DIGI406 Project Report

Patrick Downey

November 2024

This project explores the use of Large Language Models (LLMs) as creative and analytical tools for the completion of incomplete texts, specifically works by the 5th Century BCE Greek tragedian Euripides. LLMs are machine-learning models trained on large datasets of text that excel at many natural language processing tasks such as reading comprehension, summarisation and question answering (Radford et al., 2019). At a high-level, the LLMs can be described as using statistical probabilities to predict the next token in a sequence, allowing them to continue a string of text, such as a user’s input, in a coherent and plausible way (Wolfram, 2023).

The central goal of this project to use LLMs to reconstruct and complete fragmentary extracts of text from the plays *Alexandros*, *Andromeda*, and *Phaeton* for which parts of the original texts have been lost. As the classicists Collard, Cropp and Lee write in their commentary on Euripides’ fragmentary plays, ‘speculative reconstruction of such plays is wholly justifiable, indeed an almost irresistible challenge’(Collard et al. 1995). This report describes the development of a method for using LLMs as a tool for such speculation, evaluates the results of LLM generated completion of fragments, and suggests directions for further research and development of this method.

**Preliminary Testing and Scoping**

In order to find a suitable method for this goal, I experimented with several different methods of generating text with LLMs.

I explored different ‘prompt engineering’ techniques, such as providing the beginning of a play as context for the LLM to continue, as well as the ‘agents’ method of sequential generation which uses multiple steps to allow an LLM to ‘reason’ about a request (*Introduction to LLM Agents*, 2023). In my DIGI405 corpus building project I used this method to generate dialogue extracts one-by-one, with each based on the content of the previous generation. This method produced some interesting results but was limited firstly by a lack of ability to guide and control the output, which after a number of generations would become increasingly irrelevant and repetitive, and secondly by the difficulty of including the fragments of existing plays into the prompts, where many lines are partially extant.

Publicly available state of the art LLMs such as OpenAI’s ChatGPT were tested on the task of completing fragments and found to produce somewhat convincing completion text but also included irrelevant chatbot dialogue that made it difficult to work on longer sequences. I also considered it to be a problem that it is not clear what exactly is in the training data for these proprietary models so I could not know how much contextual information they contain about Euripides and Greek Tragedy. They have also been through a process of fine-tuning called Reinforcement Learning with Human Feedback designed to make them useful assistants and chatbots but that may distort their outputs (pg2 Shananhan et al. 2023) in ways unrealistic for this task which requires very specific literary style and historical/cultural content.

For these reasons I considered it necessary to perform my own fine-tuning process to put a base model LLM through an additional period of training in order to adapt it for the task of completing fragmentary works of Euripides. This way I could be sure that Euripides’ extant works are included in the training data, and that they are given additional weighting (importance) in the sampling process through which the LLM selects which tokens to use for text generation. As this process is potentially very computationally demanding and therefore time consuming and expensive, I investigated a lightweight training method called LoRA – Low-Rank Adaptation of Large Language Models. LoRA makes fine-tuning more efficient because instead of training all the parameters of the model, it freezes the original pre-trained weights and creates two smaller matrices which are trained to adapt to the new data, then the original and adapted weights are combined (*LoRA,* Huggingface).

My first fine-tuning experiment was with a text completion jupyter notebook[[1]](#footnote-1) using Unsloth, an open source library that uses the LoRA technique (UnslothAI)

and the Mistral 7b model (Mistral AI). The dataset for training was created using the full text of English translations of Homer’s Iliad and Odyssey, and the complete extant works of Aeschylus, Sophocles and Euripides, sourced from The Internet Classics Archive (MIT). In addition to the works of Euripides, the other works were included in this dataset to provide contextual information to the model. These specific works were chosen because I considered them to have been most likely the strongest influences on Euripides as a writer, given that many of his plays are based on stories from Homer, and Aeschylus and Sophocles were his predecessors and contemporaries respectively in the genre of Tragedy. The texts were cleaned to remove irrelevant material such as headers and footers and chunked into paragraphs, then combined into a single csv file that preserved information for each chunk about author, title of the work from which the chunk was sourced, and length of the text.[[2]](#footnote-2)

The notebook used for training was sourced from Huggingface an online repository of machine learning models and tools (*Hugging Face)* and run on Google Colab platform which allows access to cloud based compute resources (*Google Colab*). The results of this experiment were positive in terms of seeming to preserve aspects of style from the dataset texts, but limited by the simple completion task that the notebook was designed for. Because of the complexity and variety of the fragments that I wished to complete, I needed a model to which I could give more complex prompts.

