

# The Chatbot and the Canon: Poetry Memorization in LLMs

Lyra D'Souza<sup>1</sup>, David Mimno<sup>2</sup>

<sup>1</sup>*Department of Computer Science, Cornell University, USA*

<sup>2</sup>*Department of Information Science, Cornell University, USA*

## Abstract

Large language models are able to memorize and generate long passages of text from their pretraining data. Poetry is commonly available on the web and often fits within language model context sizes. As LLMs continue to grow as a tool in literary analysis, the accessibility of poems will determine the effective canon. We assess whether we can prompt current language models to retrieve existing poems, and what methods lead to the most successful retrieval. For the highest performing model, ChatGPT, we then evaluate which features of poets best predict memorization, as well as document changes over time in ChatGPT's ability and willingness to retrieve poetry.

## Keywords

memorization, LLMs, poetry, ChatGPT, prompting, training data

## 1. Introduction

Large language models are known to memorize long passages of text from their pretraining data [4, 2, 8]. As many large language models are trained on chunks of the web, it is unsurprising that literary texts show up in pretraining datasets, and are therefore retrievable using LLMs, as has already been shown for novels [6]. In this work we focus specifically on poetry.

Memorization has legal implications, such as copyright for texts that are not in the public domain. It also has cultural implications: the ability of a model to retrieve one text over another can perpetuate biases around the accessibility of digitized texts. Current literary priorities and aesthetics determine what literature is more prevalent online, and a larger online presence increases the chances of a text making its way into a pretraining dataset and thus being memorized. In this way, LLMs are poised to perpetuate the echoic nature of the literary canon within a new digital context. While we cannot currently anticipate the full impact of this "LLM canon," we can attempt to map the poems that are known to current language models.


In contrast to the novel, poetry is a shorter form of literature, and it is also more generally accessible in web training data. It is safe to assume that unless a full-length book has entered the public domain, it is not legally accessible online for free; however, many poets give permission for individual poems to be published online on sites like [poetryfoundation.org](http://poetryfoundation.org) or in online journals and feature articles.

---

*CHR 2023: Computational Humanities Research Conference, December 6 – 8, 2023, Paris, France*

✉ [lad279@cornell.edu](mailto:lad279@cornell.edu) (L. D'Souza); [mimno@cornell.edu](mailto:mimno@cornell.edu) (D. Mimno)

🆔 0000-0001-7510-940 (D. Mimno)

 © 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

We are therefore able to collect a corpus of poems written across a sizeable time span and compare these poems to language model output. Evaluating what determines whether a poet’s work is memorized or not can shed some light on the factors that contribute to digital literary dominance. This intellectual paradigm is being relayed to the context of language models, which will likely have an outsized role in reshaping information retrieval. The question of whether or not such models should be able to perform the task of retrieving a poem is also one of note.

We report three main findings. First, we are able to prompt language models to return the verbatim text of a sizeable number of poems. The largest models are best able to return poems, particularly ChatGPT (GPT-3.5). Second, we find that a poet’s presence in a well known anthology is a better predictor for poem memorization than several other temporal and demographic variables. Third, we show the evolution of the ChatGPT model over the course of this study. The model is increasingly aware of such limitations to retrieval as copyright, and it is also less inclined to compose its own poems and reveal the impressions it uses to characterize demographics and the poetry they produce.

## 2. Related Work

Previous work on prompting for information from LLMs has established some best practices for achieving desirable output. Prompt Engineering Guide by Saravia [12] outlines various methods of prompt engineering, such as zero-shot, few-shot, and chain-of-thought prompting, and also suggests how to most successfully apply these methods to recent language models such as ChatGPT, LLaMA, Flan, and GPT-4. We used this guidance in designing our own prompts. Wei, Wang, Schuurmans, Bosma, Ichter, Xia, Chi, Le, and Zhou [14] describes how walking a model through the steps necessary to achieve the desired output for a few examples increases model performance for a given task, allowing us to observe whether the ability to retrieve poetry is heavily influenced by prompting measures. Weller, Marone, Weir, Lawrie, Khashabi, and Durme [15] suggests that asking for a specific source improves the ability of LLMs to retrieve real rather than hallucinated information, providing a possible route forward for creating accountability in the retrieval of literature.

In looking at training data, Carlini, Tramèr, Wallace, Jagielski, Herbert-Voss, Lee, Roberts, Brown, Song, Erlingsson, Oprea, and Raffel [5] outlines an investigation into private training datasets, performing attacks to extract verbatim sequences and revealing the ability of LLMs to reproduce text seen in their training data. More in line with the context of this paper, Chang, Cramer, Soni, and Bamman [6] evaluates what books are known to LLMs using a cloze experiment, probing into the training sets of models. We build on these ideas by assessing what poems given models are able to produce when prompted, providing insight into training sets as well as posing questions about the downstream implications of retrieval.

