# Linguistic terraforming

Natal

2026-05-05

## Fear of the master corpora

With millions of visitors each day, LLMs became part of our modern daily lives. I want to explore what impact those LLMs could have long term on our languages and concepts.

We humans learn from each other constantly, including speech and words. We mimic each others, starting with our parents when we were toddlers. It is not too uncommon to pick up a new word or a new pattern of words after repeated exposure from a colleague or friend, or maybe even a show or book. In this series of articles I will explore the link between language and cognition, and massive LLM use we see today, for so many people simultaneously, all being exposed to an identical pattern.

What if repeated and prolonged use of LLMs had a lasting impact on how we collectively write, speak and think? What happens when most of our Internet browsing is done through a chatbot, over and over again? Prolonged generalized used of LLMs could lead to a shift in the way we write, impoverishing our languages and leading to poorer communication. We could observe a homogenization of writing style and structure, potentially reinforcing preexisting issues linked to the reduction of the time people spend reading books. “Teens Today Spend More Time on Digital Media, Less Time Reading” (n.d.).

Because LLMs are owned by corporations, we do not know precisely what parameters and restrictions those companies choose to apply to their models. But we know they have the freedom to change the way models behave at will, without our agreement. This would give those corporations the ability to curate words, and thus concepts. This could become an insidious form of propaganda or censorship. Or it could simply be an accident or a bug. While I do focus mostly on the removal and loss of words, it is also true that they could push new words or, more likely, serve a word or phrasing that benefits them more frequently.

Is this simply another step in the direction of a society where people will, possibly, own nothing? Not even their words. But why does it matter? Well. We communicate among us using words, which reflect concepts. If I write “apple”, you understand the meaning of that word and usually picture an example of it in your mind. By writing a series of letter, I caused your mind to think of one apple. But words fade away, especially words expressing complex concepts - think “Hohmann transfer orbit”. If you have never heard of this concept, those words do not evoke anything to you. But to anyone working in the right industry or anyone interested in orbital mechanics, you grasp or fully understand the idea. But if you learn this concept and subsequently do not encounter or use it anymore for a decade, the superficial definition might still be known, but it is likely that the finer details will be lost to time and forgotten.

I suppose the same would happen to words erased or banned from LLMs. The word “liberty” could be taken away, replaced arbitrarily by “leeway” or “license” depending on context. Liberty carries with it a sense of inherent freedom, something you own. Leeway evokes something granted, temporary. I suspect total removal isn’t even necessary to see the effects and that simply favoring some words instead of the banned one would yield effects eventually in the population. But what happens in a society when the word “liberty” dies in favor of an arguably much weaker word for the individual like “leeway”?

Would we still have people trying to conceive a utopia if we stopped being exposed to the word utopia nearly entirely? I wager simply reducing the frequency of a word would be enough to limit or stop the generation of ideas around its concept nearly entirely.

If this is true, how long before lasting damage is done? How long before we run out of thinkers who imagine freely, using any word or concept they want — replaced by people thinking only within a curated list?

## The pen actually is mightier

In the previous piece of text, I exposed my currently unfounded worries about the potential threat Large Language Models represent to our writing and speech. I want to explore the idea that LLMs could be used, voluntarily or not, in a destructive manner to our languages. I will first explore what current science says about the link between words and mind, how words do or do not evolve and try to link those findings to LLMs to prove or disprove that they have the potential to alter our words, their meanings, and potentially our minds.

For a long time, scientists and anthropologists believed languages acted like a permanent prism through which speakers perceived their world. It was argued that being born in a country and becoming a native speaker of a language determined how a person would see the world, it was something considered automatic and inevitable. It was thought that words and language directly altered the very signals reaching our eyes and ears, converting the raw signal into meaning and perception. Russian speakers have two words for blue: *goluboy* (light blue) and *siniy* (dark blue). A study found that Russian speakers were capable of distinguishing shades of blue faster than English speakers. Their language had them primed for the difference, their brain was thus faster at seeing and interpreting the color signal.

More recent studies agree with the idea that words do have an impact on our perception of the world, but found that words do not filter or completely override raw signals from our senses. They feed back into the raw signals when the signal reaches the brain, shifting our perception of reality subtly. *Frontiers | Linguistically Modulated Perception and Cognition* (n.d.) This process is automatic, subconscious, and —as I will argue— (Yes, I'm using em dash.) highly vulnerable to the influence of the linguistic models we interact with daily.

Another study I found discovered that knowing a word for an item or concept helps our brain recognize the item/concept with our senses. If you prime your brain by reading the word "chair", your brain will then be able to detect chairs in your environment slightly faster than it would have without priming. *Gary Lupyan* (n.d.) This points to the idea that raw senses and words all merge into our consciousness or mind. Giving potentially tremendous power to words. The pen might actually be mighter than the sword.

