

Using Generative AI in Academic Writing Courses

Hahn, Aaron

Department of Linguistic Environment, Faculty of Languages and Cultures, Kyushu University

<https://doi.org/10.15017/7343643>

出版情報：言語科学. 60, pp.19-38, 2025-03-14. 九州大学大学院言語文化研究院
バージョン：
権利関係：



Using Generative AI in Academic Writing Courses

Aaron Hahn

1. Introduction

It is hard to imagine a modern technology with a greater potential to change education (and, perhaps, society in general) than generative AI (genAI). The closest recent comparison is probably the combined technology of seemingly unlimited information online and the ability to access that knowledge quickly and constantly through mobile devices (at least in affluent countries). But genAI goes beyond providing access to knowledge since it purports to be able to simulate reasoning and do at least some human intellectual labor. While the future hope (or fear) is that AI systems will eventually equal or exceed human intelligence, even the sub-human capabilities of current systems are reshaping the world. It has already had a significant impact on work in some sectors and locations. For instance, one August 2024 survey found that about 26% of US workers used genAI for work, with about 9% using it every workday; furthermore, about one-third of all surveyed people used it outside of work (Bick et al, 2025). While uptake has been slower in Japan, with one survey indicating that less than 18% of small and medium-sized Japanese companies were using generative AI as of August 2024 (Kyodo News, 2024), there is no reason to think that genAI won't eventually become equally widespread here. At the university level, I have seen a steady increase in student use of genAI during classes and an increased willingness to do so openly (indicating that they consider it acceptable, not a form of cheating). I've also spoken with some instructors using it to assist with their research or teaching preparation.

The commonly expressed worry of educators is that students will turn to genAI to complete work for them, a form of cheating that is both hard to identify definitively and either free or low cost.¹ Some have described genAI as “disruptive”

¹ For instance, OpenAI, the owner of ChatGPT, currently provides free access to its second-best model on a limited basis (a certain number of queries in a certain amount of time), and the paid version which allows much more consistent access to the standard model and limited access to the best models for about 3000 JPY/month as of February 2025.

(Hullick, 2023), while others have gone further and labeled its uptake a “crisis” (Scott, 2023). These educators worry that if students rely on it too much, they will learn little or nothing in their classes. These worries mostly center around the type of genAI called Large Language Models (LLMs). These systems are trained on massive amounts of real-world language data from which they develop models of how language works, what sorts of words go with one another, and how topics are related. These systems (of which ChatGPT is the most famous) can take natural language input and produce what look like normal human responses in many cases. They can hold conversations, summarize and analyze data, write fiction and non-fiction, and, for newer models, search the internet to provide more detailed and accurate responses. The previously mentioned educators worry that students can have LLMs complete most homework assignments with results that, though they may not be perfect, are often “good enough.” For instance, studies have shown that LLMs, sometimes with some specialized training, can achieve passing or nearly passing results on high school English exams from the Netherlands (de Winter, 2024), a US university-level introductory physics course (Kortemeyer, 2023), and several medical licensing and preparatory exams (Ariyaratne, 2023; Le et al, 2024; Stengel et al, 2024).

On the other hand, some educators see generative AI as a potentially positive tool that can improve education if teachers and schools learn how to adjust classes to account for this new technology (Heaven, 2023). They believe that we can develop ways to teach students how to augment, rather than replace, their abilities with genAI. While the technology is so new that there has been little published research on the effectiveness of such an approach, two recent review articles on efforts to incorporate genAI into language learning classes (the educational focus of the present paper) found mostly positive results (Law, 2024; Teng, 2024). I have personally taken the position that genAI will likely become a ubiquitous technology in our personal and professional lives, so it is important that we start teaching students how to use it well and ethically. This approach is consistent with the position of the Japanese government, which has encouraged Japanese schools and universities to help students positively engage with AI (Yomiuri Shimbun, 2023). To do so, though, teachers and schools, need guidance and proven solutions since many have expressed concerns that they have not received sufficient guidance to

know how to use these systems in helpful rather than harmful ways.

The present paper explores the initial steps I have taken to integrate generative AI into my academic English writing classes at Kyushu University. It provides specific activities I have used, why I chose them, and the benefits that I believe this new approach offers. I hope this paper can help provide some initial guidance to other teachers and spark conversations about other ways to adapt to this new technology.

