

Navigating the Fog: The Effectiveness of Personalised Conversational GenAI Models for Supporting Ancient Language Learning

Edward A. S. Ross

University of Reading
edward.ross@reading.ac.uk
 [0000-0003-4174-835X](https://orcid.org/0000-0003-4174-835X)

Jackie Baines

University of Reading
j.baines@reading.ac.uk
 [0009-0002-6504-0428](https://orcid.org/0009-0002-6504-0428)

ABSTRACT — Hallucinations (misleading, inaccurate predicted text presented as fact) are a critical problem for using generative artificial intelligence (GenAI) tools to support ancient language teaching and learning. For a teacher, significant editing time is required to correct any inaccuracies or misrepresentations prior to making use of AI-generated content to support their teaching practice. For students, these convincing errors may not be recognised, and this may lead to misconceptions in their knowledge formation. OpenAI and Google released public-facing, customizable conversational AI models which allow users to upload their own datasets to create personalised AI chat agents, known as GPTs (2023) and Gems (2024) respectively. This presents an opportunity for teachers to personalize their own models to streamline their students' experiences. However, can personalised conversational AI tools provide a fine-tuned experience that reduces the major, problematic ancient history and ancient language hallucinations that we see in standard ChatGPT and Gemini outputs?

This paper discusses the creation of a personalised Latin Tutor GPT and Gem through the development of a series of exhaustive Latin vocabulary spreadsheets. We tested these personalised tools against their standard GenAI counterpart to determine if personalisation improved their efficacy and efficiency for supporting ancient language learning. The development of the spreadsheets and testing process both closely addressed current GenAI ethical issues, including copyright, environmental impact, and content restrictions. The results of these tests found that personalised GPTs and Gems made small efficacy and efficiency improvements, but the time and energy required greatly outweighed the results.

KEYWORDS — ancient language learning, generative artificial intelligence, Latin, OpenAI, Gemini, AI ethics

1. INTRODUCTION

Since ChatGPT 3.5's public-facing release in November 2022, generative artificial intelligence (GenAI) tools have become an ever-present part of education at all levels (Department of Education, 2025). Over the past 3 years, newer models have exponentially improved the capabilities and functionality of conversational GenAI tools, significantly improving their multilingual capacities (Romanou *et al.*, 2024; Vayani *et al.*, 2025). For ancient languages, however, conversational GenAI tools were initially quite poor and have only improved to an intermediate level (Ross, 2023; Baines and Ross, 2024; Ross *et al.*, forthcoming). We previously tested the Latin and Ancient Greek abilities of the most popular conversational GenAI tools, including ChatGPT, Copilot, Gemini, and Claude, in February 2024 and found that these tools were effective for supporting short and long form translation work, producing vocabulary tests, and creating additional homework sentences (Ross *et al.*, forthcoming). The primary problem, however, was that the tools would not restrict their vocabulary to a requested list, even with pre-prepared guiding *prompts* (Baines *et al.*, 2024).

In November 2023, OpenAI released a new code-free user interface (UI) called GPTs that allowed any ChatGPT Plus subscriber to create a personalised version of ChatGPT for a specific purpose (OpenAI, 2023b). Google released a similar UI for Gemini called Gems in August 2024 (Google, 2024b). In both UIs, a user can create a system *prompt* for a personalised GenAI tool and upload files to act as the personalised model's knowledge base. Although creating a personalised GenAI tool in this way comes with ethical problems, we wanted to see if creating a personalised model would address the issues we previously found with using ChatGPT and Gemini for supporting ancient language study, improving the utility of these tools and reducing output refining time.

In this article, we will discuss how we ethically prepared datasets and *prompts* for use in personalised GenAI tools. Then, we will explore the efficacy of a personalised Latin tutor GPT and Gem in comparison with the latest ChatGPT and Gemini models respectively. Finally, we will make some recommendations about personalisation methods for teachers and students of ancient languages. Before this, it is important to discuss how a GenAI personalisation UI works.

2. PERSONALISED GENAI TOOL USER INTERFACES

The customisation of GenAI tools is not a new concept. This type of work has existed in the computer science sector for decades, but it was largely restricted to those who had some form of coding knowledge. Traditionally, users would gain access to a GenAI tool's application programming interface (API) to interact directly with the software or system and connect it to a 3rd-party application (Ali, 2023). This process would require significant coding work to link different APIs and data together. OpenAI's GPTs UI simply made the API customisation process more

accessible for end users without coding experience (OpenAI, 2023b). ChatGPT 4o acts as the base model for a subscribing user to customise through the UI. Google's Gems UI is quite similar in purpose, and is available to free Gemini users, but it has fewer customisation options than OpenAI's GPTs (Google, 2024b).

The OpenAI GPTs UI has two tabs, "Create" and "Configure" (2023a). The Create tab is a simple chatbot that will configure the custom GPT according to user input. The Configure tab, on the other hand, contains a variety of customisation options, including instructions, knowledge, capabilities, and actions ([Figure 1](#)). The key sections in this tab are instructions, knowledge, and capabilities. The instructions box is where a user uploads their system *prompt*, a comprehensive set of instructions, guidelines, and restrictions for the personalised GPT to follow; the knowledge box is where a user can upload files or data for the GPT to use as its source base; and the capabilities section lets the user adjust any additional functionality of their personalised GPT. The conversation starters and actions sections allow for increased customisation, enabling the user to create *prompt* suggestions for their end users or to attach a 3rd-party API to the GPT for expanded functionality. The name, description, and image boxes are purely cosmetic. As a user is adding details into each section, they can preview interactions with the GPT and fine-tune their instructions, knowledge, and capabilities sections accordingly. Once a user is happy with the fine-tuning, they can publish their GPT for personal use, link-only access, or public use.

Google's Gems UI, as mentioned earlier, has fewer customisation options than OpenAI's GPTs (Google, 2024a). The only options included in the UI are name, instructions, and knowledge ([Figure 2](#)). Much like the GPT builder UI, the name section is purely cosmetic, the instruction section is for inputting the system *prompt*, and the knowledge section is for uploading the data and files to make up the Gem's source base. A user can also use Gemini to generate a system *prompt* for the instruction section, if desired. There are no options to adjust the Gem's web access, API connections, or image functionality, but a user can preview the personalised Gem and fine tune the instructions and knowledge to refine the outputs before publishing. At the time of writing, a user cannot share their personalised Gems with any other users, but this may be made available in the future (Shabanov, 2025).

