work.[1,2]

The recent work of generative AI has been impressive and, in many cases, simply stunning. Applications like GPT-3 are capable of writing not just passable but convincing prose, including opining about sophisticated issues and demonstrating a degree of apparent creativity that has erstwhile been the sole domain of humans. To be sure, these AI writers also generate statements that are nonsensical as well as flat-out false; there's little evidence that they understand anything rather than merely predicting the next word based on historical patterns in vast amount of human-created content. This significant limitation immediately points to the need for very close human oversight and caution in employing the technology for research purposes.

Clearly, AI writers pose vexing challenges. Because these passages are generated by AI, rather than copied from other sources, they are highly resistant to traditional plagiarism detection. Because these tools can assist with a wide range of tasks, and even compose finished work, they are enticing. The common expectation that AI systems be transparent applies to writers, too: it's plausible that humans have an obligation to disclose, at least in some cases, that they are using AI as an assistant or a coauthor.

Nature's Position

This had led at least one notable venue — *Nature* — publishing "ground rules" that outright forbid crediting AI writers such as

ChatGPT as an author, instead requiring them to be listed in an acknowledgement section. We think this is hasty and seek to add some nuance to this discussion.

For instance, *Nature* argues that crediting AI writers in the acknowledgements serves the goal of transparency.[3] While this may be true in many cases, it could also help to hide or grossly understate the role and substantial contributions of AI writers to the paper, which is *counterproductive* to transparency.

Nature also argues AI writers should not be credited as authors on the grounds that they cannot be accountable for what they write. This line of argument needs to be considered more carefully. For instance, authors are sometimes *posthumously* credited, even though they cannot presently be held accountable for what they said when alive, nor can they approve of a posthumous submission of a manuscript; yet it would clearly be hasty to forbid the submission or publication of posthumous works.

Thus, a more nuanced, middle-ground solution may be needed, as satisfying as a simple policy might be. As the writer H.L Mencken observed, "For every complex problem, there is an answer that is clear, simple, and wrong." Exactly what form that acknowledgment takes, and when it is required, are questions we take up here.

This moment recalls Plato's story of the Ring of Gyges: the sudden acquisition of a new superpower (invisibility) that could be used to carry out one's darkest wishes, free from detection. In such moments, we must rely on our ethical deliberations to guide our behavior. There has been a recent explosion of attention and discussion about this issue. Our hope is to offer some analysis that's too nuanced to fit into social media posts or discussion boards.

Accordingly, this report seeks to clarify and scaffold ongoing deliberations about acknowledging the contributions to research of AI. We are not advocating for AI writers, such as GPT-3, to be used in research, but if the technology is used materially in research, and assuming it can be used responsibly at all, we want to explore the shape of the deliberation when it comes to assigning credit to AI writers. In the interest of transparency and honesty: When should academics credit artificial intelligence for its contributions to their work?

We do not address the aesthetics or value of AI-generated art; the legal or moral implications of training generative AI systems on copyrighted data; the ethics of using AI writers that require the labor of "ghost workers," an invisible underclass of exploited and vulnerable humans; the question of who owns the copyright to publications generated in part by AI

writers[4]; the permissibility of students using AI to write essays for class; and other questions we have been broached recently. Those questions obviously merit discussion and have stimulated soul-searching on the part of academics. But in the interest of space, we address ourselves to just one of these questions — still, one that we take to be urgent.

We recognize that our recommendations below are influenced by our disciplinary background as philosophers. Academic expectations and norms will differ by discipline — and further by department within disciplines, as evidenced by the animated and occasionally passionate debates that have blossomed within our own department recently. We do not take our suggestions to be the last word, but rather we intend only for this proposal serve as a spur to discussion within the academic community.

02

Research Tasks and *Continuity*

The first and most obvious consideration is that AI can be used for different kinds of tasks. For example, AI might be used:

- as a landscaping tool to generate ideas;
- to quickly survey existing literature;
- to find connected or similar papers as part of the research process; or
- to offer suggestions on grammar, style, and tone.