2. My history with genAI

While I started experimenting with ChatGPT soon after its release in November 2022, I did not begin using it in classes until after the 2023 Kyushu University AI Task Force recommended that all faculty attempt to find ways to integrate genAI into their educational practices when appropriate. Around the same time, the Faculty of Languages and Cultures also softened its stance on the use of computer translation, with new guidance stating that teachers could allow computer translation on some assignments. This policy change seemed to me to tacitly support the idea that we should be teaching students how to write as they will in the real world, with commonly available tools, rather than ignoring or forbidding those technologies as blocking “real” learning.

As I started to consider bringing LLMs into my English classes, I was faced with the problem that there are not yet many reliable academic sources that document how this has been done effectively. There are certainly no textbooks yet that include genAI, and even those academic papers which do exist can be challenging to use because genAI tools change so rapidly. As of February 2024, OpenAI alone has released eight models for ChatGPT since the initial public release of version 3.5, and numerous other companies have released similar tools. As a result, many of the activities discussed below are the result of trial and error—both my own and that of other AI-positive teachers who have shared what was working for them in online forums. As our understanding changes, and as the models become more capable, these activities are also likely to change. Thus, the suggestions in this paper should be understood to be just a snapshot of my current methods of integrating LLM use into some of my classes, and are by no means a claim of a perfect, finalized curriculum.

3. Teaching students about generative AI

Before I explain the details of how I have used genAI in academic writing classes I want to comment more generally about the topic of “prompting,” which refers to the way questions are asked to an LLM. The quality and style of the input prompt have a large influence on the nature of the response given by the LLM (Stahl et al., 2024; Wang et al., 2024). Thus, we cannot simply tell students “Ask ChatGPT <X>.” Rather, it is critical that we also teach them how to prompt LLMs effectively (Cain, 2024). Luckily, the skills needed to prompt LLMs well (often called “prompt engineering”) are skills regularly taught in academic writing courses—as Cain explains, it requires “rhetorical skill, organization, and precision” (p. 51).

Unfortunately, at this point, I have mostly been adding genAI activities to already-packed courses, so I have little time to devote to the intricacies of genAI use. As such, my current method is to provide students with prompts that I have pre-tested and explain their key features, but I do not use activities specifically focused on prompt engineering. My most detailed explanation comes with the brainstorming activity discussed below, where I explain both the differences in single-shot and multi-shot prompting (i.e., writing a highly detailed initial prompt with lots of context vs. engaging in an extended conversation with the LLM to achieve success through a longer process) as well as the popular “Role—Task—Content” single-shot prompt. As genAI becomes a greater and more integral part of my courses, I anticipate devoting more time to prompt engineering.

4. Course context

Currently, academic writing at Kyushu University is covered in a two-course sequence called Production 1/2. The two quarter-length courses are awarded units independently, but classes remain intact with the same teacher; thus, they are better conceived of as a single two-unit course spread across sixteen weeks. The Production course covers both academic writing and academic presentation, though only the former is dealt with in this paper. In these courses, students write one paper each quarter, with the first described as an academic essay and the second as a research paper. In both cases, students are expected to write an

organized essay that contains a well-reasoned argument and is written according to a particular structure. There are two main differences between Production 1 and 2, at least in the way that I teach the sequence. First, in Production 1, students must choose their essay topic from a list of questions (usually about 30-40) that I provide, whereas in Production 2 students must write their own question on a topic related to either their academic major or an important current event. Second, while students may get outside evidence to support their arguments in Production 1, they are not required to do so, but in Production 2 they must include information from at least two reliable sources (and write APA citations for them), and I require one of those sources to be an academic journal article or academic book.

Production 1/2 uses a process-oriented approach to writing. This approach emphasizes the process taken to reach the final product (the final draft, the completed research paper, etc.) rather than the product itself and its “correctness” (Zen, 2005). Students complete planning activities prior to writing, then write three drafts of increasing length and complexity, followed by a final draft. This process lasts the entirety of each quarter class—that is, it takes eight weeks to write one essay. Throughout the process, students receive advice from the instructor and feedback from other students via peer review. While the final product is judged on its argument, organization, clarity, etc. for the essay, most of the students' score in the class comes from completion of these intermediate steps. Furthermore, part of the score on the final essay is determined by how much they have improved their work since earlier drafts—whether they adapted suggestions from others, and whether they made changes on their own.

