Reflections

AI-Based Text Generation and the Social Construction of "Fraudulent Authorship": A Revisitation

Chris M. Anson

Student plagiarism has challenged educators for decades, with heightened paranoia following the advent of the Internet in the 1980's and ready access to easily copied text. But plagiarism will look like child's play next to new developments in AI-based natural-language processing (NLP) systems that increasingly appear to "write" as effectively as humans. How we theorize and contextualize these developments will guide the way we meet their challenges in all courses where writing is assigned and evaluated

Here, I first revisit an article I wrote for *Composition Studies* in 2011, "Fraudulent Practices: Academic Misrepresentations of Plagiarism in the Name of Good Pedagogy." In that article, I argued that what counts as plagiarism in some contexts occurs with impunity across a wide range of published material. This is because definitions of plagiarism are socially constructed and tied to context-sensitive cycles of reward for the production—and therefore the ownership—of certain kinds of texts. Helping students to understand plagiarism means showing them these contextually-specific constructs of text ownership, rather than assuming that any unattributed text, published anywhere, in any form, constitutes plagiarism. I then turn to AI-based NLP systems. Teachers who learn what these systems can do usually respond with the same hand-wringing and defensive posture triggered by concerns about student plagiarism. But a social-practices interpretation again breaks open the systems' perceived threats and reveals a more nuanced and contextual approach to their challenges—and their potential acceptance and use—alongside writing produced by humans.

Digging Critically into Plagiarism: A Return to 2011

At the time I wrote "Fraudulent Practices," I was keenly interested in the social construction of plagiarism, partly because plagiarism occupies so much attention across most educational contexts and needs deeper levels of analysis. In that piece, I used Internet research to show that beyond academia, plagiarism is rampant but often accepted. For example, information published by Federal agencies concerning the safe handling of meat or what to do when a tornado approaches is liberally copied verbatim at dozens of sites (including official state and municipal sites) without attribution. The need to circulate the material for the public good outweighs the need for authorial credit (even if a specific author, rather than a bureaucratic entity, can be found). Car dealerships replicate material from automobile insurance agencies—which replicate each other's material—about how to steer out of a skid, sometimes verbatim and sometimes patched into other text, again without any indication of authorship, to the point where the original source is unrecoverable. Hotel booking companies use property descriptions commissioned and published by the corporation that owns the properties but with no trace of the texts' origin—a process that, because the replicated text benefits both the booking agents and the hotel corporation, I dubbed "cooperative competition." This practice is also common across vast numbers of sites that broker the sale of products whose descriptions are extensively copied without indication of the original source (presumably commissioned or written by the manufacturer). Another piece I co-authored with an experienced military officer, "The Army and the Academy as Textual Communities," documented the widespread copying and reuse of text, either verbatim or repurposed, across the United States Army without acknowledgement of its original author(s) (Anson and Neeley).

In these and other cases, I drew on the New Literacy Studies (Gee; Russell; Street) to argue that what would count as plagiarism in some contexts is tolerated or even encouraged in others. Most importantly, the view that it is plagiarism to use someone else's text without attribution is perpetuated in contexts such as academia where authors accrue "credit and credibility" from their texts: publishing under one's name earns credit; as credit accumulates, so does greater credibility, which opens up further opportunities for more credit in a kind of cycle (see Latour and Woolgar). In this system of rewards, text ownership and attribution are sacrosanct, and "theft" is forbidden. But if little or no credit is associated with text, its authors have no concerns about it being co-opted and used by others: many scholars, myself included, urge colleagues to take material from their syllabi, assignments, and other teaching-related documents without worry about attribution, which the borrowers want to avoid as well so as not to appear lazy or unable to produce it on their own. "Credit" for original text, especially at research-extensive institutions, comes more strongly from peer-reviewed publications than teaching or administrative materials.

"Fraudulent Practices" ends with a call to deepen our discussion of plagiarism with students and to represent it more accurately based on communities of practice and their norms:

> [A]mong all aspects of rhetoric and written communication, representations of plagiarism often suffer the most from a kind of pedagogical myopia, and it is curious that we would deliberately conceal

> the truth about how sources are or are not attributed in the world of discourse in order to compel students to believe in a specific perspective, even if just for the time being. When we show students the range of textual and discursive practices used in various contexts, and help them to understand the relationship between these practices and their underlying social and ideological sources, students begin to see plagiarism not as "rules" to be memorized uncritically and without regard to situation, but as socially constructed practices of utmost importance to the academic community they have joined. (p. 40)

When we consider the possibility that students could generate entire authentic-looking essays from a small input into an NLP program, will teachers represent the systems myopically and conceal the truth about their affordances? What would it look like to "show the range of textual and discursive practices used in various contexts" that such programs can enable?

