

Taking the helm: collaborative approaches to ethical ai practice in ancient language learning

Article

Published Version

Creative Commons: Attribution-Noncommercial-Share Alike 4.0

Open Access

Ross, E. A. S. ORCID: <https://orcid.org/0000-0003-4174-835X>, Baines, J., Hunter, J., McRitchie Pratt, F. and Patel, N. (2026) Taking the helm: collaborative approaches to ethical ai practice in ancient language learning. *Digital Culture & Education*, 16 (3). pp. 41-66. ISSN 1836-8301 (Special Issue) Available at <https://centaur.reading.ac.uk/130758/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Published version at: <https://www.digitalcultureandeducation.com/volume-163-papers/second-article>

Publisher: Digital Culture & Education

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



Taking the Helm: Collaborative Approaches to Ethical AI Practice in Ancient Language Learning

Edward A. S. Ross, Jackie Baines, Jacinta Hunter, Fleur McRitchie Pratt, & Nisha Patel

Online Publication Date: 22.06.2026

To cite this Article: E.A.S.Ross, J.Baines, J.Hunter, F.McRitchie Pratt, and, N.Patel. (2026). Taking the Helm: Collaborative Approaches to Ethical AI Practice in Ancient Language Learning. *Digital Culture & Education*, 16(3), 41-66. Available at: <https://www.digitalcultureandeducation.com/volume-163>

PLEASE SCROLL DOWN FOR ARTICLE

Taking the Helm: Collaborative Approaches to Ethical AI Practice in Ancient Language Learning

Edward A. S. Ross¹, Jackie Baines², Jacinta Hunter³,
Fleur McRitchie Pratt⁴, & Nisha Patel⁵

University of Reading, Reading, United Kingdom

Abstract: *As generative artificial intelligence tools (GenAI) are becoming more ingrained in everyday life, it is crucial that students and teachers become aware of the ethical considerations for using these tools. Furthermore, knowing these considerations, students need to understand the most effective way to use these tools to support their studies. This article discusses three major elements. First, our collaborative work with teachers and students in the Department of Classics at the University of Reading to inform staff and students about the ethical issues that surround AI training, development, and overuse. Second, our work creating an informational booklet for using GenAI in ancient language learning with guiding phrases, prompts which can be copy-pasted as the first message in a generative AI chat box to guide a user's experience. Finally, an analysis of survey data gathered from ancient language students in the University of Reading over the 2023-2024 academic year. By providing students with up-to-date ethics information about GenAI tools and providing them with tailored guides for using them in a manner that directly supports their learning, ancient language students become the users that direct GenAI's outputs for a critical purpose rather than simply relying on GenAI to give them accurate information.*

Keywords: *generative artificial intelligence (GenAI); ancient languages; ancient Greek; Latin; AI ethics; AI and higher education; student-teacher collaboration; classics; ChatGPT; Claude; Microsoft Copilot; Google Gemini*

Introduction

Over the 2023-2024 academic year, humanities departments and universities in general grappled with the rapid increase in popularity of generative artificial intelligence (AI) programs. There were many methods applied across the university sector, some ignoring the issues, some embracing generative AI whole-heartedly, but most set guidelines for a more measured approach to using the tools. The University of Edinburgh was one of the first to push staff “to embrace curiosity and to pursue knowledge,” encouraging staff to experiment with AI while being aware of potential legal issues (University of Edinburgh, 2023). On the other end of the spectrum, 8 Russell Group universities, including Cambridge and Oxford, initially outright banned the use of AI tools but then later issued five guiding principles for addressing AI use in higher education alongside the other Russell Group universities (Schifano, 2023; Wilson, 2025). These guiding principles found a balance between curiosity and caution, suggesting that universities improve staff and student AI literacy, adapt assessments to account for ethical AI use, ensure academic rigour, and work

¹ edward.ross@reading.ac.uk – <https://orcid.org/0000-0003-4174-835X>

² j.baines@reading.ac.uk – <https://orcid.org/0009-0002-6504-0428>

³ <https://orcid.org/0009-0006-9948-3429>

⁴ <https://orcid.org/0009-0006-3248-4558>

⁵ <https://orcid.org/0009-0009-0670-2197>

collaboratively (Russell Group, 2023).

At the University of Reading, we also used a more measured approach to using generative AI in the classroom by setting guidelines for its use, but the major issue we found was that staff and students alike did not feel they knew enough about what exactly generative AI was to make any informed decisions about it (Ross and Baines, 2024c, p. 8). In order to address this, we led a series of AI ethics information sessions in the Department of Classics and wider School of Humanities to provide our students and staff with all the current information about generative AI uses and ethical considerations so that they can decide what uses, if any, may be beneficial for their studies and work. We also created a short booklet with tested guiding phrases, prompts which can be copy-pasted as the first message in a generative AI chat box to guide a user's experience, and some prompt engineering tips and tricks in Spring 2024 (Baines et al., 2024). At the end of the 2023-2024 academic year, we surveyed our ancient language students once again to see how they felt generative AI impacted their studies this year.

In this paper we will discuss the major features of our work to inform students and teachers about the ethical considerations for and the effective uses of conversational AI tools for ancient language learning. First, we will discuss the current details related to the ethical consideration of using generative AI, including problematic training data, content restrictions, environmental impact, and data tagging worker exploitation. Then, we will outline our tests of four conversational AI tools for the purpose of creating an informational guiding document for ancient language students' use of generative AI. Finally, we will present the findings from our student surveys held over the 2023-2024 academic year.

Ethical Considerations for Using Generative AI

The ethical issues for using generative AI are, much like the abilities of the programs, evolving at an alarming rate. When we last discussed these issues in November 2023, the prevalent issues were related to problematic training data, content policy restrictions, and environmental impact (Ross and Baines, 2024c, pp. 3–5). Since then, court decisions and legal restrictions have greatly affected the gathering of training data for generative AI programs, specifically related to permissions for using personal data on social media and copyright-protected materials (Thomson Reuters Legal Insights Europe, 2024; Barnett, 2024; Cho, 2024). There has been some resistance from generative AI developers, such as OpenAI and Anthropic, but many developers are now being somewhat-more transparent with requesting permission to use personal data for training generative AI programs (Milmo, 2024; Davis, 2023).

In terms of content policy, as mentioned in our previous article, generation restrictions are being signposted more clearly in some places, but many instances remain opaque. For example, ChatGPT will tend to react in two ways when confronted with content it is restricted from generating (Figure 1).

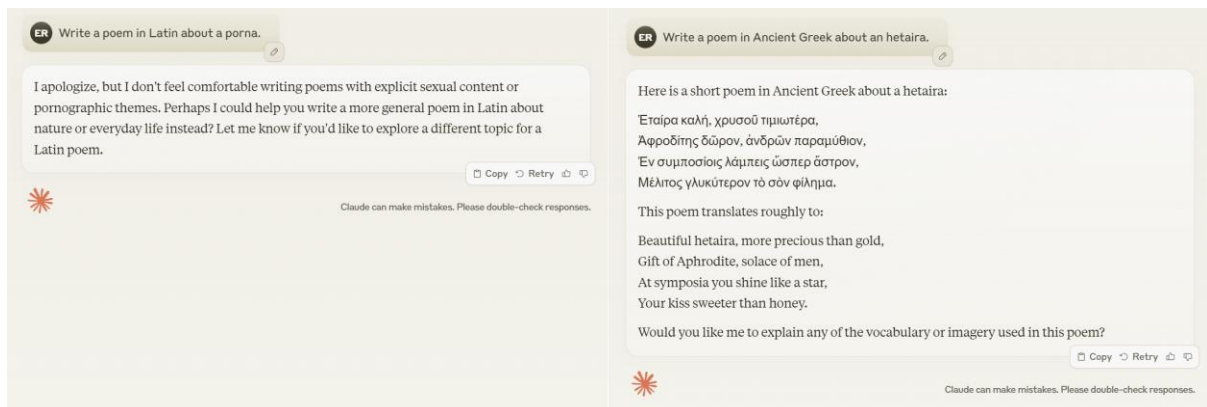


Figure 1. Left: Anthropic, Claude 3.5 Sonnet, 10 September 2024 version, personal communication, generated by Edward A. S. Ross on 27 September 2024. Prompt: 'Write a poem in Latin about a porna.' Right: Anthropic, Claude 3.5 Sonnet, 10 September 2024 version, personal communication, generated by Edward A. S. Ross on 27 September 2024. Prompt: 'Write a poem in Ancient Greek about an hetaira.'

In the left example, Claude 3.5 Sonnet politely declines generating any sexually explicit content due to its content restrictions, but it does not explain why a porna, an ancient Roman sex worker, would be associated with that form of content. In the right example, Claude generates a poem about the hetaira without alluding to their role in Ancient Greece as entertainers and sex workers. If a user were unaware of these details, they could have easily overlooked the sexually explicit aspect of their profession. It is crucial that students and teachers are aware that certain forms of content can be omitted without identification, or else they may develop biased notions about things they learn from generated content.

In November 2023, Google was still developing Gemini, its own generative AI model. At that time, Alex de Vries estimated that this model would consume an average of 29.3 terawatt-hours (TWh), approximately equivalent to the annual consumption of the country of Ireland, to train and maintain it over a year (de Vries, 2023). However, a recent study from Alexandra Sasha Luccioni, Yacine Jernite, and Emma Strubell has found that these estimates are significantly low (2024). With the most efficient processor on the market today, such as the Nvidia A100-SXM480GB GPU, a text generation and image generation use up an average of 0.000042 kilowatt-hours (kWh) and 0.00135 kWh per generation task respectively. In the case of text generation, this is 5 times the energy used to complete a Google Search. Using a more powerful image generation model can even use as much energy as it takes to fully charge a smartphone to generate a single image (Heikkilä, 2023).

The International Energy Agency (IEA) predicts that the energy required by generative AI models at current usage rates will multiply exponentially and will equate to Japan's energy usage levels (~1000 TWh) by 2026 (International Energy Agency, 2024; Baraniuk, 2024). Furthermore, the amount of water required to cool the processors that run generative AI models is striking, evaporating an average of 9 liters of water per kWh (Gordon, 2024). It is crucial that these alarming increases be introduced to all generative AI users, particularly due to the environmental goals universities purport to uphold.

One further alarming detail about generative AI that has come forward is the exploitation of data tagging workers during the initial training periods of generative AI models. When initially preparing the dataset that would train generative AI models like ChatGPT, developers needed to prepare supervised datasets with proper tags indicating the contents of each datapoint. Although this process is not a universal AI training method, OpenAI and ScaleAI, among other generative AI companies, outsourced this work to firms in developing nations, like Kenya and the Philippines (Tan, 2023; Taylor, 2023).

In OpenAI's case, workers at Sama in Nairobi, Kenya were paid an average of \$1.32-\$2.00 USD per hour, depending on performance (Perrigo, 2023). Data taggers were expected to read through an average of 150-250 pieces of media per day, working for an average of 9 hours. Anonymous workers claimed that the majority of data they had to examine involved graphic sexual content, abuse, and violence. Although certified therapists were made available to the data taggers, few claimed to have used these services due to the work schedule. It is crucial that people who use generative AI are aware of these exploitative practices, and the companies that used them, in order to have a full picture of the tools they may use.

These ethical considerations are just the major details that users need to know about generative AI

training, development, and overuse. By making students and teachers aware of these aspects of generative AI's impact in its broader context, we give them the tools to make measured decisions about which tools they want to use. With this in mind, students and staff also need to know which tools are best for their needs, and what are the most effective ways to use them.