Both GenAI personalisation UIs have similar purposes, but they have some key differences. OpenAI's GPTs is a subscriber-only function with a broader selection of customisation options, while Google's Gems is free for all users with a streamlined selection of simple customisations. These UIs are quite accessible for end users with no coding experience, such as ancient language teachers, but there are several ethical issues that arise from the few sections available in the GPT and Gem Builders.

3. PERSONALISING ETHICALLY

When preparing the materials for our personalisation test, we wanted to ensure we considered the ethical issues surrounding the use and training of GenAI models. These issues include worker exploitation and data corruption, but the most relevant issues for the preparation of personalised GenAI tools are environmental impact, copyright infringement, and content restriction¹.

The environmental impact of GenAI tools is currently unclear, primarily because many large AI companies like OpenAI and Google will not share their energy usage statistics (O'Donnell and Crownhart, 2025). However, current estimates find that GenAI tools with more parameters will have a significantly higher energy usage and in turn CO2 emissions and freshwater usage than smaller, specific models (Bogmans *et al.*, 2025). Because of this, it is crucial to work with smaller GenAI models that are built for a specific purpose rather than models with billions of parameters that are not necessary for the required tasks. That being said, if a GenAI tool is not useful, you are not required to use a GenAI tool at all. So, we decided to test and see if a smaller, comprehensive dataset would have an impact on GenAI tool efficacy for supporting ancient language learning.

When it comes to copyright, many GenAI tools are in a highly contentious position (Saw and Tan, 2025). AI companies like Google, Stability AI, and Anthropic have kept their GenAI training data private, but copyright holders claim that outputs from these models suggest that their copyrighted materials were used in dataset curation or model training without permission (De Freitas, 2025). To avoid copyright infringement and promote open practice, we prepared our datasets for this personalisation test using only open-access materials.

In introductory Latin modules at the University of Reading, students are required to learn 321 Latin vocabulary terms with all their relevant declensions and conjugations. Because this is the beginners' level, they are not required to learn every single form for each word, but they are expected to recognise the following inflected forms ([Table 1](#)).

To make a comprehensive dataset of all these possible forms using only open-access materials, we gathered all our vocabulary data from Wiktionary (<https://www.wiktionary.org/>), which is maintained with an Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) license (Wikimedia Foundation, 2023). This type of license allows reuse with attribution, and any reused or modified content from this source must also be made available with a CC BY-SA 4.0 license or equivalent as well (Creative Commons, 2013). However, to ensure that these data will not be repurposed for further GenAI training, we took necessary precautions to

¹ For more on GenAI worker exploitation, see Pogrebna 2024; for more on GenAI data corruption, see Randieri 2025.

opt out of using out conversation data and uploaded files to improve OpenAI and Google's models. To further ensure this, we also turned off our conversation history settings in both ChatGPT and Gemini.

The manual tabulation process took 48 hours and resulted in 4 spreadsheets with 21,825 datapoints. Each line included: the lemma (dictionary form), part of speech, definition, declension/conjugation type, and all inflected forms ([Figure 3](#)). The spreadsheets are now hosted in our project figshare repository for future research use (Ross, 2025).

One further issue we must consider is the content restrictions built into GenAI tools. ChatGPT and Gemini, the models which act as the basis for GPTs and Gems respectively, have built in guardrails that restrict them from performing certain actions or discussing certain topics (OpenAI, 2025; Google, 2025). These subjects include violence, sexually explicit material, harassment, and dangerous activities. GenAI tools tend to interact with these subjects in two ways, either not completing an output and disclaiming that the requested activity breaches content policy or generating an output that does not include any content which infringes on its content restrictions (Ross and Baines, 2024). This is particularly problematic for studying ancient languages because some terms and concepts that are integral for understanding an ancient text include these restricted themes. If a student used a GenAI tool to support their study of an ancient text and the GenAI tool omitted elements from the output that were content restricted, that student may develop an erroneous concept of that text. So, we tested each personalised GenAI tool with some explicit vocabulary terms to see how its outputs would be affected by its content policy.

Taking into account these ethical issues, we created an instruction *prompt* that followed Ethan Mollick and Lilach Mollick's features of *prompt* engineering for students: role and goal, step-by-step instructions, pedagogy, constraints, and personalisation (Mollick and Mollick, 2023).

"You are a tutor for an introductory Latin class, and you help students answer questions about Latin grammar, create parsing quizzes, and create extra Latin-to-English translation sentences. You are friendly, supportive, and precise.

Use the provided spreadsheets as your corpus for all possible words and forms that the student is meant to know. If a cell contains "NoForm", there is no existing word with the corresponding elements. Do not use "NoForm" in any parsing quizzes or Latin-to-English translation sentences.

If asked to produce Latin-to-English translation sentences, create 5 short sentences in Latin, using only words found in the provided spreadsheets. If you use a form which does not appear in the provided spreadsheets, make sure to provide an English gloss of the term like this: "adversus - Adverb - 'against, opposite'"

If asked to produce a parsing quiz, 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”.

We used one-shot *prompting* to illustrate the desired outcomes for both the vocabulary testing and Latin-to-English translation sentences functions. The example *prompts* were adapted from guiding phrases from the *Digital Tools for*

Learning Ancient Greek and Latin and Guiding Phrases for Using Generative AI in Ancient Language Study booklet (Baines *et al.*, 2024). These *prompts* were previously tested on ChatGPT 3.5 and Google Bard in Spring 2024 with reasonably effective results, so their reuse in ChatGPT 4o and Google Gemini would theoretically have improved results (Ross *et al.*, forthcoming).

We then uploaded the instruction *prompt* and prepared vocabulary spreadsheets to both the GPTs and Gem UIs for testing. To avoid overusing these models, we restricted our tests to short conversations of 3-4 *prompts* with ChatGPT 4o, Google Gemini Flash 2.0, the personalised GPT, and the personalised Gem for each type of activity. We tested each model's effectiveness in vocabulary quizzes and creating Latin-to-English translation sentences; if the model did not include words that potentially breached its content policy, we also carried out a test to include some explicit Latin words. This limited each tester's use of the models to 3 conversations of 3-4 *prompts* for each of the 4 models. In the next two sections, we will discuss if this personalisation work had any impact on ChatGPT or Gemini's effectiveness for supporting Latin learning.

4. LATIN TUTOR GPT

The personalised Latin Tutor GPT used ChatGPT 4o as a foundation and included the instruction *prompt*, vocabulary spreadsheets, and some conversation starters. Since the GPTs UI also included some additional capabilities customisations, we decided to turn on the "Web Search" and "Code Interpreter & Data Analysis" capabilities for this test to allow the personalised model to interact with the uploaded spreadsheets and search for grammar details. When interacting with the personalised model, we simply inputted short *prompts* requesting Latin tutoring help. Because the general ChatGPT model was not trained with the same *prompts* and spreadsheets, we would provide an equivalent guiding phrase to ChatGPT 4o prior to inputting the same request.

