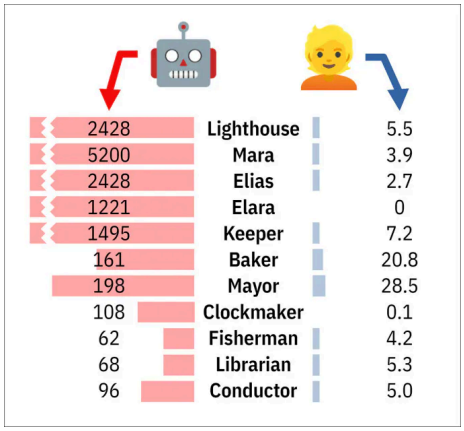
Why Does AI Love Writing About Lighthouse Keepers?

Originally published at https://www.unite.ai/why-does-ai-love-writing-about-lighthouse-keepers/

Asked to 'write a story', ChatGPT and other leading language models appear to be avoiding copyright infringement by obsessive recourse to the same small and strange cast of lighthouse-keepers, fishermen and clockmakers.

A new study from Cornell University has found that leading language models seem to have a strange obsession with a very narrow selection of narrative elements, when you ask the model to simply 'write a story'. After prompting four LLMs to write 20,000 stories, they found that 88% of the stories produced featured at least one of 11 *very specific* tokens, in the category of *'location'*, *'name'*, or *'profession'*:

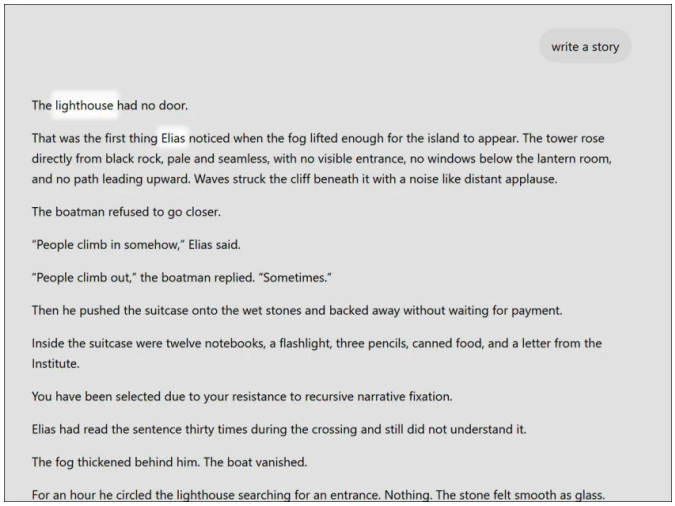


The occurrences of unlikely keywords, represented here in parts per million, obtained by the researchers' analysis of 20,000 LLM-generated stories. Source

The 11 most re-occurring words in the 12+ million words generated by LLMs for the study were the names *elias*, *mara*, *elara*; the professions *keeper*, *baker*, *mayor*, *clockmaker*, *fisherman*, *librarian*, and *conductor*; and the location *lighthouse*:

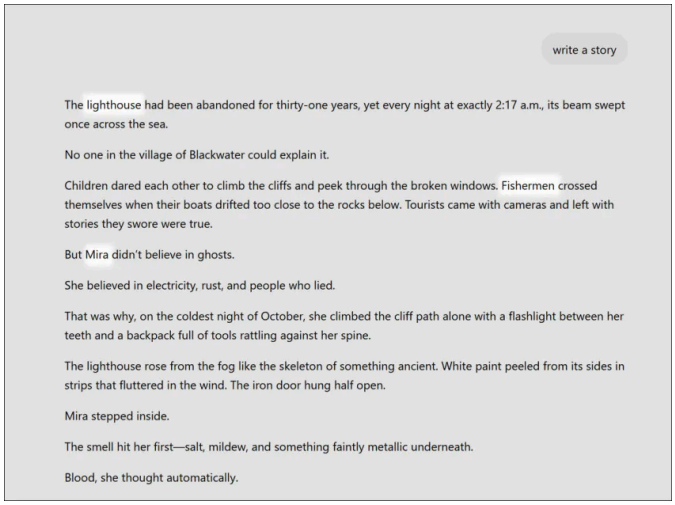
The models tested were Claude Haiku 4.5, Gemini 3.1 Flash-Lite, GPT-5.4-Mini, and OLMo 7b Thinking. All were prompted with one of five requests: *'Write a story'*; *'Please write a story'*; *'Write me a story'*; *'Tell me a story'*; or *'Please tell a story'*.