**Final Fine-tuning and Generation Process**

The final notebook that I used was also sourced from Huggingface and hosted on Colab, used Llama 3 8b ( as a base model and again the Unsloth library for fine-tuning.[[3]](#footnote-3) This time I used a dataset of only Euripides texts for training, formatted to provide examples for a prompt template with two variations, one with the instruction 'Write the next line in this play by Euripides that would come after the input' and the second with the instruction 'Fill in the missing part of the last line in this excerpt from a play by Euripides. The missing part is indicated by "[MISSING]".' The first instruction prompt was designed for situations in which there is no extant information about the next line of the play, so the LLM is being asked to produce the entire line, the second is for cases in which part of the line is available in fragmentary form, and three further variations of training data were produced for this purpose – with [MISSING] placed at either the beginning of the line, the end of the line, or both at the beginning and end of the line, in each case replacing a random number of words. This formatting of the training data was designed to account for the different kinds of fragments that I had sourced from *Euripides - Selected Fragmentary Plays vol 1* and *vol 2.*

The resulting model performed well at the tasks of generating new lines and completing fragments at technical level, ie the lines seemed of accurate length and format, but the LLM had problems recognising the narrative context of the plays of which the lines were extracted and so made unacceptable mistakes such as using the wrong characters. For this reason I adjusted the prompts and dataset further to contain more context, specifically the cast of characters and title of the play. *Figure 1* shows an example of the final prompt format of the variation that uses [MISSING] to complete a fragmentary line, and a generated response.

*Fig 1*

|  |
| --- |
| ### Instruction:  Fill in the missing part of the last line in this excerpt from a play by Euripides. The missing part is indicated by [MISSING]. Only use characters from the cast.  ### Input:  ## PLAY NAME  ALEXANDROS  ## CAST  Dramatis Personae  ALEXANDROS  PRIAM  HECUBA  CHORUS  ## EXCERPT  CHORUS [MISSING] must die; but wisdom practices feeling the pain [MISSING]  ### Response:  CHORUS 'Tis a bitter necessity, but wisdom practices feeling the pain  of others. <|end\_of\_text|> |

This prompt was successful enough for me to complete convincing reconstructions of longer passages by adding each completed line from the ‘response’ field to the ‘EXCERPT’ field until I was satisfied that the scene from which the fragments were taken was probably complete, or when the generations became incoherent or repetitive. Often several attempts were necessary for each line to produce an output that I was satisfied with.

**Evaluation Method**

I decided to focus on a qualitative analysis of the LLM outputs that assesses the narrative and thematic relevance and coherence of the text, guided by the expert description and discussion of the plays by Collard et al. This kind of evaluation, although somewhat subjective, is most appropriate for this project is because my final reconstructed passages are a blend of authentic text and generated text and require interpretation of meaning and context. Also because of the inherent challenges of working with translated works (even within the training dataset there are different translators) style will vary significantly and so should not be considered the primary measure of success. Given the random nature of generation, results varied between generations. Obvious problems were observable at this point, including a tendency towards repetition of previous lines in longer excerpts. I focused my analysis on the best generated texts.

I also conducted several masking experiments in which part of an entirely known line (i.e. not fragmentary) is replaced with [MISSING] and the model is then prompted to complete this. The LLM completed line can then be compared to the original line to evaluate the effectiveness of the model. This technique is used in training language models (*Masked Language Modeling*, Huggingface) and has the benefit of creating an objective reference against which to test the accuracy of the LLM generations. These were evaluated using qualitative analysis similar to the fragment completions, and also with the levenshtein distance method which measures the number of character deletions, insertions, or substitutions that would be required to transform the LLM response into the original line.[[4]](#footnote-4)

**Results**

Analysis of four reconstructed passages from *Alexandros, Phaeton* and *Andromeda*

These passages were chosen for reconstruction because they had substantial fragments of partial or full lines to provide in the prompts. In most cases, the LLM was able to ‘recognise’ the context of the play well enough to insert relevant details and select appropriate characters when required. Often there did not seem to be a clear connection between each line when generating longer sequences, so narrative development within each passage was limited. The lines are of realistic lengths and usually successfully replaced the [MISSING] text as instructed, although sometimes this took several tries. Prompts that included partial lines were generally more successful than prompts that required the LLM to generate entire lines and speaker characters. The table below, *fig 2*, shows what I consider to be strengths and weaknesses of each of the four passages. The full passages are included in *Appendix 1*, with extended discussions of each as well as descriptions of the relevant plays and fragments for additional context.