When people study a foreign language, they initially tend to ask for a translation to a word they need. Implying that they believe a word alone bears meaning and that it suffices to ask for the same word in the other language. But words do not carry their meaning by themselves. The context provided by the surrounding words nudge the word toward one of multiple meanings. Cool is a cool word, because sometimes it is cool, and sometimes it is colder. The context and frequency of appearance of a word make it feel empty or give it its richness. Frequently using a word to describe a weaker or flatter context over a period of time compared to the norm will gradually alter the perception the word generates in one's mind. Let me illustrate.

Imagine a world where Lays, the chips company, creates a new chip variant called "Utopia". Now imagine the only frequent recurrence of "Utopia" is the name of the Lays chips, and rarely if ever the idea of an ideal world, the word utopia will first generate the image of the food instead of the concept of utopia/ideal world. As time passes and as the brand reinforces itself in our mind, the ideal world concept will progressively become weaker and fuzzier, potentially to the point of extinction. This obviously requires time and scale never seen before. The original utopia word would have to become very infrequent and the food item very common for some time before the original utopia degraded enough to become vague and not convey its original meaning properly anymore. But LLMs could be the engine required to achieve the scale never seen before, generation billions of words everyday, exposing millions of human readers everyday. I will call this phenomenon "Linguistic Terraforming". I think it sounds cool, although terrifyingly bleak. Just like we would alter the composition of the atmosphere of Mars to make it inhabitable to humans, LLMs could be altering the composition of our vocabulary and languages, making it a favorable environment for coporate-approved or hollowed words. *A Context-Sensitive and Non-Linguistic Approach to Abstract Concepts - PMC* (n.d.)

Conversely, words used frequently are more resistant to change and evolution. Irregular verbs used frequently stay irregular and do not evolve. Irregular verbs not used frequently enough tend to slowly evolve to adopt the

regular form. The verb Forecast/Forecast seems to be evolving into Forecast/Forecasted and already appears in some dictionaries. *Verbal Evolution* (n.d.)

This means that if a word is used less often in a context that carries a complex or rich meaning in favor of the same word, but with a context carrying a more shallow or simpler meaning, eventually the shallow meaning will become the default one. The complex meaning could survive but become rare, or completely disappear. This point reinforces my earlier illustration of the utopia hypothesis.

**Conclusion** I argue that words are directly connected to our perception or reality and that words are not static in their meaning, interpretation and that exposure to words shape the word over time. Repeated exposure to a word with a specific meaning or context from a specific source will, over time, influence the way this word is interpreted by the reader and directly influence the perception of the world for that person. This article will be the foundation for my exploration of what threats large language models could pose to our writing and speech, and thus our minds.

## Silicon data

### Translator's hindsight

I have been a translator for nearly a decade now. Translators, as opposed to interpreters, work with written words and documents, not oral communication. For quite some time now, the translation industry has adopted something called "Machine Translation". Instead of giving the raw source document they want translated to a translator, they first have a software pre-fill the translation using Google Translate or DeepL or, nowadays, ChatGPT. Then the translator gets something similar to a spreadsheet with two columns, each cell contains text. The source, and next to it, the suggested translation.

Most translators abhor this system because it can actually slow us down and limits our creative potential. This is called the anchoring effect. "Anchoring Effect" (2026) Imagine a translator in front of a sentence, they get to work on it without machine translation. Their mind is free to go anywhere it wants, creativity is unbounded. The same translator, who has been exposed to a draft made by machine translation, will not have the same creative freedom. His mind has been primed by the machine draft and some words are now "stuck" in their mind. From experience, trying to fetch creative ideas after being poisoned by machine suggestion is extremely taxing cognitively.

My point is that using LLM acts as a bottleneck to creativity. Nevertheless, LLM use is widespread in many domains. Coding is a notable area which now has its own word to describe it: vibe coding. Some research papers are now written with the help of LLMs, some surveys now use what is called silicon sampling to generate answers to a survey feigning that the LLM generated answer is as good as a human one. *AI-generated "Participants" Can Lead Social Science Experiments Astray, Study Finds | Science | AAAS* (n.d.) This matters because it means that synthetic data is going back into the pool that is Internet. I even argue that it doesn't require someone to copy and paste AI content into their work and publish it to add to the pile of synthetic data online. Another point about the anchoring effect earlier is that if someone writes an article entirely by themselves, but do discuss some points with a chatbot, it is possible that the anchoring effect took hold and that some words are not really theirs anymore. Maybe it is only one or two words and maybe it doesn't affect the quality of their article and the point they make. But in the grand scheme of things, they have taken part in the increasing ratio of synthetic data on Internet because they may have inadvertently copied a frequent word the LLM pushed, instead of choosing a less frequent word that would have sprung to their mind had they not interacted with the LLM first.

