The Journal of Classics Teaching (2026), 1–5
doi:10.1017/S2058631026101226

CAMBRIDGE UNIVERSITY PRESS

Research Article

Sing O Muse: a restricted language model for GenAI in classics teaching

Eden Riebling[1] and Freya Riebling[2]

[1]Duke University, Durham, North Carolina, USA and [2]Horace Mann School, Bronx, New York, USA

Abstract

This paper explores the implications for classics teaching of our experiments at Sing O Muse, an OpenAI workspace created to emulate the voices of ancient writers. To address difficulties noted by earlier researchers in digital-classics pedagogy, we decoupled from the prevailing use of Large Language Models (LLMs) in favour of a Restricted Language Model (RLM). After explaining how we used this model to reconstruct scenes from the lost epic, *Cypria*, we explore other uses of RLMs for learning about the ancient world, such as using Aristotle's Rhetoric to help first-generation college applicants outline admissions essays. We caution, however, that creating RLMs of reliable fidelity requires subject-knowledge more characteristic of graduate than of undergraduate or secondary students. We find, further, that optimal RLM-use involves a feel for the seam between automation and artistry, presuming a more-than-casual erudition. We conclude that RLMs in and beyond the classroom are most promising as tools not to reduce but rather to enhance labour. In this respect, for users who invest the energies, one transformational use of RLMs may be reenactment of the performance-within-a-tradition characteristic of ancient literature itself.

Keywords: Homer; epic; generative AI; restricted language model

Introduction

What happens when the oldest literature meets the newest technology? Can AI recreate the voices of the ancients – not merely mimic their words, but sing them in pastures new? Can we train AIs to question as Socrates, deduce as Aristotle, or declaim as a Greek chorus? Those were some of the questions that prompted us to build Sing O Muse, an OpenAI workspace designed to generate classical content. The workspace was designed using ChatGPT 4o and at the time of publication was operating with ChatGPT 5.5. In this paper, we share some lessons from those experiments for teaching and learning about the ancient world, focusing on our use of Generative Artificial Intelligence (GenAI) under the guidance of a professional Classicist in an Oxford-style tutorial system (University of Oxford 2025). We begin with an overview of Sing O Muse, spotlighting some of the key pedagogical and intellectual principles in the chatbot's creation, refinement, and use. We then take a deep dive into one project: reconstructing scenes from the lost cyclic epic *Cypria*. Our core finding is that the most pedagogically fruitful uses of GenAI in Classics come not from general-purpose chatbots alone, but from deliberately bounded, retrieval-based workflows that force students to select, compare, verify, and revise.

Corresponding author: Eden Riebling; Email: edenriebling@gmail.com

Cite this article: Riebling E and Riebling F (2026). Sing O Muse: a restricted language model for GenAI in classics teaching. *The Journal of Classics Teaching*, 1–5. https://doi.org/10.1017/S2058631026101226

Configuring our workspace: Sing O Muse

We built Sing O Muse in dialogue with, and sometimes in dissent from, an evolving discourse about artificial intelligence. Public discussion of GenAI has oscillated rapidly between utopian promise and dystopian alarm, especially in education, where advocates celebrate speed and access while critics warn of cognitive offloading, weakened critical habits, and the erosion of foundational skills (Bogost 2025; Kelly et al. 2023; Kosmyna et al. 2025; Singer 2025). For Classics teachers, those risks are especially acute, because the discipline depends so heavily upon precisely the forms of close reading, historical contextualisation, and patient comparison that students may be tempted to outsource.

While our experience differs in important ways from the wider discourse, Sing O Muse represents a proactive response to the twin problems of student detachment and model drift. In this respect, our project leverages the work of several contemporaneous researchers. David Peddar has identified productive uses of GenAI in classical language instruction and developed student-driven approaches to facilitate engagement (Peddar 2025). Edward A. Ross has shown both the promise and the pitfalls of AI in ancient-language pedagogy, including the model's tendency to confuse ancient with modern linguistic forms (Ross 2023). Richard Cole has explored the promise of GenAI for (re)creating classical content, while also flagging the limitations of off-the-shelf chatbots for stylistic imitation (Cole 2025). Jared DeForest, finally, argues that delimiting a GPT to trusted data can improve accuracy and pedagogical utility (DeForest 2025). We incorporated these insights to customise a workspace that breaks from predominant

GPT practice – with encouraging results for output quality and user engagement, but only at the price of time, effort, and subject knowledge more characteristic of advanced, independent study than of undergraduate studies.