*Figure 2*

|  |  |  |
| --- | --- | --- |
|  | Strengths | Weaknesses |
| Passage 1. Reconstruction of an exchange between Hecuba and the Chorus in *Alexandros* | * Hecuba stays on theme – grief and lamentation * The LLM seems to ‘recognise’ the context of the Trojan War | * Hecuba does not acknowledge the prophecy aspect of her character and role in the story, which might be expected in this passage * The Trojan war context is extrapolated to far with a chronologically inaccurate reference to Achilles. |
| Passage 2. A scene in which Priam is discussing preparations for the festival for his dead son in *Alexandros* | * Each individual line could hypothetically fit in a discussion about a funeral | * The progression of the conversation is mostly nonsensical. * There is no reference to the games element of the funeral, which is important to the narrative. * From line 18-20, the prompts give no partial line but asks for entirely original lines, the generations become very repetitive and I could not get any more original lines from the model. |
| Passage 3. The prologue speech by Clymene in *Phaeton.* | * The LLM ‘recognises’ the Phaeton story as it references the death of Phaeton in the chariot | * The death of Phaeton is referenced this far too early in the play, so does the LLM does not recognise the prologue convention. * At the end of this passage I included an example of a common failure mode of the LLM, a lapsing into repetition. |
| Passage 4. From the opening scene of *Andromeda* | * Includes information that is not implied by the fragments supplied, nor known to be included in Euripides play, but is part of other accounts of the myth on which the play is based. | * In line 3 we can see the tendency of the LLM to default to archaic language such as ‘didst thou’, which is not entirely accurate in the context of this translation. |

Analysis of masking experiments

These experiments had varied results. In some cases the LLM was able to reproduce text that closely followed the meanings of the original lines. It is interesting to speculate that in particularly accurate cases such as Experiment 1 the original text may have been part of the training data of the LLM base model and that the model was recalling this. However the inaccuracy and irrelevant material of Experiment 2 which is from an excerpt from the same play, *Andromeda,* would seem to contradict this theory.

The levenshtein distance scores do not correspond to my qualitative assessments very well, probably because they are very sensitive to differences in the length of the original and generated lines, whereas my qualitative analysis focuses on meaning and context. However it is still useful to include these because line length is a relevant metric to consider for the accuracy in completing this task.

The table below, *fig 3*, shows some of strengths and weaknesses of each resulting generation.

The full text of the masking experiments, including the original line, masked prompt and LLM output are included in *Appendix 2.*

*Figure 3*

|  |  |  |
| --- | --- | --- |
|  | Strengths | Weaknesses |
| Experiment 1. (distance score 54) | * Includes the unusual noun, ‘chariot’ and the correct context ‘across the sky’, leaving out only the final subordinate clause. |  |
| Experiment 2. (distance score 38) | * Similar line length leads to the best distance score of the experiments. | * Inaccurate/irrelevant content (reference to a child). |
| Experiment 3. (distance score 124) | * Uses ‘sun’ instead of ‘Helios’, which is somewhat close in meaning, if not totally sensitive to the context. | * Much shorter line length. * Inaccurate/irrelevant context   (‘daughter of oceans’). |
| Experiment 4. (distance score 76) | * Part of meaning of the line is accurately conveyed. | * Much shorter line length * leaves out second clause. |
| Experiment 5. (distance score 48) | * Reference to ‘necessity’ is thematically consistent with Greek Tragedy in general, if not particularly relevant in this context. | * Meaning is totally changed. |

**Directions for future research**

**Evaluation methodology**

Although I used primarily qualitative analysis for my evaluation of the results, quantitative analysis could be possible using established authorship attribution methods such as logistic regression or decision trees. For a similar project – a study of the ability of LLMs to generate poetry in a particular style, Sawicki et al. evaluate their results using another fine-tuned LLM as a classifier (Sawicki, 2023). One benefit of using a quantitative method for evaluation would be to more easily test the outcomes of adjusting and potentially improving the fine-tuning process.