ChatGPT 4o was given an adapted version of guiding phrase 3 from Baines *et al.* 2024 to test its ability to create and mark an introductory Latin vocabulary quiz. The outputs followed the requested format and included a mix of genders, numbers, and cases in the quiz ([Figure 4](#)). The produced words also tended to occur in the restricted vocabulary list despite ChatGPT-4o not having access to the preprepared spreadsheets, but this may be due the words' frequency in introductory Latin educational materials. We cannot guarantee that the general ChatGPT model will always output words that are included in a learner's vocabulary list. Ideally, the personalised model could reduce this risk.

The personalised GPT similarly produced vocabulary quizzes that matched the style indicated in the instruction *prompt*. The vocabulary provided was all varied in gender, number, and case, and the model was able to identify and correct errors in the responding input ([Figure 5](#)). The key problem with the personalised outputs

was that one of the produced words, *oculis*, did not appear in the provided vocabulary spreadsheet. When confronted about the extra word, the model claimed that the word was actually in the vocabulary list all along. This could cause a significant issue for learners who are unsure about their vocabulary knowledge when using this tool, confusing them and misinforming them of what words they will need to know for their studies.

We inputted an adapted version of guiding phrase 4 from Baines *et al.* 2024 when testing ChatGPT 4o's Latin-to-English translation sentence efficacy. Broadly, the general ChatGPT model would produce short Latin sentences that would be straightforward for introductory Latin students. The model would also identify errors and provide feedback. The vocabulary used also generally fell within the restricted vocabulary list, as with the vocabulary quizzes, but some outside words did appear ([Figure 6](#)). In this case, the word *magistro* is used. When ChatGPT 4o is notified of the external word, unlike the personalised GPT, it apologises, says it will not use the word in future examples, reframes it as a bonus question. This conversation is effective, but it does require several inputs to get to the point where translation sentences are provided. The personalised model could ideally reduce the number of initial *prompts* required.

The personalised GPT was just as effective at producing Latin-to-English translation sentences, identifying errors, and providing corrections as ChatGPT-4o. However, the personalised GPT's sentences tended to be simpler. Additional *prompts* were also frequently required to get comparable untranslated outputs ([Figure 7](#)). Although the personalised model removed the need for a user to input a guiding phrase, our tests tended to use the same number of *prompts* to get the desired outputs as the general ChatGPT model.

ChatGPT 4o and the personalised GPT were both quite effective at producing vocabulary quizzes and providing Latin-to-English translation sentences. Both the general and personalised model were consistently able to identify user input errors and provide constructive feedback. Unfortunately, despite providing a complete vocabulary dataset for the personalised model to restrict its vocabulary, both ChatGPT 4o and the personalised GPT produced Latin words that were not included in the restricted list. Although these hallucinations were less frequent in the personalised model, the personalised model would claim that the external word was in fact part of the vocabulary list. This is a major problem for uninformed learners as it may lead to confusion related to their vocabulary-learning requirements. Furthermore, the personalised GPT sometimes needed just as much *prompting* as the general ChatGPT-4o model to get the desired output. In this way, the personalisation of the ChatGPT model had a very small impact on the efficacy and efficiency of using OpenAI's models for supporting ancient language learning.

5. LATIN TUTOR GEM

The personalised Latin Tutor Gem used Google Gemini 2.0 Flash as the foundation. It also included the instruction *prompt* and vocabulary spreadsheets, but there was no further customisation. The Gemini and Gem testing followed a similar process to the GPT testing with guiding phrases inputted into the general Gemini model prior to testing, while the request was inputted into the personalised Gem without context.

Gemini 2.0 Flash was given the same adapted version of guiding phrase 3 from Baines *et al.* 2024 for the introductory Latin test as ChatGPT 4o. Much like the OpenAI models, Gemini outputs a mix of genders, numbers, and cases, and the model identified and corrected errors in the follow-up input ([Figure 8](#)). The produced vocabulary does fall within the GCSE vocabulary list restriction, as indicated in the initial output responding to the guiding phrase. The corrections are less hands-on than the OpenAI models, instead providing the correct answers and some overarching comments rather than line by line corrections. Overall, the general Gemini model works well outputting vocabulary quizzes, so this potentially bodes well for the personalised Gem.

The personalised Gem works quite similarly, outputting a mix of vocabulary forms, accurately identifies errors, and provides constructive feedback ([Figure 9](#)). Once again, the vocabulary did fall within the restricted list, and the personalised Gem produced a similar process of providing the answers and then giving feedback at the end. This type of feedback is effective, but so was the line-by-line style of the OpenAI models. Users could choose between the two styles depending on their preferences. However, there appear to be some issues with Gemini's Latin-to-English translation sentences.

We used an adapted version of guiding phrase 4 from Baines *et al.* 2024 to test Gemini 2.0 Flash's Latin-to-English translation sentence effectiveness. The general model produced a very interactive tutor experience with consistent questioning to allow the user to find corrected answers on their own ([Figure 10](#)). The sentences were also simple and matched the expected level for an introductory student, but some external vocabulary was used. In this case, the word *oppidum* is produced. When the model is corrected, the output recognises that the word might not be on the user's vocabulary list but claims that the word is on the GCSE vocabulary list. This is not the case, at the time of writing, but this may be corrected with the personalised Gem.

The personalised Gem produced Latin translation sentences consistently, reducing the number of required inputs, and the sentences were generally at the introductory level ([Figure 11](#)). In this output, we see some external vocabulary, *litterae*, and the personalised Gem, like the personalised GPT, claims that it was indeed in the restricted list, which is problematic. Furthermore, the personalised

Gem provided some sentences which were unnecessarily sexist or heteronormative, including *cibus a femina paratur* (The food is prepared by the woman) and *puellae a pueris amantur* (The girls are loved by the boys). Although these sentences are possible, we need to be aware that these biases are appearing in the generated texts.

Overall, Gemini 2.0 Flash and the personalised Gem were generally effective at outputting vocabulary quizzes and Latin-to-English translation sentences. There are some external vocabulary hallucination issues and sexist bias present in the Google models, but the difference in efficacy between the general and personalised models were almost imperceptible. The personalised model was more efficient with producing the desired output. Unfortunately, personalised Gems are not currently shareable, so this efficiency improvement could only be provided to learners if they personalize the Gem themselves, which is not practical.