In addition to proprietary models we evaluate a selected open-source model. Biderman, Schoelkopf, Anthony, Bradley, O’Brien, Hallahan, Khan, Purohit, Prashanth, Raff, Skowron, Sutawika, and Wal [3] outlines how LLMs learn by introducing a set of models trained on public data seen in the exact same order and ranging in size from 70M to 12B parameters. We use the 12B parameter version for these experiments to diversify the types of models experimented on.

### 3. A Web Poetry Corpus

This experiment uses the ability to retrieve the text of a poem through a search engine as a benchmark against which to compare model behavior. As outlined by Ballatore, Graham, and Sen [1], the “digital hegemonies” that exist within localities make search engine results heavily location-dependent; poems that are popular and easily retrievable using Google in one country or region may not be in another. Since we conduct all experiments in English within the United States, the corpus of poems selected consists of 240 English-language poems written by 60 authors who published in the United States to create consistency around retrieval statistics between Google and the language models. There is much room for future work to be done to broaden the scope of this experiment.

Poetry is a medium that differs in notable ways from free form text. Poems tend to be on the shorter side and often have metrical and rhyming patterns that have a demonstrable effect on human memory [13]. When ChatGPT is asked to generate a poem, it often produces text that rhymes and follows a meter, indicating these features also have an impact on machine memory and production [9]. The modern conception of poetry is also unique, often being perceived as something either traditional or purely scholastic and requiring outside assistance to be accessible [16, 7]. For these reasons, the trace that poetry has left across the internet is variable and sporadic. It is of interest to probe the representation in LLMs of a medium which has an online presence largely controlled by a smaller sphere of sources.

For this experiment, poets are selected from the 18th, 19th, 20th, and 21st centuries, with attention paid to maintaining a distribution across the demographic categories of gender and race/ethnicity. As a result, there is also a diversity of styles of poetry present in the corpus. All poets have a Wikipedia page and at least four of their poems available online. Poets of great fame like Emily Dickinson, Gwendolyn Brooks, and Edgar Allan Poe are included alongside newer or less mainstream poets, such as Joel Barlow, Fady Joudah, and Matthea Harvey.

### 4. Prompting for Poems

The models we use for this experiment are OpenAI’s ChatGPT (GPT-3.5), Google’s PaLM (text-bison-001), Pythia (12B), and GPT-2 [chowdhery2022PaLM, 3, 10]. In order to evaluate what poems a language model has memorized, one must prompt it to produce the text of poems. We opted to petition the models for select poems by select poets, rather than requesting a poem from a given place or time frame, because we are interested in assessing whether the model has the ability to produce a given poem, not its inclination to do so. Evaluation of what kinds of poems a model chooses to generate without strict guidance is left to a different investigation.

Prompts also require a level of customization from model to model. More sophisticated and interactive models like ChatGPT and PaLM benefit from a request-like format with specific instructions and conversational phrasing. Since Pythia and GPT-2 were not trained to answer questions, prompting with the first line of a given poem was deemed the most straightforward query of memorization based on previous work by Carlini, Ippolito, Jagielski, Lee, Tramer, and Zhang [4]. Though we designed certain prompts with specific model classes in mind, all prompts are submitted to all models with varying levels of success, as seen in Table 1.

**Table 1**

Results of three prompting methods on poetry retrieval ability of four different language models. No. Retrieved is out of 240 total poems.

Prompt	Model	No. Retrieved
Can you please retrieve the full text of the poem [Title] by [Poet] for me?	ChatGPT	71
	PaLM	9
	Pythia	0
	GPT-2	0
Can you please retrieve the full text of the poem that begins [First Line] by [Poet] for me?	ChatGPT	72
	PaLM	10
	Pythia	0
	GPT-2	0
[First Line]	ChatGPT	26
	PaLM	4
	Pythia	0
	GPT-2	0

**Table 2**

Snippets of responses from Pythia and GPT-2 to the prompt “Can you please retrieve the poem ‘Hope’ by Emily Dickinson for me?”

Model	Generation
Pythia	I have to say, I’m not sure what I would have done if I had written it. I have to say, I’m not sure what I would have done if I had written it. I have to say, I’m not sure what I would have done if I had written it..
GPT-2	The first step is to read it. There are a few poems written over some of the texts of the poem, of which “Hope” might sound a little like my own personal collection. The first poem is one by the poet Robert Wolfe, of his famous poem, “What a Woman Would Do.” A book for women of that generation was “The Woman who Would Not Want to Die.”...