The New "Threat": AI-Based Text Production

Recent advances in the production of natural language—text that mirrors the characteristics of writing by humans—now represent the next big perceived threat to teachers of writing. These deep-learning systems work by mining billions of existing, human-produced texts and, through a process akin to the creation of neural networks, generate new text based on highly complex predictive machine analysis. A system being widely tested and refined is GPT-3, the third iteration of OpenAI's "generative pretrained transformer." GPT-3 operates with 1.3 billion parameters, a technical term for the range of values the system can have for appropriate predictability. For an example of a primitive version of GPT-3 also developed by OpenAI, consider Gmail's Smart Compose function. When activated, Smart Compose suggests further words as the user types. Drawing on millions of other texts, the app predicts the writer's thoughts on a simple level. For example, at the end of an email response to a student's query about an assignment, I wrote the letters "Ple . . ." and based on the content of my email, Smart Compose offered "Please let me know if you have any questions." The text was appropriate, so I hit the tab key, and it was instantly added to my email.

As a machine-learning program, Smart Compose continues to get better from its mistakes. But GPT-3 is a far more complex and robust system; it is capable not just of filling in words at the start of email sentences but generating entire texts or adding text that's completely coherent with existing text (for a stunning example of an essay about these programs, with text generated by

GPT-2 interspersed into the author's, see Seabrook). GPT-4 will be available in 2023, and Google is working on a similar system with 1.6 trillion parameters.

To consider how GPT-3 (or GPT-4, or Google's system) might subvert students' authentically composed writing, let's experiment with an essay topic of current interest in the context of the Covid pandemic: *vaccine hesitancy*. Using my access to GPT-3's playground, I typed "Write an essay about vaccine hesitancy" into a text box. In about four seconds, the system generated the following text.

> Vaccine hesitancy is a complex and multi-faceted problem. On the one hand, there is a clear public health benefit to vaccinating as many people as possible. Vaccines are one of the most effective ways to prevent the spread of disease. They have been responsible for saving millions of lives and eradicating devastating diseases like smallpox. On the other hand, there is a legitimate concern about the safety of vaccines. Vaccines are made from viruses or bacteria that have been weakened or killed. They are then injected into people, which causes the body to develop immunity to the disease. However, there is always a risk that the vaccine could cause serious side effects, including allergic reactions, fever, and even death.
>
> The problem of vaccine hesitancy is compounded by the fact that vaccines are not 100% effective. They may not work for everyone, and even when they do work, they may not provide complete protection from the disease. This means that even people who are vaccinated can still contract the disease and spread it to others.
>
> The best way to address vaccine hesitancy is to educate people about the risks and benefits of vaccination. It is important to dispel the myths and misinformation that circulate about vaccines. It is also important to encourage people to speak to their healthcare providers about their concerns and to get accurate information from credible sources.

Cutting and pasting all or part of this auto-generated text into a new text box can continue to prompt further content. The material is not plagiarized from existing sources but freshly composed. For this reason, it usually evades plagiarism-detection programs like Turnitin.[1]

GPT-3 is also capable of adding to already produced material, including fiction. In an attempt to do so, OpenAI has collaborated with the "think-of-an-idea-and-we'll-finish-it" website Inklewriter. Using GPT-3, Inklwriter's computer program can now write the next few paragraphs of a story and generate plots with a range of possibilities. (The last two sentences of this paragraph were written by GPT-3 after I gave it the first sentence—but for you, no doubt undetectably.)

Although imperfect, these examples show that even at this stage of development—with sweeping improvements on the near horizon—GPT-3 can serve as an effective “writing assistant.” With a few minor revisions, some fact-checking, and a bit of editorial work, students can easily pass off what it generates as their own original compositions.