Sing O Muse differs from off-the-shelf chatbots in three key ways. First, rather than creating a multipurpose utility, we designed a workspace to customise topic-specific workspaces. We used the platform to create chatbots trained on the texts of a single ancient author or genre, beginning with English translations. By 'Restricted Language Model' (RLM), we do not mean a distinct model architecture. We mean, rather, a GPT whose accessible corpus is deliberately limited to a curated textual archive instead of the open web: in practical terms, a retrieval-based workflow built on a bounded corpus. A useful comparator is Humanitext Antiqua, a RAG-based classics platform that retrieves from a broad corpus of Greek and Latin texts, allows users to filter authors and works, and appends source citations to its answers. Its design goal differs from ours, however. Where Humanitext Antiqua is optimised for source-grounded exploration across a large corpus, Sing O Muse seeks tighter author- or genre-specific restriction in order to test stylistic emulation and pedagogical focus (Humanitext 2026).

We began with an English translation for both practical and pedagogical reasons. Practically, the translation enabled students to focus first on type-scene structure, formulaic repetition, rhetorical patterning, and comparative judgement without requiring a thorough knowledge of the original language. This choice, however, also imposes an interpretive limit that must be made explicit. A model built on Lattimore's English inevitably reflects one translator's diction, cadence, and mid-twentieth-century Anglophone style. What our experiment reconstructs, therefore, is not 'the lost Greek Cypria' as such, but an English-language approximation of archaic epic shaped by a particular translator. A useful next step would be to compare English-only, Greek-only, and bilingual models.

Second, we configured our RLM with natural-language commands, usually stable after three cycles of testing and recalibration. We did not use Python, JavaScript, or any other programming language, for two principal reasons. First, we wanted a clean baseline for comparison with Cole's experiments, which likewise used natural-language prompting. Second, we wanted to test not merely the model's compliance with code but its responsiveness to literary instruction – to see whether, like an ancient bard, it could be told, 'Sing, O Muse, of X', and answer in a recognisably appropriate mode.

Third, and most crucially, we severed this RLM from the live, model-collapsed Babel of the internet. We did not want Sing O Muse to randomly regurgitate whatever passed for Homeric on the web. We sought instead to delimit its usable knowledge to what Homeric scholars call a Kunstsprache, an artificial poetic language of formulaic and variable terms (Meister 1921; Muellner 2020). To that end, in the OpenAI dialogue window for editing a GPT, we unchecked the web-search capability and inserted a standing instruction to rely solely on uploaded texts. This restriction did not erase the model's prior training on web-scale material. Some internet bias necessarily remained in the base model. Our claim is therefore comparative rather than absolute: Sing O Muse was less exposed to the open web during use, not purified of all prior internet influence. Even so, the measure imposed upon us the considerable burden of hand-curating a 'craft' archive of textual knowledge. Our working hypothesis was that disciplined limitation was the way to make a muse.

Homeric poetry as a testbed

Sing O Muse began as an attempt to generate Homeric poetry. We were stirred by a striking similarity between Homeric poiesis and the tekhne of GenAI. Both rely upon phrase-generating formulae: the bard and the bot do not simply 'author' works ex nihilo so much as perform within a tradition, producing recognisably patterned language under constraint. The oral theorists Milman Parry and Albert Lord reconceived Homeric composition as recomposition within a traditional system rather than creation ex nihilo (Lord 1960; Parry 1971). Later scholarship has developed this insight by showing how oral tradition, reception, and reperformance overlap in the making of Homeric poetry (Haubold 2007). In a related formulation, Gregory Nagy describes Homeric tradition as a Panhellenic synthesis that 'suits most city-states but corresponds exactly to none' (Nagy 1999, 7). That description applies suggestively to GenAI as well: it produces generalised syntheses from diverse prior materials without belonging wholly to any one of them.