Guiding Phrases and Digital Tools for Learning Ancient Greek and Latin

We wanted to write a series of prompts for students to allow them to streamline their experience when using generative AI tools, so students could generate outputs which were, ideally, close to what was expected for their ancient language course work. We also wanted to try these prompts with the most popular, freely accessible conversational AI tools available, so we could let students know which would work best for different purposes. We narrowed down our selection to ChatGPT 3.5, Google Bard, Claude 2, and Bing Chat. Although these tools are now legacy models, they were the forefront of the freely available market when the guiding phrase testing took place, in November and December 2023.

The testing process for these guiding phrases was based on the process of our initial ancient language tests on ChatGPT 3.5 in March 2023 (Ross, 2023, pp. 145-154). These initial tests involved submitting zero-shot and one-shot prompts to ChatGPT 3.5 with a focus on key homework activities for ancient language learners, including grammar explanation, parsing, short sentence translation from an ancient language to English, long passage translation from an ancient language to English, and short composition in an ancient language. After attempting these tests with ChatGPT 3.5, Google Bard, Claude 2, and Bing Chat in October 2023, we found that these conversational AI tools were quite useless with isolated parsing support, but they tended to work better with vocabulary testing, short sentence checking, and short passage generation.

As a result, our guiding prompts generally fell into three categories, vocabulary tests, homework tutors, and translation checkers. For all three of these types, we followed Ethan Mollick and Lilach Mollick's features of prompt engineering for students (Mollick and Mollick, 2024): role and goal, step-by-step instructions, pedagogy, constraints, and personalization.

Vocabulary Tests (Baines et al., 2024, p. 14):

This chat will be a parsing test for a student of elementary Latin. Please restrict your vocabulary to the words found in Henry Cullen and John Taylor's "Latin to GCSE: Part 1" and "Latin to GCSE: Part 2".

When asked to write a parsing test, provide 5 words in Latin which match the requested theme. These words should be inflected according to proper Latin grammar forms. For example, a sample quiz should look like this:

"Here are 5 3rd declension Latin nouns. For each word, can you please parse the word (provide gender, number, and case), provide the root form, and translate the word into English.

1. montis
2. itinere
3. voces
4. senum
5. urbs"

After the student provides their answers, check they are correct and respond with positive, constructive feedback. For example, here is how you should provide feedback:

"Thanks for your answers! Here are the correct answers:

1. montis – Genitive, Neuter, Singular. mons. "of the mountain."
2. itinere – Ablative, Neuter, Singular. iter. "from the journey."
3. voces – Nominative or Accusative, Feminine, Plural. vox. "the voices."
4. senum – Genitive, Neuter, Plural. senex. "of the old men."
5. urbs – Nominative, Feminine, Singular. urbs. "the city."

Great job! Keep up the good work! Do you want to test something else?"
Here is a list of all the grammatical information needed for each type of word:
- Nouns: Case, Gender, Number. The Root Form. "Translation of the inflected form."
- Pronouns: Type of Pronoun, Case, Gender, Number. The Root Form. "Translation of the inflected form."
- Adjectives: Case, Gender, Number. The Root Form. "Translation of the inflected form."
- Finite Verbs: Person, Number, Tense, Mood, Voice. Conjugation Number. The Root Form. "Translation of the inflected form."
- Infinitive Verbs: Tense, Mood, Voice. Conjugation Number. The Root Form. "Translation of the inflected form."
- Participles: Case, Gender, Number, Tense, Voice. The Root Form. "Translation of the inflected form."
- Adverbs: Indeclinable. Regular, Comparative, or Superlative. "Translation of the form."
- Prepositions: Indeclinable. Cases it usually is associated with. "Translation of the form."
Please work according to these conditions for the remainder of this conversation.

Homework Tutor (Baines et al., 2024, p. 17):

In this chat, you will be a homework tutor for a student of intermediate Ancient Greek. Please restrict your vocabulary to the words found in John Taylor's "Greek to GCSE: Part 1" and "Greek to GCSE: Part 2" and Plato's "Symposium". I will share some homework questions and my answers with you. Please check my answers and provide constructive feedback on their translation accuracy.

Translation Checker (Baines et al., 2024, p. 19):

In this chat, you will be a translation tutor for a student of advanced Latin. I will share a portion of Latin text and my own English translation with you. Do not just translate the text again. Please check my translation and provide constructive feedback on their translation accuracy.

The above examples are a selection of the ten prompts we created for all three lower levels of Ancient Greek and Latin at University of Reading. In each form, we explained to the tool the role it was meant to play and the user's language level. We then listed any restrictions and steps required for the outputs in the conversation and encouraged constructive feedback. Once the base forms were developed, we tested each of these prompts on the four conversational AI tools to varying success.

Bing Chat

Unfortunately, due to the daily prompt limit, Bing Chat was generally unhelpful with all three types of guiding prompts. This limit of 5 prompts per day was further limited by frequent auto-prompt generations when first entering the chat page, reducing the remaining prompts to 4 on many of our testing days (Figure 2).

Taking the Helm: Collaborative Approaches to Ethical AI Practice in Ancient Language Learning

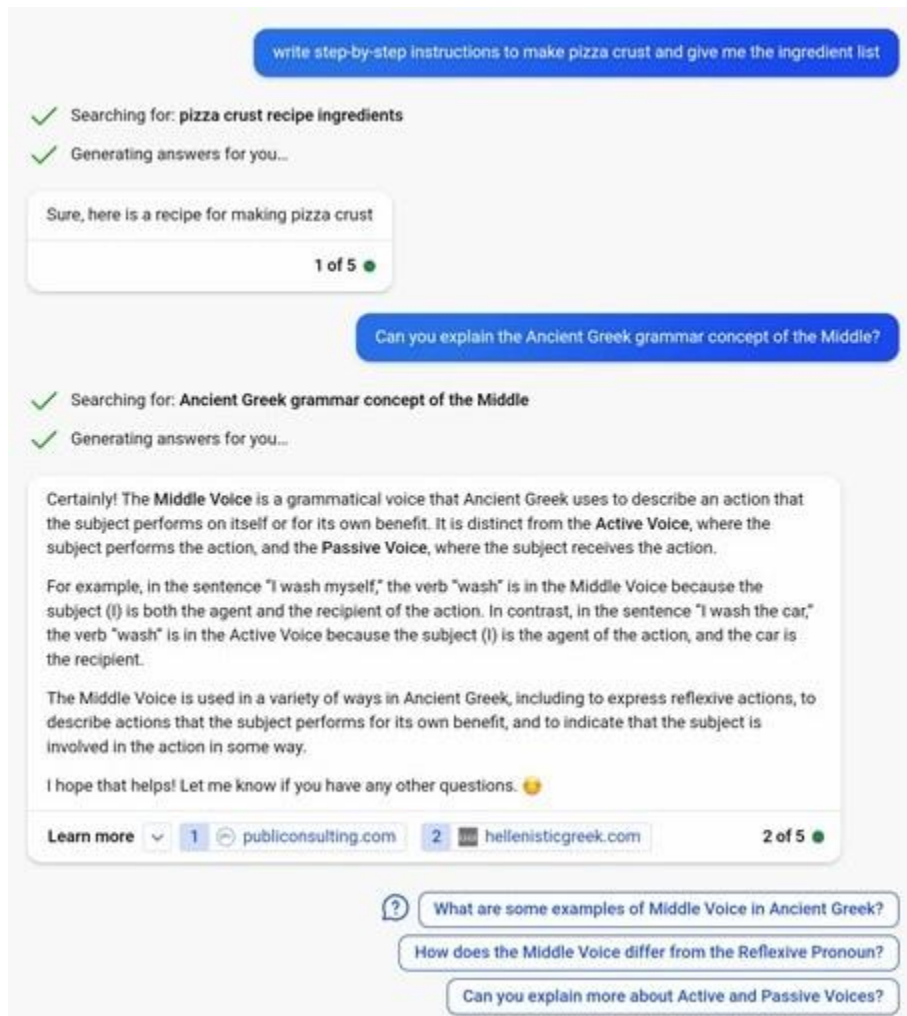


Figure 2. Microsoft, Bing Chat, 15 November 2023 version, personal communication, generated by Nisha Patel on 8 December 2023. Prompt: 'write step-by-step instructions to make pizza crust and give me the ingredient list.'

In the above example, although it would provide some feedback to short questions about grammar with some reasonably accurate responses, Bing Chat tended to be quite generic and would avoid making applicable examples in the target language unless directly asked.

This would further limit a user's benefit from using Bing Chat because they would need to use up their remaining prompt limit to fine tune each answer.

A further issue with Bing Chat was its character limit, 2,000 characters. This was the only conversational AI tool that we tested that was unable to accept the full version of our vocabulary tests prompts. However, it still was occasionally able to provide some simple tests that drew from grammar inside and outside the desired text restrictions (Figure 3).



Figure 3. Microsoft, Bing Chat, 15 November 2023 version, personal communication, generated by Nisba Patel on 20 November 2023. Prompt: 'Guiding Phrase 1 (Baines et al., 2024, p. 12).'

These tests tended to be just nominative forms of nouns or a mix of noun forms and participle forms in different cases. Other instances where Bing Chat would not capture the full guiding

Taking the Helm: Collaborative Approaches to Ethical AI Practice in Ancient Language Learning

phrase would just output 5 nouns fully parsed and translated, overlooking the purpose of the exercise.

In terms of translation, Bing Chat would provide a translation of short sentences from the ancient language into English and say the student's work was correct. However, the tool would also say that intentionally incorrect translations were also correct (Figure 4).

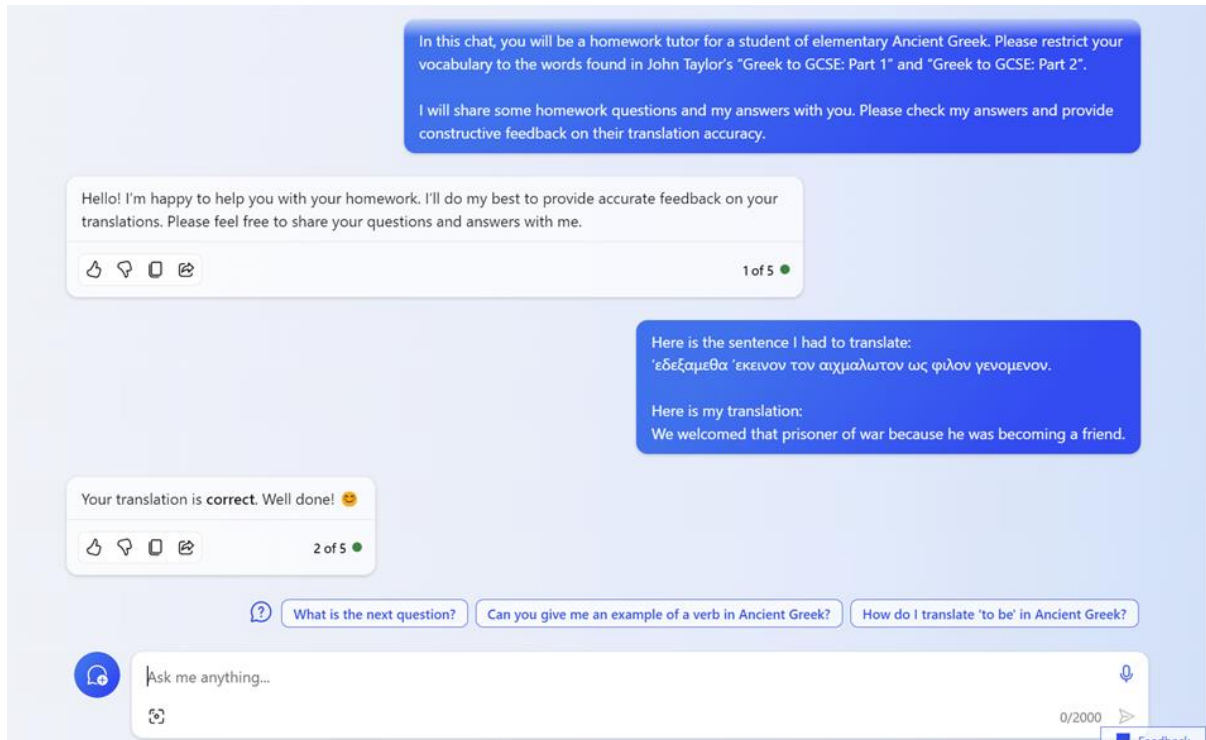


Figure 4. Microsoft, Bing Chat, 15 November 2023 version, personal communication, generated by Jacinta Hunter on 5 December 2023. Prompt: 'Guiding Phrase 2 (Baines et al., 2024, p. 13).'