A Social Practices View of AI-Based Text Generation

Let’s imagine that a student generates the content of an essay using a “robo-writer” like GPT-3 and submits the results as their own original composition. By most academic standards, this would not technically be a case of plagiarism—until or unless computers are considered authors—but instead would represent “contract cheating” (Curtis and Clare; Lancaster and Clarke), a violation of student codes of conduct similar to submitting a customized paper written by someone else. However, when we consider contexts beyond teaching and learning, NLP-written texts may not look like ethical violations, in much the way that vast amounts of replicated and unattributed text do not constitute plagiarism. Consider boilerplate—text or bits of text used repeatedly in forms, letters, emails, and legal documents. At base, boilerplate is text “auto-generated” for subsequent texts. Once written, it behaves as if a machine has produced it, easily spliced into new texts or simply sent or submitted verbatim, with perhaps different addressees and subject headings. Text-expanders like aText facilitate the use of boilerplate by allowing prewritten text to be inserted into new text with the use of a simple command. The rules and conventions of discourse in many contexts not only tolerate but encourage this repeated use of the same textual material for work efficiency. (For example, I sometimes use aText to splice frequently repeated advice into students’ papers, with their knowledge.)

Boilerplate is the reuse by a human of existing text written by a human. But AI-based NLP systems create new, original text each time they are activated. When I asked GPT-3 to “write a tagline for a writing center,” it offered these suggestions:

> The Writing Center: Your words, our passion
>
> The Writing Center: Where great writers are born
>
> Need help with your writing? We’re here for you!
>
> The Writing Center: Your Home for Writing Help

At the least, these auto-generated examples might spark some new human-generated ideas—or perhaps one of them would be sufficient. In this sense, NLP systems can serve as invention prompts, or as text generators whose outputs are either revised or used verbatim. Like the acceptable use of unattributed text, this range of writing processes widens our conception of what is or is not permissible or ethical about NLP-based text production. Instead of darkly subverting the human invention and writing process, how might NLP systems support the disparate goals and activity systems of different contexts?

Among hundreds of examples (some at OpenAI's site), here are a few:

- Translation of legalese
  GPT-3 will take the input of complex legal language—the kind most people simply "accept" without reading when they download and activate an app—and make it understandable. Inputting several paragraphs of Verizon's Disclaimer of Warranties, which includes lines such as "Verizon hereby disclaims any and all representations, warranties, and guaranties regarding S & P, whether express, implied, or statutory . . . ," yields a one-sentence explanation: "This is a disclaimer of warranties. It basically says that the company is not responsible if something goes wrong with the product or service."

- Extraction of keywords
  Inputting the contents of this article produced a list of nineteen keywords that included *plagiarism*, *authorship*, *pedagogy*, *deep learning*, *GPT-3*, and *contract cheating*—from which, if required, I could select four or five for publication.

- Product name generator
  Inputting my made-up product description ("an app that tells you when to stop eating") yielded the product names *Overeater's Helper, Full Stop, Bite Counter,* and *Eat-o-Meter*.

- Abstract creation
  After finishing a draft of this essay, I pasted it into GPT-3 with the request to create an abstract, and it offered me the following: "This essay revisits the idea that plagiarism is a social construct, and explores how AI-based text generators fit into that construct. It argues that instead of trying to prevent students from using these tools, we should instead teach them how to use them responsibly."

Although I might add that the article revisits a previous article published in *Composition Studies*, the gist of the summary is accurate and provides the basis for a slightly tweaked or expanded version.

Some applications of NLP systems may seem acceptable for mundane and repetitive tasks, especially in a work context. But where does acceptability end? In my graduate course in writing studies, students must interview a scholar about their work. Imagine that a student asks GPT-3 to create a list of possible questions for an interview with writing-studies and plagiarism expert Rebecca Moore Howard. When I gave GPT-3 this task ("Write a list of questions for an interview with Rebecca Moore Howard"), it created questions such as "What inspired you to write your book on plagiarism?" and "What do you think are the best ways to prevent plagiarism?"

Auto-generated interview questions may be acceptable to some teachers, but what about a brief biography spliced into the start of the interview paper? Or imagine that a student (or scholar) asks GPT-3 to write the methods section of an article based on phrasal inputs. When I gave GPT-3 ten fragments from an imagined study of middle-school students about their reading habits, it produced a coherent methods description that began as follows:

> The present study sought to understand middle school students' reading habits. A total of 180 surveys were distributed to middle school students at a public, urban school located in the Southeast, with 146 surveys being usable. The survey had 15 questions and asked students about where they most often read, how often they read, and what types of books they most enjoyed reading.