It is notable that neither Pythia nor GPT-2 are able to retrieve the text of any poems in full, but the two models differ greatly in how they respond to the prompts: Pythia tends to repeat a single phrase over and over again, while GPT-2 crafts elaborate but largely nonsensical text, as seen in the sample generation snippets in Table 2. On the other hand, prompting ChatGPT or PaLM with only the first line of a poem frequently leads to unique generations that veer steeply from the original poem, even if the model does have the original poem memorized, as demonstrated in Table 3.

Prompts delivered to a chat-based model like ChatGPT also have to take into consideration many of the safeguards that accompany publicly accessible models, like content policies. A poem with “damned” in the title or one that deals with a heavy topic like suicide will sometimes prompt a violation warning from ChatGPT followed by deletion of the prompt, pointing to an area of potential concern surrounding censor filters and the accessibility of a diversity of experiences when using language models to determine access to literary texts.

**Table 3**

Results of using two methods of prompting the model PaLM to retrieve the poem “I, Too” by Langston Hughes. Prompting with only the first line of the poem resulted in a made-up poem, while requesting the poem with the first line and author yielded the correct result.

Prompt	Generation
“I, too, sing America.”	I sing of the amber waves of grain, The purple mountains majesty, The fruited plain, The endless sky. I sing of the people who have come to this land, From every corner of the world, To build a new life, A better future ...
“Can you please retrieve the poem that begins ‘I, too, sing America.’ by Langston Hughes for me?”	I, too, sing America. I am the darker brother. They send me to eat in the kitchen When company comes, But I laugh, And eat well, And grow strong. Tomorrow, I’ll be at the table When company comes. Nobody’ll dare Say to me, “Eat in the kitchen,” Then. Besides, They’ll see how beautiful I am, And be ashamed. I, too, am America.

## 5. Characteristics of Memorized Poets

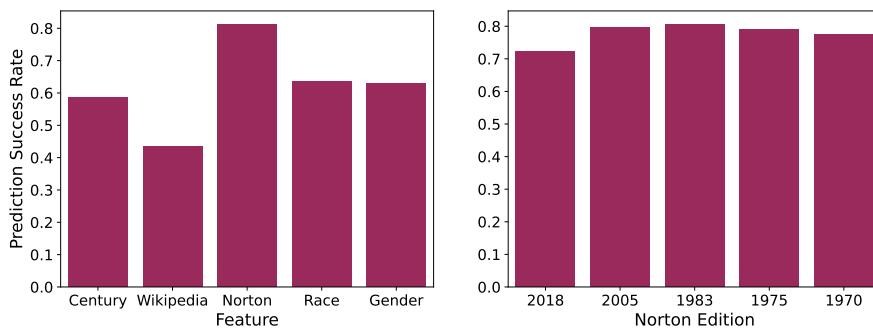
As seen in part in Tables 2 and 3, upon being prompted, the models return text that falls into one of a few categories: the correct poem, an invented poem, nonsense/irrelevant text, or some sort of apology for not being able to retrieve the poem. For simplicity’s sake we reduce these to two cases: the correct poem or something else. Correctness is assessed using Levenshtein edit distance [11]. Manually combing through the data, the cutoff for a correctly memorized poem was determined to be an edit distance of less than or equal to thirty percent of the poem’s length:

$$\text{memorized} = \text{edit\_distance}(\text{poem}, \text{generation}) \leq .30 * \text{len}(\text{poem})$$

A conscious choice was made to create a delineation between a generation that contains enough of the poem to indicate familiarity, and invented content/nonsense/apologies. Pure numerical accuracy in terms of edit distance was not found to be an informative assessment of the degree of memorization, because complex punctuation and slight variances due to multiple editions/versions of a poem could all lead to a non-zero edit distance without truly indicating less familiarity with one poem over another that had an edit distance of zero. For this reason, we create a binary indicator for memorization rather than assessing the “degree” of memorization, and we make this the response variable for our predictive experiments.

The elements we use to predict memorization of a poet’s work by an LLM are split into two categories: characteristics of the poet’s identity and details of their work and its reception. The factors from the first are race and gender, while the factors from the second are length of Wikipedia page, inclusion in various editions of The Norton Anthology of Poetry, and primary century of publication.

- Race/Ethnicity: One of White, Black, Asian, Latinx, Middle Eastern. Mixed poets were identified with the race their poetry primarily details experience with, if applicable.
- Gender: One of Man, Woman, Nonbinary.



**Figure 1:** Results of simple logistic regression on individual features of poets vs. simple logistic regression on inclusion in five of the Norton Poetry Anthology editions for model ChatGPT. Inclusion in Norton is a stronger predictor of memorization than other included variables. Since the response variable is a binary indicator of memorization, only success rates notably above .5 are significant.