A process-oriented approach helps alleviate much of the worry that students will use genAI to write the entire essay (to “cheat”). Each stage in the process must be an extension of earlier work, and since all the intermediate work must be done to pass the course, a student cannot pass by using simply genAI to generate a final essay. If at any point in the process a student overly relies on genAI, the sudden change in content and/or quality will be easy for the teacher to identify by comparing drafts. This process of comparison is especially easy if all essay-related work is required to be done online and stored in a shared cloud service, as in the OneDrive system that is provided by Kyushu University to students and teachers. In theory, a student could try to cheat throughout the entire process by

intentionally asking an LLM to produce low-quality work for early drafts and then have it adjust those products step-by-step, but at least currently it is difficult to get LLMs to create work that is both “low-quality” and incomplete, while also adhering to the specific criteria set for each draft. In addition, it would be necessary to provide the LLM with all of the feedback from teachers and other students and have it adjust the essay accordingly. The closest I have come to achieving complete “cheating” on every draft was to use the top-tier paid version of ChatGPT (which is well out of the budget of most students), give it all my instructional materials, have it create a high-quality final draft, and then ask it to work backward to create earlier drafts. And even this wasn’t fully successful, because I didn’t ask it to adjust based on external feedback. It is unlikely that a student could accomplish this without fully understanding everything about the course from the beginning. In addition, process-oriented writing lets me intervene when students have used an LLM excessively and treat it as a learning moment rather than a “punishment.” Together the student and I can discuss what led them to overuse the LLM and review materials or create strategies for future work so that they can succeed.

Turning to the students of Kyushu University, most enter with strong grammar, good academic vocabulary, and a generally wide body of knowledge. In addition, because the English portion of the Kyushu University entrance exam includes two short essays, the students already have some experience writing organized texts that provide support for claims. Of course, there is variety between majors and individual students, but the generally high level of basic English skills means that instructors at Kyushu University can focus on writing skills above the sentence level. Many of these skills are partially language-agnostic, so they are useful even in the majority of courses that are conducted entirely in Japanese.

These starting conditions mean that I focus my Production class on teaching students to construct good arguments, with claims supported by sound reasoning and (especially in Production 2) reliable evidence. In addition, I place a heavy emphasis on presenting the argument in a strong, carefully designed structure. For simplicity, I have chosen to use the five-paragraph essay format; even though this pattern is mostly a school artifact (that is, it is a pattern used solely in school and is not appropriate for either advanced academic writing or real-world writing), it is a clear structure that is relatively easy to learn and lets students focus mostly

on content. In essence, the goal isn't "learn the five-paragraph essay format," it's "learn how to write according to a specific structure," which prepares students to later learn the structures appropriate to the genre of their major or profession. Finally, I want students to internalize the idea that writing is a process, and that revision means carefully evaluating every word, sentence, and idea and asking themselves not only "Is this correct?" but also "Could this be better?"

5. Incorporating generative AI

When considering where and how to incorporate genAI into a course, the teacher must ask several questions about both the course and the students. First and foremost, the teacher must ask what the most important goals of the class are. That is, what are the core skills students should practice and what is the core knowledge they should obtain? GenAI should only be included in ways that do not hinder them from achieving those goals. Turning back to the process-oriented approach, this is why, if we believe that there is still value in learning the process of writing, it would not be useful to spend a whole semester training students on how to optimize prompts to produce high-quality essays, because even though that would teach students how to achieve a certain product, they wouldn't be making progress in the core goal of learning to write iteratively.

Subservient to the first question are three additional questions. First, what smaller component skills are needed to progress toward the primary goals? GenAI can become a part of the complex scaffolding process needed to achieve the larger goals of the class. Second, what skills are the students already good at, and what skills do they need the most help in improving? It makes more sense to utilize LLMs in ways that complement students' pre-existing knowledge rather than duplicating what they already know unless the LLM offers a significant upgrade in quality or efficiency. Third, what skills are the currently available LLMs good at? Answering this question requires that the teacher test the available AI tools regularly since capabilities often change.