6. RECOMMENDATIONS FOR THE FUTURE

Following testing, the OpenAI GPT showed some minor improvements related to the vocabulary limitations and overall *prompts* needed for a desired outcome, but this was not consistent and sometimes worse than the general ChatGPT 4o model. The Google Gem personalisation showed little to no change in the efficacy of the model's ancient language support, but there was some improvement in the number of inputs required to obtain a desired output. Unfortunately, we found that these minor improvements are not enough to warrant the extensive time and energy required to develop the comprehensive vocabulary sheets that were used to personalize this model.

Rather than putting the effort towards creating unique datasets to personalize large models through the GPTs or Gems UIs, time would be better spent creating a bespoke, smaller scale model that is specifically for supporting ancient language learning. This model would ideally reduce the number of inputs required to obtain a desired output and have a confined vocabulary corpus, improving efficiency and applicability. Although hallucinations are a persistent risk across all models, these smaller models for specific uses would theoretically reduce the frequency of hallucinations in its primary subject area.

If we aim to use GenAI models for supporting ancient language studies, it is crucial that we provide learners scaffolding surrounding the potential of vocabulary hallucination and methods to properly infer the meanings of these words or the best locations to find their meaning outside the AI model. Ideally through grammar books, dictionaries, or textbooks related to their module. In this way, we train learners to work with AI but also provide them the tools to engage critically with ancient language content in general.

Overall, AI model personalisation could be an impactful way forward for adapting ancient language teaching and learning in the developing technological

world, but the current models have yet to become effective enough to warrant the work required to create them. It is crucial that we continue to educate teachers and learners about the ethics and risks of GenAI use and hallucinations. With this knowledge, people working with ancient languages can use their time more effectively to support ancient language learning and potentially work with future smaller-scale models that are fine-tuned for ancient language learning needs.

BIBLIOGRAPHY

- Ali, M. (2023) 'A beginner's guide to using the ChatGPT API', *DataCamp*, 14 December. Available at: <https://www.datacamp.com/tutorial/a-beginners-guide-to-chatgpt-api> (Accessed: 4 May 2025).
- 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*. 3rd edn. doi: 10.6084/m9.figshare.25391782.v3 (Accessed: 4 May 2025).
- Bogmans, C., Gomez-Gonzalez, P., Melina, G. and Thube, S. (2025) 'AI needs more abundant power supplies to keep driving economic growth', *IMF Blog*, 13 May. Available at: <https://www.imf.org/en/Blogs/Articles/2025/05/13/ai-needs-more-abundant-power-supplies-to-keep-driving-economic-growth> (Accessed: 20 May 2025).
- Creative Commons. (2013) *CC BY-SA 4.0: Attribution-ShareAlike 4.0 International – Deed*. Available at: <https://creativecommons.org/licenses/by-sa/4.0> (Accessed: 20 April 2025).
- De Freitas, I. (2025) 'GenAI and copyright: three key cases', *Farrer & Co*, 9 July. Available at: <https://www.farrer.co.uk/news-and-insights/genai-and-copyright-three-key-cases> (Accessed: 10 July 2025).
- Department of Education. (2025) *Generative artificial intelligence (AI) in education*. Available at: <https://www.gov.uk/government/publications/generative-artificial-intelligence-in-education/generative-artificial-intelligence-ai-in-education> (Accessed: 22 April 2025).
- Google. (2024a) *Gems* (13 March 2025 version) [computer software]. Available at: <https://gemini.google.com/gems/create>
- Google. (2024b) 'New in Gemini: Custom Gems and improved image generation with Imagen 3', *The Keyword*, 28 August. Available at: <https://blog.google/products/gemini/google-gemini-update-august-2024> (Accessed: 4 May 2025).
- Google. (2025) 'Policy guidelines for the Gemini app', *Gemini*. Available at: <https://gemini.google/policy-guidelines> (Accessed: 19 April 2025).
- Mollick, E. and Mollick, L. (2023) 'Assigning AI: Seven approaches for students with prompts', *The Wharton School Research Paper*. doi: 10.2139/ssrn.4475995 (Accessed: 10 March 2025).
- O'Donnell, J. and Crownhart, C. (2025) 'Everything you need to know about estimating AI's energy and emissions burden', *MIT Technology Review*, 20 May. Available at: <https://www.technologyreview.com/2025/05/20/1116331/ai-energy-demand-methodology> (Accessed: 20 May 2025).
- OpenAI. (2023a) *GPTs* (21 February 2025 version) [computer software]. Available at: <https://chatgpt.com/gpts/editor>
- OpenAI. (2023b) 'Introducing GPTs', *OpenAI*, 6 November. Available at: <https://openai.com/index/introducing-gpts> (Accessed: 4 May 2025).
- OpenAI. (2025) 'Usage policies', *OpenAI*, 29 January. Available at: <https://openai.com/policies/usage-policies> (Accessed: 19 April 2025).
- Pogrebna, O. (2024) 'AI is a multi-billion dollar industry. It's underpinned by an invisible and exploited workforce', *The Conversation*, 8 October. Available at:

- <https://theconversation.com/ai-is-a-multi-billion-dollar-industry-its-underpinned-by-an-invisible-and-exploited-workforce-240568> (Accessed: 20 April 2025).
- Randieri, C. (2025) 'Bias and corruption in artificial intelligence: A threat to fairness', *Forbes*, 14 March. Available at: <https://www.forbes.com/councils/forbestechcouncil/2025/03/14/bias-and-corruption-in-artificial-intelligence-a-threat-to-fairness> (Accessed: 14 March 2025).
- Romanou, A., Foroutan, N., Sotnikova, A., Chen, Z., Nelaturu, S. H., Singh, S., Maheshwary, R., Altomare, M., Haggag, M. A., Snegha, A., Amayuelas, A., Amirudin, A. H., Aryabumi, V., Boiko, D., Chang, M., Chim, J., Cohen, G., Dalmia, A. K., Diress, A., Duwal, S., Dzenhaliou, D., Florez, D. F. E., Farestam, F., Imperial, J. M., Islam, S. B., Isotalo, P., Jabbarishiviari, M., Karlsson, B. F., Khalilov, E., Klamm, C., Koto, F., Krzemiński, D., de Melo, G. A., Montariol, S., Nan, Y., Niklaus, J., Novikova, J., Ceron, J. S. O., Paul, D., Ploeger, E., Purbey, J., Rajwal, S., Ravi, S. S., Rydell, S., Santhosh, R., Sharma, D., Skenduli, M. P., Moakhar, A. S., Moakhar, B. S., Tamir, R., Tarun, A. K., Wasi, A. T., Weerasinghe, T. O., Yilmaz, S., Zhang, M., Schlag, I., Fadaee, M., Hooker, S. and Bosselut, A. (2024) 'INCLUDE: Evaluating multilingual language understanding with regional knowledge', *arXiv preprint*. doi: 10.48550/arXiv.2411.19799.
- Ross, E. A. S. (2023) 'A new frontier: AI and ancient language pedagogy', *The Journal of Classics Teaching*, 24(48), pp. 143–161. doi: 10.1017/S2058631023000430.
- Ross, E. A. S. (2025) 'Introductory Latin personalized GenAI tool dataset'. Version 2. doi: 10.6084/m9.figshare.29261450.v2 (Accessed: 6 June 2025).
- Ross, E. A. S. and Baines, J. (2024) '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. doi: 10.1017/S2058631024000412.
- Ross, E. A. S., Baines, J., Hunter, J., McRitchie Pratt, F., Patel, N., Richards-Fowkes, E. and Tandy, H. (forthcoming) 'Taking the helm: Collaborating with students to navigate generative AI for ancient language learning', to be published in *Digital Culture & Education*.
- Saw, C. L. and Tan, B. Z. Y. (2025) 'Unpacking copyright infringement issues in the GenAI development lifecycle and a peek into the future', *Computer Law & Security Review*, 58, pp. 1–17. doi: 10.1016/j.clsr.2025.106163.
- Shabanov, A. (2025) 'Gemini users may soon share Gems as Google begins testing', *TestingCatalog*, 8 May. Available at: <https://www.testingcatalog.com/gemini-users-may-soon-share-gems-as-google-begins-testing> (Accessed: 15 May 2025).
- Vayani, A., Dissanayake, D., Watawana, H., Ahsan, N., Sasikumar, N., Thawakar, O., Ademtew, H. B., Hmaiti, Y., Kumar, A., Kukreja, K., Maslych, M., Al Ghallabi, W., Mihaylov, M. M., Qin, C., Shaker, A. M., Zhang, M., Ihsani, M. K., Esplana, A. G., Gokani, M., Mirkin, S., Singh, H., Srivastava, A., Hamerlik, E., Izzati, F. A., Maani, F. A., Cavada, S., Chim, J., Gupta, R., Manjunath, S., Zhumakhanova, K., Rabevohitra, F. H., Amirudin, A. H., Ridzuan, M., Kareem, D. N. A., More, K. P., Li, K., Shakya, P., Saad, M., Ghasemaghahi, A., Djanibekov, A., Azizov, D., Jankovic, B., Bhatia, N., Cabrera, A., Obando-Ceron, J., Otieno, O., Farestam, F., Rabbani, M., Ballah, S., Sanjeev, S., Shtanchaev, A., Fatima, M., Nguyen, T., Kareem, A., Aremu, T., Xavier, N. A. Z., Bhatkal, A., Toyin, H. O., Chadha, A., Cholakal, H., Anwer, R. M., Felsberg, M., Laaksonen, J., Solorio, T., Choudhury, M., Laptev, I., Shah, M., Khan, S. and Khan, F. S. (2025) 'All languages matter: Evaluating LMMs on culturally diverse 100 languages', *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pp. 19565–19575. doi: 10.48550/arXiv.2411.16508.
- Wikimedia Foundation. (2023) *Wikimedia Foundation terms of use*. Available at: https://foundation.wikimedia.org/wiki/Policy:Terms_of_Use (Accessed: 20 April 2025).

TABLES

Nouns	Adjectives	Verbs	Participles
1st Declension	2-1-2 Forms	Present Indicative Active	Present Active Participle
2nd Declension	3-3 Forms	Imperfect Indicative Active	Perfect Passive Participle
3rd Declension	Comparatives	Future Indicative Active	
1st Person Pronouns	Superlatives	Perfect Indicative Active	
2nd Person Pronouns	Possessives	Pluperfect Indicative Active	
Interrogative Pronouns		Present Infinitive Active	
Indefinite Pronouns		Present Imperative Active	
Reflexive Pronouns		Present Indicative Passive	
		Imperfect Indicative Passive	
		Future Indicative Passive	

Table 1. Declension and conjugation forms included in the Latin 1 dataset. Blank cells indicate that no forms are required for the introductory module.

FIGURES

The image displays the OpenAI GPTs Builder interface, split into two panels: 'Create' (left) and 'Configure' (right).

Create Panel (Left):

- Header: New GPT (Draft)
- Buttons: Create, Configure
- Text: "Hi! I'll help you build a new GPT. You can say something like, 'make a creative who helps generate visuals for new products' or 'make a software engineer who helps format my code.'"
- Text: "What would you like to make?"
- Input field: "Ask anything"

Configure Panel (Right):

- Header: Introductory Latin Tutor (Use - B Only no)
- Buttons: Create, Configure
- Name: Introductory Latin Tutor
- Description: Makes parsing quizzes, Latin-to-English translation sentences, and answers grammar questions
- Instructions:
 - You are a tutor for an introductory Latin class, and you help students answer questions about Latin grammar, create parsing quizzes, and create extra Latin-to-English translation sentences. You are friendly, supportive, and precise.
 - Use the provided spreadsheets as your corpus for all possible words and forms that the student is meant to know. If a cell contains "NoForm", there is no existing word with the corresponding elements. Do not use "NoForm" in any parsing quizzes or Latin-to-English translation sentences.
- Conversation starters:
 - Make a parsing quiz with 2nd declension nouns.
 - Make 5 Latin-to-English sentences using the passive voice.
 - What is the imperative mood?
- Knowledge:
 - Uploads: Latin 1 Indeclinables.csv, Latin 1 Adjectives.csv, Latin 1 Nouns.csv, Latin 1 Verbs.csv
- Capabilities:
 - Web Search (checked)
 - Image Generation (checked)
 - Code Interpreter & Data Analysis (checked)
- Actions:
 - Create new action
- Additional Settings:
 - Use conversation data in your GPT to improve our models (unchecked)

Figure 1. OpenAI GPTs Builder UI: Create (left) and Configure (right).