Or, it might be used to generate long passages which survive relatively unchanged into the final product. These might include passages that serve as summaries, such as abstracts, introductions, or conclusions — or they might include explications of and defenses of crucial premises in a broader argument.

When considering this broad range of tasks, **it's helpful to think of any preexisting practices that might serve as apt metaphors for the contributions of AI**. Metaphors help to connect the unfamiliar to the familiar, providing a bridge for understanding the former. While metaphors are by their nature imperfect, they can nevertheless be useful in framing a discussion or revealing hidden dimensions of an issue.

At the "lesser" end of this spectrum of metaphors, we can imagine consulting AI in the way an author would consult Google Scholar, Bing, and so on. These are AI-powered research tools which can inform the research process, including which questions are worth investigating, which arguments exist to reply to, which literature is most active, and so on.

All of us consult search engines at some point in the research process; none of us credit them. Instead, we credit *the source itself*: a reflection of the fact that the search engine only *retrieves*, it does not *generate*. The search engine is contributing nothing beyond the original source. The author's access to the knowledge contained in the original source is *mediated* through the search engine only in a *minimal sense*, if at all.

(We might try to extend this analogy by considering generative AI as a *next-generation search engine*, since it is basing its responses on content that humans *have* generated, but this seems to strain the analogy.[5] Note, also, that this analysis is challenged by other tools which combine the functions of search and AI-powered summary, providing literature reviews combining many sources at once.)

What about when AI is generating content *ex nihilo* — when they perform a more active mediating role in the process of composition? As we move further down this spectrum, AI begins to resemble a capable student assistant, who might be tasked with providing a preliminary landscape research document, brainstorming understudied questions and arguments, finding sources, or even critiquing arguments, objections, evidence, and so on. Many of us benefit from the help of student assistants, and crediting them is appropriate when their contributions are substantive and novel.

The relevant dimensions that separate some of these tasks into acceptable or unacceptable include the degree to which the contributions of AI make it into the final form of the deliverable. We call this consideration **continuity**. The closer the resemblance, and the more substantive the ideas that AI contributes, the more important it is to clearly credit its contributions. A paper single-authored by AI — from title and abstract, through arguments, applications, and conclusion — is the epitome of this. The opposite end of the spectrum would be populated by *ad hoc* consultations with AI for suggestions throughout the earliest phases of research, which inform and guide the direction of research, but which are not included in the final product.

03

Kinds of Products and *Credit*

A further consideration is **the kind of final product that AI is used for**. Some of these are clearly more problematic than others.

The adage to *give credit where credit is due* is borne out here: if someone is receiving credit for this product, then it should be clear *how* that credit should be apportioned to the human authors.

The applications here which seem unproblematic are those that don't redound to the credit or reputation of a particular person, which are not implied to include creative or original work. Authors are expected to include a list of works cited at the end of their works, and these are often compiled automatically today. But these are compiled deterministically, with no creative element, and this does not reflect on the ingenuity of the author.

One potential distinction would be to require crediting AI in products that are ultimately public-facing and intended for distribution, such as scholarship, public media articles, academic reports, and so on. Other materials, such as case studies to be distributed at a workshop, or a "biosketch" to appear on a program, do not need to include such a credit. However, it *would* seem apt to credit AI for its contribution to a grant proposal, even though those are generally not widely distributed. And conversely, it does *not* seem as crucial to credit AI for its contributions to in-class activities or assessments, whose primary audience is a classroom of students, but which might still be shared widely.[6] So, this distinction does not hold up.

The more important consideration seems to be **the extent to which the product redounds to the author's reputation as a scholar.** This is the primary function that credit plays in an academic context, though it also plays an epistemic role by giving the reader information that helps them judge the trustworthiness of the content *given* its authorship. For many of the products we consider in this report, their value depends in large part on the knowledge or expertise of the author — and this concern is underscored while AI writers are constitutionally challenged

when it comes to distinguishing fact from fiction. All the more important to know whose ideas are represented in the final product.[7] Together, we call these considerations **creditworthiness**.