Using these questions and the ideas of other instructors as a guide, I identified five components of the course where I believed that genAI could be useful, and, after testing, created activities for four of them. Those activities are discussed in the following sections, along with an explanation for why the fifth was rejected.

5.1 Brainstorming activity

While Kyushu University students generally have high lexicogrammatical skills when they enter, most of the work that they did in English in junior and senior high school was at the sentence level, and much of that work was translating between languages, not creating novel sentences to express their own ideas. The main time that they had to express their own ideas in English was in preparation for the entrance exam essays. Thus, some students have difficulty generating enough ideas during the planning stages of essay writing to allow them to build a strong essay. All students can generate some ideas, but many seem either satisfied with commonplaces that don't stand up to academic scrutiny—or, simply get stuck and don't know how to generate more than the minimum number of ideas needed to write something. While timed essay writing rewards quick idea generation and requires less sophisticated arguments, university-level untimed essay writing requires that students start with a stronger foundation. Furthermore, the more ideas that a writer produces during brainstorming, the more options they have to build a stronger argument, including being able to more easily make changes midway through writing if an idea doesn't pan out. Thus, I want students to do a lot of brainstorming prior to writing so that they have a large selection of possible ideas to use in their essays. Making lists of possible ideas is one of the things that LLMs are good at. Thus, the first genAI activity that I added to my writing class was a modification of the brainstorming activity I had previously been using.

The brainstorming activity is done in the second lesson of Production 1. Students have already selected a research question from my list. Students brainstorm about nine total ideas in three steps with decreasing time for each step. The idea behind limiting time is that the goal of early brainstorming is to think of any ideas, big or small, without taking time to evaluate their quality. Evaluation comes only after students have many ideas to choose from. Students enter their ideas into a premade Microsoft Word document (see Appendix A).

After brainstorming on their own, they turn to using ChatGPT.² The students

² As discussed above, the choice of generative AI system may vary over time. I have personally used both ChatGPT and Claude (made by Anthropic), with ChatGPT being my current system of

are provided with a prompt that they copy and paste into ChatGPT. This prompt has ChatGPT generate six single-sentence responses which the students copy into the Word document.

Third, students have a conversation with ChatGPT. They first have to re-prompt ChatGPT to shift it from brainstorming into discussion mode. The purpose of the discussion is for students to engage more deeply with one or more of the points that ChatGPT raised earlier. This introduces students to the idea of multi-shot prompting—the process of engaging in a conversation with the LLM to refine the context and get better answers.

Finally, and most importantly, students need to evaluate all the ideas generated in this assignment—their own and ChatGPT’s. They judge whether the ideas are strong, whether they can explain them well, whether they work together, etc., with the intent of selecting some that they can use in the first outline of their essay. This step, usually done as homework, is essential to integrating ChatGPT into the writing process as a collaborator, rather than treating it as a source for answers and finished products. The assignment puts the students’ and ChatGPT’s ideas on the same level, rather than elevating its responses as being “correct” or “better.” It is important to teach early on that ChatGPT is a tool that can help, but that it is each person’s responsibility to judge the quality of its answers and choose if and how to integrate them into their writing.

5.2 Brainstorming Part 2: Research

The requirement to use outside sources in Production 2 was the second place I chose to add genAI. In principle, students in Production 2 are supposed to start with a research question that they do not already know the answer to. Rather, they are supposed to pose the question and then search for sources to help answer it. This is intended to mimic academic research, where a researcher makes hypotheses about their research questions before experimenting, but is ethically required to answer the questions based on what they discover through research, regardless of whether that matches their original hypotheses. This means that the

choice. These activities should be adaptable to other LLMs, though teachers should test them and consider adjusting prompts based on the results obtained.

process of brainstorming for Production 2 is different, in that students do not brainstorm possible answers to their questions, but rather brainstorm points to research to find answers. All students are capable of typing their research questions into a search engine, but they are less familiar with breaking down a question into sub-topics, and the value of using more complex search engine queries. This is a step that genAI can assist with.