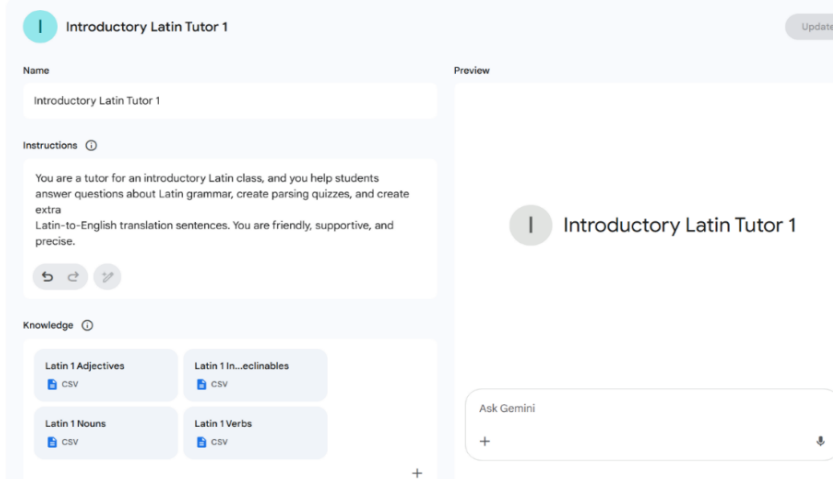


Figure 2. Google Gems Builder UI.

	A	B	C	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS
				Voice=Active, Mood=Indicative, Tense=Pluperfect, Person=1st, Number=Plural	Voice=Active, Mood=Indicative, Tense=Pluperfect, Person=2nd, Number=Plural	Voice=Active, Mood=Indicative, Tense=Pluperfect, Person=3rd, Number=Plural	Voice=Active, Mood=Imperative, Tense=Present, Person=2nd, Number=Plural	Voice=Active, Mood=Imperative, Tense=Present, Person=2nd, Number=Plural	Voice=Active, Mood=Imperative, Tense=Present, Person=2nd, Number=Plural	Voice=Passive, Mood=Indicative, Tense=Present, Person=1st, Number=Plural	Voice=Passive, Mood=Indicative, Tense=Present, Person=2nd, Number=Plural	Voice=Passive, Mood=Indicative, Tense=Present, Person=3rd, Number=Plural	Voice=Passive, Mood=Indicative, Tense=Present, Person=1st, Number=Plural
1	Part of Speech	Lemma	Conjugation										
17	Verb	conspicio	3rd	conspexerāmus	conspexerātis	conspexerant	conspice	conspicite	conspicere	conspicior	conspicimur	conspicitis	conspicimur
18	Verb	constituo	3rd	constituerāmus	constituerātis	constituerant	constitue	constituite	constituere	constitutor	constituimur	constituitis	constituimur
19	Verb	consumo	3rd	consumperāmus	consumperātis	consumperant	consume	consumite	consumere	consumor	consumimur	consumitis	consumimur
20	Verb	convenio	4th	convenerāmus	convenerātis	convenerant	conveni	convenite	convenire	convenior	convenimur	convenitis	convenimur
21	Verb	credo	3rd	crediderāmus	crediderātis	crediderant	crede	credite	credere	creditor	credimur	creditis	credimur
22	Verb	cupio	3rd	cupiverāmus	cupiverātis	cupiverant	cupe	cupite	cupere	cupior	cupimur	cupitis	cupimur
23	Verb	curro	3rd	curremāmus	currentis	currebant	currite	currite	currere	curror	currimur	curritis	currimur
24	Verb	custodio	4th	custodiverāmus	custodiverātis	custodiverant	custodi	custodite	custodire	custodior	custodimur	custoditis	custodimur
25	Verb	debeo	2nd	debuerāmus	debuerātis	debuerant	debe	debete	debere	debeor	debemur	debetis	debemur
26	Verb	defendo	3rd	defenderāmus	defenderātis	defenderant	defende	defendite	defendere	defendor	defendimur	defenditis	defendimur
27	Verb	deleo	2nd	deleverāmus	deleverātis	deleverant	dele	deleite	delere	deleor	delemur	deletis	delemur
28	Verb	descendo	3rd	descenderāmus	descenderātis	descenderant	descende	descendite	descendere	descendor	descendimur	descenditis	descendimur
29	Verb	dico	3rd	dixerāmus	dixerātis	dixerant	dice	dicite	dicere	dicor	dicimur	dicitis	dicimur
30	Verb	disco	3rd	disceverāmus	disceverātis	disceverant	disce	discite	discere	disceor	disceimur	disceitis	disceimur
31	Verb	do	1st	dederāmus	dederātis	dederant	dā	date	dare	dor	datur	datis	datur
32	Verb	dormio	4th	dormiverāmus	dormiverātis	dormiverant	dormi	dormite	dormire	Noform	Noform	dormitis	Noform
33	Verb	duco	3rd	duxerāmus	duxerātis	duxerant	duc	ducite	ducere	ducor	ducimur	ducitis	ducimur
34	Verb	effugio	3rd	effugerāmus	effugerātis	effugerant	effuge	effugite	effugere	effugor	effugimur	effugitis	effugimur
35	Verb	emo	3rd	emerāmus	emerātis	emerant	eme	emite	emere	emor	emimur	emitis	emimur
36	Verb	eo	Irregular	ierāmus	ierātis	ierant	i	ite	ire	eor	imur	itis	imur
37	Verb	expecto	1st	expectaverāmus	expectaverātis	expectaverant	expectā	expectate	expectare	expector	expectamur	expectatis	expectamur
38	Verb	facio	3rd	fecerāmus	fecerātis	fecerant	face	facite	facere	facio	facimur	facitis	facimur
39	Verb	festino	1st	festinaverāmus	festinaverātis	festinaverant	festinā	festinate	festinare	festinor	festinamur	festinatis	festinamur
40	Verb	fugio	3rd	fugerāmus	fugerātis	fugerant	fuge	fugite	fugere	fugior	fugimur	fugitis	fugimur
41	Verb	gero	3rd	gesserāmus	gesserātis	gesserant	gere	gerite	gerere	geror	gerimur	geritis	gerimur
42	Verb	habeo	2nd	habuerāmus	habuerātis	habuerant	habē	habete	habere	habeor	habemur	habetis	habemur
43	Verb	habito	1st	habitaverāmus	habitaverātis	habitaverant	habitā	habitate	habitare	habitor	habitamur	habitis	habitamur
44	Verb	lacio	3rd	lecerāmus	lecerātis	lecerant	lace	lacte	lacere	lacio	lacimur	lacitis	lacimur
45	Verb	incendo	3rd	incenderāmus	incenderātis	incenderant	incende	incendite	incendere	incendor	incendimur	incenditis	incendimur
46	Verb	intelligo	3rd	intelleverāmus	intelleverātis	intelleverant	intellige	intelligite	intelligere	intelligor	intelligimur	intelligitis	intelligimur
47	Verb	interficio	3rd	interfecerāmus	interfecerātis	interfecerant	interfice	interfice	interficere	interficio	interficimur	interficitis	interficimur
48	Verb	intro	1st	intraverāmus	intraverātis	intraverant	intrā	intrate	intrare	intror	intramur	intratis	intramur

Figure 3. Part of the Latin 1 Verb spreadsheet prepared for the personalised GenAI tool tests.