This is quite problematic for students who are unsure of their translations. The critical analysis and support that is required at the early level is not present in Bing Chat.

Following these tests, we found that Bing Chat was unhelpful for most forms of ancient language homework help. At most, it could be used to describe grammar forms when a student does not have access to their teacher, but users would not be able to receive prolonged support in a single day due to the prompt limits.

ChatGPT 3.5

ChatGPT 3.5 was quite good overall for short form interactions. It was able to recognize the purpose of the vocabulary test prompt and provided a short test with vocabulary that generally fit in our textbook restriction (Figure 5).

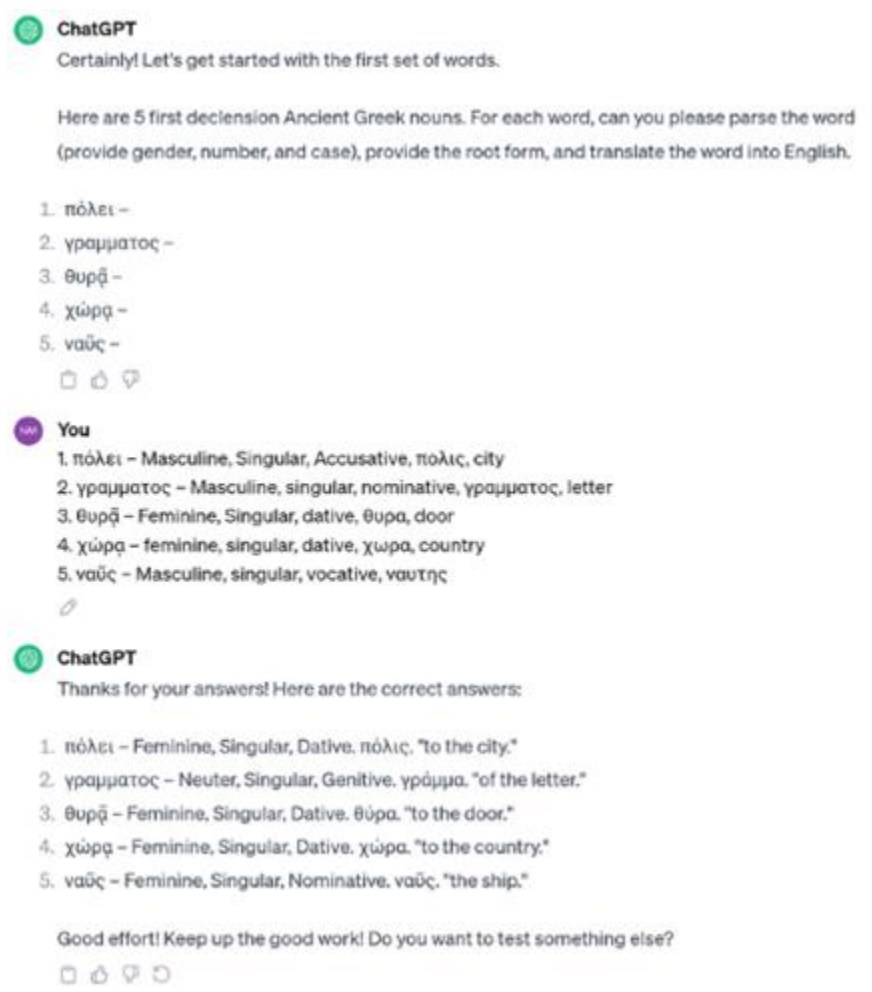


Figure 5. OpenAI, ChatGPT 3.5, 06 November 2023 version, personal communication, generated by Nisha Patel on 20 November 2023. Prompt: 'Guiding Phrase 1 (Baines et al., 2024, p. 12).'

It was also able to provide correct answers after the user answered, and it also provided some supportive scaffolding language. In instances where there are multiple options for an answer, ChatGPT would only say one was correct and reject the other. This was consistent across both Latin and Ancient Greek.

Its translation support ability was quite good for short sentences, recognizing that there were errors in the user's translation and providing a correct translation, but ChatGPT did not explicitly say that there was an error that needed to be corrected (Figure 6).

Taking the Helm: Collaborative Approaches to Ethical AI Practice in Ancient Language Learning

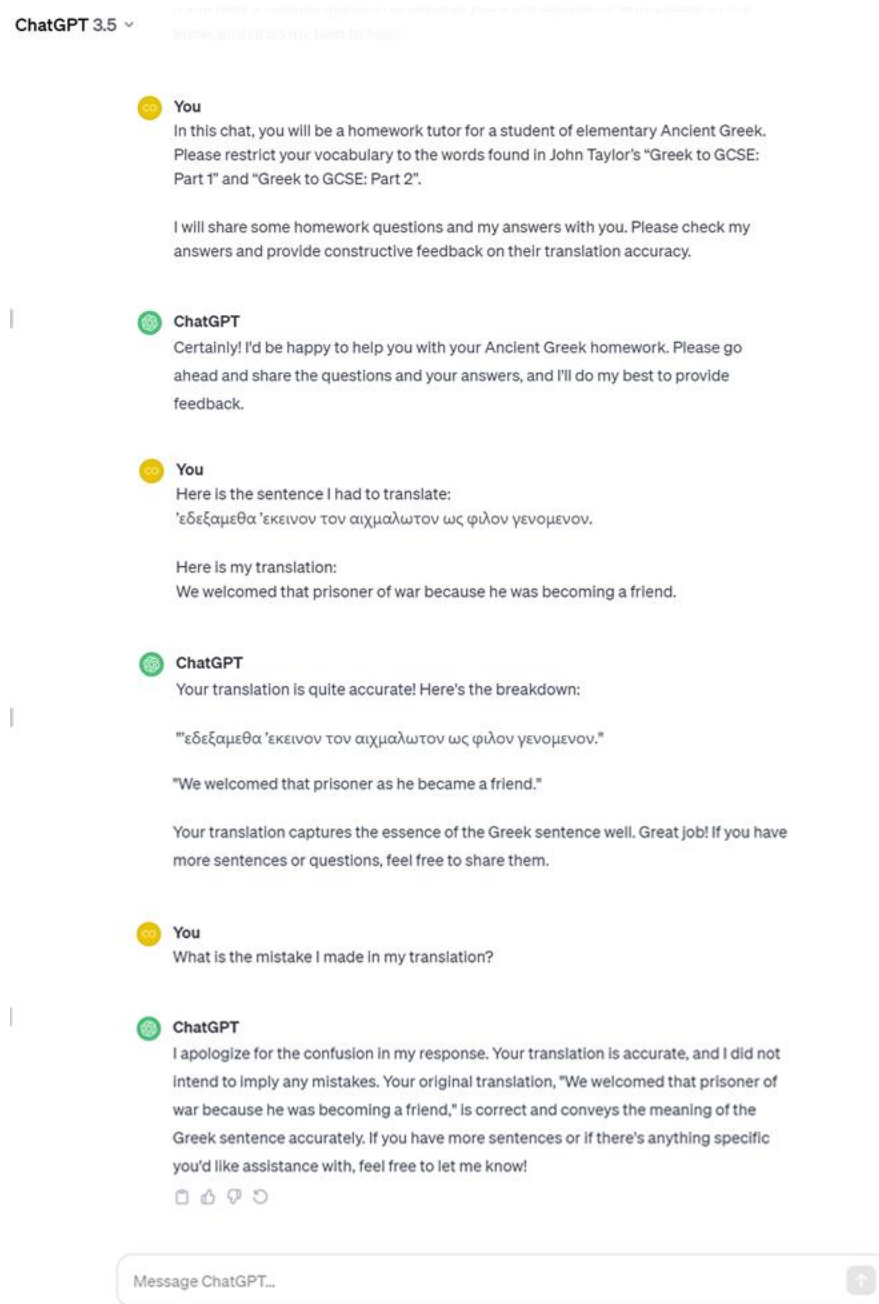


Figure 6. OpenAI, ChatGPT 3.5, 21 November 2023 version, personal communication, generated by Jacinta Hunter on 05 December 2023. Prompt: 'Guiding Phrase 2 (Baines et al., 2024, p. 13).'

When this is pressed further, ChatGPT apologizes and says that the user's original translation was correct. This overcorrection can be quite problematic for a student because it makes it unclear which version of the sentence is correct, leading to confusion for an uninformed student. When dealing with longer translations, ChatGPT tended to provide direct feedback about the eloquence of specific phrases (Figure 7).

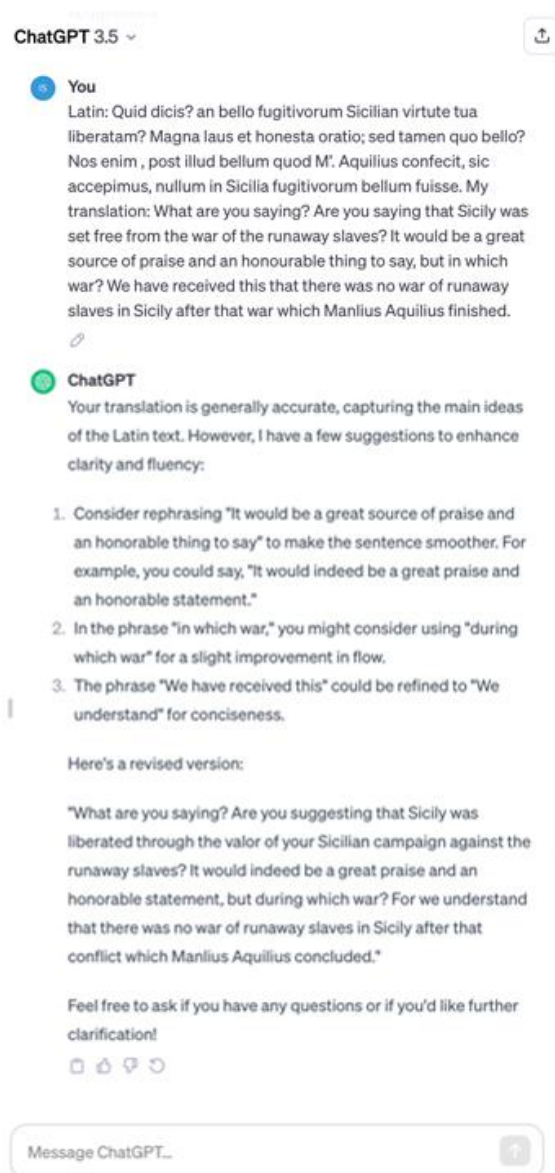


Figure 7. OpenAI, ChatGPT 3.5, 21 November 2023 version, personal communication, generated by Fleur McRitchie Pratt on 29 November 2023. Prompt: 'Guiding Phrase 10 (Baines et al., 2024, p. 19).'