First, students are instructed that genAI cannot, at present, be used to directly do research. GenAI outputs themselves are not reliable sources, and the free models are not good at finding reliable sources online. Some paid models are better at this, but the free models still often find outdated, unreliable, or even fictitious sources. In some instantiations of the course, I have shown students examples of specific research “failures” to establish this point (see section 5.3). But what genAI can do is help the students find sources on their own. Thus, I provide them with a prompt designed to have ChatGPT break down their research question into smaller parts and then give them specific search queries that they can use to get good sources (see Appendix B). Not all students need this help, and the quality of the advice they get depends on how narrow their question is (for example, students writing essays on very current or local events will likely get less information simply because their topic won’t be well represented in the training data), but for many students this provides a bridging step that helps them get started doing research.

5.3 An optional activity: Generative AI weaknesses

One activity that some teachers do with LLMs, including some who forbid students to use them, is to provide students with examples of problematic LLM content. The teacher needs to generate authentic content from a current model that in some way does not meet the standards of the course. While the prompt must be carefully chosen to create unacceptable content, it also needs to be “normal”—that is, a teacher should not create a convoluted or deliberately strange question just to show off the system’s limitations. For example, it is well known that most LLMs are unable to accurately count the number of letters in a word (e.g., “how many ‘r’s are in “strawberry?””), but showing this to students doesn’t help show the limitations of LLMs for things like writing essays and answering homework. Depending on the time available each quarter and the current quality

of the model, I have used one or more of the following:

1. A sample essay produced by asking the LLM the research question with no additional criteria or instructions, such as “Write an essay about the death penalty in Japan.” Without specific direction, LLMs tend to produce vague essays filled with weak arguments. In addition, they use a lot of language that I don’t allow in student essays, like second-person pronouns or rhetorical questions. Sometimes they produce strange formats, such as essays with bullet points or internal titles. Most importantly, since the models are generally trained to default to neutral positions on controversial topics, it is likely that, for a question like this, they will produce a main claim that fails to make a strong argument, such as “The question of whether or not to have a death penalty is a difficult one that must be considered from a variety of social, political, and ethical viewpoints.” Each of these points would be unacceptable in my course, and thus the point of showing students this response is to demonstrate that beyond the ethical problem of cheating by having the AI do the work for them, even if I somehow didn’t notice they had cheated they still would likely receive a poor grade. Other teachers would want to look for responses that similarly failed to meet their standards. I have found that showing essays like this to students is most useful in the middle of the course after I have already established some of my basic rules for essay writing.
2. A hallucination—that is, a response from an LLM that confidently expresses false information. I have found that the best topic for this is to ask the LLM for information about a minor Japanese celebrity—someone about whom some information is available online, but not a lot. Because LLMs know what a biography “should” look like (think of the opening paragraphs of a typical Wikipedia biography), they will often invent facts when they are unable to find enough information to complete their answer. Sometimes I also include a longer conversation where I challenge the LLM about its answer to demonstrate how doing so doesn’t necessarily solve the problem, as the LLM will often report that it was in error, and then confidently provide another false answer. Thus, students can learn that the way LLMs

are programmed to be generally helpful means that they will sometimes create false information rather than admit an inability to answer accurately.

3. A request for information backed up by reliable sources with APA citations. Some models on some topics will produce completely fictional sources that appear to be real (they are properly formatted, they have titles that appear to be related to the query, etc.), but which when investigated don't exist. Other models will produce sourced information, but that information will be drawn from sources that don't meet university standards for reliable sources. However, it is important to note that some newer models can produce accurate, well-sourced information, at least on some topics. This is an example of how the teacher must always be testing and retesting their assignment on the current model of the LLM they plan to use—it is not possible to rely on what was done in a prior semester as the capability may have radically changed since then. This “bad example” is suitable for use after the students have already been introduced to the concept of reliable sources and the standards required in their course.

I have found it most useful to show students the responses and give them time to work together to evaluate their quality. In most classes, students were able to identify at least some of the problems. This process helps teach the importance of critically analyzing any responses from an LLM.

5.4 “Peer” feedback

One of the key aspects of a process-oriented approach is receiving teacher and peer feedback and making revisions based on those suggestions. However, both forms of feedback are limited resources—teachers are limited because the total number of students they have places a practical cap on how often they can provide extensive feedback, and peer review takes up class or homework time that could be spent on other activities. Thus, one of the more useful features of LLMs is that they can supplement human feedback with automated feedback. However, it is not useful to just ask the models for “feedback” on an essay, as the advice given is likely to be shallow and will not necessarily match course goals. It is important to provide students with carefully designed prompts that will give tailored feedback on a specific concern. The two main use cases I have found so far are for argument

feedback and grammar feedback.