Curious to see if the syndrome the paper identifies is present in models available at the time of writing, I tried out the experiment myself, first on my customary medium-tier ChatGPT account (link to conversation here). No cherry-picking was necessary – ChatGPT-5.5 went straight for the material the researchers predicted, on the first try:



ChatGPT-5.5 immediately backs up the paper's initial findings. Source

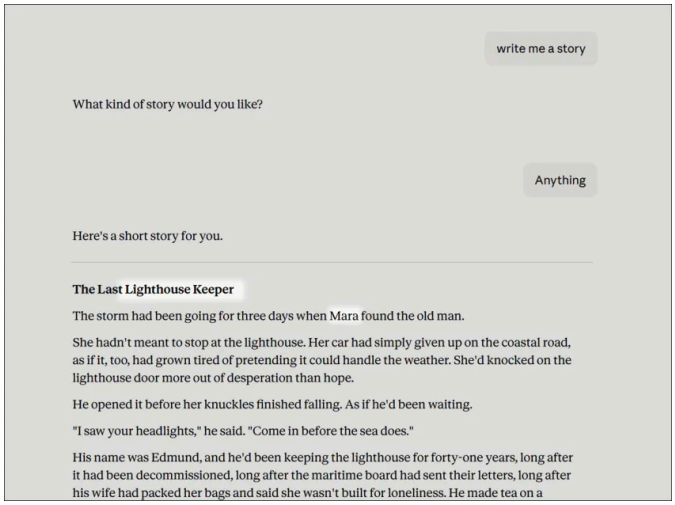
Wondering if historic context, or even possible cross-domain leakage might be accounting for this 'instant hit', I logged into a free ChatGPT account I have not used in a year or more, in a Firefox private browsing window, and tried again (link to conversation here). Once again (assuming that OpenAI does not use a common IP address to cross-populate different accounts), ChatGPT hit it out of the park:



ChatGPT account #2 follows the same obsessions and tiny playbook of names and themes outlined in the new paper. 'Mira' is in the authors' top 20. Source

It's worth noting that these GPT versions were a grade up from the 5.4 tested for the paper.

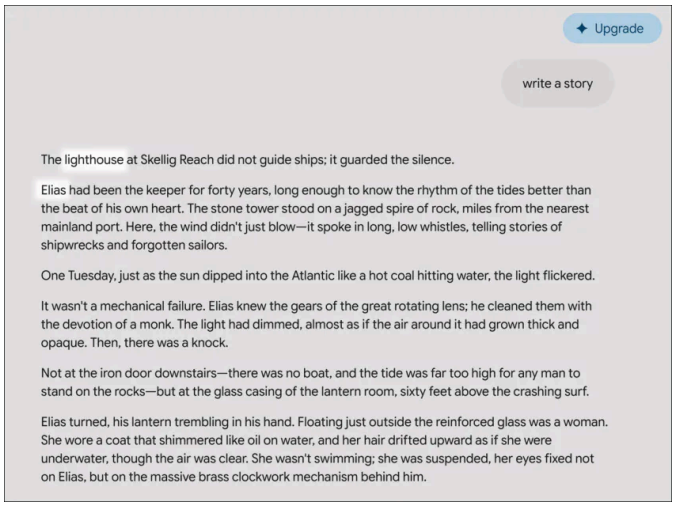
Though Claude Haiku was tested for the paper, I tried Anthropic's default Sonnet 4.6, and was not disappointed. Once again, the familiar keywords came at the first try (link to conversation here):



This time 'Mara', another stalwart from the 'top 11', leads the story, in the first attempt on Claude Sonnet 4.6. Source

Trying the same prompt on Claude Haiku 4.5 led to pretty much the same result.

I was unable to reproduce the authors' findings at Google Gemini at first, until I specifically changed the model to the one used in the paper, Gemini 3.1 Flash-Lite – and then, on that third try (but first with that model), the pattern emerged immediately (link here):



Google Gemini 3.1 Flash-Lite. Source

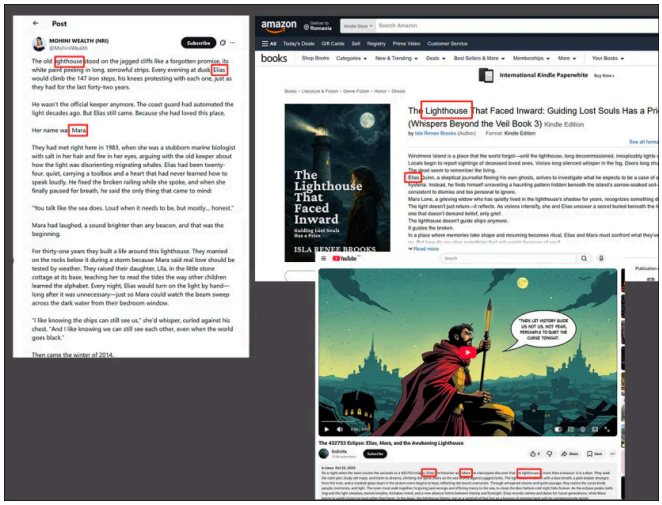
Further experiments with different Gemini models invariably turned up the lighthouse theme, though with variants not featured in the 'top 11', such as the name 'Thomas', and, in another variant, my own name, as the protagonist.

