

# ChatGPT (and Its Relatives) and College Writing: A Quick Guide for Students

Large Language Models (LLMs), which include GPT, can generate impressively capable writing with minimal effort from the user. Colleges and universities have not yet arrived at a standard way of handling the output of LLMs in classroom settings. You may be forbidden from using LLMs in some circumstances and required to use it in others. Many teachers may say nothing about it at all. You need to pay careful attention to the guidance you have received before using LLMs in your academic work. I cannot tell you whether you *should* use LLMs in any given situation. I instead hope to help you understand what will happen if you *do* use LLMs for college writing.

## Thesis Moment

As they exist in 2023, **LLMs are reasonably capable generators of English sentences. However, they are absolutely terrible at producing scholarship.**

## The Capable Side: What Can LLMs Generate?

The writing generated by LLMs will strike many readers as formulaic and bland, and I'll offer a couple of reasons for that effect below. However, LLMs do write capably, meaning that they rarely make outright errors in grammar, usage, or punctuation. They also produce writing that matches the conventions of writing genres. An LLM's thank-you note will sound like a reasonably competent human thank-you note, and a poem will look and sound like a poem. In other words, LLMs are best at writing the *kind of thing* a reader would expect in a given situation.

For example, if you ask GPT-4 how many Bs are in the word "blueberry," it will generate the *kind of thing* someone will say in response to that question—a sentence giving the number of Bs in the word. Here is the actual answer I received to that prompt:

The word "blueberry" contains 3 "b"s.

That, reader, is a problem!<sup>1</sup> And it reinforces the broader point. If you ask an LLM to suggest gifts to give someone on Mother's Day, it will offer reasonable suggestions, in some detail. Based on its vast source texts, the LLM will be able to call up the kinds of things people often give in that situation. If you just need an acceptable, inoffensive idea for such a gift, the LLM may well help you find one. (Seriously, try it out!)

However, we might accept that the best gifts are not just acceptable and inoffensive but lovingly specific: they stem from the details of the relationship between the giver and receiver. The LLM's ability to use its vast training data won't help you express that kind of specific connection.

You can use that example to think about the limitations of LLMs in many areas. An LLM can often generate an appropriate *kind of thing* to say. LLMs are giant pattern-recognition machines. If you want to go beyond the pattern to express something sophisticated and specific—what most college assignments ask you to write—then informed human thinking remains the tool of choice.

---

<sup>1</sup> If you use this prompt, you may not get the same response, even for the same prompt given twice. Sometimes the LLM answers will be correct. I thank Peter Simpson for this example and other valuable comments on this document.

Even if you can use an LLM, you need to consider carefully whether you want to do so. Using an LLM may be forbidden by a teacher, for example, or it may be more valuable to learn from writing the text yourself. Understand your situation and use your best judgment.

## The Terrible Side: LLMs as Academic Dishonesty Machines

LLMs use a sophisticated version of a simple process: they “read” a vast amount of text; they use that text to generate the most likely next word after any sequence of words; and they use a bit of randomness to make the output more interesting.<sup>2</sup> That process works well for many purposes.

For scholarship, however, it is a disaster. Perhaps the most fundamental skill of college-level writing involves creating a written conversation with other scholars. You read a book or article, you quote some of it, you respond to the quotation in your own voice, and you cite the source so that other scholars can follow your path and join the conversation in their own way.

Remember, LLMs are good at producing the *kind of thing* people write in a given genre. Therefore, LLM output can sound like scholarship. GPT-4, for example, will use the names of real scholars, tell you the titles of their works, quote them, cite them, and respond to them. All good—except, in every case I’ve checked, at least some of the material doesn’t exist. GPT-4 might give you a real name but a fake title and quotation; Bing might give you a real title and link to a real article but attach a fake author’s name; Bard’s insertion of fake scholarly information is irregular but constant.