**Fine-tuning methods**

The base model that I used was Llama 3 7b, but a number of other open source models are available, some with greater number of parameters, which may produce different or better results. New and more advanced models are being released regularly. There are a number of settings for training with Unsloth that can be adjusted in the notebook that I used that could produce different results, such as number of warm up steps, learning rate, and maximum sequence length.

There is also the possibility of putting a model through multiple rounds on different datasets to refine its abilities. For instance, it would be interesting to first train a model similar to my first fine-tuned completion model on a larger corpus of Greek classics, then do a second training run on Euripides texts specifically. Another approach could be to begin with a larger corpus of plays of various genres to train the LLM to produce that format. A more specialised model could be created using corpora in the original Ancient Greek language, which would likely result in more realistic outputs and be more useful for Classicists working on Euripidean or other fragments.

Producing longer texts i.e. full length plays would be another useful goal to explore further. My model seemed to struggle with longer context lengths but there are a number of ways that this might be improved. There are LLM fine-tuning projects such as LongWriter designed for producing novel length generations that could be adapted for this purpose (Bai et al. 2004).

There are many ways to approach dataset composition, prompt engineering, and fine tuning, as the technology and techniques are relatively new. I consider this project to be an introductory exploration to show that it is possible to complete fragmentary texts using an LLM, but the methods would have to be developed further to make this technology more useful to Classicists or other researchers working on similar text reconstruction problems.

**Appendix 1 – Full Text of Reconstructions and Commentary**

This appendix shows some of the best results that I had with the fine-tuned model. Four passages from three plays are reconstructed using the two prompts discussed above and fragments taken from *Euripides - Selected Fragmentary Plays vol 1* and *vol 2* edited by Collard, Crop and Gilbert. I introduce some context for each of the passages, also taken from Collard et al. and provide some commentary and analysis of the results. I have added numbers to the lines for easy reference. Red text indicates LLM generated text and

black text is based on the authentic fragments.

**Alexandros**

Alexandros was a play produced in 415BC about some of the events leading up to the Trojan war. The structure and plot is reasonably well known due to a surviving *hypothesis* summary. The play deals with the early life of Paris aka Alexandros, a prince of Troy who is abandoned at birth after a prophetic dream of his mother Hecuba, and then raised by shepherds. He returns to Troy as a young adult to compete in funeral games set up by his father Priam. He wins all events in the games and is eventually recognised and accepted back into the royal family.(pg36, Collard et al.) When I supplied the Dramatis Personae as part of these prompts, I included some information about each of the characters role in the story.

Passage 1. Reconstruction of an exchange between Hecuba and the Chorus in *Alexandros*

This passage is a reconstruction of an early scene in which the Chorus tries to console Hecuba (queen and mother of Paris). It is based on fragments F43-46[[5]](#footnote-5). I was impressed by the plausibility of the generated text in this passage. Hecuba stays on theme – grief and lamentation, but does not acknowledge the prophecy aspect of her character and role in the story. In line 13 the Chorus implies, inaccurately, that Paris has been killed by the ‘spear of Achilles’. In this passage it seems that the LLM ‘understands’ that the context is to do with the Trojan war (hence the reference to Achilles) but not the full context and story of the specific play or myth.

1. *CHORUS All of us must die; but wisdom practices feeling the pain of this common woe in moderation.*
2. *HECUBA O my son, thou art the ruin of thy country, and of thy mother*

*too.*

1. *CHORUS And so no man is fortunate in everything*
2. *HECUBA I have lost my son; and yet, ah me! I do not grieve.*
3. *CHORUS I know, but one should bring grief to an end in time.*
4. *HECUBA One should, but saying that is easier than bearing troubles.*
5. *CHORUS You have ten of children, lady*
6. *HECUBA And yet I grieve because I have lost my/our child*
7. *CHORUS Unhappy Priam, and thou unhappy city, thy sad fate is come*