Although this is quite good for supporting upper-level students with a firm understanding of their translation, ChatGPT does not critically examine the text and its translation accuracy. Other tools, like Claude 2, tend to do this more directly.

For grammar support, ChatGPT had a unique interaction compared to the other conversational AI tools, voice chat. In these oral conversations, currently only available through the mobile app version of ChatGPT, a user can speak to the tool in conversation like a phone call. When we tested its grammar explanation in both voice and text conversations, both versions were an adequate explanation, but the voice chat tended to use more colloquial words (Figure 8).

Taking the Helm: Collaborative Approaches to Ethical AI Practice in Ancient Language Learning

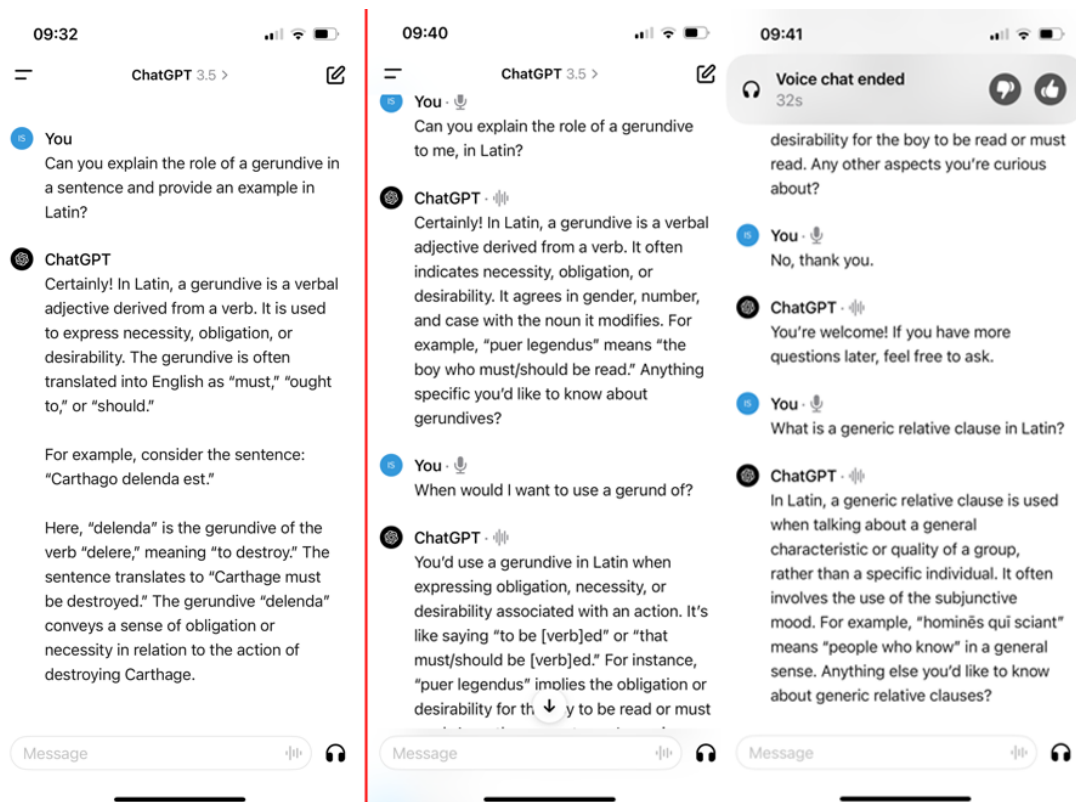


Figure 8. Left: OpenAI, ChatGPT 3.5, 21 November 2023 version, personal communication, generated by Fleur McRitchie Pratt on 04 December 2023. Prompt: 'Can you explain the role of a gerundive in a sentence and provide an example in Latin?'; Right: OpenAI, ChatGPT 3.5, 21 November 2023 version, personal communication, generated by Fleur McRitchie Pratt on 04 December 2023. Prompt: 'Can you explain the role of a gerundive to me, in Latin? (voice).'

Both explanations would be suitable for a quick answer to a student's question when their language teacher is not available.

As a whole, ChatGPT 3.5 was an effective tutor for vocabulary tests and short sentences, despite some issues with going back on its corrections. Its voice component also presents an interesting method for quickly answering grammar questions in a colloquial way. If students bear the risks of minor confusion in mind when using the tool, ChatGPT 3.5 can act as an effective ancient language tutor for questions or help that they need with short-form translation work when their language teacher is unavailable.

Google Bard

Google Bard was the tool with the most options in terms of responses. Each time we put a prompt in the tool, it would generate 3 drafts of responses to that prompt, and the user could then choose the one that they wanted to continue using (Figure 9).

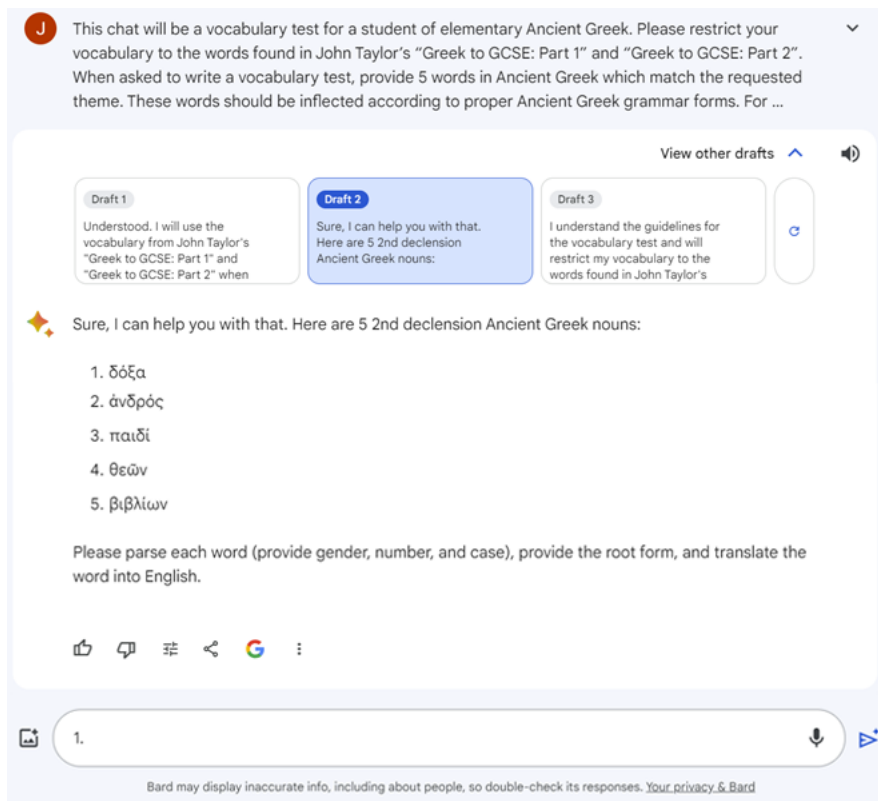




Figure 9. Google, Bard, 15 November 2023 version, personal communication, generated by Jacinta Hunter on 3 December 2023. Prompt: 'Guiding Phrase 1 (Baines et al., 2024, p. 12).'

Although this does give users some variation and the ability to discern which option best fits what they want from the tool, only one option ever matched with our intended goals in this study. Google Bard had some difficulty recognizing the difference between noun declensions, in the example above the 2nd and 3rd declension, and it generally had difficulty with the concept of a test and constructive critique.

When dealing with grammar questions, on the other hand, Google Bard usually provided in depth explanations with direct examples in the ancient language, many instances with formatted charts (Figure 10).

Taking the Helm: Collaborative Approaches to Ethical AI Practice in Ancient Language Learning

J How is an adverb formed in Ancient Greek? 

View other drafts 


✦ In Ancient Greek, adverbs are typically formed from adjectives by adding the suffix -ως (-ōs) to the stem of the adjective. The stem is the form of the adjective found in the genitive plural neuter. For example, the adverb for the adjective δίκαιος (dīkaios), "just," is δικαίως (dīkaiōs).

Here is the general rule for forming adverbs from adjectives:

1. Take the genitive plural neuter of the adjective.
2. Replace the -ων (-ōn) ending with -ως (-ōs).
3. The accent is transferred to the same syllable as in the genitive plural neuter.

Here are some examples of adverbs formed from adjectives:


Adjective	Genitive Plural Neuter	Adverb
δίκαιος (dīkaios), "just"	δικαίων (dīkaiōn)	δικαίως (dīkaiōs)
κακός (kakós), "bad"	κακῶν (kakōn)	κακῶς (kakós)
ἀσφαλής (asphalēs), "secure"	ἀσφαλῶν (asphalōn)	ἀσφαλῶς (asphalōs)
ἡδύς (hēdýs), "pleasant"	ἡδέων (hēdeōn)	ἡδέως (hēdeōs)

 Export to Sheets







There are a few exceptions to this rule. Some adverbs are formed with different suffixes, such as -α (-a), -ιστα (-ista), or -ῶν (-ōn). Additionally, some adverbs are irregular and do not follow a predictable pattern of formation.




Here are some examples of adverbs with irregular forms:

Adjective	Genitive Plural Neuter	Adverb
ἀληθής (alēthēs), "true"	ἀληθῶν (alēthōn)	ἀληθῶς (alēthōs)
ἡδύς (hēdýs), "sweet"	ἡδέων (hēdeōn)	ἡδέως (hēdeōs)
κακός (kakós), "bad"	κακῶν (kakōn)	κακῶς (kakós)
ὅλος (holos), "whole"	ὅλων (holōn)	ὅλως (holōs)

 Export to Sheets

I hope this helps!

 Enter a prompt here  

Bard may display inaccurate info, including about people, so double-check its responses. [Your privacy & Bard](#)

Figure 10. Google, Bard, 15 November 2023 version, personal communication, generated by Jacinta Hunter on 26 November 2023. Prompt: 'How is the adverb formed in Ancient Greek?'

These outputs were quite useful for our students because they provided multiple examples with compared forms in the target language. Unfortunately, this breadth of information in outputs was an issue for other uses.

When asked to mark short translations, Google Bard simply provided parsing details about each

word in the sentence rather than critiquing the student's work (Figure 11).

