# RTTP AND AI LARGE LANGUAGE MODELS

*Last revised June 30 2023*

The arrival of Artificial Intelligence (AI)-generated text poses special challenges for instructors who use Reacting to the Past. It is less a question of whether students will use utilities such as Chat GPT, but when. This document begins with some notes about how to draw players into the game. This is not new advice, but large language models make it even more important to consider than in the past.

This is followed by some notes about how to detect the use of ChatGPT and how to construct countermeasures against this use. It concludes with some advice about how to promote the thoughtful use of AI with Reacting.

## Player Behavior

For many students, the biggest intellectual challenge in playing a Reacting game is adopting the persona of their role. They often struggle to figure out what they are supposed to think about the ideas and situations that they encounter. This cognitive overload might mean that they resort to ChatGPT as a shortcut rather than wrestling with these ideas. Consequently, it is important to help your players to ease into the game. This has always been important. Today, for a variety of reasons, it may be more important than ever.

Take time to introduce players to the historical context, their roles, the documents, and the main ideas in the game. Also make sure that they understand your expectations regarding citations and the disclosure and description of any interactions they have with AI and other people as part of their work on the game’s assignments.

## Incorporation

Simply prohibiting the use of programs like ChatGPT is likely to become increasingly difficult, as the technology improves and students become more accustomed to its use. A more productive approach is to guide students toward using it as a tool that helps them to create quality work, and not a utility that does it for them.

**Encourage caution.** Given the difficulties that ChatGPT has with historicity, documents, and point of view, ChatGPT is not particularly useful when dealing with the *details* of any game. When players struggle to understand the situation, they may rely on ChatGPT to summarize documents, outline essays, or provide historical context. In all these cases. ChatGPT can reword and attempt to explain difficult passages and concepts, which can provide basic ideas, but it often misses complexity, nuance, and period details. This issue compounds as the game continues because of the inability to track the ways in which the reality of the game can diverge from the historical record. Despite these shortcomings, it can provide *general* feedback that can be helpful. It is even more helpful if students understand its limitations.

**Anticipating objections.** If you enter a draft of a written assignment that advocates a certain course of action and ask ChatGPT to provide critical feedback, it will. Often the feedback suffers from the 21st century biases described above; it is more or less like asking your roommate to review something for you.

 **Revising drafts.** ChatGPT can shorten or expand any piece of writing. When shortening, it sometimes makes errors by omitting important points. When lengthening, it can become repetitive and discursive. Occasionally, it starts hallucinating.

**Sharpening prose.** ChatGPT is very good at cleaning up grammar and spelling, but it sometimes changes correct proper nouns into words that are more common in its training data.

**Engaging in dialogue.** The patient helpfulness of ChatGPT means that some students feel free to ask it questions that they never would in class or on a discussion board. If you help them with prompt engineering, the answers are more likely to be reliable.

**Engineering prompts.** Well-designed prompts greatly improve output. Multi-sentence prompts are much more likely to give you something that you can use. Experiment with the sentences below. Start by filling in the blanks for the first sentence and use it as a prompt. Then fill out the second and see if it improves. Continue adding additional parameters. As you do, ask yourself, is this a writing process?

As a [role] write a [description] for [audience].

The purpose of this piece is [purpose].

Present it in [format] and maintain a [tone].

Consider the context [context],etc.

## Prevention

**Provide clarity.** At present, academic approaches to the use of AI are all over the map. Take time in your setup sessions to make sure that everyone understands what your class policies are. Also help students to understand the learning objectives for the game and the related assignments. Also explain your expectations regarding the use of AI. It may be helpful to require students to submit a screenshot or hardcopy of their interaction as well as a short piece of writing that describes the ways in which they used AI. Please note that disciplinary expectations for citing AI are in flux.

**Ask for specificity.** As is noted above, ChatGPT struggles with historical points of view and the degree to which issues and ideas may change once a certain turning point is reached. Consequently, you may want to emphasize specific elements of both in your assignments.

**Require document use.** Understanding and using documents is central to the intellectual engagement of Reacting games, but players sometimes attempt to skirt around deep engagement. ChatGPT can judiciously sprinkle some quotations, but they rarely connect to arguments. Instead, they are “chrome” that superficially fulfill any expectations that you have laid out for their use. Here, the size of ChatGPT’s database combined with its tendency to reproduce conventional wisdom in terms of interpretation make it relatively easy to detect.

The bespoke historical context essays and curated document sets that appear in Reacting gamebooks are a sort of walled garden. Documents are often edited for brevity and clarity. Many are also in translation. If asked to include information about documents, ChatGPT will not be able to distinguish between the versions of those documents that appear in the gamebook and other, similar things that are in its database. Consequently, if you require information from the gamebook and prohibit information that is not in the gamebook, players using ChatGPT for their quotations and interpretations will encounter a great deal of difficulty.