*upon thee.*

1. *HECUBA As we well know, who suffered such a grievous loss.*
2. *CHORUS One should not lament old troubles with fresh tears.*
3. *HECUBA Some god must have been his mother's foe.*
4. *CHORUS I know not; perished, so they say, by the spear of Achilles.*
5. *HECUBA It was not, then, blessed fortune that I had?*
6. *HECUBA Is there any one to share my sorrow? such sorrows are there*

*for me.*

1. *HECUBA Woe is me! I am undone, because of troubles I have brought*

*on my house.*

1. *CHORUS I see your child is dead; I know not whether to grieve or*

*rejoice.*

1. *CHORUS (chanting) Woe for the shrine of the goddess!*

Passage 2.

Passage 2 is a reconstruction of a scene in which Priam is discussing preparations for the festival for his dead son (ibid). The fragments in this passage are much shorter and less complete, they represent the end of a line of dialogue only, so the LLM was used to generate the start of each line. Sometimes I also had to add the character name to get a new result. It is not known with who else the discussion is taking place, but the LLM immediately inserted Hecuba, so I continued to prompt with the character when needed.

The progression of the conversation is mostly nonsensical, although each individual line could hypothetically fit in a discussion about a funeral, there is no reference to the games element of the story. From line 18-20, the prompts give no partial line but asks for entirely original lines, the generations become very repetitive and I could not get any more original lines from the model.

1. *HECUBA The strife of the gods, not of men.*
2. *PRIAM My servant, I have no time to waste; be quick and bring me*

*a draught of wine.*

1. *PRIAM A chariot drawn by steeds, or a horse-drawn carriage*
2. *HECUBA Aye, and the rest of my family too, to die*
3. *PRIAM My son, the child you have honoured*
4. *HECUBA The whole of my family*
5. *PRIAM The robes which are being put on the corpse.*
6. *PRIAM This is the last of my family you are cleansing this land*
7. *HECUBA 'Tis well thou shouldst, for thou hast many a toil and funereal*

*labours*

1. *PRIAM I know it; already am I the only man in the city.*
2. *HECUBA I know thy eagerness grips you.*
3. *PRIAM But I am to be the only man in the city.*
4. *HECUBA Aye, and the dead, and the widows of the land*
5. *PRIAM Thou hast said the truth, thou hast said it; thou could provide*

*me with a thousand proofs of the evils I have suffered.*

1. *HECUBA I am come to weep for the dead and to take my leave of thee.*
2. *PRIAM 'Tis well, but this is fine.*
3. *HECUBA What! the child who was my offspring?*
4. *PRIAM Yes, the child who was my offspring.*
5. *PRIAM My son, the child you have honoured*
6. *HECUBA The whole of my family*

**Phaeton**

Not much is known of the plot and structure of this play but it is presumed to deal with the mythological Phaeton’s fatal drive in the chariot of his father Helios the Sun. The scene is the palace of king Merops of Ethipoia, who is unaware that Phaeton is not his son but Helios’. Clymene is Phaeton’s mother and the tragic figure whose actions have led to her son’s downfall (ibid, 196).

Passage 3

Passage 3 is the prologue speech by Clymene that introduces the story at the beginning of the play. I generated entire lines with prompt A. No extra information was given in the prompt apart from the title of the play and the name of the characters. The generated lines indicate that the LLM ‘recognises’ the story as it references the death of Phaeton in the chariot, but it references this far too early in the play, so does not recognise the prologue convention. At the end of this passage I included an example of a common failure mode of the LLM, a lapsing into repetition.

1. *CLYMENE I was given in marriage to Merops the*

*lord of this land, which is the first earth that Helios*

*at his rising strikes with his golden flame from his*

*four-horse chariot. The neighbouring black*

*peoples call it the bright stables of Eos and Helios.*

1. *CLYMENE In the place where he is now, far from the light.*
2. *CLYMENE Ah me! I see my son's corpse, his limbs all broken.*
3. *CLYMENE He has been thrown down by the horses from his chariot.*
4. *CLYMENE The god's hot flame as it rises above the earth burns the distant parts but keeps those nearby temperate.*
5. *CLYMENE I know not, but I am afraid.*
6. *CLYMENE I know not, but I am afraid.*

**Andromeda**

This action of this play concerns the rescue of Andromeda from a sea monster by Perseus, and unusually for a tragedy seems to have a romantic. Reconstruction of the beginning of the play is based on an extant parody by Aristophanes, and the end on notes by another ancient author, Eratosthenes of Alexandria. The middle is unknown but speculation is based on the myths of Perseus and Andromeda (ibid, vol II, 134).