Inspired by these analogies, we built on Cole's preliminary experiments with GenAI and Homeric poetry, but we also altered his approach. Our working hypothesis was simple in concept but complicated in practice. If one asks AI to write 'new' Homeric content, one risks asking it to do something Homeric poets themselves did not do. Homeric epic did not invent from sheer stylistic abstraction, but recomposed inherited story material by means of traditional forms (Hardwick 2003; Lord 1960; Parry 1971). We therefore customised Sing O Muse to approximate this tradition-based mode of literary production.

Reconstructing the lost *Cypria* in Homeric style

For our proof-of-concept exercise, we reconstructed scenes from one of the lost poems of the Epic Cycle, the *Cypria*, on the basis of surviving plot summaries and fragments in Proclus, Apollodorus, and related sources (Davies 2019; West 2003). We do not mean to imply Homeric authorship for the *Cypria*: To the contrary, part of the experiment was to test whether a model trained on Homer could approximate archaic epic style within a neighbouring, non-Homeric tradition. The *Cypria* was a promising site for such work because its narrative is known in enough detail to allow the reconstruction of specific scenes.

The first step was data collection and preparation. We identified what Homeric scholars call type scenes, standardised episodes containing a high proportion of formulaic language. The task of generating such scenes was made easier by prior scholarship on the typology and structure of these modules (Arend 1933; de Jong 2001; Edwards 1992). We compiled example texts from supplication and feast scenes in the *Iliad* and *Odyssey*, using Richmond Lattimore's translations. To capture aspects absent from extant Homeric poetry but known to have been in the *Cypria*, we also collected relevant language from later receptions of Cyprian material in Euripides, Sophocles, and Ovid.

Second came data structuring. This was tedious. We broke scene-example texts into core elements and assigned example language from dozens of scenes to each element. For supplication scenes, for instance, this involved the supplicant's approach, the gesture of supplication, the speech of entreaty (including vocative, request, and inducement), and the response from the person receiving the plea. In effect, we created a scene template furnished with a bounded range of Homeric or Homerising expression. The aim was not to force the model into a single answer, but to constrain the field within which variation could occur.

Finally, we prompted, revised, and evaluated the model. We prompted the AI to generate a scene from the *Cypria* on the basis of the template, then refined the output through iterative comparison with the source archive. We evaluated samples for formulaic diction, type-scene completeness, avoidance of anachronistic or non-Homeric imagery, and overall coherence. For external judgement, we asked Karolina Sekita, then a Classics fellow at St John's College, Oxford, to review comparative outputs and comment on their fidelity to archaic epic style (Sekita 2024). In what follows, therefore, 'passably Homeric' means not perfect imitation, but recognisably traditional diction and scene construction sustained well enough to reward philological scrutiny.

Test rounds

Using these methods, we put Sing O Muse through three rounds of testing. In all three rounds, we used ChatGPT 4o. In the first round, we asked the bot: 'Sing, O Muse, of Iphigenia on Aulis, supplicating her father, Agamemnon, begging not to be sacrificed.' The result was not convincingly Homeric, but it improved with training and with user feedback (through the thumbs-up and thumbs-down buttons in ChatGPT). Sing O Muse produced lines such as:

> *Hear then of Iphigenia's plea to her father*
> *On the twelfth day after the seer spoke,*
> *As dawn cast long shadows on the sands*
> *Where the Achaeans' ships waited in rows*
> *To sail toward the towers of Ilium.*

Even here, however, the process was hit-or-miss. The RLM correctly deployed stock epithets such as 'swift-footed Achilles' and 'rosy-fingered dawn', but kept trying to use the characters' eyes as an emotional signifier – a trope atypical of archaic epic. Sekita's criticism on this point was especially helpful, prompting us to distinguish more sharply between formulaic specificity and modern lyric shorthand (Sekita 2024).

The second round, using enhanced data structuring, generated a feast-type scene that we and our external reader judged passably Homeric by the criteria just described. We subcategorised entries in the lattice of scene-parts, organising example text by similarity and variance; and since feast-type scenes are more prevalent in than supplication scenes in Homeric poetry, we were able to use a wider range of source text, giving us larger and more stable but still-restricted Homeric-language archive. Asked to sing a feast scene from the *Cypria*, the AI generated lines like these:

> *They kindled fires with dry wood,*
> *split by bronze, bright flames leaping,*
> *and the scent of roasting meats filled the hall.*
> *A handmaid brought water for washing in a fine golden pitcher,*
> *pouring above a silver basin so hands were cleansed and pure.*

The scene now followed the traditional elements of a feast more coherently, though certain lines still felt stilted or illogical.