**Require citations.** ChatGPT struggles to provide accurate citations. This is particularly true if you are drawing exclusively from a Reacting gamebook. If you are using paper copies, you should definitely require page numbers. Ebooks are trickier because they do not have uniform page numbers.

**Include game events.** Unless users input a significant amount of information about game events, ChatGPT has a difficult time incorporating them into the documents that it produces. Even then, it tends to lean toward historical outcomes because that is what is in its training data. Requiring players to address and respond to game events makes it difficult for them to rely on ChatGPT. This is less true toward the beginning of the game, which usually aligns fairly closely with the historical record, than the latter part of the game, which may diverge significantly.

## Emphasize the Writing Process

Rather than fixating on the artifacts that are created by student writing, it can be worthwhile to consider focusing on the process of writing. If the artifact is what you are assessing, students have a good reason to take shortcuts through the writing process. ChatGPT can be one of those shortcuts. This can be time-consuming, but it helps students recognize that they learn by writing. Consider the following approaches:

**Ask for prewriting.** Asking players to do some pre-game stream of consciousness writing can help to open them to the possibilities. It also provides a paper trail for the development of their ideas.

**Set up meetings.** Interacting with players about their assignments is helpful for them. Teaching Assistants, preceptors, and campus Writing Centers can also fulfill this function. If you delegate this work, consider asking for some notes that reflect on the conversation.

**Provide scaffolding.** Breaking assignments down into pieces that build upon one another helps players to develop their ideas over time. Like prewriting, this also provides a paper trail.

**Emphasize speaking.** Real-time assistance from AI is extremely difficult to manage. In addition, it is difficult to promote the ideas in an essay that you did not write.

 **Require reflection.** Asking players to reflect on their experiences is cognitively useful. It is also very difficult to conjure with ChatGPT.

## Detection

AI technology is changing rapidly, so the approaches described below may or may not work. In particular, it is unwise to rely upon plagiarism checkers.

**Cadence.** Spend some time with a free AI like ChatGPT 3.5 or Bing, which combines ChatGPT with a search function. See how it responds to different prompts. Soon, you will be able to identify its generic voice, which is banal and unnaturally balanced. Some liken it to the “voice of Wikipedia.” This can be modified by clever or experienced users, but most users continue to rely upon basic prompts.

**Perfect prose.** Unless instructed to make errors, ChatGPT provides output with perfect grammar and spelling. Even if you instruct it to make occasional errors, they tend to be random and sprinkled evenly through the output.

**Ahistoricism.** As a large language model, ChatGPT does not have an internal map of reality. This means that it has great difficulty putting historical elements in chronological order unless it is creating something like a timeline. Since almost all the data that trained it was in English and generated in the 21st century, it has great difficulty with historicity and cultural specificity. In other words, if you ask it about the government of Athens, it will provide you with contemporary information. If you ask about *classical* Athens, it will provide a generally accurate historical answer, but will not distinguish between different periods of history during this era. If you ask it to comment on the direct democracy of classical Athens, it will do so from the perspective of the 21st century US.

**Hallucination.** ChatGPT will provide an answer to just about any prompt, but for many of the historical situations modeled in Reacting games these tend to be “confidently incorrect.” Sometimes the errors are significant. Other times, they are so general that they do not apply to the situation that players face in the game. Infamously, these errors sometimes include citations for sources that do not exist.

**Point of View.** ChatGPT has trouble handling the historical specificity of the situation, but it is capable of taking on a point of view, but only if you prompt it to do so. The more detailed the prompt, the better. It is not intimidated by length. Cutting and pasting the biography and assignment from your role sheet can produce a capable essay. It can also incorporate in-game events, but only if you prompt it to do so. This requires a significant amount of work, so most of the time it produces very general points of view. Unless prompted to do otherwise, it also tries to present balanced arguments. It has a difficult time playing a zealot.

**Prohibited content.** Prohibitions on producing certain outputs are programmed into most AIs. Curious users sometimes try to “jailbreak” AIs to circumvent this programming, but this is becoming increasingly difficult. At present, several arguments that sometimes appear in Reacting games are forbidden. These include incitements to violence, anti-Muslim language, and information about how to perform a variety of illegal acts. This is not to say that you should encourage these arguments in your assignments, but note that a speech given by a leader on the Fourth Crusade written by ChatGPT will scrupulously avoid anti-Muslim content.

**Seeking Reconciliation.** Similarly, ChatGPT strives toward reconciliation and empathy. This may be a desirable trait, but it flattens output for roles that have sharp positions. It blunts nationalism and refrains from the endorsement of violence. Above all, it seeks to reconcile.

**LMS detection.** Some learning management systems inform the instructor of the probability that students are using outside sources while they are writing. This function is not always accurate, but this may be a tool worth investigating..