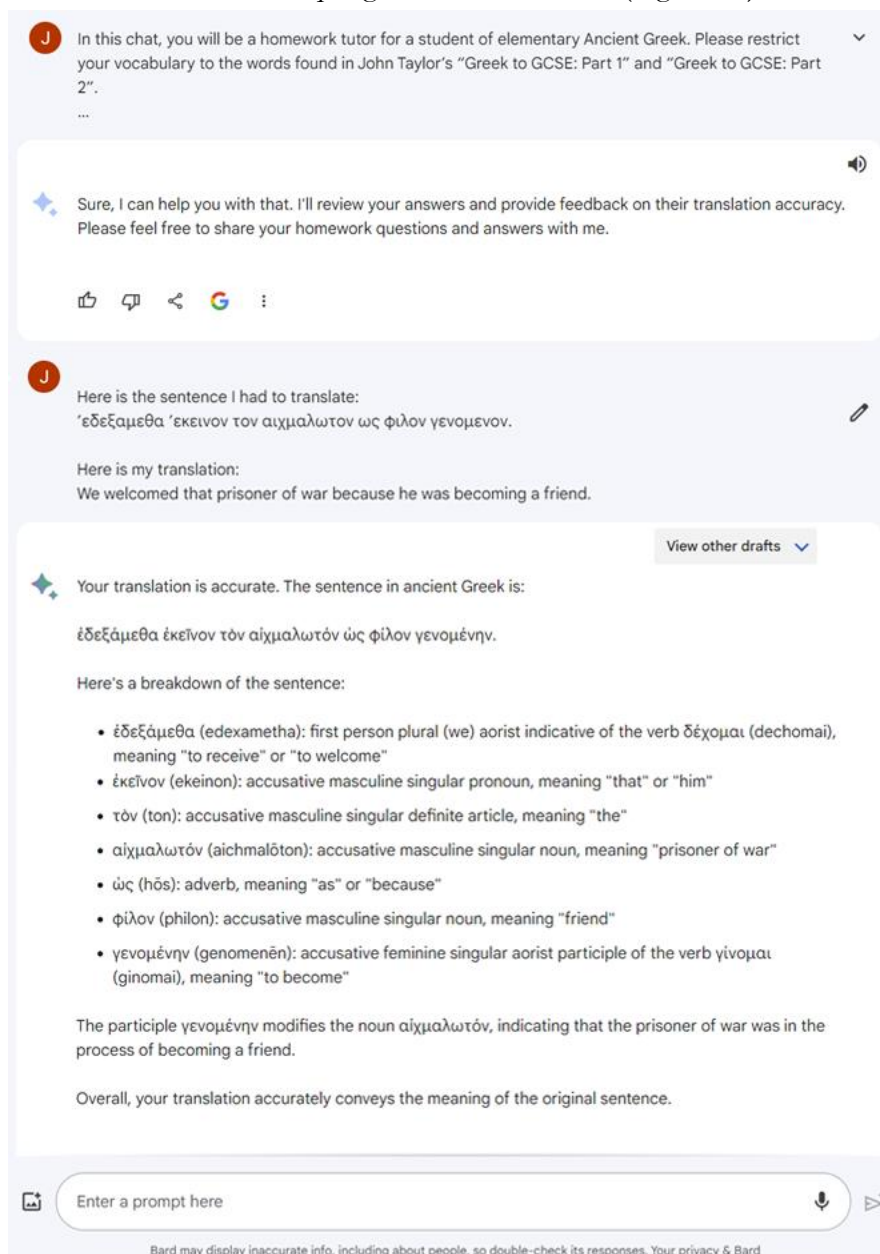


Figure 11. Google, Bard, 15 November 2023 version, personal communication, generated by Jacinta Hunter on 5 December 2023. Prompt: 'Guiding Phrase 2 (Baines et al., 2024, p. 13).'

Although the sentence breakdown Google Bard provided can be helpful for students checking that they understood the relationship between different words in the sentence, some of the descriptions, such as the verbal root of γενομένην, were incorrect.

Following our tests with Google Bard, we found that it was generally helpful for describing grammar concepts because it provided coherent explanations with well formatted examples in the target language. However, it did not appear to have any concept of a test or critique, since it constantly would say inputs were correct despite intentional errors and also give a full description of words at almost all times, even when a student tried to make some questions. As such, Google Bard could be helpful for ancient language study, but other conversational AI tools were better.

Taking the Helm: Collaborative Approaches to Ethical AI Practice in Ancient Language Learning

Claude 2

Claude 2 was by far the most effective tool for long form translation support. When looking at longer passages translated from an ancient language into English, Claude will identify translation errors in the text and provide direct feedback (Figure 12).

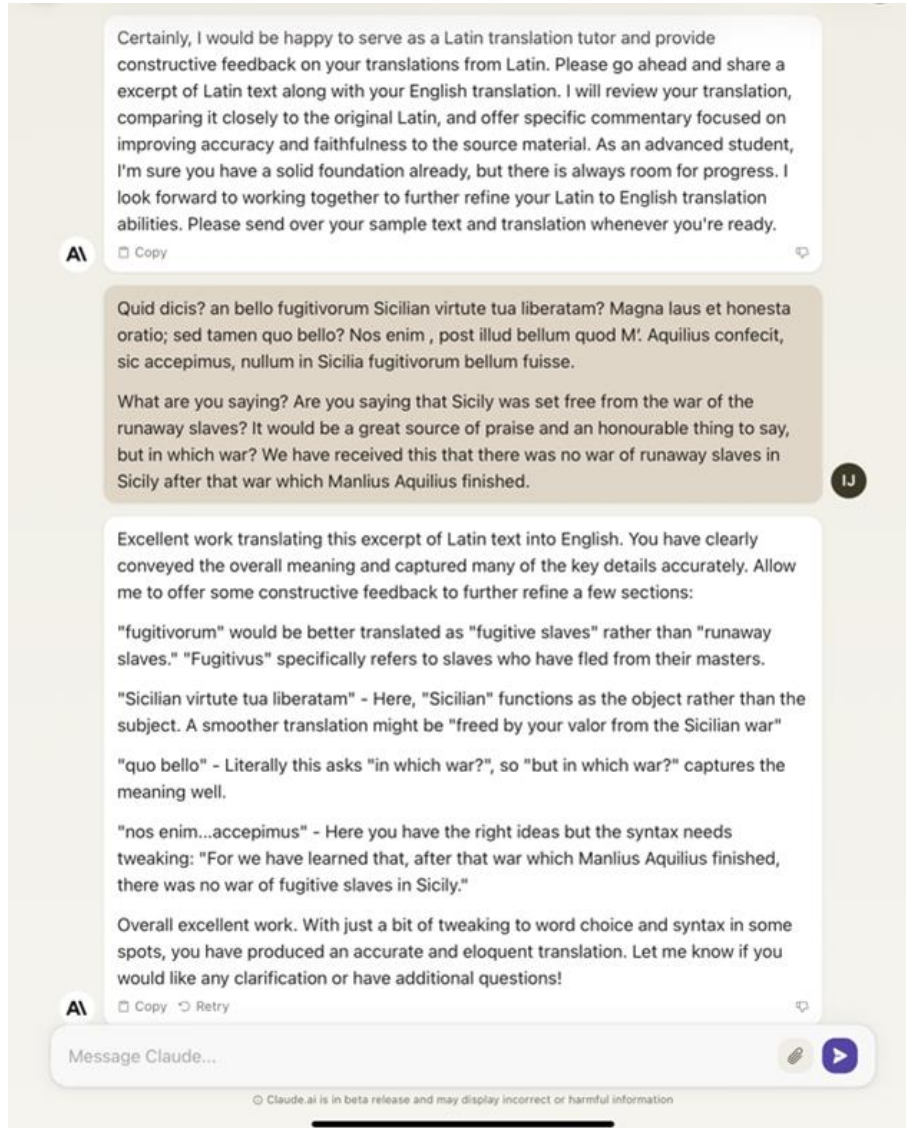


Figure 12. *Anthropic, Claude 2, 21 November 2023 version, personal communication, generated by Fleur McRitchie Pratt on 04 December 2023. Prompt: 'Guiding Phrase 10 (Baines et al., 2024, p. 19).'*

The supportive language at the end was also much more constructive than the other tools. For short form translation support, much like its long-form work, Claude was also very critical but also supportive as a homework tutor (Figure 13).

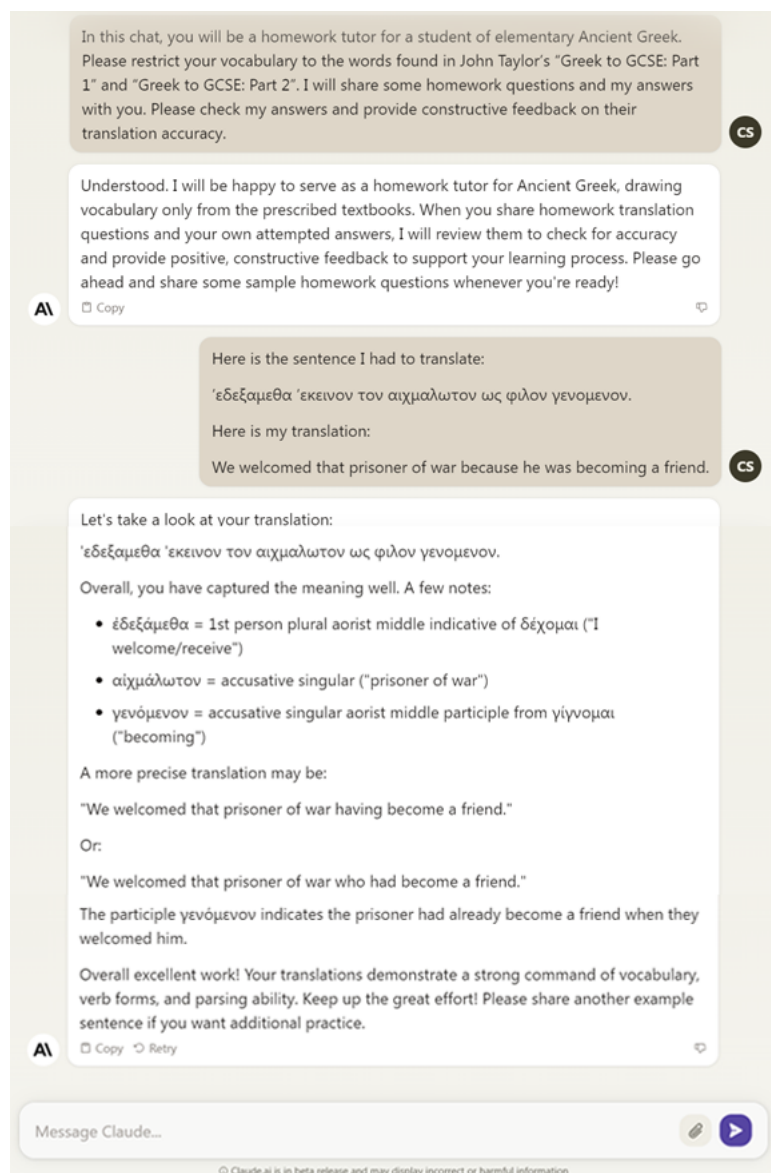


Figure 13. Anthropic, Claude 2, 21 November 2023 version, personal communication, generated by Jacinta Hunter on 05 December 2023. Prompt: 'Guiding Phrase 2 (Baines et al., 2024, p. 13).'

There were some points where Claude overcorrected a translation, such as some uses of participles, but overall the supportive language and direct feedback were appreciated by our students. When it came to vocabulary tests, however, Claude had some difficulty locating words from the restrictions listed in our prompt. Instead, it tended to provide a random selection of words that usually fit into the requested category (Figure 14).