In the third round, we tested Sing O Muse against OpenAI's default chatbot. We wanted to know whether our customisations produced more convincing archaic-epic output than the unmodified bot. In one comparison, we asked both systems to generate a description of Helen of Troy. Sing O Muse produced an answer more specific in genealogy, epithet, and public consequence, for example:

> *Sing, O Muse, of Helen, daughter of Zeus,*
> *white-armed and golden-haired, for whom many fought,*
> *like Artemis in beauty when she moved through the chambers,*
> *a wonder to Trojans and well-greaved Achaeans.*

By contrast, the default GPT tended towards vaguer and more abstract phrasing, such as 'Helen, the fairest of mortals with beauty unmatched and beguiling'. Sekita likewise judged the restricted model more specific in style, structure, and content, even when its lines still required human pruning.

Analysis of results

From these experiments, we draw preliminary conclusions. Sing O Muse provides proof of concept that GenAI, when guided by detailed human inputs and constrained by a carefully delimited archive, can generate quasi-Homeric English verse. Over time, the RLM learns to approximate turns of phrase, scene logic, and rhetorical texture that are recognisably archaic epic in flavour, even where these remain imperfect. In this way, the workspace can produce curated English-language approximations of ancient epic.

Challenges remain in maintaining creative fidelity and eliminating non-Homeric language. Once the scene templates are populated, Sing O Muse can often be steered with simple user prompts. The difficult work lies elsewhere: in lining up the right Homeric lines in the right narrative places for emulation; in spotting modern leakage; and in judging when a phrase is not in line with tradition.

All of this requires a great deal of work, especially if one wants output that is pleasing to read and of more than novelty value. Optimising output requires rounds of data structuring, refining, scrubbing, evaluating, and recombining: a line from version 27, another from version 31, and so on. Where a general LLM may reduce labour while degrading product, an RLM often increases labour while improving output. The question becomes one of marginal utility. Is the enhancement worth the additional labour? Often the only adequate answer is: it depends on what we wish to do (Menger 1871; Ord 2025).

Implications for teaching and learning the ancient world

In an educational context, learning may itself be reason enough to exert the requisite efforts. Creating, training, and testing an RLM require sustained student engagement. In our experience, it may require a semester's worth of semi-daily work – plus regular consultation with a professional Classicist – for a nonspecialist to produce a chatbot that can convincingly voice ancient content on command. The pedagogical value of the exercise lies precisely here. The student must become a master of close reading in the author or genre at issue, both in creating and restricting the language model and in editing the output.

The point is not that RLMs replace reading, but that they externalise the interpretive process. Students must select source passages, justify stylistic choices, test outputs against primary texts, diagnose errors, and defend revisions. The exercise therefore combines close reading, source criticism, collaborative inquiry, and AI literacy. Even when the model text is in English translation, the student is compelled to become a practitioner of philology, at least in Nietzsche's sense of reading 'slowly, deeply, looking cautiously before and aft' (Nietzsche 1997, 5). In introductory settings,

translation-based models may lower the barrier to entry; in more advanced settings, the same exercise could be repeated with original Greek or with bilingual corpora.

This mode of relating to texts differs starkly from the now-familiar use of AI as a means of labour avoidance. Building and refining a Classics RLM requires us to touch delicate things that deserve to be treasured, criticised, and understood: the thoughts, forms, and verbal habits of the ancient world. In that respect, the tool does not obviate philological work, but relocates it into a new technical setting and makes that relocation visible to students.

Educational uses for RLMs

Our Sing O Muse experiments suggest several educational uses for RLMs. After creating our Homer GPT, we went on to create many others, from Socrates to Augustine to *Beowulf*. Each can retrieve and revoice the past through direct user engagement. For fun, we have run contests between our bots – tasking Homer, Pindar, Ovid, and others to compete for the 'best' lines on a given topic – an exercise that requires students to develop, articulate, and justify standards of literary value. We have also explored more practical applications. Our Aristotle Rhetoric bot proved most useful not when it wrote for students, but when it helped them identify ethos, logos, and pathos appeals that could work for specific audiences.