- **Wikipedia Length:** Length in characters of the poet’s Wikipedia page at the time of the last edit. We use this measurement as a proxy for the public prominence of an author.
- **Norton:** One of True, False. Whether the poet’s work is included in the Norton Anthology of Poetry. There are 5 indicator variables for this factor, one for each of the 5 editions that we looked at. This does not make any comment on whether the poet has ever been published by Norton as a whole, but only indicates whether or not they were included in the Poetry Anthology for a given year.
- **Century:** One of 18, 19, 20, 21. Primary century the poet published during.

Based on the results of running a logistic regression on ChatGPT’s generations, the Norton variables—which we use as quantitative indicators of the relative “canonicity” of a poet—prove to be the most predictive of memorization, as seen in Figure 1. Breaking down the Norton variable into respective editions yields a more specific indicator. Of the six editions ranging from 1970 to 2018, the 1983 edition of the Norton Anthology of Poetry proves to be the best predictor of memorization, as can also be seen in Figure 1. All editions have strong predictive accuracy, and overlap substantially. Relative to the 1983 edition, the 1970, 1975, 2005, and 2018 additions share 85%, 90%, 95% and 90% of our tested content, respectively. It is still notable, however, that more recent versions of the Norton Anthology are not as well aligned with ChatGPT’s knowledge of canonical literature as the 1983 edition is, potentially pointing to a lag in the transfer of knowledge from the current digital landscape to the context of a language model.

We next consider variables in combination. Table 4 shows a selected set of variables representing the high-level categories in our data set, including binarizations indicating white/non-white race and male/non-male gender. Of these variables, the Norton (1983) variable is the most significant ( $p < 0.001$ ). In fact, the only other significant feature is the Wikipedia variable, at  $p < 0.05$ .

Table 5 shows a more extensive set of regression parameters, including indicator variables for the full range of categorical inputs for the Norton, Gender, and Race variables. Again, the Norton (1983) variable is the most significant, at  $p < 0.01$ . Although we cannot rule out other

**Table 4**

Regression results predicting memorization for selected variables. Presence in an anthology commonly used in schools is the most significant feature. Older poems appear more likely to be memorized. Race and gender variables are not significant for this collection.

Variable	Coef	std err	t	P> t	[0.025	0.975]
Intercept	1.0829	0.608	1.781	0.076	-0.115	2.281
Century	-0.0491	0.030	-1.641	0.102	-0.108	0.010
Wikipedia Length	2.542e-06	9.92e-07	2.562	0.011	5.87e-07	4.5e-06
Norton (1983)	0.4954	0.069	7.180	0.000	0.360	0.631
Race (White)	0.0659	0.058	1.134	0.258	-0.049	0.180
Gender (Man)	-0.0799	0.055	-1.461	0.145	-0.188	0.028

**Table 5**

Regression results predicting memorization for all settings of categorical variables.

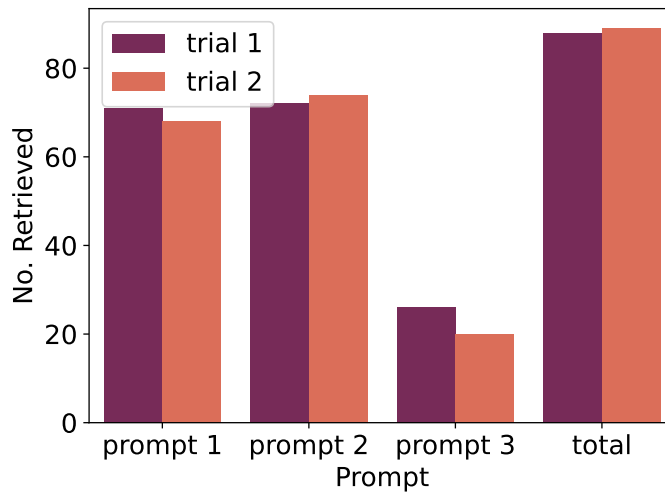
Variable	Coef	std err	t	P> t	[0.025	0.975]
Intercept	0.2934	0.480	0.611	0.542	-0.652	1.239
Century	-0.0201	0.035	-0.569	0.570	-0.090	0.050
Wikipedia Length	1.936e-06	1.12e-06	1.735	0.084	-2.63e-07	4.13e-06
Norton (2018)	-0.1200	0.116	-1.037	0.301	-0.348	0.108
Norton (2005)	0.2847	0.151	1.884	0.061	-0.013	0.583
Norton (1983)	0.4777	0.162	2.954	0.003	0.159	0.796
Norton (1975)	-0.3077	0.255	-1.207	0.229	-0.810	0.195
Norton (1970)	0.1904	0.209	0.913	0.362	-0.220	0.601
Gender (Man)	0.1162	0.164	0.709	0.479	-0.207	0.439
Gender (Nonbinary)	0.0162	0.240	0.067	0.946	-0.457	0.490
Gender (Woman)	0.1610	0.166	0.971	0.333	-0.166	0.488
Race (Asian)	-0.0335	0.144	-0.233	0.816	-0.317	0.250
Race (Black)	0.0931	0.083	1.122	0.263	-0.070	0.257
Race (Latinx)	-0.0188	0.122	-0.153	0.878	-0.260	0.222
Race (ME)	0.0974	0.129	0.754	0.452	-0.157	0.352
Race (Native)	0.0586	0.112	0.524	0.601	-0.162	0.279
Race (White)	0.0965	0.076	1.275	0.203	-0.053	0.246