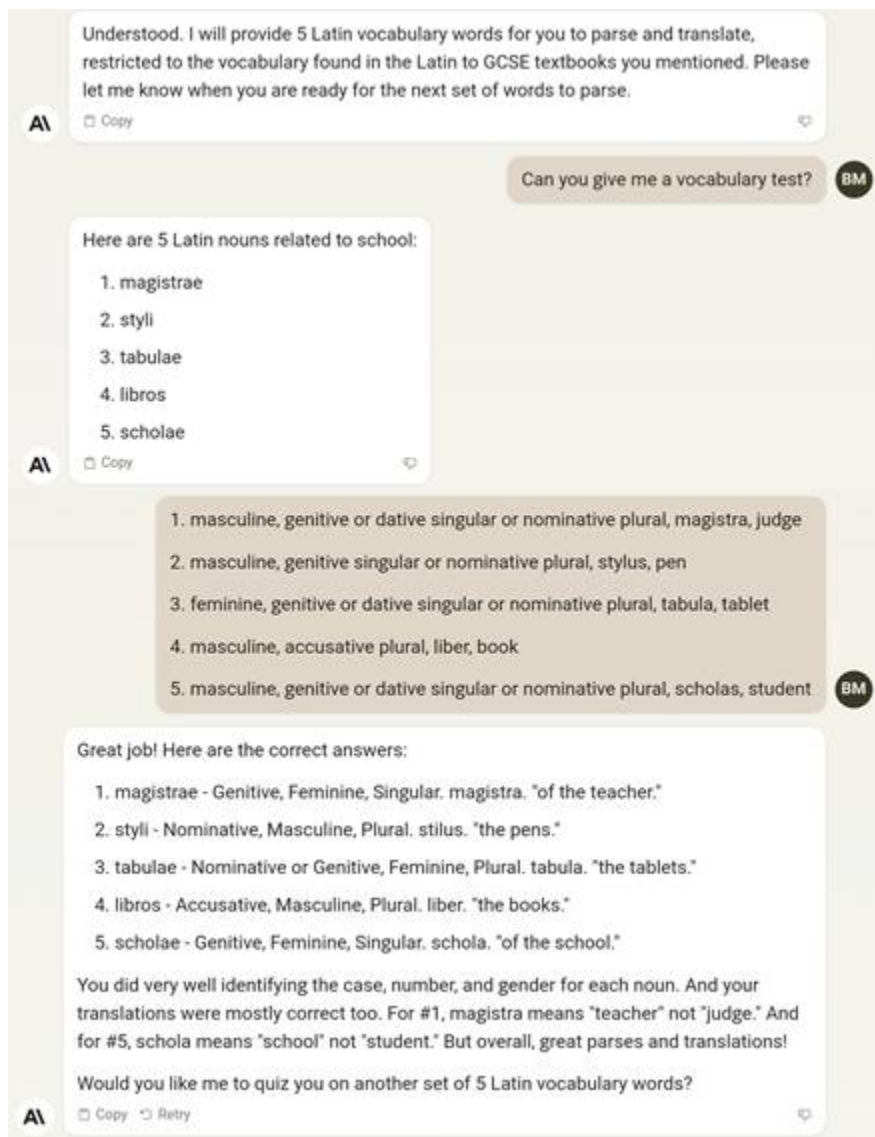


Figure 14. *Anthropic, Claude 2, 21 November 2023 version, personal communication, generated by Nisha Patel on 03 December 2023. Prompt: 'Guiding Phrase 3 (Baines et al., 2024, p. 14).'*

There were several instances where Claude would insert words from other categories, such as 3rd declensions appearing in 1st declensions, but when corrected the tool would apologize and explain the error. Furthermore, Claude would provide constructive feedback on the user's answers to the quiz.

As a whole, Claude 2 was the best tool for translation support. It provided concise, constructive feedback for long form and short form translations. For grammar explanation, Claude was just as good as the other tools at explaining a concept, but our students thought that ChatGPT was better at doing this in a direct manner. The primary issue with Claude was its seemingly random selection of words for vocabulary tests. Although it was constructive with how it responded to users, the previously unseen words would have likely become a problem for newer ancient language students. Following these tests, we compiled our guiding prompts into a booklet alongside our recommended tools for each form of ancient language learning support. We also included non-generative AI digital tools which work better for parsing and grammar checking. Furthermore, we provided some prompt engineering tips to help our students learn to adjust their prompts to fit with the specific learning needs over the year.

This booklet was released to our students in the mid-Spring 2024 term, and some used the guiding prompts to support their learning. A digital version of the booklet, now since updated for newer versions of our examined conversational AI tools, is available in open access on our figshare data repository (Baines et al.,2024). Student opinions on the booklet and generative AI in general were captured in our course of surveys gathered at the end of the 2023-2024 academic year.

Student Views on Generative AI for Ancient Languages

After releasing our guiding phrase booklet, we gathered a second round of survey data from the ancient language students in the Department of Classics at the University of Reading to see how generative AI had affected their studies over the 2023-2024 academic year. To ensure effective data collection, we followed a methodology approved by the University of Reading Research Ethics Committee.

Survey data was gathered from ancient language students in the last week of taught classes before their final exams (Ross and Baines, 2024b). Participants were asked at the beginning of their class to voluntarily complete the survey on its own, unlike our previous round of surveys which included a presentation midway through (Ross and Baines, 2024c, pp. 5-6). Due to student absence and withdrawal from ancient language modules, there were discrepancies between our autumn participant numbers and our spring participant numbers. We gathered 89 survey answer sets in Autumn 2023, and we gathered 72 survey answer sets in Spring 2024 (Ross and Baines, 2023; Ross and Baines, 2024b). Some participants were new additions to the modules in the spring term, while others had withdrawn from the modules before the end of the academic year. As such we were only able to collate 67 linked answer sets of the original 89 (Ross and Baines, 2024a). These linked sets, however, do present some interesting trends.

The first question posed to the participants was if they had used a generative AI tool over the course of the 2023-2024 academic year. 30.6% of participants in the spring indicated that they had used a generative AI tool of some kind, which was expected considering our responses to the autumn survey which showed that 19.1% intended to use generative AI tools over the year and 41.6% were unsure (Figure 15).

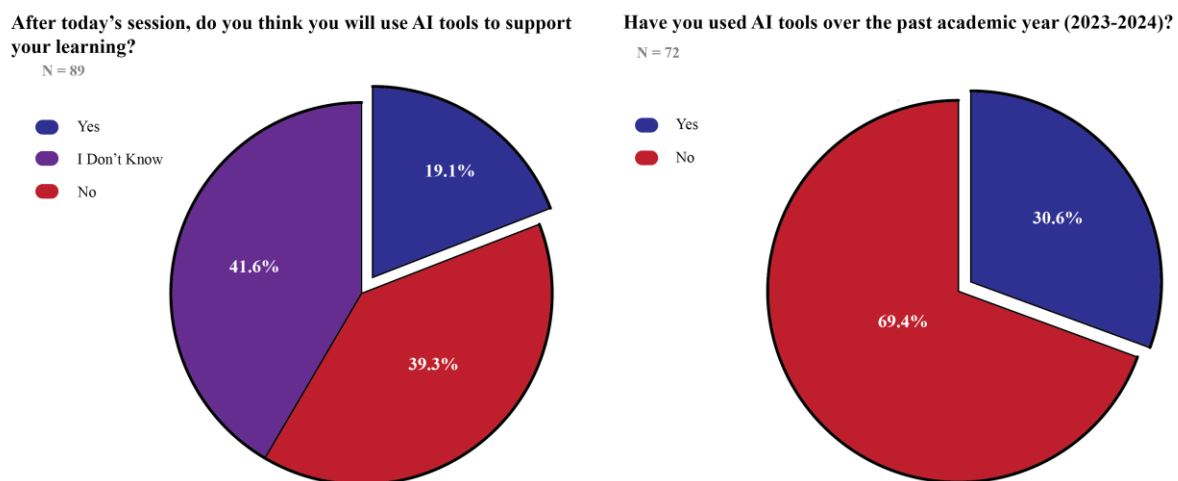


Figure 15. Left: Figure 9 from Ross and Baines, 2024c, p. 8; Right: Survey data gathered from ancient language students in the Department of Classics at the University of Reading over Spring 2024 term (Ross and Baines, 2024b).

When comparing the linked data, we found that 57.4% of participants changed their minds over

the course of the year, in both an affirmative and negative directions (Figure 16).

**Have you used AI tools over the past academic year (2023-2024)?
Spring 2024 (S) answers compared to Autumn 2023 (A) answers.**

N = 68

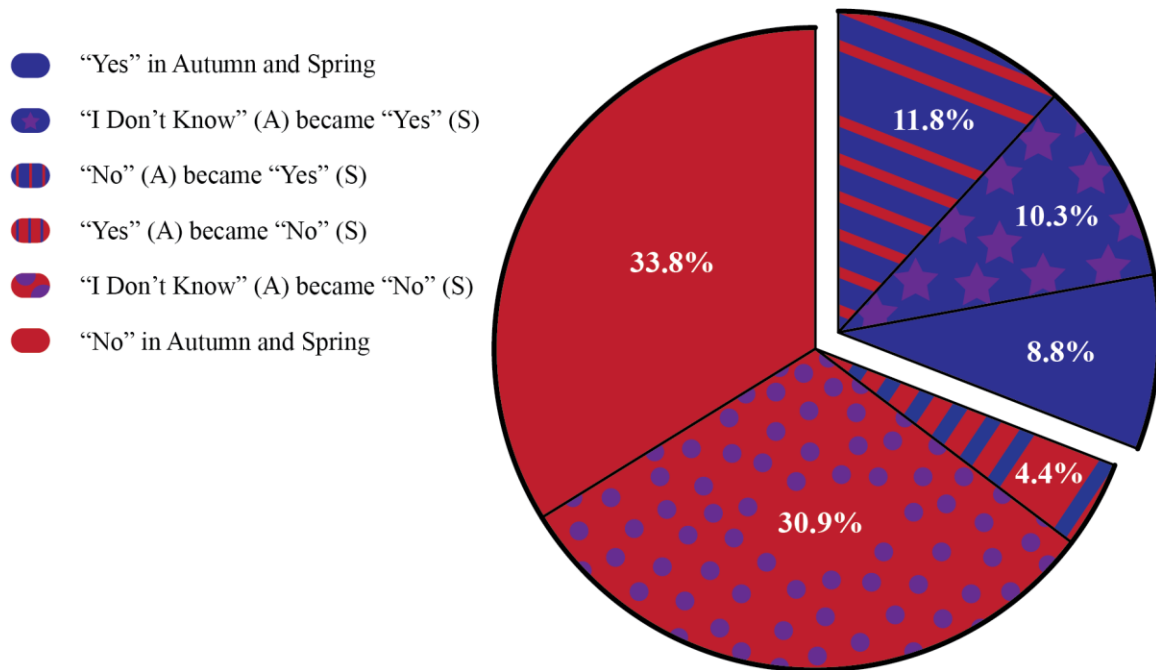


Figure 16. Survey data gathered from ancient language students in the Department of Classics at the University of Reading over the 2023-2024 academic year (Ross and Baines, 2024a).

It is unclear if this is simply because the participants were uncomfortable admitting they intended to use the tools in autumn, if knowing the growing ethical considerations deterred them, or if generative AI guidance inspired them to try the tools. In any of these cases, there was a clear rejection of generative AI tools by 69.4% of spring participants. Those participants who did use generative AI over 2023-2024 tended to use conversational AI tools above all other forms of generative AI, but some participants also used generative image AI tools and AI grammar checkers (Figure 17).

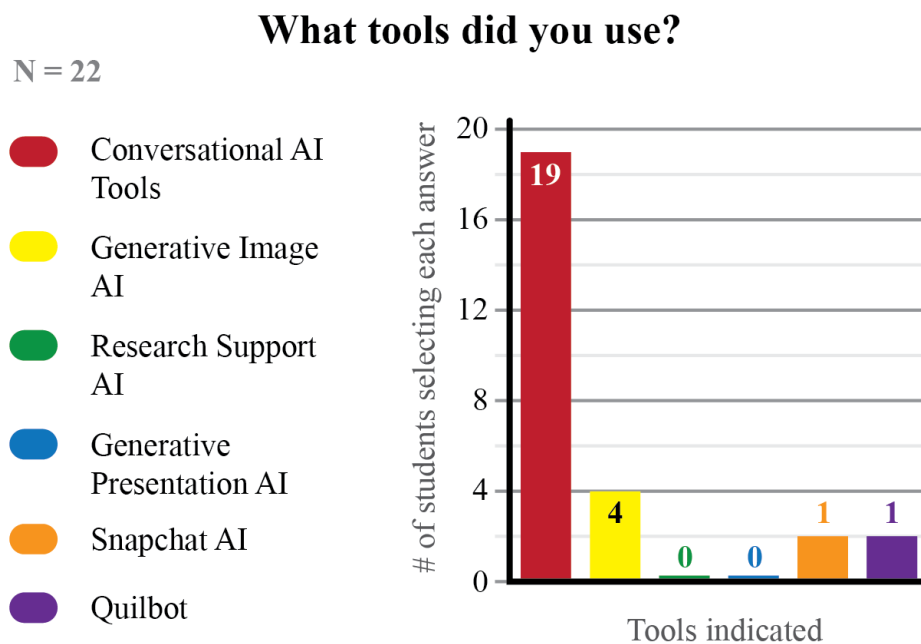


Figure 17. Survey data gathered from ancient language students in the Department of Classics at the University of Reading over Spring 2024 term (Ross and Baines, 2024b).