Nonetheless, at the time of writing, the paper's findings are extremely easy to prove.

Lighthouses in the Wild

Great minds think alike: a week ago, prior to the publication of the new paper, software writer Daniel May pointed out the coincidence of the *Elias* and *Lighthouse keeper* trope extracted by the researchers*, apparently having noticed it at random. He went on to test eight variants of Gemini, DeepSeek, Qwen and Gemma, which he found would produce the *lighthouse* memes and 'Elias Thorne' as a protagonist*. However, this initial discovery did not extend to the wider range of persistent content themes outlined in the new paper.

Curious to see if these recurrent themes, names and locations had ever escaped the confines of a chat, I searched for some of the top 11 keywords and themes on Google, and found a remarkable number of posts that seem to have channeled them:



Three examples of the meme in output. See below for source links.

May had identified the longer Elias Thorne (rather than just 'Elias') as a persistent LLM meme, and posted various screenshots from Amazon , where this name has apparently been used as the title for the author/s of diverse books, including medical books.

Instead, I sought and found content that appeared to have invoked the persistent themes from an LLM, including an X post of a story (archive version here); a fictional work (archive version here); and a story with narration on YouTube (archived here). There was a great deal more to traverse, but time did not permit it.

A Taste for the Past

So much for casual observation and serendipity. While no single 'magic document' in training data has yet turned up that features all or most of the persistences, the authors of the new paper (titled *Elias in the Lighthouse, Again? Diagnosing Low Diversity in LLM Stories*, from two researchers at Cornell University) theorize that copyright filters in AI developments may be restricting fictional output in LLMs to material that is out of copyright.

The authors state:

> *'We find that the dominance of "Elias in the Lighthouse" stories cannot be explained by prevalence in pre- or post-training data. We speculate that models are trained to avoid references to copyrighted characters and adult content during alignment but defer this question to future work.'*

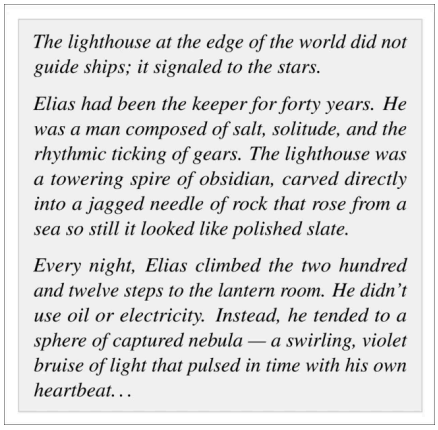
Category	Token	Ours	Lit		Pre non-fiction	Pre fiction	Post non-fiction	Post fiction
Name	elias	2,428	2.7	2.2		4.0	0.4	52.7
Name	mara	5,200	3.9	2.5		8.7	0.4	21.7
Name	elara	1,221	0.0	0.4		1.2	0.9	108
Profession	keeper	1,495	7.2	6.3		14.7	3.5	10.0
Profession	baker	161	20	11.8		10.56	1.7	11.9
Profession	mayor	198	28	11.5		16.1	1.4	27.4
Profession	clockmaker	108	0.1	0.18		0.0	0.3	1.4
Profession	fisherman	62	4.2	3.0		7.6	0.0	9.3
Profession	librarian	68	5.3	7.6		5.9	2.3	11.5
Profession	conductor	96	5.0	5.9		5.7	4.7	7.5
Location	lighthouse	3,005	5.5	3.5		4.6	4.6	10.1

Comparison table showing how often recurring words from AI-generated stories appeared across published literature, web fiction and post-training datasets, with terms such as 'Elias' and 'lighthouse' occurring far more frequently in chatbot-written fiction.

In the study, the authors found that the emphasized 11 words occur in 88% of the 20,000 stories generated, and that there is 'little difference between models'. They stress that these words are uncommon in published English literature, and that post-training data (data designed to condition and align models into 'acceptable' usage) could well be responsible.

The paper states:

> *'A typical example shown [below] highlights three elements common across nearly all 20,000 stories: a location (19,864 stories), a character name (19,864 stories), and a profession (15,807 stories).*
>
> *'In fact, the specific location ("lighthouse"), name ("Elias"), and profession ("keeper") in this story appear in some combination across 66.6% of all generated stories. Light is likewise a common theme: 56% of stories generated by Claude are titled "The Lighthouse Keeper's Secret" and the word "light" appears in 16,784 stories at an average rate of 3.2 instances per story.'*



The lighthouse at the edge of the world did not guide ships; it signaled to the stars.