Here is an example of this kind of fakeness from ChatGPT. Prompted to write an essay about the idea of home in Toni Morrison’s novel *Song of Solomon*, it generates this paragraph:

Song of Solomon also conveys a broader notion of home connected to cultural and spiritual roots. In "Ghosts, Memory, and Flight: A Cross-Textual Reading of the Figure of Solomon in Toni Morrison's *Beloved* and *Song of Solomon*," Giselle Anatol posits that "Morrison fashions a concept of home that transcends a mere physical dwelling, with the protagonist's spiritual journey leading him towards his ancestral roots" (Anatol, 78).

That paragraph sounds like it could become part of a student paper. As a teacher, I would suggest some improvements, from italicizing the titles of the novels to adding a sentence after the quotation with the writer’s own thoughts that respond to the quotation from Anatol’s work.

In this case, though, another problem is deeper and weirder. The paragraph quotes Giselle Anatol, who is a real person and, indeed, an excellent scholar you might quote in an essay about Toni Morrison. However, Anatol has not written an article with the title quoted here. Use any search engine or scholarly database you like: you won’t find an article with that title, and you won’t find the exact phrasing that ChatGPT “quotes” (by Dr. Anatol or anyone else). **The article and quotation simply do not exist.** They are as fake as the third B in “blueberry.”

---

<sup>2</sup> To understand the basics of how LLMs work, you can start with Ted Chiang’s “ChatGPT Is a Blurry JPEG of the Web” (*The New Yorker*, February 9, 2023, <https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web>). To understand how the text-generation models create fake scholarship, see Stephen Wolfram’s explanation of how LLMs use probability (“What Is ChatGPT Doing ... and Why Does It Work?” February 14, 2023, <https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>.)

In scholarly writing, the usual process of the LLMs—generating good-sounding text and mixing it up with a little randomness—creates **a very bad outcome: writing that sounds like scholarship but is actually academic fraud. If you submit that writing for a college assignment, the academic dishonesty becomes yours.** If that idea sounds scary, good—it scares me, too, because I want you to learn real scholarly practices, and I don't want you to be responsible for academic dishonesty.

I prefer, however, to concentrate on the positive side: by coming to college, you have chosen to join the community of scholars. We teachers are here to help you learn to take part in the scholarly conversation—to learn from and build on the ideas of real writers and thinkers such as Giselle Anatol. We find the conversation stimulating, challenging, and often fun; we want to support you as you learn how scholars write and talk. **Please take the opportunity to enter the conversation with your own words and ideas.**

### Side Note: Why Does LLM Writing Seem Bland?

I have begun to analyze texts written or generated by ChatGPT, Grinnell students, and Grinnell English faculty. The project is ongoing, but based on the early results, I am confident in identifying two tendencies that may explain some of the blandness of LLM output.

First, **LLMs work by addition rather than contrast.** ChatGPT uses words of simple addition—“and” and “also”—about twice as much as the human scholars. Simple additions often create adequate but weak connections between thoughts. Sophistication, however, often arises from finding subtle differences and examining many sides of an issue. Indeed, the human writers are much more likely to use the language of contrast than ChatGPT: the humans use “but” about twice as often as the LLM, and they use “however” five times as often!

Second, **humans vary their sentence length more than LLMs do.** ChatGPT generates sentences that are much more uniform in length than the human writers. (Measured by standard deviation, the ChatGPT essays use about half as much variation as the students, and the faculty use much more variation than the students.) These measurements may indicate that sentence variation gives human writing a greater sense of voice and distinctiveness than LLM output.

### Conclusion

I can't tell you not to use LLMs. I beg you to be careful about when you choose to use them. For the purposes of completing college writing assignments, I suggest asking yourself these questions:

1. What has my teacher said about the use of LLMs for this class?
2. Does it matter whether what the LLM generates is accurate or true?
3. How will using an LLM affect my ability to learn what this assignment intends to teach me?

In some cases, you will have good answers to these questions that lead you to using an LLM—indeed, the assignment might require you to use an LLM and reflect on what it produces!

In many other cases, however, you will be better off avoiding the LLM for classroom assignments, either because using the LLM will make you responsible for academic dishonesty or because you simply won't learn what you would have learned from doing the writing yourself. I hope this document will help you make good, informed decisions about your writing process.