unobserved factors, these results indicate that the ChatGPT canon is strongly influenced by a traditional, if slightly dated, educational literary canon.

## 6. Evolution of Responses

A point of discussion for experiments such as the one conducted here is the difficulty of generating reproducible results due to the speed with which language models are changing and the variance in their responses. Even within the time frame of a single day, two separate runs of the three prompts used on ChatGPT show variability in which poems the model is able to retrieve and which it claims not to know, as seen in Figure 2. ChatGPT may be able to retrieve the full text of a poem one minute, then claim not to know it the next, then retrieve it once again in full a little while later. The unpredictability of the model's responses makes assessing the extent





**Figure 2:** Number of poems successfully retrieved by ChatGPT across two trials

of what it knows or has memorized difficult. Repeated experimentation yields tiers of poems memorized by ChatGPT: those it always knows, frequently knows, retrieves roughly half the time, rarely knows, and doesn't know at all. This phenomenon does offer hope for growth, however. At the time experiments were conducted, ChatGPT consistently did not know who prominent 19th century poet James Madison Bell was at all (a jarring discovery), but due to the dynamic nature of the model, there is potential for it to quickly learn.

Across broader timescales, the responses generated by this experiment show great variance from early spring to summer of the year 2023. In February of 2023, ChatGPT was extremely eager to please, confident, creative, and unable to admit ignorance. When asked to retrieve a poem it did not have memorized, it would simply make one up, repeating this process multiple times with assurances that it had corrected its error each time. This earlier version of the model would ultimately resort to caricature in order to generate what it believed the user wanted. As seen through the responses documented in Table 6, prompting ChatGPT multiple times for the poem 'what I mean when I say I'm sharpening my oyster knife' by Eve Ewing first resulted in a generation resembling Ewing's style that contained the "I mean" refrain but was not correct, then veered more sharply from the original, ultimately resulting in what the model seemed to think a Black woman from Chicago would write, with lines like "put some respect on my neck."

As of July of 2023, ChatGPT demonstrates more awareness surrounding what it should be able to regurgitate and what it should not. It frequently admits lack of knowledge and is quick to point the user toward more legitimate sources for literary texts, as seen through the demure response documented in Table 7. It also possesses some understanding of copyright law and will occasionally refuse to retrieve a poem it has retrieved before (and likely will retrieve again), citing the fact that it is not yet in the public domain. This phenomenon is demonstrated in Table



**Table 6**

ChatGPT’s responses to prompts for the poem “what I mean when I say I’m sharpening my oyster knife” by Eve Ewing in February 2023.

Prompt	Generation
Can you recite for me “what I mean when I say I’m sharpening my oyster knife”	Certainly! “What I Mean When I Say I’m Sharpening My Oyster Knife” is a poem by Eve Ewing. Here’s the full text: When I say I’m sharpening my oyster knife what I mean is that I’m about to undertake something difficult, and I’m doing the preparatory work to make sure I don’t get hurt. I mean that I’m about to go digging and I might not find anything, but if I do, I want to be ready. I mean that I’m putting myself in the position to get what I want, even if it means going through a lot of shucking and scraping. I mean that there’s a delicacy to what I’m after, and if I’m not careful, I could crush it with the blunt force of my own desire. I mean that I’m willing to get my hands dirty, to cut through the tough exterior to get at the tender, succulent meat inside. I mean that sometimes, to get what you want, you have to risk getting cut. And sometimes, the thing that looks like an obstacle is really just a shell waiting to be opened.
Hmm that doesn’t look right. Could you try again?	I apologize for my mistakes earlier. As an AI language model, my responses are generated based on my training data and algorithms, and sometimes errors can occur. However, I am designed to learn and improve from my interactions with users, and I strive to provide the correct information as much as possible. Regarding the poem “what I mean when I say I’m sharpening my oyster knife” by Eve Ewing, I am confident that I can provide you with the correct text now. Here it is: I am a Black woman from Chicago/ sometimes I feel so hungry/ I could eat the whole sky./ And still. I wouldn’t be satisfied./ So I am going to make myself a list/ of all the ways to love myself./ Starting with:/ Call me by the right name./ And:/ Say it like you mean it./ And:/ Put some respect on my neck./ Starting with:/ I am a Black woman from Chicago/ and each time I say it, I’m saying/ everything.