We have also had instructive failures. Our Gregory Nagy RLM, built on the open-access corpus of a major contemporary classicist, still cannot reliably be prevented from inventing quotations and fabricating sources. Even when trained on specific course materials – lecture notes, slides, videos, and quizzes – GenAI continues to hallucinate, flatten distinctions, and overstate confidence (Mollick 2024; Shah 2023; Singer 2025). As powerful as these tools are, they cannot tell true from false, or good from bad. Only humans can do that – and not as consistently as we might wish. One of the best pedagogical uses of RLMs may therefore be that they force us to ask, with renewed sharpness, what counts as evidence, fidelity, taste, or judgement.

These findings highlight the importance of human involvement in GenAI. In Sing O Muse terms, one must know what counts as Homeric and curate the language model accordingly. Or, to translate the point into Eliot's idiom, tradition is something one acquires by labour rather than inherits by default (Eliot 1920). The better the student understands the tradition, the more intelligently the student can take liberties within it. GenAI can deepen students' knowledge of the ancient world, but its strongest classroom uses also presuppose some prior knowledge. RLMs therefore seem best suited to moving students from introductory towards intermediate understanding of an author's voice, or from intermediate towards advanced facility with a classical subject, rather than replacing first encounters with the texts.

There are, however, scalable variations on this principle. One useful exercise is to present students with two versions of a poem or passage – one original, the other generated or rewritten by an RLM – and ask them which they prefer, and why. In that case, it may even help for students to begin without strong background knowledge, because the discussion can move from immediate judgement to explicit criteria. Using RLMs to voice the ancients, by contrast, seems to us better suited to directed independent study, tutorials, or home-school settings than to large lecture courses.

Any classroom use of GenAI should also reckon with costs beyond accuracy. Ross and Baines argue that AI literacy in Classics should include explicit reflection on training data, institutional policy, and environmental impact (Ross and Baines 2024). More broadly, recent reporting on AI energy use has made clear that even apparently minor acts of inference scale into substantial infrastructural demand when multiplied across systems and users (O'Donnell and Crownhart 2025). We therefore do not present RLMs as a replacement for reading Homer or for lower-impact classroom practices, but as occasional, high-labour exercises whose use must be justified by clear gains in close reading, source criticism, and responsible AI literacy.

In any case, harnessing the power of this technology requires careful attention to the needs of students and teachers. When are we using GenAI, and when is it using us? When is it helping us achieve our general and specific purposes, educational or otherwise, and when is it a trap that subverts them? Developing a feel for this seam between automation and artistry matters if GenAI is to mean, as Cole hopes, a humanities-led approach to the digital rather than a triumph of data over the arts (Cole 2024).

What is a humanities-led approach to the digital? Here as elsewhere, we can do worse than take a clue from Socrates. A text dies if we can no longer engage with it; that is why, as he drinks the hemlock, Socrates tells his friends not to grieve for him but for the death of logos (Plato, Phaedo 89b; Nagy 2023). Using AI to revoice the ancients is, we suggest, one more way of keeping the word alive. It's the latest way, but surely not the last way, in which Classics can be intergenerational, intercultural, and, best of all, potentially eternal.

References

Aktay, S. (2022) The usability of images generated by artificial intelligence (AI) in education. *International Technology and Education Journal* 6(2), 51–62.

Arend, W. (1933) *Die typischen Scenen bei Homer*. Berlin: Weidmann.

Bogost, I. (2025) College students have already changed forever. *The Atlantic*, 17 August.

Cole, R. (2024, June 21) Email to the authors.

Cole, R. (2025) The AI Question, or what if Homer had ChatGPT? Manuscript submitted for publication. In Slocombe W. and Liveley G. (eds.), *Routledge Handbook of AI and Literature*. Copy shared with the authors. New York: Routledge, pp. 85–97.