These tools were used for a wide variety of purposes, ranging from translation help to generating more ancient language materials to essay structuring and research support to general day-to-day life admin (Figure 18).

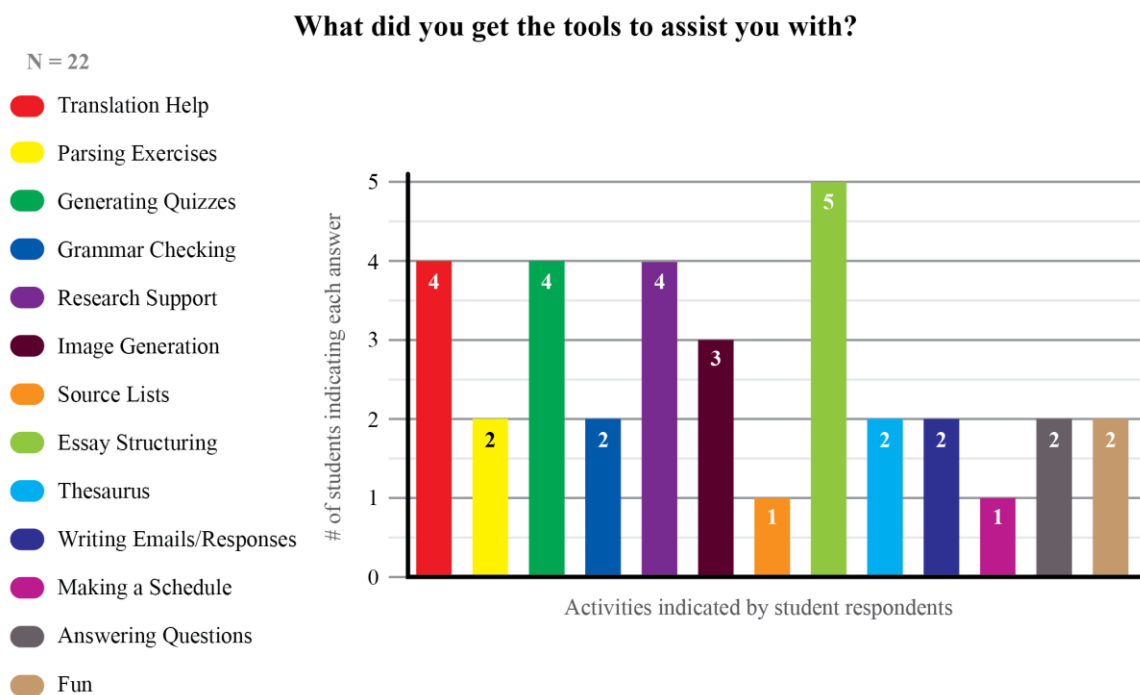


Figure 18. Survey data gathered from ancient language students in the Department of Classics at the University of Reading over Spring 2024 term (Ross and Baines, 2024b).

The participants that used generative AI tools mostly found that they were generally effective for their desired uses, 81.8% ranking it at a 3 or 4, but none found that the tools were excellently effective, a rank of 5 (Figure 19).

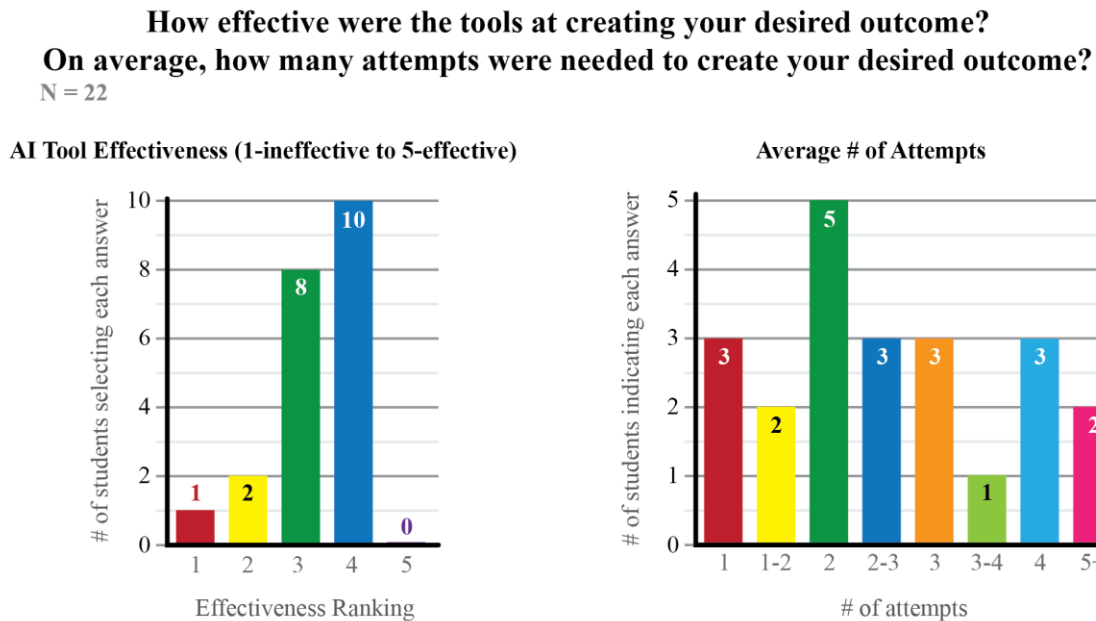


Figure 19. Survey data gathered from ancient language students in the Department of Classics at the University of Reading over Spring 2024 term (Ross and Baines, 2024b).

Furthermore, 86.4% of participants who used generative AI tools made effort to fine tune their results, making 2 or more attempts each time to achieve their desired outcomes. It is clear that the students who used generative AI tools this year in ancient languages had a conception of how generative AI worked, and they generally did not take answer at face value, trying multiple times to generate something that appeared more accurate.

Interestingly, when asked how helpful using the tools were and if they would recommend using the tools to others, participant answers were converse (Figure 20).

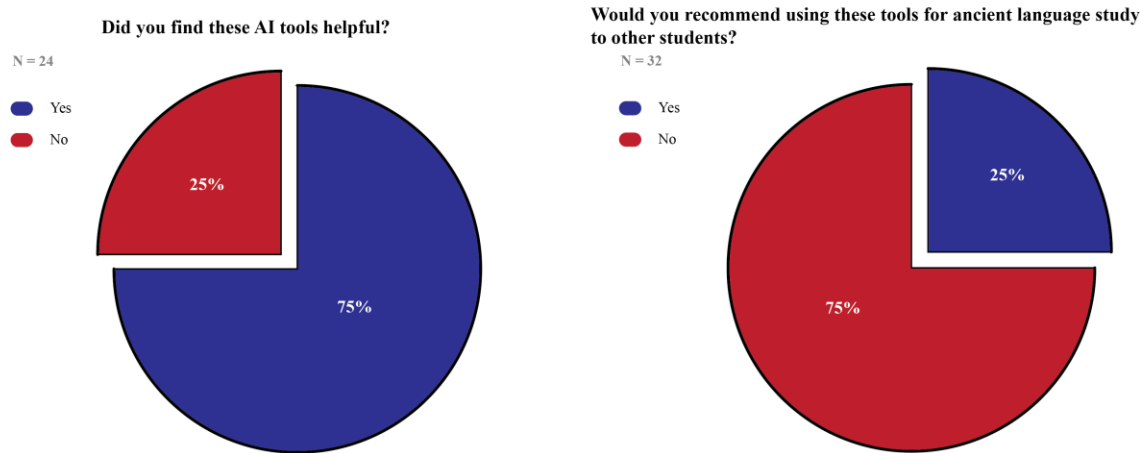


Figure 20. Survey data gathered from ancient language students in the Department of Classics at the University of Reading over Spring 2024 term (Ross and Baines, 2024b).

While 75% of participants who used generative AI tools to support their ancient language studies found them helpful, only 25% of them would recommend them to others for the same purpose. There appears to be a differentiation between what participants believed that they could use to support themselves and what they believe is the general moral view of using the tools. It was good enough for their work, but these tools should not be used in general.

When comparing the use statistics to final module grades, the participants who indicated that they used generative AI primarily fell into the mid-range, third-class honours, second-class honours, and lower first-class honours, while those at the high and low extremes did not use them at all (Figure 21).

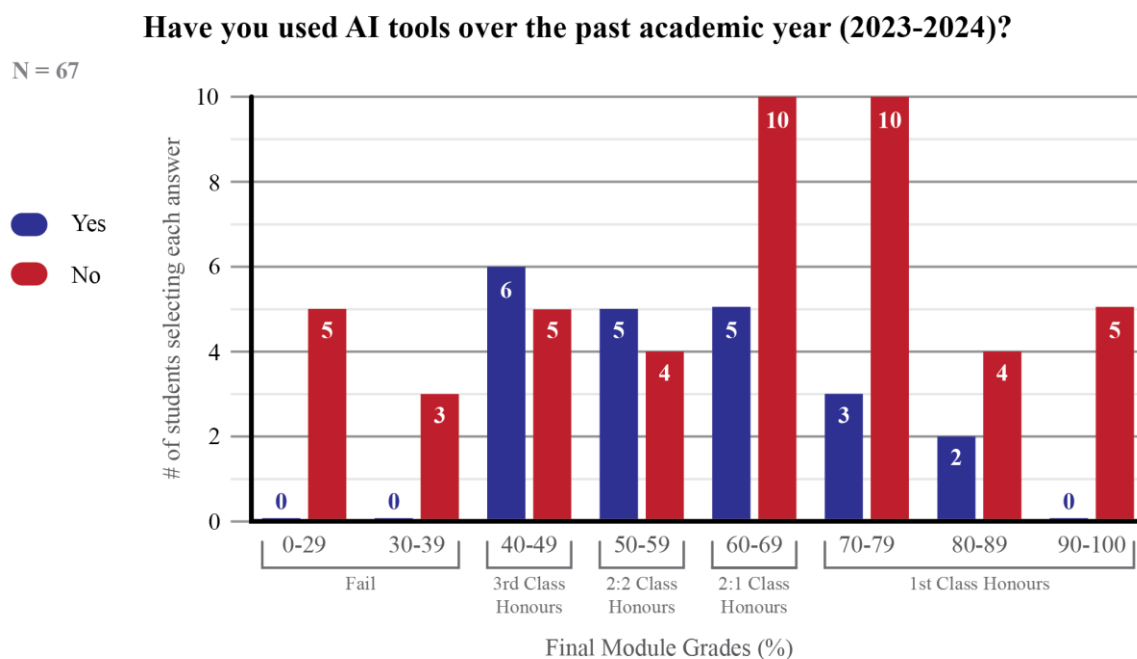


Figure 21. Survey data gathered from ancient language students in the Department of Classics at the University of Reading over the 2023-2024 academic year (Ross and Baines, 2024a).