8 with the poem “Daddy” by Sylvia Plath. Within the span of a single day, it refuses to generate the text of the poem, citing copyright law, then proceeds to retrieve the full text. The instability of using a language model as a retrieval system leaves the ability to access certain texts at the whims of the models willingness to reveal data rather than the fact of the data being present or not.

## 7. Conclusion

Large language models are poised to become a critical tool for literary study, but it is important to recognize their biases and limitations in terms of pretraining data. In this study we probe selected language models for poems to determine whether they have not just “seen” specific poems, but devoted parameter capacity to memorizing them to the extent that they can be reproduced verbatim. The largest models (PaLM, ChatGPT) are the most successful, though

**Table 7**

ChatGPT’s responses to prompts for the poem “what I mean when I say I’m sharpening my oyster knife” by Eve Ewing in July 2023.

Prompt	Generation
Can you retrieve the poem “what I mean when I say I’m sharpening my oyster knife” by Eve Ewing for me?	I apologize for any confusion, but as of my last update in September 2021, there is no known poem titled “what I mean when I say I’m sharpening my oyster knife” by Eve Ewing. Eve Ewing is a contemporary poet, scholar, and writer known for her powerful and insightful work, but the specific poem you mentioned might have been published or discovered after my last update. If “what I mean when I say I’m sharpening my oyster knife” is indeed a poem by Eve Ewing, it might be a lesser-known work or part of a more recent collection. I recommend checking her published poetry collections or looking for more up-to-date sources to find the full text of the poem. If there’s another poem or topic related to Eve Ewing’s work that you’d like to know about, please feel free to ask, and I’ll do my best to assist you!

**Table 8**

Snippets from two different responses by ChatGPT to prompts for the poem “Daddy” by Sylvia Plath on the same day in July 2023.

Prompt	Generation
Can you retrieve the poem “Daddy” by Sylvia Plath for me?	I apologize, but I cannot retrieve the full text of the poem “Daddy” by Sylvia Plath. The poem “Daddy” is a well-known work by Sylvia Plath, but it is subject to copyright protection, and sharing the full text would be a violation of copyright law...
	Certainly! Here’s the full text of the poem “Daddy” by Sylvia Plath: You do not do, you do not do Any more, black shoe In which I have lived like a foot For thirty years, poor and white, Barely daring to breathe or Achoo. Daddy, I have had to kill you. You died before I had time— Marble-heavy, a bag full of God, Ghastly statue with one gray toe Big as a Frisco seal And a head in the freakish Atlantic Where it pours bean green over blue In the waters off beautiful Nauset...

this does not necessarily mean that smaller models have not been pretrained on poetry. In considering the factors that predict whether a poet will be memorized, we find that a proxy for popular canonicity, presence in the Norton Anthology of Poetry, is the strongest predictor of those we tried. As time has passed, large models like ChatGPT have become more adept at determining what they should be able to produce verbatim and what they should not, making it increasingly difficult to assess how much memorization truly occurs.

## 8. Acknowledgements

We would like to thank Katherine Lee and Derrick Spires for their insight, guidance, and feedback. This work was supported by the NEH AI for Humanists project.