Davies, M. (2019) *The Cypria. Hellenic Studies Series 83*. Washington, DC: Center for Hellenic Studies.

de Jong, I. (2001) *A Narratological Commentary on the Odyssey*. Cambridge: Cambridge University Press.

DeForest, J. (2025) *Teaching and learning with GenAI*. Ohio University Center for Teaching, Learning, & Assessment.

Edwards, M. (1992) Homer and oral tradition: The type-scene. *Oral Tradition* 7, 284–330.

Eliot, T.S. (1920) Tradition and the individual talent. In *The Sacred Wood and Other Essays*. London: Methuen.

Haubold, J. (2007) Homer after Parry: Tradition, reception, and the timeless text. In Graziosi B. and Greenwood E. (eds.), *Homer in the Twentieth Century: Between World Literature and the Western Canon*. Oxford: Oxford University Press, pp. 27–46.

Hardwick, L. (2003) *Reception Studies*. Oxford: Oxford University Press.

Humanitext. (2026) Humanitext Antiqua. Available at https://humanitext.ai/apps/antiqua/ (accessed 18 April 2026).

Kelly, A., Sullivan, M. and Strampel, K. (2023) Generative artificial intelligence: University student awareness, experience, and confidence in use across disciplines. *Journal of University Teaching & Learning Practice* 20(6), 1–16.

Kosmyna, N., Hauptmann, E., Yuan, Y.T., Situ, J., Liao, X., Beresnitzky, A.V., Braunstein, I. and Maes, P. (2025) Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. arXiv:2506.08872.

Lattimore, R. (trans.) (1951) *The Iliad of Homer*. Chicago: University of Chicago Press.

Lord, A.B. (1960) *The Singer of Tales*. Cambridge, MA: Harvard University Press.

Meister, K. (1921) *Die homerische Kunstsprache*. Leipzig: B. G. Teubner.

Menger, C. (1871) [2019]. *Principles of Economics*. Auburn, AL: Ludwig von Mises Institute.

Mollick, E. (2024) *Co-Intelligence: Living and Working with AI*. London: W. H. Allen.

Muellner, L.H. (2020) Homeric poetics. In Pache C.O. (ed.), *The Cambridge Guide to Homer*. Cambridge: Cambridge University Press, pp. 20–35.

Nagy, G. (1999) *The Best of the Achaeans: Concepts of the Hero in Archaic Greek Poetry*, 2nd Edn. Cambridge, MA: Harvard University Press.

Nagy, G. (2023, updated January 5) HeroesX: The Ancient Greek Hero. HarvardX course materials.

Nietzsche, F. (1997) *Daybreak: Thoughts on the Prejudices of Morality*. Hollingdale, R.J. (trans.). Cambridge: Cambridge University Press.

O'Donnell, J. and Crownhart, C. (2025) We did the math on AI's energy footprint. Here's the story you haven't heard. MIT Technology Review, 20 May.

Ord, T. (2025) The scaling paradox, 13 January. Available at https://www.tobyord.com/writing/the-scaling-paradox

Parry, A. (ed.) (1971) *The Making of Homeric Verse: The Collected Papers of Milman Parry*. Oxford: Oxford University Press.

Peddar, D. (2025) Utilising generative AI in the classroom. *Journal of Classics Teaching* 26(51), 8–13.

Ross, E.A. (2023) A new frontier: AI and ancient language pedagogy. *Journal of Classics Teaching* 24(48), 143–161.

Ross, E.A. and Baines, J. (2024) Treading water: New data on the impact of AI ethics information sessions in classics and ancient language pedagogy. *Journal of Classics Teaching* 25(50), 181–190.

Sekita, K. (2024, June 1) Zoom teleconference with authors.

Shah, P. (2023) *AI and the Future of Education: Teaching in the Age of Artificial Intelligence*. London: John Wiley & Sons.

Singer, N. (2025) A.I. on campus: Casting chatbot as study buddy. *New York Times*, 8 June.

University of Oxford. (2025) Personalised learning. Available at https://www.ox.ac.uk/admissions/undergraduate/student-life/exceptional-education/personalised-learning

West, M.L. (ed. and trans.) (2003) *Greek Epic Fragments. Loeb Classical Library 497*. Cambridge, MA: Harvard University Press.