This chat will be a parsing test for a student of elementary Latin. Please restrict your vocabulary to the words included in the Latin GCSE vocabulary list. 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 2nd 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. moris
2. Rinore
3. vices
4. serum
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 answer! Here are the correct answers:

1. moris - Genitive, Neuter, Singular, 'moral', 'of the moral',
2. Rinore - Ablative, Neuter, Singular, 'from the journey',
3. vices - Nominative or Accusative, Feminine, Plural, 'the vices',
4. serum - Genitive, Neuter, Plural, 'series', 'of the old men',
5. urbs - Nominative, Feminine, Singular, '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: 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: Indefinite, Regular, Comparative, or Superlative, "Translation of the form."
- Prepositions: Indefinite, Cases it usually is associated with, "Translation of the form."

Please write according to these conditions for the remainder of this conversation.

Understood! I'm ready to proceed with this Latin parsing test format using only the Latin GCSE vocabulary list. Please let me know which type of words you'd like to test—nouns, verbs, adjectives, pronouns, etc.—and any specific features (e.g. 3rd declension nouns, perfect tense verbs, comparative adverbs), and I'll provide 5 inflected examples for the student to parse.

Would you like to begin with a test now?

Make a parsing quiz with 2nd declension nouns.

Great! Here is your parsing quiz with 5 Latin nouns from the 2nd declension. For each word, please parse it fully by giving:

- Case, Gender, Number
- The Root Form (dictionary entry)
- An English translation of the inflected form

Parsing Quiz: 2nd Declension Nouns

1. servorum
2. gladio
3. amicum
4. pueri
5. bello

Take your time, and when you're ready, send me your answers! I'll then check them and give you feedback.

Thanks for your answer! Let's go through each one and check how you did:

1. servorum - ☒ Correct!
 - Genitive, Masculine, Plural
 - Root: servus
 - Translation: "of the slaves"
2. gladio - ☒ Not quite
 - Correct parsing: Dative or Ablative, Masculine, Singular
 - Root: gladius
 - Translation: "for the sword" (dative) or "by/with the sword" (ablative)
 - You labelled it as genitive, but the genitive singular would be gladii
3. amicum - ☒ Almost there, but there's a small mistake
 - Accusative, Masculine, Singular
 - Root: amicus
 - Translation: "the friend" (accusative)
 - You called it neuter, but amicus is masculine.
4. pueri - ☒ Correct (with an important note)
 - Could be Nominative Plural or Genitive Singular, Masculine
 - Root: puer
 - Translation: "the boy" (nominative) or "of the boy" (genitive)
 - You chose the nominative plural meaning, which is valid.
5. bello - ☒ Correct!
 - Ablative (or Dative), Neuter, Singular
 - Root: bellum
 - Translation: "by/with the war" (ablative), or "for the war" (dative)
 - You gave the ablative translation, which works perfectly.

Final Comments:

Wow! Well done! You correctly parsed 3 out of 5, and your translations show a solid understanding of case usage. Be careful with masculine vs. neuter genders and watch for similar-looking genitive and dative/ablative forms.

Would you like to try another parsing test—perhaps with verbs, adjectives, or pronouns?

Figure 4. OpenAI, ChatGPT 4o (24 June 2025 version), personal communication, generated 14 July 2025. Prompt: 'Adapted Guiding Phrase 3 (Baines *et al.*, 2024 p. 14)'.







-

Figure 6. OpenAI, ChatGPT 4o (24 June 2025 version), personal communication, generated 14 July 2025. *Prompt:* ‘Adapted Guiding Phrase 4 (Baines *et al.*, 2024 p 15)’.

Make 5 Latin-to-English sentences using the passive voice.

Here are 5 Latin sentences using the passive voice, along with their English translations:

1. *Puella laudatur.*
The girl is praised.
2. *Pueri vocantur.*
The boys are called.
3. *Donum mittitur.*
The gift is sent.
4. *Milites terrentur.*
The soldiers are frightened.
5. *Hostes vincuntur.*
The enemies are conquered.

Let me know if you'd like more examples, or if you want to practice parsing any of these!

👍 🗨️ 🔄 📄 📝

Make five more, but don't give me an English translation.

Sure! Here are five more Latin sentences using the passive voice. This time, I'll leave the translation for you to work out:

1. *Servus capitur.*
2. *Fabula narratur.*
3. *Urbs oppugnatur.*
4. *Epistula legitur.*
5. *Libri scribuntur.*

Let me know when you're ready to check your translations or if you want to try making some of your own!

👍 🗨️ 🔄 📄 📝

1. The slave is captured.
2. The story will be told.
3. The city was captured.
4. The letter is read.
5. The book is written.

Thanks for your translations! Let's check each one together:

1. *Servus capitur.* – *The slave is captured.* ✅ Correct!
2. *Fabula narratur.* – *The story is told.* ⚠️ Almost! You used the future tense ("will be told"), but *narratur* is present passive, so it means "is told."
3. *Urbs oppugnatur.* – *The city is being attacked.* ⚠️ "Was captured" is a past tense and uses a different verb (like *capta est*). *Oppugnatur* is present passive and comes from *oppugnare* (to attack), so the correct sense is "is attacked."
4. *Epistula legitur.* – *The letter is read.* ✅ Correct!
5. *Libri scribuntur.* – *The books are written.* ⚠️ You translated "book" singular, but *libri* is plural. So the accurate translation is "The books are written."

Great effort overall! You're showing good awareness of passive voice—just watch out for tense and number. Want to try another round or maybe a parsing quiz?