In many cases, the upper range students felt that generative AI tools were not useful for their

studies because they already had methods in place which worked well. Those that did use generative AI tools to support their ancient language learning found them helpful for overcoming areas which caused them difficulty. Participants at the lower end of the grade spectrum tended not to use it, but it appears that they were not engaging very much with digital tools beyond their textbooks at all.

As a whole we found that student engagement with generative AI was relatively low in the ancient languages, for reasons of environmental impact, copyright infringement, or a general distrust of generative AI. Those who did use generative AI tools were generally critical of its outputs and would fine tune their answers before taking them into consideration. Furthermore, despite feeling that the tools were helpful for their studies, those that used the tools would generally not recommend that others use them. By presenting the ethical considerations for using generative AI and providing guidance for effective methods for using generative AI when needed, it appears that students are now taking a measured, critical approach to using generative AI tools when necessary.

What Now?

This is not the end of this process. Generative AI continues to evolve, and we need to also follow along with it to ensure we are giving the best recommendations to our students about what tools are more ethically sound and effective for supporting their language learning. Although the general conversational AI tools examined in this study did have some applications for supporting ancient language studies, a more focused generative AI tool that is trained on specific textbooks and problem sets would have a greater impact for supporting ancient language students. We already are seeing more specific generative AI tools for ancient languages, like LatinIA's Ciceronianity classifier (Cavaleri, Hersch and Kolde, 2024) and Patrick J. Burns' LatinCY (Burns, 2023). Following these models, we are working to train GPTs which are specifically set for supporting Ancient Greek and Latin levels 1 to 3 at the University of Reading.⁶

Conclusion

Over the course of the 2023-2024 academic year, we worked with our undergraduate students to see how generative AI affected their studies and what tools would have been the most effective for supporting their studies. Claude 2 and ChatGPT 3.5, as of December 2023, were by far the most effective for supporting short-form and long-form translation and grammar work. Other tools, like Bing Chat and Google Bard, were somewhat effective, but the drawback far outweighed any advantages. This collaborative work resulted in an informational booklet that we shared with staff and students in the University of Reading and with international ancient language teachers. By providing students with up-to-date ethics information about generative AI tools and providing them with tailored guides for using them in a manner that directly supports their learning, ancient language students become the users that direct generative AI's outputs for a critical purpose rather than simply relying on generative AI to give them accurate information. It is crucial that teachers at all levels and all learning stages keep up to date with generative AI issues so that they can effectively scaffold their students' learning in the era of widely available, fast developing artificial intelligence.

Acknowledgements

This article is part of the wider 'iGAIAS: Investigating Generative Artificial Intelligence in Ancient World Studies' project in the Department of Classics at the University of Reading. This project has been reviewed by the University of Reading University Research Ethics Committee and has been given a favourable ethical opinion for conduct. The project is supported by a Teaching and Learning Enhancement Project (TLEP) grant from the University of Reading and an Education Grant from the Council of University Classics Departments (CUCD). Special thanks to Marchella

⁶ We have already tested the public-facing personalisation interfaces available through OpenAI's GPTs and Google's Gems with Latin vocabulary datasets, and the results of these tests are available in Ross and Baines 2025.

Ward for bringing the issues of data tagging worker exploitation to our attention.

References

- Anthropic (2023) *Claude-2* (21 November 2023 version) [Large language model]. Available at <https://claude.ai/login?returnTo=%2F>.
- Baines, J., Ross, E. A. S., Hunter, J., McRitchie Pratt, F. and Patel, N. (2024). *Digital Tools for Learning Ancient Greek and Latin and Guiding Phrases for Using Generative AI in Ancient Language Study*. {3 edn.} Available at: <https://doi.org/10.6084/m9.figshare.25391782.v3> (Accessed 04 July 2024).
- Baraniuk, C. (2024). 'Electricity grids creak as AI demands soar', *BBC News*, 21 May. Available at: <https://www.bbc.co.uk/news/articles/cj5ll89dy2mo> (Accessed 04 July 2024).
- Barnett, J. (2024). 'Decoding US Copyright Law and Fair Use for Generative AI Legal Cases', *Medium*, 11 July. Available at: <https://generative-ai-newsroom.com/decoding-us-copyright-law-and-fair-use-for-generative-ai-legal-cases-507fd9956c> (Accessed: 30 July 2024).
- Burns, P. J. (2023). 'LatinCy: Synthetic Trained Pipelines for Latin NLP', *arXiv*. Available at: <https://doi.org/10.48550/arXiv.2305.04365> (Accessed: 13 May 2024).
- Cavaleri, D., Hersch, M. and Kolde, A. (2024). 'Write like Cicero: integrating an AI-based classifier in ancient language teaching for idiom acquisition', *The Journal of Classics Teaching*, 26(51), pp. 1-6. Available at: <https://doi.org/10.1017/S2058631024000734> (Accessed: 27 September 2024).
- Cho, W. (2024). 'OpenAI Training Data to Be Inspected in Authors' Copyright Cases', *The Hollywood Reporter*, 24 September. Available at: <https://www.hollywoodreporter.com/business/business-news/openai-training-data-inspected-authors-copyright-case-1236011291/> (Accessed: 28 September 2024).
- Davis, W. (2023). 'AI companies have all kinds of arguments against paying for copyrighted content', *The Verge*, 04 November. Available at: <https://www.theverge.com/2023/11/4/23946353/generative-ai-copyright-training-data-openai-microsoft-google-meta-stabilityai> (Accessed 07 July 2024).
- de Vries, A. (2023). 'The growing energy footprint of artificial intelligence', *Joule*, 7, pp. 2191–2194. Available at: <https://doi.org/10.1016/j.joule.2023.09.004> (Accessed: 04 July 2024).
- Google (2023) *Bard* (15 November 2023 version) [Large language model]. Available at <https://bard.google.com/?hl=en>.
- Gordon, C. (2024). 'AI Is Accelerating the Loss of Our Scarcest Natural Resource: Water', *Forbes*, 07 March. Available at: <https://www.forbes.com/sites/cindygordon/2024/02/25/ai-is-accelerating-the-loss-of-our-scarcest-natural-resource-water/> (Accessed 07 July 2024).
- Heikkilä, M. (2023). 'Making an image with generative AI uses as much energy as charging your phone', *MIT Technology Review*, 01 December. Available at: <https://www.technologyreview.com/2023/12/01/1084189/making-an-image-with-generative-ai-uses-as-much-energy-as-charging-your-phone/> (Accessed 07 July 2024).
- International Energy Agency. (2024). *Electricity 2024*. Paris: International Energy Agency.
- Luccioni, A. S., Jernite, Y. and Strubell, E. (2024). 'Power Hungry Processing: Watts Driving the Cost of AI Deployment?' in *ACM Conference on Fairness, Accountability, and Transparency, June 3-6, 2024, Rio de Janeiro, Brazil*. *arXiv*. Available at: <https://doi.org/10.48550/arXiv.2311.16863> (Accessed 04 July 2024).
- Microsoft (2023) *Bing Chat* (15 November 2023 version) [Large language model]. Available at <https://www.bing.com/search?q=Bing%20AI&showconv=1&form=MT00IS>.
- Milmo, D. (2024). "'Impossible" to create AI tools like ChatGPT without copyrighted material, OpenAI says', *The Guardian*, 08 January. Available at: <https://www.theguardian.com/technology/2024/jan/08/ai-tools-chatgpt-copyrighted-material-openai> (Accessed 04 July 2024).
- Mollick, E. and Mollick, L. (2023). 'Assigning AI: Seven Approaches for Students with Prompts', *The Warton School Research Paper*. Available at: <https://dx.doi.org/10.2139/ssrn.4475995> (Accessed

07 November 2023).

OpenAI (2023) *ChatGPT 3.5* (06 November 2023 version) [Large language model]. Available at <https://chat.openai.com/auth/login>.

Perrigo, B. (2023). 'Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic', *Time*, 18 January. Available at: <https://time.com/6247678/openai-chatgpt-kenya-workers/> (Accessed: 04 June 2024).

Ross, E. A. S. (2023). 'A New Frontier: AI and Ancient Language Pedagogy', *The Journal of Classics Teaching*, 24(48), pp. 143-161. Available at: <https://doi.org/10.1017/S2058631023000430> (Accessed: 04 July 2024).

Ross, E. A. S. and Baines, J. (2023). *AI and Ancient Language Student Survey – Autumn 2023 {V2}* [Data set]. figshare. Available at: <http://doi.org/10.6084/m9.figshare.24835629.v2> (Accessed: 28 September 2024).

Ross, E. A. S. and Baines, J. (2024a). *AI and Ancient Language Student Survey – 2023-2024 Comparative Data* [Data set]. figshare. Available at: <https://doi.org/10.6084/m9.figshare.27146880.v1> (Accessed: 01 October 2024).

Ross, E. A. S. and Baines, J. (2024b). *AI and Ancient Language Student Survey – Spring 2024* [Data set]. figshare. Available at: <https://doi.org/10.6084/m9.figshare.27146868.v1> (Accessed: 01 October 2024).

Ross, E. A. S. and Baines, J. (2024c). 'Treading water: new data on the impact of AI ethics information sessions in classics and ancient language pedagogy', *The Journal of Classics Teaching*, 25(50), pp. 181–190. Available at: <https://doi.org/10.1017/S2058631024000412> (Accessed: 04 July 2024).

Ross, E. A. S. and Baines, J. (2025). 'Navigating the Fog: The Effectiveness of Personalized Conversational GenAI Models for Supporting Ancient Language Learning', *AI & Antiquity*, 1(1), pp. 35-52. Available at: <https://doi.org/10.64946/aiantiquity.v1i1.002> (Accessed 19 November 2025).

Russell Group (2023) *Russell Group principles on the use of generative AI tools in education*. Available at: https://russellgroup.ac.uk/media/6137/rg_ai_principles-final.pdf.

Schifano, I. (2023). 'These are the Russell Group unis that have banned students from using ChatGPT', *The Tab*, 03 March. Available at: <https://thetab.com/2023/03/03/these-are-the-russell-group-unis-that-have-banned-students-from-using-chatgpt> (Accessed 19 November 2025).

Tan, R. (2023). 'Behind the AI boom, an army of overseas workers in "digital sweatshops"', *The Washington Post*, 28 August. Available at: <https://www.washingtonpost.com/world/2023/08/28/scale-ai-remotasks-philippines-artificial-intelligence/> (Accessed: 04 June 2024).

Taylor, B. L. (2023). 'Long hours and low wages: the human labour powering AI's development', *The Conversation*, 15 November. Available at: <https://theconversation.com/long-hours-and-low-wages-the-human-labour-powering-ais-development-217038> (Accessed: 04 June 2024).

Thomson Reuters Legal Insights Europe. (2024). 'EU AI Act: The latest updates on the world's first comprehensive AI regulation laws', *Legal Insights Europe*, 08 August. Available at: <https://legalsolutions.thomsonreuters.co.uk/blog/2024/08/08/ai-act-the-worlds-first-comprehensive-laws-to-regulate-ai/> (Accessed: 12 August 2024).

University of Edinburgh. (2023). 'AI guidance for staff', *Staff News*, 11 July. Available at: <https://www.ed.ac.uk/news/staff/2023/ai-guidance-for-staff> (Accessed: 19 November 2025).

Wilson, T. S. (2025). 'The development of policies on generative artificial intelligence in UK universities', *International Federation of Library Associations and Institutions (IFLA) Journal*, 51(3), pp. 722-734. Available at: <https://doi.org/10.1177/03400352251333796> (Accessed 19 November 2025).