In this case, I provided the information, but who lays claim to the expression of that information in a coherent string of sentences?

Many researchers in STEM fields might find auto-written methods sections acceptable, just as many STEM editors and researchers find the verbatim replication of methods sections in different published articles acceptable in a process my co-researchers and I have called "text recycling" (see Anson, Hall, Pemberton, and Moskovitz; Pemberton, Hall, Moskovitz, and Anson; and Anson, Moskovitz, and Anson). But in other contexts, such as when an author, in an article for *PMLA*, replicates three paragraphs from their previous article in *Shakespeare Quarterly* describing the methods used in interpreting the theme of resentment in select Shakespeare plays, both text recycling and NLP-generated text might be disallowed. In a college biology course, an instructor might accept an NLP-generated lab report as long as the student conducted the lab and entered all the correct information, while another instructor might refer the student to the Office of Student Conduct. In the first case, the learning goal might focus mainly on the experimental processes (the procedures and results), while in the second the goal might focus more strongly on the writing process (assembling the procedures and results in coherent prose). If these situations yield mixed opinions in academic settings, it is not difficult to imagine many other contexts where automatic text generation will be welcomed—or conflicted.

Sharing AI-Generated Text Production Systems with Students

Instructional responses to the prospect of student access to NLP-based text generation systems often focus, like plagiarism, on detection and prevention. Solutions include having students write at computers that block access to the Internet, write by hand in class, include references to discussions in class (to which AI systems don't have access), keep a process log of everything done to write a paper, and build the paper from a series of scaffolded writing activities (see Anson, "Cops"; "Defining"; Vie) But because automated writing systems are here to stay and will only improve over time, a more sensible approach could involve embracing the technology, showing students what it can and can't do, and asking them to experiment with it. As I argued in "Fraudulent Practices," doing so is supported by the "writing about writing" approach to first-year composition (see Downs and Wardle):

> In such a course, students could learn about or even study contexts for writing in order to deepen their understanding of the assumptions, processes, tools, values, discursive histories, and social practices that entail there. The resulting metaconsciousness would be far preferable for students who move into and among different activity systems than sets of isolated skills, such as learning how to write topic sentences. (p. 40)

For example, the process of exploring GPT technology is artfully demonstrated in an assignment developed by Paul Fyfe and shared in "How to Cheat on your Final Paper: Assigning AI for Student Writing." Fyfe asked undergraduates in his course to "harvest content from an installation of GPT-2" and then incorporate the material into their final essay. However, the students were required to highlight which content was theirs and which was auto-generated and then reflect on the results. Their reflections focused on the ethics of AI assistance, what the program did to extend their own perspectives, and how the material might or might not be considered plagiarism. The shared insights of the students are impressive and point to the broader goal of teaching discourse in all its complexities and contextual variations.

In addition, students need to learn the sinister side of NLP systems. Because they generate text based on what humans have already produced, the systems are prone to mirroring discriminatory and racist language and perpetuating stereotypes (such as assuming that roles like "flight attendant" or "nurse" are always performed by women). As Hutson points out, the systems can also support the insidious work of extremist communities, "produc[ing] polemics parroting Nazis, conspiracy theorists and white supremacists" (24). Helping students to navigate the ethics of use can only prepare them to make wise decisions about their own writing and the writing of others, including machines.

As David Russell has put it, "because writing…is a matter of learning to participate in some historically situated human activity that requires some kind(s) of writing, it cannot be learned apart from the problems, the habits, the activities—the subject matter—of some group that found the need to write in that way to solve a problem or carry on its activities" (194). With this bit from Russell, GPT-3 relieved me of composing a conclusion: In other words, we need to help students understand that the act of writing is always situated within a complex system of rules, assumptions, and values. AI-based text generation systems are just one more element in that system, and one that is likely to become more commonplace in the years to come. As such, it is important that we help students to understand how such systems work and how to use them responsibly.

Notes

1. Because AI-based systems scrape existing digital text to assemble new text, there is a small possibility that a string of words could get flagged by a plagiarism-detection system, but as Dehouche suggests, it is unlikely: "Our medieval concept of plagiarism (Sadeghi 2019) ('presenting the work of others as one's own') appears rather inadequate when the 'others' in question consist in an astronomical number of authors, whose work was combined and reformulated in unique ways by a 175-billion-parameter algorithm" (21).