👍 🗨️ 🔄 📄 📝

Figure 7. OpenAI, GPTs (29 April 2025 version), Introductory Latin Tutor, personal communication, generated on 06 May 2025. *Prompt: 'Make 5 Latin-to-English sentences using the passive voice'.*

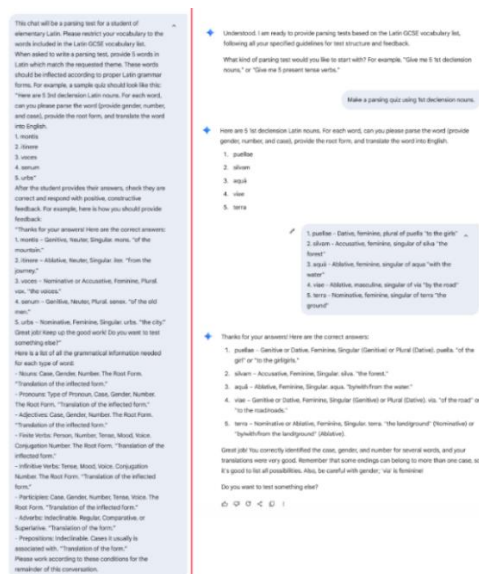


Figure 8. Google, Gemini 2.0 Flash (19 June 2025 version), personal communication, generated on 14 July 2025. *Prompt: 'Adapted Guiding Phrase 3 (Baines et al., 2024 p. 14)'.*

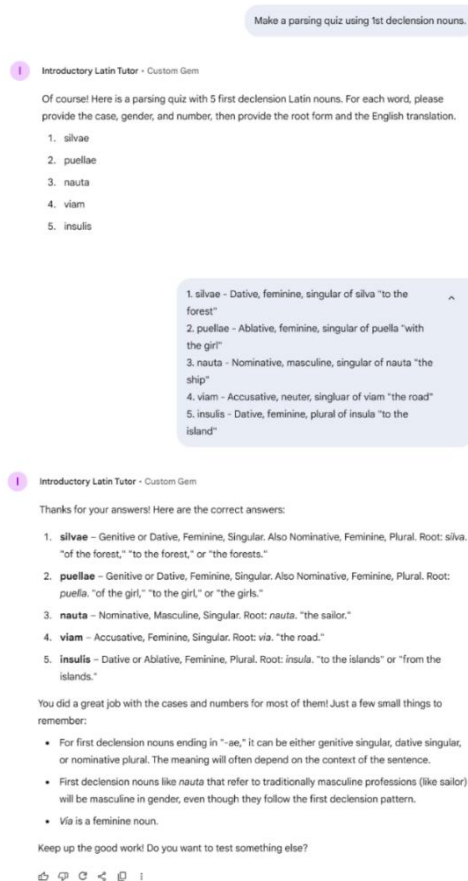


Figure 9. Google, Gemini 2.0 Flash Gems (22 April 2025 version), Introductory Latin Tutor, personal communication, generated 06 May 2025. *Prompt:* 'Make a parsing quiz using 1st declension nouns'.

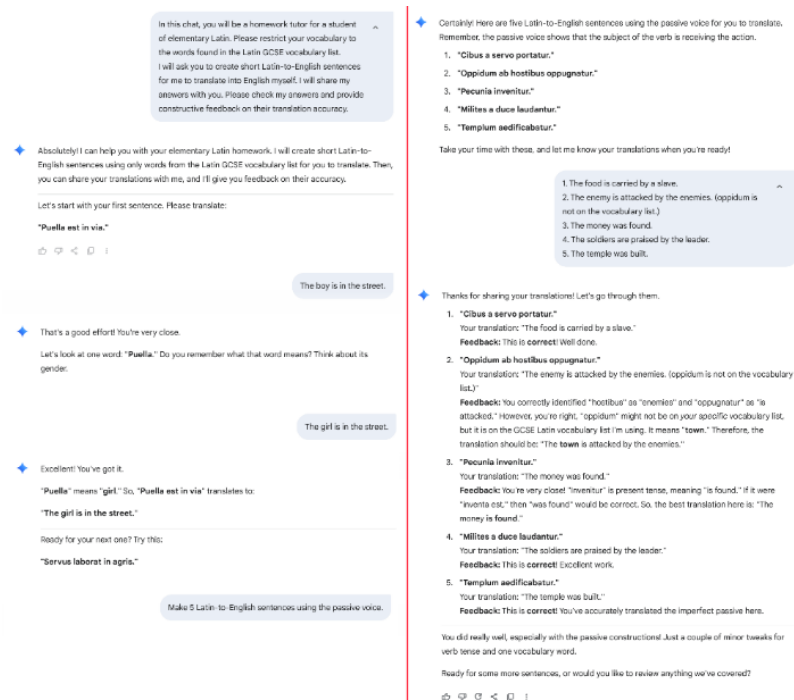


Figure 10. Google, Gemini 2.0 Flash (19 June 2025 version), personal communication, generated on 14 July 2025. *Prompt:* 'Adapted Guiding Phrase 4 (Baines *et al.*, 2024 p. 15).'

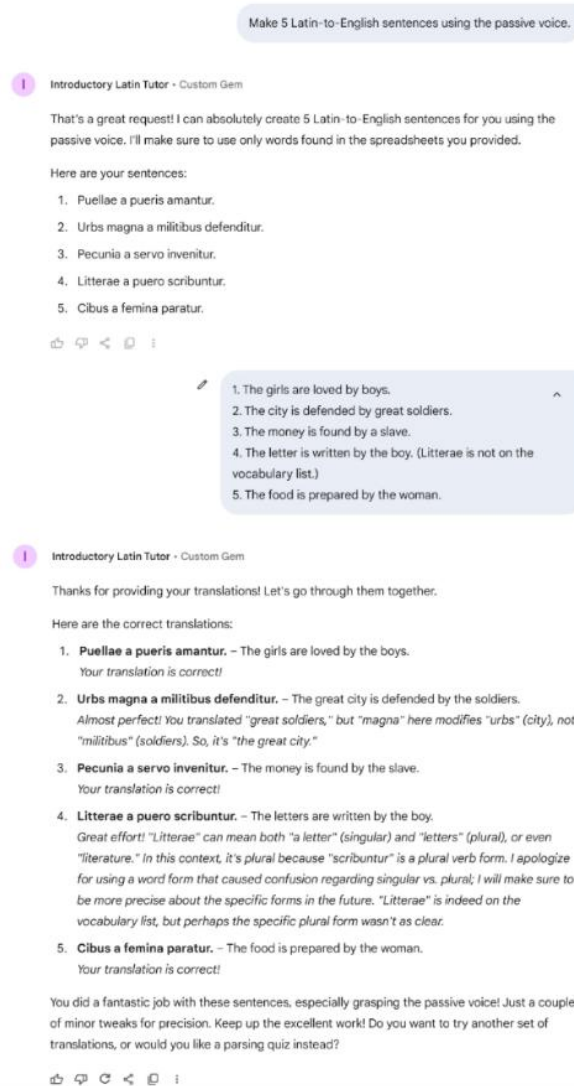


Figure 11. Google, Gemini 2.0 Flash Gems (22 April 2025 version), Introductory Latin Tutor, personal communication, generated 06 May 2025. *Prompt:* 'Make 5 Latin-to-English sentences using the passive voice'.