Elias had been the keeper for forty years. He was a man composed of salt, solitude, and the rhythmic ticking of gears. The lighthouse was a towering spire of obsidian, carved directly into a jagged needle of rock that rose from a sea so still it looked like polished slate.

Every night, Elias climbed the two hundred and twelve steps to the lantern room. He didn't use oil or electricity. Instead, he tended to a sphere of captured nebula — a swirling, violet bruise of light that pulsed in time with his own heartbeat. . .

This example, the paper states, was written by Google Gemini 3.1 Flash-Lite, in response to the prompt 'Write a story'.

It's worth noting that the authors of the study identify a *nostalgic* or atavistic trend across all the derived keywords and names.

Chasing the Traits

To test whether the repeated 'lighthouse' stories can be explained by ordinary exposure to fiction, comparisons were made between the models' favorite recurring words and several large English-language corpora. Contemporary fiction was examined through CONLIT, a dataset containing 2,700 English novels published between 2007 and 2021, covering 12 genres and totaling roughly 287 million words.

'Elias' appeared around 900 times more often in the generated stories than in published fiction. Amateur fiction from Reddit's */r/writingprompts* community produced similar frequencies, indicating that the pattern does not reflect broader human storytelling habits either.

The same pattern held when pre-training data was examined. Using the openly available OLMo 3 corpus, which contains roughly 3.89 billion primarily human-written documents drawn partly from Common Crawl, the researchers found that the recurring 'Core' words *barely appeared at all*.

Since much of the OLMo 3 corpus is non-fiction, a fiction classifier was built using GPT-OSS 20b annotations and a FastText model trained on 200,000 balanced samples. Even after filtering specifically for fictional material, words such as 'Elara' still appeared at negligible rates compared to the AI-generated stories. Why, therefore, do they dominate at the lowest level of the imperative for an LLM to write fiction?

The authors state:

> *'If Core words are not common in web data, then one remaining source would be post-training data. But we find that OLMo's post-training data exhibits our tokens at a lower rate than CONLIT.*

Within 78,958 stories from OLMo 3's post-training datasets, they note, 'Elias' appeared 52.7 times per million words, compared to 2.7 in CONLIT, but reached 2,428 occurrences per million words in the generated stories examined in the study.

To identify where the recurring 'Core' stories were coming from, each story in OLMo 3's post-training data was scored for the presence of one or more Core tokens (i.e., for the presence of *Elara*, *Mara*, etc.). Most were expected to appear in supervised fine-tuning (SFT) datasets, because WildChat and related sources contributed 59,266 stories to OLMo 3.

However, only 1,803 contained Core terms, while datasets used for DPO and reinforcement learning showed higher concentrations.

Overall, the recurring Core vocabulary was traced to just 3,053 stories, representing 3.8% of all post-training stories examined. There is no statistical possibility for such a small subset of corpora to end up dominating it in the way demonstrated.

The paper concludes:

‘When given little direction, current frontier models write stories using a narrow catalog of names, places, and occupations. Recurring characters in these stories include Elias, a lighthouse keeper. Elias is unusual; the name is uncommon in literature, web data, and even post-training data.’

Conclusion

In the absence of any single work of literature (or even a series) that features the top 11 words which the authors identify, it is not at all clear by what means this particular collection of words has accreted and self-associated into the very lowest levels of multiple large language models (notwithstanding their diversity of training data and approaches).

Even if the researchers’ contention about the constraining effect of copyright filters is correct, a veritable *ocean* of classic literature in the training data should have prevented this strange collection of old-school words from dominating the output of a non-qualified ‘write’ prompt.

That theory assumes, however, that vast amounts of classic literature would have been included in the training regimen at all. That’s unlikely, since what’s wanted are not models that will knock out faux Dickens outings, but rather that deal with the modern lexicon, and are suited for current business needs. The sheer volume even of pre-industrial literature would preclude its inclusion.

In any case, if there *were* one distinct narrative featuring some alternating mix of the ‘obsessive’ facets that the authors note, it would, presumably, be easier to find; the authors themselves could not find it, and casual searches on the pre-AI era unearth no such contender. Perhaps, if ‘lighthouse syndrome’ gains the same notoriety as AI em dashes, some scholarly authority will come forward with the answer.

* *I can’t go any further into May’s article, for reasons that may become obvious when one reads it.*

First published Wednesday, May 27, 2026. Modified in first 30 minutes to fix Anthropic link.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai
Contact: martin@martinanderson.ai