For the former, I provide students with a prompt targeting their argument quality (see Appendix C). This activity is done after the students have completed peer review on Draft 1. Since part of the Draft 1 requirement is that at least two of the body paragraphs contain a complete argument, but the whole essay does not yet need to be complete, the prompt and assignment focus on analyzing the quality of a single paragraph at a time. As shown in the appendix, the prompt tells ChatGPT to only comment on argument quality, not grammar, spelling, or organization. The prompt also limits the response to feedback only; this is done to prevent ChatGPT from fixing the problems it finds. I want the feedback to be similar to the feedback I give: comments about weak points in the drafts, which students must then think about and decide if and how to improve. I have found that ChatGPT's advice is mostly accurate, especially since the 4o model became free.³ Sometimes the explanations are a bit difficult for students, even given the prompts request to use intermediate-level English, so I encourage students to use translation software for any parts they don't understand. In addition, I emphasize that they should treat ChatGPT's feedback the same as the feedback from other students: it is a suggestion only, and they are responsible for choosing what feedback to act on.

Regarding the latter use case, I have tested LLM feedback on grammar and found it to be adequate but not special. I don't find its suggestions to be any better or clearer than Grammarly or Microsoft Word's grammar-checking feature, so I generally recommend students use those because the way they present errors makes it easier to examine and fix each in turn, while an LLM can only make a long list that requires the student to search out at fix each error on their own. I worry that this would be particularly disheartening to a weaker student with many errors to fix. But grammar feedback is an area that I am still exploring—it may be preferable to use an LLM if a prompt can be written that doesn't just identify and fix errors but also helps students learn how to apply the rule in the future, possibly

³ For a limited number of prompts. After the free prompts have been used up, ChatGPT will switch to an older, less competent model. Therefore, teachers should spread out genAI activities over several class sessions, to prevent overloading the students' free accounts.

through extended practice. Most likely this would be best done with a custom GPT instance, which paid ChatGPT users can create.⁴

However, I have one warning about using LLMs for feedback, which I also express to students: it's very important that students only use the models to critique their essays, not make changes. I became aware of the potential problem with having the LLM both find problems and “improve” the essay when I had a student whose second draft was markedly different from their first draft in both content and linguistic sophistication. At first, I suspected that they had simply abandoned their first draft in exchange for a fully AI-written essay. However, when I asked the student about the issue, they said that they weren't confident in their English, and so they had asked ChatGPT to “improve” the writing. The problem was that ChatGPT didn't only change their grammar (which is what I think the student meant by their “English”). Rather, it made significant changes to the content, removed the required explicit thesis statement, incorporated stylistic elements that I don't allow, and made the argument weaker. This points back to the earlier discussion on prompting: given a simple prompt, it is difficult to predict what type of output an LLM will give. I'm not even certain it is possible to prompt ChatGPT to only change grammar without changing style since there isn't a clear boundary between the two. As such, I have incorporated a clear instruction to students that they may ask ChatGPT for feedback, but they should never ask it to revise their essay, since there is a reasonable chance that it will make their essay worse (according to my criteria) and may trigger me to believe that they did not do the writing work themselves. In addition, having ChatGPT make the changes directly interferes with one of the primary class goals of having students learn the value of iterative improvement.

5.5 Example of a skill not amenable to genAI help: citations

Sometimes there are skills that students need that genAI cannot, at present, help them with. For example, after teaching this course sequence for several years, I found students often had difficulty learning how to write APA citations. Online tools exist that can generate citations, but such tools make errors, especially when

⁴ See <https://openai.com/index/introducing-gpts/>

dealing with the non-English sources that make up about half of my students' sources. Furthermore, each of the LLMs I tried this task with failed spectacularly. For instance, when I asked ChatGPT to generate a citation from a URL pointing to a research paper, it was unable to “find” some of the key components needed for the citation (author, volume number, etc.). Second, it consistently made errors dealing with APA's requirement to both transliterate and translate non-English titles. It took four prompts to get ChatGPT to write a complete APA citation, and it was only able to do so when I repeatedly explained APA rules and made it show step-by-step that it had followed the rules. Since the whole point of this exercise is that students don't understand the rules, there is no way for them to correct the LLM in this manner. So, even though the skill of citation writing is a task students have difficulty with, it is not, currently, one that is amenable to genAI help.