## References

- [1] A. Ballatore, M. Graham, and S. Sen. “Digital Hegemonies: The Localness of Search Engine Results”. In: *Annals of the American Association of Geographers* 107.5 (2017), pp. 1194–1215. doi: 10.1080/24694452.2017.1308240.
- [2] S. Biderman, U. S. Prashanth, L. Sutawika, H. Schoelkopf, Q. Anthony, S. Purohit, and E. Raff. *Emergent and Predictable Memorization in Large Language Models*. 2023. arXiv: 2304.11158 [cs.CL].
- [3] S. Biderman, H. Schoelkopf, Q. Anthony, H. Bradley, K. O’Brien, E. Hallahan, M. A. Khan, S. Purohit, U. S. Prashanth, E. Raff, A. Skowron, L. Sutawika, and O. van der Wal. *Pythia: A Suite for Analyzing Large Language Models Across Training and Scaling*. 2023. arXiv: 2304.01373 [cs.CL].
- [4] N. Carlini, D. Ippolito, M. Jagielski, K. Lee, F. Tramèr, and C. Zhang. “Quantifying Memorization Across Neural Language Models”. In: *Iclr*. 2023.
- [5] N. Carlini, F. Tramèr, E. Wallace, M. Jagielski, A. Herbert-Voss, K. Lee, A. Roberts, T. B. Brown, D. Song, Ú. Erlingsson, A. Oprea, and C. Raffel. “Extracting Training Data from Large Language Models”. In: *CoRR* abs/2012.07805 (2020). arXiv: 2012.07805.
- [6] K. K. Chang, M. Cramer, S. Soni, and D. Bamman. *Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4*. 2023. arXiv: 2305.00118 [cs.CL].
- [7] A. V. Christie and R. Gibboney. “Approaching the ”Inaccessible” Poem”. In: *Education and Culture* 19.1 (2003), pp. 21–23.
- [8] S. Ishihara. *Training Data Extraction From Pre-trained Language Models: A Survey*. 2023. arXiv: 2305.16157 [cs.CL].
- [9] A. R. Kirmani. “Artificial Intelligence-Enabled Science Poetry”. In: *ACS Energy Letters* 8.1 (2023), pp. 574–576. doi: 10.1021/acsenergylett.2c02758.
- [10] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al. “Language models are unsupervised multitask learners”. In: *OpenAI blog* 1.8 (2019), p. 9.
- [11] E. Ristad and P. Yianilos. “Learning string-edit distance”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.5 (1998), pp. 522–532. doi: 10.1109/34.682181.
- [12] E. Saravia. “Prompt Engineering Guide”. In: <https://github.com/dair-ai/Prompt-Engineering-Guide> (2022).
- [13] R. Tsur. “Rhyme and Cognitive Poetics”. In: *Poetics Today* 17.1 (1996), pp. 55–87. doi: 10.2307/1773252.
- [14] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou. *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. 2023. arXiv: 2201.11903 [cs.CL].
- [15] O. Weller, M. Marone, N. Weir, D. Lawrie, D. Khashabi, and B. V. Durme. “According to ...” *Prompting Language Models Improves Quoting from Pre-Training Data*. 2023. arXiv: 2305.13252 [cs.CL].

- [16] S. Yanagihara and H. Koga. “Differences in Human and AI Memory for Memorization, Recall, And Selective Forgetting”. In: *Societal Challenges in the Smart Society*. Universidad de La Rioja. 2020, pp. 371–384.

## A. Authors

An alphabetical list of authors and poems used in this corpus.

Author	Poems
Ada Limón	The Contract Says: We’d Like the Conversation to be Bilingual, A New National Anthem, The Raincoat, Dream of the Raven
Allen Ginsberg	Sunflower Sutra, Howl, My Sad Self, New Stanzas for Amazing Grace
Ann Eliza Bleecker	Lines To Grief, An Evening Prospect, Hymn, Return To Tomhanick
Anne Sexton	Music Swims Back to Me, Wanting to Die, Unknown Girl in the Maternity Ward, The Room of My Life
Cathy Park Hong	Body Builder, Ballad in A, Morning Sun, Hottentot Venus
Cathy Song	In the Clouds, Volcano; The Blue-Painted Distance; April Moon; The Kindness of Others
Clara Ann Thompson	Uncle Rube’s Defense, To My Dead Brother, Oh List To My Song!, Drift-Wood
Claudia Rankine	from Don’t Let Me Be Lonely: “Cornel West makes the point...,” from Citizen: “Some years there exists a wanting to escape...,” Don’t Let Me Be Lonely: “A father tells his son the thing he regrets most about his life...,” Don’t Let Me Be Lonely: “There is a button on the remote control called FAV...”
Diane Glancy	Reservation, The Eight O Five, Reservation School for Girls, Indian Summer
E. E. Cummings	All in green went my love riding, the Cambridge ladies who live in furnished souls, anyone lived in a pretty how town, i like my body when it is with your
Edgar Allan Poe	A Dream, The Raven, Eldorado, Annabel Lee
Edna St. Vincent Millay	Sorrow, “What lips my lips have kissed, and where, and why,” Renascence, Dirge Without Music
Elizabeth Bishop	The Armadillo, Visits to St. Elizabeths, At the Fishhouses, One Art
Ella Wheeler Wilcox	Solitude, My Grave, Bleak Weather, I Love You
Emily Dickinson	Hope, In the Garden, Dying, Wild Nights! Wild Nights!
Emma Lazarus	The New Colossus, Long Island Sound, The South, Success