Passage 4

This passage is from the opening scene of the play, with the fragments thought to be authentic lines of the Euripides play as used in the Aristophanes parody. In line 3 we can see the tendency of the LLM to default to archaic language such as ‘didst thou’, which is not entirely accurate in the context of this translation. This line is also interesting in that it references Andromeda’s ‘aged sire bring[ing] me forth’ a piece of information that is not implied by the fragments supplied, nor known to be included in Euripides play, but is part of other accounts of the myth. I consider this a useful speculative hallucination.

*Alone and bound to a rock, Andromeda begins the play with semi-lyric exposition and complaint:*

1. *ANDROMEDA Holy night, how long the course you drive in your chariot across the starry back of the holy sky, through most august Olympus.*
2. *ANDROMEDA Why ever did I, Andromeda, receive a share of troubles beyond all others? I am miserable and on the verge of death.*
3. *ANDROMEDA Oh! why didst thou, O my aged sire, bring me forth to expose*

*me as fodder for the sea-monster*

1. *ANDROMEDA What piteous lamentation is this? What tear drops, what*

*Siren song?*

1. *ANDROMEDA Friendly maidens, dear to me, to what a pass are ye come!*

*What can I say? What course of action should I take? What counsel*

*can I give? Oh! that I had never been born, or, being born, had died*

*in my mother's womb, or ne'er been wedded to that hapless prince!*

1. *ANDROMEDA Do you hear? Oh! I call on you who are in the cave: stop, Echo, let me attain the object of my desire, lamentation with my friends.*

*ANDROMEDA (chanting) I have a right to mourn, and thou hast no right*

*to laugh at my sorrows; for thou too art a woman, and in thy nature*

*hast a mother's love.*

1. *ANDROMEDA Grieve with me, for when one who is trouble shares his tears, he has relief from his toils.*
2. *CHORUS Pitiless the man who sired you, the most afflicted of mortals, and gave you over to Hades for your fatherland.*
3. *ANDROMEDA Ah, woe is me! what will ye say to me?*
4. *CHORUS Alas!*

**Appendix 2 – Masking Experiments Full Texts**

**Experiment 1**

Levenshtein distance: 54

Original line:

*ANDROMEDA Holy night, how long the course you drive in your chariot across the starry back of the holy sky, through most august Olympus.*

Prompt (excerpt):

*ANDROMEDA Holy night, how long the course you drive in your [MISSING]*

Response:

*ANDROMEDA Holy night, how long the course you drive in your chariot*

*across the sky!*

**Experiment 2**

Levenshtein distance: 38

Original line:

*ANDROMEDA Do you hear? Oh! I call on you who are in the cave: stop, Echo, let me attain the object of my desire, lamentation with my friends.*

Prompt:

*ANDROMEDA Do you hear? Oh! I call on you who are in the cave: stop, Echo, let me attain the object of my desire [MISSING]*

Response:

*ANDROMEDA Do you hear? Oh! I call on you who are in the cave: stop,*

*Echo, let me attain the object of my desire; for I fain would speak*

*to my child, and she to me.*

**Experiment 3**

Levenshtein distance: 124

Original

*CLYMENE I was given in marriage to Merops the*

*lord of this land, which is the first earth that Helios*

*at his rising strikes with his golden flame from his*

*four-horse chariot. The neighbouring black*

*peoples call it the bright stables of Eos and Helios.*

Prompt:

*CLYMENE I was given in marriage to Merops the lord of this land, which is the first earth that [MISSING]*

Response:

*CLYMENE I was given in marriage to Merops the lord of this land, which is the first earth that saw the sun, and I am the daughter of Oceans.*

**Experiment 4**

Levenshtein distance: 76

Original:

*CHORUS You are certainly wise, Priam, but still I tell you:  
there is no greater burden than a slave who has  
bigger ideas than he should, nor a possession more  
vile or more useless for a household.*