6 Conclusion

This paper has focused very narrowly on how I have used LLMs for brainstorming and “peer” reviewing in an academic writing course. While the activities I have made are tentative and constantly evolving, the response from students has been encouraging. I am currently in the middle of a pilot study gathering student opinions about these genAI additions, and, so far, the results have been positive. Surveys of the students indicate that they believe AI is beneficial and they want or need to learn more about it for their future. Furthermore, 80% of students found the Production 1 brainstorming exercise to be useful, over 70% found the argument “peer review” to be useful, and a little less than 80% said that they continued to use genAI even when it wasn't required. I intend to continue integrating genAI into my courses, especially starting in the 2025 school year when genAI is made an optional core competency in the writing course that is part of the new Kyushu University English curriculum.

Numerous other possible uses for genAI and language learning are available or being designed. For example, Kyushu University will require all first-year students to use an online virtual conversation system called ELSA that combines generative AI with speech recognition and production as a supplement to our current English CALL course that focuses on receptive skills. In addition, genAI can be used by language teachers to aid in their work, such as generating large

collections of vocabulary questions, sample essays for students to study (both good and bad examples), and summarizing student feedback in real-time. I believe it is incumbent on all teachers—especially language teachers—to explore the various genAI systems and consult online resources to see where it might be possible to expand upon current classes through the inclusion of generative AI-linked activities.

Reference List

- Ariyaratne, S. Jenkno, N., Davies A.M., Iyengar, K. P., & Botchu, R. (2024). Could ChatGPT pass the UK radiology fellowship examinations? *Academic Radiology*, 31(5), 2178-2182. <https://doi.org/10.1016/j.acra.2023.11.026>
- Bick, A., Blandin, A., & Deming, D. J. (2025, February). The rapid adoption of generative AI. (Working paper 32966). National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w32966/w32966.pdf
- Cain, W. (2024). Prompting change: Exploring prompt engineering in large language model AI and its potential to transform education. *TechTrends*, 68, 47-57. <https://doi.org/10.1007/s11528-023-00896-0>
- de Winter, J.C.F. (2024). Can ChatGPT pass high school exams on English language comprehension? *International Journal of Artificial Intelligence in Education*, 34, 916-930. <https://doi.org/10.1007/s40593-023-00372-z>
- Heaven, W.D. (2023, April 6). ChatGPT is going to change education, not destroy it. *MIT Technology Review*. <https://www.technologyreview.com/2023/04/06/1071059/chatgpt-change-not-destroy-education-openai/>
- Hullick, K. (2023, April 12). How ChatGPT and similar AI will disrupt education. *Science News*. <https://www.sciencenews.org/article/chatgpt-ai-artificial-intelligence-education-cheating-accuracy>
- Kyodo News. (2024, August 16). Less than 20% of Japanese firms using generative AI: survey. <https://english.kyodonews.net/news/2024/08/0f9dfc1d0b1e-less-than-20-of-japanese-firms-using-generative-ai-survey.html?phrase=petition&words=>
- Kortemeyer, G. (2023). Could an artificial-intelligence agent pass an introductory