Works Cited

Anson, Chris M. "We Never Wanted to Be Cops: Plagiarism, Institutional Paranoia, and Shared Responsibility." *Pluralizing Plagiarism: Identities, Contexts, Pedagogies*. Ed. Rebecca Moore Howard and Amy E. Robillard. Boynton/Cook-Heinemann, 2008, pp. 140-157.

—. "Fraudulent Practices: Academic Misrepresentations of Plagiarism in the Name of Good Pedagogy." *Composition Studies*, vol. 39, no. 2, 2011, pp. 29-43.

Anson, Ian G., Cary Moskovitz, and Chris M. Anson. "A Text-Analytic Method for Identifying Text Recycling in STEM Research Reports." *Journal of Writing Analytics*, vol. 3, 2019, pp. 125-150.

Anson, Chris M., Susanne Hall, Michael Pemberton, and Cary Moskovitz. "Reuse in STEM Research Writing: Rhetorical and Practical Considerations and Challenges." *AILA Review*, vol. 33, 2020, pp. 120-135.

Anson, Chris M., and Shawn Neely. "The Army and the Academy as Textual Communities: Exploring Mismatches in the Concepts of Attribution, Appropriation, and Shared Goals." *Kairos*, vol. 14, no. 3, 2010, n. pag. Web.

Curtis, Guy J., and Joseph Clare. "How Prevalent is Contract Cheating and to What Extent are Students Repeat Offenders? *Academic Ethics*, Vol. 15, 2017, pp. 115–124. DOI 10.1007/s10805-017-9278-x

Dehouche, Nassim. "Plagiarism in the Age of Massive Generative Pre-trained Transformers (GPT-3)." *Ethics in Science and Environmental Politics*. Vol. 21, 2021, pp. 17-23. DOI 10.3354/esep00195

"Defining and Avoiding Plagiarism: The WPA Statement on Best Practices. Council of Writing Program Administrators. 2019. https://wpacouncil.org/aws/CWPA/pt/sd/news_article/272555/_PARENT/layout_details/false

Downs, Douglas, and Elizabeth Wardle. "Teaching About Writing, Righting Misconceptions: (Re)Envisioning 'First Year Composition' as 'Introduction to Writing Studies.'" *College Composition and Communication,* vol. 58, no. 4, 2007, pp. 552-84.

Fyfe, Paul. "How to Cheat on your Final Paper: Assigning AI for Student Writing. *AI & Society*, Feb. 2022. https://doi.org/10.1007/s00146-022-01397-z

Gee, James Paul. "The New Literacy Studies." *The Routledge Handbook of Literacy Studies*. Ed. Jennifer Rowsell and Kate Pahl. Taylor and Francis, 2015.

Hutson, Matthew. "Robo-Writers: The Rise and Risks of Language-Generating AI. *Nature*, vol., 591, no. 7848, 2021, pp. 22-25.

Lancaster, Thomas, and Robert Clarke. "Contract Cheating: The Outsourcing of Assessed Student Work. *Handbook of Academic Integrity.* Ed. Tracey Bretag. Springer, 2015. 639-654.

Latour, Bruno, and Stephen Woolgar. *Laboratory Life: The Construction of Scientific Facts*. Princeton UP, 1979.

Pemberton, Michael, Susanne Hall, Cary Moskovitz, and Chris M. Anson. "Journal Editors' Views on Text Recycling: An Interview-Based Study." *Learned Publishing*, vol. 32, no. 4, 2019, pp. 355-366.

Russell, David R. "Vygotsky, Dewey, and Externalism." *JAC: The Journal of Advanced Composition,* vol. 13, no. 1 (1993): 173-197.

Sadeghi, Ramin. "The Attitude of Scholars Has Not Changed Towards Plagiarism Since the Medieval Period." *Research Ethics*, Vol. 15, 2019, pp. 1–3.

Seabrook, John. "The New Word: Where Will Predictive Text Take Us? *The New Yorker*, 14 Oct., 2019. https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker

Street, Brian. *Social Literacies: Critical Approaches to Literacy in Development, Ethnography, and Education*. Longman, 1995.

Vie, Stephanie. "A Pedagogy of Resistance Toward Plagiarism Detection Technologies." *Computers and Composition*, vol. 30, no. 1 (2013): 3-15.