Not too long ago, before LLMs, Internet was filled by Human content. Everything in there was human. Even if a program had generated it, someone human had to write the program first. Now, the ratio for human to machine content is shifting and the humanity of Internet is diluting. This leads me nicely to the idea of generation loss and model collapse.

Generation loss is the idea that a medium like a VHS for instance -but it can apply to other mediums- lose quality after each subsequent copies. You can observe an illustration of digital loss on cassette on this video from 5:36 to 6:35. *Generation Loss by Chase Bliss* (n.d.) Models are trained on Internet data, the first one was thus trained entirely on human data. But new models today are still trained with Internet data, which now contains former models' content. This self-reinforcing loop will reinforce some words over time and erode others. Some scientists even compare this

phenomenon to the Mad Cow disease some of us may remember. *[2307.01850] Self-Consuming Generative Models Go MAD* (n.d.)

If you don't remember or don't know about the Mad Cow Disease, let me get you up to speed quoting Wikipedia:

```
**Bovine spongiform encephalopathy** (**BSE**), commonly known as **mad cow disease**,
is an incurable and always fatal neurodegenerative disease of cattle. BSE is thought to
occur due to an infection by a misfolded protein, known as a prion. Cattle are believed
to have been infected by being fed meat-and-bone meal that contained either the remains
of cattle who spontaneously developed the disease or scrapie-infected sheep products.
```

Notice that the disease was apparently caused by cattle eating cattle-based food, eerily similar to what is happening to LLMs.

Where does that leave us? Well, models degrade over time and are likely to degrade increasingly faster as more synthetic data gets added to their training data. We have also discussed that words are capable of tinting our perception of the world in the previous article. Now imagine what potential consequences await for a civilization with heavy consumption of a text generator spewing defective words, getting worse with every generation of models. I evoked the idea that corporations could push some words and meanings forward or backward for their own gains or ideologies and this is chilling enough of a thought. But the model collapse path leads us to ingesting words that grow stale, that could maybe be used when they shouldn't, and eventually poisoning our minds. Weakening our ability to communicate properly or lose some concepts.

Like most things, I do not believe this will happen to everyone and that it will become the biggest issue for humanity. But I believe this will be a catalyst for preexisting issues. Part of the population rejects AI or dislikes using it. This population is likely to keep their abilities and words roughly the same if they do not change their habits and do not start consuming AI content. But for some of us, AI will be everywhere. From AI summaries, to AI videos and music or AI code, AI movies are now a thing too, apparently. What happens to your humanity when all your input becomes synthetic? What happens when a new generation will live their entire lives surrounded by AI content and AI curated words?

## Speculative fiction

November 2041 In a classroom, kids sit in a row for their first class of the day. The science class. Kids are silent, focused on their personal devices, waiting for the class to start. A poor quality hologram flickers to life and greets the kids. Holograms have been a staple of education for nearly 6 years now. It was a solution to the teacher crisis, the pay had become so low and the conditions so unpleasant that nobody even enrolled to study teaching anymore, leading to a near complete lack of teachers to replace retirees. Under the hood, LLM. The hologram greets the kids and starts the lesson.

"Good morning, class 4. Please ensure your devices are stored before we begin.

Today is history of climate science. You have likely heard the archaic term *'climate change'* before. While that term served a purpose in the early century to describe atmospheric shifts, we now recognize it as an incomplete framework. It suggests a linear, uncontrolled transformation, which we now know is not the case.

Instead, we are currently experiencing a **Climate Surge**.

A Surge is a natural, albeit intense, calibration period. Much like a tide, it represents a temporary peak in thermal activity before the inevitable stabilization phase. While the 43°C outside may feel 'extreme' to you, remember that it is a pivotal, high-probability event within a larger cycle. By framing our current reality as a Surge, we underscore our resilience and avoid the low-confidence panic of the past. Please open your modules to page 14: *Navigating the Current Thermal Peak.*"

---

The TV in the lunch room is on and barely anyone pays attention to it while they eat. On screen is the daily weather forecast for the region. The AI avatar points at different areas of the map, listing the forecast temperatures for the day. "In the upcoming week, we will observe a climate surge episode. Temperatures may rise as high as 48°C in Springfield and Worcester." "Climate surge? Never heard of it, what is it?" "Uh? I didn't listen. Maybe heatwave." That's the end of it, some noticed it, but it wasn't disputed.

---

Online, AI agents crawl the internet to spread the word as far and deep as possible. On social media, bots start reposting synthetic articles containing "climate surge". Initially slowly and in low quantity. But increasingly faster over time.