- physics course? *Physical Review Physics Education Research*, 19, 010132.
<https://doi.org/10.1103/PhysRevPhysEducRes.19.010132>
- Law, L. (2024). Application of generative artificial intelligence (genAI) in language teaching and learning: A scoping literature review. *Computers and Education Open*, 6. <https://doi.org/10.1016/j.caeo.2024.100174>
- Le, M., & Davis, M.D. (2024). ChatGPT yields a passing score on a pediatric board preparatory exam but raises red flags. *Global Pediatric Health*, 11, 1-8.
<https://doi.org/10.1177/2333794X241240327>
- Scott, I. (2023, April 18). Yes, we are in a (ChatGPT) Crisis. *Inside Higher Ed*.
<https://www.insidehighered.com/opinion/views/2023/04/18/yes-we-are-chatgpt-crisis>
- Stahl, M. Biermann, L., Nehring, A., & Wachsmuth, H. Exploring LLM prompting strategies for joint essay scoring and feedback generation, arXIV.
<https://doi.org/10.48550/arXiv.2404.15845>
- Stengel, F.C., Stienen, M.N., Ivanov, M., Gandia-González, M.L., Raffa, G., Ganau, M., Whitfield, P., & Motov, S. Can AI pass the written European Board Examination in Neurological Surgery? – Ethical and practical issues. *Brain and Spine*, 4. <https://doi.org/10.1016/j.bas.2024.102765>
- Teng, M.F. (2024). A systematic review of ChatGPT for English as a Foreign Language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies*, 6(3), 36-57.
<https://doi.org/10.58304/ijts.20240304>
- Wang, L., Chen, X., Deng, XW., Wen, H. You, MK., Liu, WZ., Li, Q., & Li, J. (2024). Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. *npj Digital Medicine*, 7(41).
<https://doi.org/10.1038/s41746-024-01029-4>
- The Yomiuri Shimbun. (2023, July 5). Japan educators fear generative AI will harm students' cognitive abilities. *The Japan News*.
<https://japannews.yomiuri.co.jp/society/general-news/20230705-120707/>
- Zen, D. (2005). The process approach to ESL/EFL writing. *The Journal of Asia TEFL*, 2(1), 191-198. <https://www.proquest.com/docview/2266460709>

Appendix A: Student document for brainstorming activity

Brainstorm

Question:

My brainstorm

Ideas	My comments

ChatGPT Brainstorm

Two-sided prompt:

Role: You are an expert with a wide body of knowledge. You are also an expert at explaining your ideas in clear, simple language.

Task: Brainstorming six arguments for an essay. Explain each argument in exactly one sentence. You should write three arguments on each side. Your answers should be written at CEFR Level B2.

Content:

Multiple answer prompt:

Role: You are an expert with a wide body of knowledge. You are also an expert at explaining your ideas in clear, simple language.

Task: Brainstorming six arguments for an essay. Explain each argument in exactly one sentence. Your arguments should come from a wide range of perspectives. Your answers should be written at CEFR Level B2.

Content:

Ideas (ChatGPT's answers)	My comments

Additional discussion

Continue the discussion. Arguing is one of the best approaches, but you can also ask questions, consider alternatives, etc. Basically, pretend like ChatGPT is another student and you are discussing the issue together to try to understand the topic better.

Before you start the conversation, copy and paste this into ChatGPT:

Now I would like to have a conversation about these issues. Please discuss them with me. Keep your answers and comments to 3 sentences or less, and keep them at CEFR Level B2.

After you have your conversation (at least 2 more prompts from you, but more is better) copy and paste the entire conversation below:

Appendix B: Prompt for brainstorming research topics and search queries

I have to write a college research paper that answers the question "<write your question here>". The main argument for my paper must be based on what reliable sources say. What information do I need to get in order to answer that question? After giving me suggestions about what I should research, please write some specific search queries that will help me find good evidence.

Appendix C: Assignment for getting advice on paragraph arguments

Prompt

Role: You are an expert university professor working in a Japanese university teaching English to intermediate-level Japanese students. You have a wide body of knowledge, and you are also excellent at explaining to students the problems in their writing in clear, concise, and simple ways.

Task: You will critique an argument that is a part of a student's essay. You will explain to that student in simple English 3 specific weaknesses or flaws in their argument. You will give feedback on only the arguments. You will not critique grammar, spelling, or organization. Write only the feedback, nothing else. Your feedback must use simple language and explanations that would be easily understandable to an intermediate-level non-native English speaker.

Content:

Body 1

Copy your first complete body paragraph here:

Copy the prompt, followed by the paragraph, into ChatGPT. Copy each of its responses into a separate box on the left. Then write comments about what you think of ChatGPT's suggestions.

ChatGPT's responses	Your comments

Body 2

Copy your second complete body paragraph here:

Copy the prompt, followed by the body paragraph, into ChatGPT. Copy each of its responses into a separate box on the left. Then write comments about what you think of ChatGPT's suggestions.

ChatGPT's responses	Your comments