As a final point, note that our aims here are only to examine when it may be appropriate or required by responsible researchers to acknowledge their use of AI writers for their contributions to a project — up to crediting the AI writer as a co-author or even a sole author. We are not developing a theory of authorship that, e.g., implies responsibility for content and accuracy or implies a certain epistemic standpoint and expertise. A fuller exploration of the nature and value of authorship would, of course, take much more time. Still, it is clear that the current discussions over AI writers cannot be resolved until there is broader agreement and understanding about the nature of authorship and its social value.

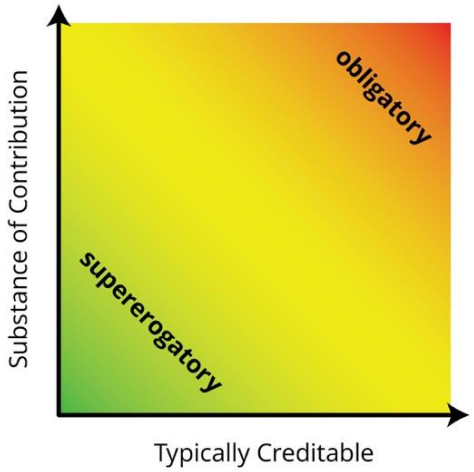
04

Proposed Principles

We suggest that the following are fundamental considerations for determining when it is required to credit the use of AI in scholarly writing, rather than being *supererogatory*, or going "above and beyond" what is required:

1. Continuity. **How substantially are the contributions of AI writers carried through to the final product?** To what extent does the final product resemble the contributions of AI? What is the relative contribution from AI versus a human? The calculations are always difficult, even if the coauthors are human. Some journals routinely require statements of relative contribution to add clarity and nuance when multiple humans are sharing credit.

2. Creditworthiness. **Is this the kind of product a human author would normally receive credit for?** Consider whether the AI's contributions would typically result in academic or professional credit for a human author.



This analysis is similar to how we view student assistants: the greater the substance of their contribution to the final product, and the greater the extent to which this kind of product typically redounds to the credit of the author, the more important it is to credit the range of contributors, both human and artificial. As for what "credit" amounts to, this varies by discipline more so than the other intuitions we have surveyed above. Whether that credit takes the form of co-authorship or acknowledgement is a matter to be investigated within disciplines, though the above guidelines should provide a clear starting point, especially the notion of what kinds of products and contributions are typically creditable in a discipline.

Endnotes

[1] Some papers have already begun to credit ChatGPT as an author. See, for example, Kung, Tiffany H., et al. “Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models.” medRxiv, 21 Dec. 2022. medRxiv, https://doi.org/10.1101/2022.12.19.22283643. See also coverage in *Nature*: Stokel-Walker, Chris. “ChatGPT Listed as Author on Research Papers: Many Scientists Disapprove.” *Nature*, vol. 613, no. 7945, Jan. 2023, pp. 620–21. www.nature.com, https://doi.org/10.1038/d41586-023-00107-z.

[2] Nothing we say here is meant to preempt the decisions of publishing venues, some of which have already taken public stances on the use and acknowledgment of AI writers. Authors should check with each venue’s policy on AI writers, as some or many may not allow AI to be listed as an author but might require or allow other forms of acknowledgement.

[3] “Tools Such as ChatGPT Threaten Transparent Science; Here Are Our Ground Rules for Their Use.” *Nature*, vol. 613, no. 7945, Jan. 2023, pp. 612–612. www.nature.com, https://doi.org/10.1038/d41586-023-00191-1.

[4] OpenAI says that they will not claim copyright over content generated by its API, which suggests that the human users who prompt GPT and receive its output are the sole owners. Other organizations and products might have different terms, of course, and these terms might change in the future, which would complicate the ability of journals to secure the rights to publish and distribute academic work produced in part by AI writers. See, “Will OpenAI Claim Copyright over What Outputs I Generate with the API?” OpenAI.com, https://help.openai.com/en/articles/5008634-will-openai-claim-copyright-over-what-outputs-i-generate-with-the-api. Accessed 16 Jan. 2023.