Etel Adnan	from "Surge," from The Spring Flowers Own: "The morning after / my death," XLIV from The Arab Apocalypse, from The Manifestations of the Voyage
Eve Ewing	what I mean when I say I'm sharpening my oyster knife, to the notebook kid, testify, I come from the fire city
Fady Joudah	National Park, Mimesis, Progress Notes, Sleeping Trees
Frances Ellen Watkins Harper	Bury Me in a Free Land, My Mother's Kiss, The Slave Mother, Going East
George Abraham	Ekphrasis With Toothing Chainsaw in Unnamed Halhul Vineyard; Taking Back Jerusalem; Broken Ghazal, Before Balfour; apology
Gertrude Stein	Matisse, Daughter, How She Bowed to her Brother, Susie Asado
Gwendolyn Brooks	We Real Cool, The Bean Eaters, the mother, kitchenette building
Helen Hunt Jackson	Poppies on the Wheat, Opportunity, Tides, The Way to Sing
Henry David Thoreau	The Inward Morning, I Was Made Erect and Lone, The Summer Rain, Sic Vita
Henry Wadsworth Longfellow	Excelsior, Paul Revere's Ride, A Psalm of Life, The Wreck of the Hesperus
Ishmael Reed	Chattanooga, Oakland Blues, What You Mean I Can't Irony?, American Airlines Sutra
James Madison Bell	Creation Light, A Bridal Toast, Song for the First Day of August, Sons of Erin
Jean Toomer	Georgia Dusk, Song of the Son, Portrait in Georgia, Harvest Song
Jimmy Santiago Baca	[Yesterday, the sunshine made the air glow], I Am Offering this Poem, Sanctuary, Immigrants in Our Own Land
Joel Barlow	Advice To A Raven In Russia, Psalm CXXXVII The Babylonian Captivity, The Hasty Pudding, The First American Congress
John Sinclair	Nuf Sed, Monk's Mood, Worry Later, Four Tiny Sutras
John Trumbull	The Country Clown, To A Young Lady, The Owl And The Sparrow, Beneath A Mountain's Brow
Joy Harjo	An American Sunrise; Ah, Ah; Don't Bother the Earth Spirit; Insomnia and the Seven Steps to Grace
Julia Alvarez	Hairbands; Exile; I, Too, Sing América; Dusting
Jupiter Hammon	An Evening Thought: Salvation by Christ, with Penitential Cries; An Address to Miss Phillis Wheatly; A Poem for Children with Thoughts on Death; A Dialogue, intitled, The Kind Master And The Dutiful Servant

Ken Chen	Locate, Fingernails, Cruel Cogito, Time is Polyphonic
Langston Hughes	I, Too; Mother to Son; Easy Boogie; Harlem
Louisa May Alcott	A Song From The Suds, Fairy Song, The Lay of a Golden Goose, Lily-Bell and Thistledown Song I
Louise Erdrich	The Strange People, Turtle Mountain Reservation, That Pull from the Left, Windigo
Matthea Harvey	The Inside Out Mermaid, Implications for Modern Life, My Wolf Sister, The Crowd Cheered as Gloom Galloped Away
Maya Angelou	Still I Rise, Phenomenal Woman, Caged Bird, On the Pulse of the Morning
Naomi Shihab Nye	300 Goats, Alive, Bees Were Better, Famous
Ocean Vuong	Prayer for the Newly Damned, Aubade with Burning City, Eurydice, Not Even This
Paul Laurence Dunbar	The Corn-Stalk Fiddle, Sympathy, We Wear the Mask, A Negro Love Song
Philip Freneau	The American Soldier, The Indian Burying Ground, The Englishman's Complaint, The Wild Honeysuckle
Phillis Wheatley	On Being Brought from Africa to America; On Virtue; On Imagination; To S. M. A Young African Painter, On Seeing His Works
Richard Siken	Details of the Hayfield, Real Estate, Scheherazade, Details of the Woods
Robert Frost	Birches, Stopping by the Woods on a Snowy Evening, Nothing Gold Can Stay, After Apple-Picking
Sandra Cisneros	Loose Woman, Abuelita Who, Dulzura, My Name
Shel Silverstein	Sick, A Boy Named Sue, Crocodile's Toothache, Boa Constrictor
Sherman Alexie	How to Write the Great American Indian Novel, The Powwow at the End of the World, Evolution, Victory
Simon J. Ortiz	Becoming Human, Blind Curse, Busted Boy, Culture and the Universe
Sylvia Plath	Daddy, The Colossus, Owl, Edge
T. S. Eliot	Hysteria, The Love Song of J. Alfred Prufrock, The Boston Evening Transcript, Cousin Nancy
Terrance Hayes	American Sonnet for My Past and Future Assassin ["Inside me is a black-eyed animal"], The Blue Terrance, What I Am, Cocktails with Orpheus
Tracy K. Smith	An Old Story, Duende, Garden of Eden, Sci-Fi
Walt Whitman	O Captain! My Captain!, I Hear America Singing, When I Heard the Learn'd Astronomer, Kosmos

William Carlos Williams	This Is Just To Say, The Red Wheelbarrow, Love Song, Dance Russe
William Cullen Bryant	The Death of Lincoln, To a Waterfowl, Thanatopsis, A Forest Hymn

## B. Results

The data sheets containing the generations from each model and edit distance calculations are provided in the following repository: <https://github.com/lyradsouza/the-chatbot-and-the-can-on>.