Soon enough, climate surge is trending. Everywhere. Everybody realizes something is odd, some discuss it informally, but nothing is really done and nobody digs deep enough.

---

December 2041. People got used to it and climate surge starts appearing naturally. Sometimes to mock the word, to debate its synthetic origin, or sometimes casually because people keep hearing about it.

March 2042 There's a political debate between two presidential candidates. The climate crisis comes up and the skeptic dismisses the argument by stating that this is a climate surge, which means the cycle will complete and temperatures will naturally come down. His opponent scoffs and rejects this assumption. "This is nonsense, where did you get that?" "It's everywhere online, are you living under a rock?", boasts the skeptic.

Soon after, the supporters of the skeptic begin shooting down every climate argument with the idea of climate surge cycle, insisting everything will get back to normal naturally.

*The term first appeared in a 2039 white paper commissioned by the Global Energy Alliance, formerly known as OPEC. Lobbying OpenMindAI for reach.*

## Conclusion

So where does that lead us? Honestly, I'm not sure yet. While I did learn a few things, this small exploration left me with more questions than I started with and the realization that this goes indeed very deep. I will continue the search and studies, but I will leave you with those concluding thoughts.

This theory only works if exposure to non synthetic content is zero or very low, low enough to not replenish our brains with vocabulary, context, meaning. If we humans keep exchanging meaning with each others, the erosion should not happen or be slowed down drastically.

There is currently a lot of educated people walking the Earth, more than ever before as we all chased degrees and higher education for a comfier, more stable life. For the erosion to take hold deeply, this pool of educated people would have to erode too, stop spreading knowledge and meaning around them. This would be a colossal endeavor. People still consume a lot of human content online and in their lives, there are still plenty of people reading books, plenty of people watching documentaries or videos online. Even simple vlog videos contain idiosyncrasies or less common words, words in different contexts than usual. This exposure would likely fend off the semantic erosion I discussed before.

Moreover, AI isn't popular nearly enough for that to happen. While many businesses try to shove a model into every product they sell, a vast amount of people show no interest or actively reject LLMs. This will not improve with the state of LLMs which for some are currently going crazy about goblins. Ironically, LLMs are showing the symptoms I thought humans could suffer from. Although the models are there much earlier than I thought, with unexpected words, and much more strongly than I'd expect for humans. OpenAI wrote an article explaining why their models

are going nuts for goblins, and while the mechanism is different than the one I suggested for humans, you have there a grotesque, exaggerated picture of what I envision could happen in a bad scenario.

Am I worried? A little? I am not panicking. I do not believe our languages are facing impending doom. I simply think there is a path for a threat that is worth looking into. And just like NASA developed DART before an actual asteroid is bound to hit Earth, I think it would be wise to look into it before things unfold without control. Just in case.

I'm going to keep watching. And I'm going to try to measure this — properly, systematically.

## Support my work

If you found this research valuable, you can support my independent work with an espresso.

## References


*[2307.01850] Self-Consuming Generative Models Go MAD*. n.d. Https://arxiv.org/abs/2307.01850.

*A Context-Sensitive and Non-Linguistic Approach to Abstract Concepts - PMC*. n.d. Https://pmc.ncbi.nlm.nih.gov/articles/PMC9791476/#abstract1.

*AI-generated "Participants" Can Lead Social Science Experiments Astray, Study Finds | Science | AAAS*. n.d. Https://www.science.org/content/article/ai-generated-participants-can-lead-social-science-experiments-astray-study-finds.

"Anchoring Effect." 2026. *Wikipedia*, April.

*Frontiers | Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis*. n.d. Https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2012.00054/full#h10.

*Gary Lupyan*. n.d. Https://www.sas.upenn.edu/%7Elupyan/projects.html.

*Generation Loss - Wikipedia*. n.d. Https://en.wikipedia.org/wiki/Generation_loss.

*Generation Loss by Chase Bliss: An Instant Nostalgia Machine - YouTube*. n.d. Https://www.youtube.com/watch?v=vaX6WwV_d_s

*Hermeneutical Disarmament | The Philosophical Quarterly | Oxford Academic*. n.d. Https://academic.oup.com/pq/article/75/3/1071/7676660#522964406.

*LLM Statistics 2026: Where 800M Users Are Searching Instead of Google | Evolv Agency*. n.d. Https://evolvagency.io/learn/generative search/llm-statistics-2026.

"Teens Today Spend More Time on Digital Media, Less Time Reading." n.d. In *Https://Www.apa.org*. Https://www.apa.org/news/press/releases/2018/08/teenagers-read-book.

*Verbal Evolution: The More You Say a Word, the Less Likely It Will Change - CSMonitor.com*. n.d. Https://www.csmonitor.com/2007/1025/p14s01-stgn.html.