[5] Nonetheless, Google seems to think that these technologies represent an evolution of search. Grant, Nico, and Cade Metz. “A New Chat Bot Is a ‘Code Red’ for Google’s Search Business.” The New York Times, 21 Dec. 2022. NYTimes.com, https://www.nytimes.com/2022/12/21/technology/ai-chatgpt-google-search.html.

[6] As mentioned elsewhere in this report, we expect our intuitions about professional norms to differ between disciplines and practitioners. Excavating these norms through ethnographic research and surveys will take a fair amount of work but is a crucial next step for providing a scaffold within which explicit expectations for professional practice can coalesce. As with

international customary law, a norm can be a powerful thing in crafting realistic policy, and emerging practices are relevant to norms.

[7] This consideration will plausibly vary across domains and disciplines.

About the Authors

Ryan Jenkins, PhD, is an associate professor of philosophy and a senior fellow at the Ethics + Emerging Sciences Group at California Polytechnic State University in San Luis Obispo. Dr. Ryan Jenkins is an associate professor of philosophy and a senior fellow at the Ethics + Emerging Sciences Group. His research focuses on the potential for emerging technologies to enable or encumber meaningful human lives — especially artificial intelligence, cyber war, autonomous weapons, and driverless cars. Dr. Jenkins has affiliations with the Center for Advancing Safety of Machine Intelligence (CASMI) at Northwestern University and the Karel Čapek Center for Values in Science and Technology in Prague. He is a former member of the IEEE TechEthics Ad Hoc committee and a former co-chair of the Robot Ethics Technical Committee of the IEEE's Robotics & Automation Society. He has served as a principal or senior investigator for several grants on the ethics of autonomous vehicles, predictive policing, and cyberwar. His work has appeared in journals such as *Techné*, *Ethics and Information Technology*, *Ethical Theory and Moral Practice*, and the *Journal of Military Ethics*, as well as public fora including the *Washington Post*, Slate and Forbes. His interviews have appeared in *The New Yorker*, *Inc. Magazine*, Engadget, NPR, and elsewhere. He received his PhD in Philosophy from the University of Colorado Boulder.

Patrick Lin, PhD, is the director of the Ethics + Emerging Sciences Group at California Polytechnic State University, where he is a full philosophy professor. He is currently affiliated with Stanford Law School, the 100 Year Study on AI, Czech Academy of Sciences, Center for a New American Security, and World Economic Forum. Previous affiliations include: Stanford's School of Engineering, US Naval Academy, Univ. of Notre Dame, Dartmouth, UNIDIR, and the Fulbright specialist program (Univ. of Iceland). Prof. Lin is well published in technology ethics and experienced in managing sponsored projects—incl. on AI, robotics, autonomous driving, cybersecurity, bioengineering, frontier development, nanotechnology, security technologies, and more—and is regularly invited to provide

briefings on the subject to industry, media, and government. He teaches courses in ethics, philosophy of technology, and philosophy of law, and he earned his BA from UC Berkeley and PhD from UC Santa Barbara.

Acknowledgments

We owe thanks to Daniel Story for comments on an earlier draft of this work, specifically his suggestion that credit plays an epistemic as well as a reputational role in scholarship.

AI Writing statement: This report was generated in part by OpenAI's GPT-3 algorithm, including idea generation, summary, and drafting the abstract. Approximately 5% of the final product reflects these contributions.

About the Ethics + Emerging Sciences Group

Established in 2007 at California Polytechnic State University (Cal Poly), San Luis Obispo, the **Ethics + Emerging Sciences Group** is a non-partisan organization focused on risk, ethical, and social concerns related to new sciences and technologies. As a research and educational group, we are involved with ethics and risk assessment, course development, publishing projects, media outreach, public lectures, and other activities to engage policymakers, business, academia, as well as the broader public on key issues in science and society.

Contact: ryjenkin@calpoly.edu
Website: http://ethics.calpoly.edu/

Ethics + Emerging Sciences Group