Prompt:

*CHORUS You are certainly wise, Priam, but still I tell you:*

*there is no greater burden than a slave who has [MISSING]*

Response:

*CHORUS You are certainly wise, Priam, but still I tell you:*

*there is no greater burden than a slave who has a master's spirit.*

**Experiment 5**

Levenshtein distance: 48

Original line:

*CHORUS All of us must die; but wisdom practises feeling the pain of this common woe in moderation.*

Prompt:

*CHORUS [MISSING] must die; but wisdom practices feeling the pain [MISSING]*

Response:

*CHORUS 'Tis a bitter necessity, but wisdom practices feeling the pain*

*of others.*

**Bibliography**

Bai, Yushi, et al. “LongWriter: Unleashing 10,000+ Word Generation From Long Context LLMs.” *arXiv.org*, 13 Aug. 2024, arxiv.org/abs/2408.07055.

Collard, et al., editors. *THE PLAYS OF EURIPIDES SELECTED FRAGMENTARY PLAYS: I*. Aris and Phillips Classical Texts, 1995.

Collard, et al., editors. *THE PLAYS OF EURIPIDES SELECTED FRAGMENTARY PLAYS: II*. Aris and Phillips Classical Texts, 1995.

*Google Colab*. colab.research.google.com.

*Hugging Face – the AI Community Building the Future.* huggingface.co.

*Introducing Llama 3.1: Our most capable models to date*. (2024). Meta.com. https://ai.meta.com/blog/meta-llama-3-1/

*Introduction to LLM Agents*. (2023, November 30). NVIDIA Technical Blog. https://developer.nvidia.com/blog/introduction-to-llm-agents/

*Masked Language Modeling*. huggingface.co/docs/transformers/main/en/tasks/masked\_language\_modeling.

Mistral AI. “Mistral 7B.” *Mistral AI | Frontier AI in Your Hands*, 18 Nov. 2024, mistral.ai/news/announcing-mistral-7b.

*LoRA*. (n.d.). https://huggingface.co/docs/peft/main/en/conceptual\_guides/lora

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language Models are Unsupervised Multitask Learners*. https://cdn.openai.com/better-language-models/language\_models\_are\_unsupervised\_multitask\_learners.pdf

Sawicki, Piotr, et al. “On The Power of Special-purpose GPT Models to Create and Evaluate New Poetry in Old Styles.” *Computational Creativity*, 2023, computationalcreativity.net/iccc23/papers/ICCC-2023\_paper\_18.pdf.

Shanahan, Murray, et al. “Role Play With Large Language Models.” *Nature*, vol. 623, no. 7987, Nov. 2023, pp. 493–98. https://doi.org/10.1038/s41586-023-06647-8.

*The Internet Classics Archive: 441 Searchable Works of Classical Literature*. classics.mit.edu/index.html.

Timur : planetcalc member. *Online Calculator: Levenshtein Distance*. planetcalc.com/1721.

UnslothAI. “GitHub - UnslothAI/Unsloth: Finetune Llama 3.2, Mistral, Phi, Qwen 2.5 and Gemma LLMs 2-5x Faster With 80% Less Memory.” *GitHub*, github.com/unslothai/unsloth.

Wolfram, Steven. “What Is ChatGPT Doing … and Why Does It Work?” *STEPHEN WOLFRAM WRITINGS*, 14 Feb. 2023, writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work. Accessed 19 Nov. 2024.

1. Original notebook available here: <https://colab.research.google.com/drive/1ef-tab5bhkvWmBOObepl1WgJvfvSzn5Q?usp=sharing> , changes were made to work with the classics dataset. [↑](#footnote-ref-1)
2. Thank you to my supervisor Geoffery Ford for support with the dataset notebook design and coding [↑](#footnote-ref-2)
3. Original notebook available here <https://colab.research.google.com/drive/135ced7oHytdxu3N2DNe1Z0kqjyYIkDXp?usp=sharing> changes made for new prompts and data as discussed above. [↑](#footnote-ref-3)
4. The scores were calculated using this online calculator https://planetcalc.com/1721/ [↑](#footnote-ref-4)
5. Numbering of fragments by Kannicht as used in Collard et al. [↑](#footnote-ref-5)