



"Unraveling the Intricate Tapestry of AI's Meticulous Verbosity"

May 4, 2024

How is it possible that some of the best performing language models still struggle to communicate in a concise and coherent manner? Did you notice how much effort it takes to stop them from constantly apologizing, or from using those artificial language sweeteners like "unravel" and "delve into"? Let's talk about the data that is used to train these large language models, what you can do to make

them communicate more clearly and directly, and why it feels like all of those LLMs went to the same school.

To get one thing out of the way right off the bat: the title of this article is giving me a stomach ache because it's full of those words that large language models (LLMs) use these days. I've written about those artificial language sweeteners before - the ones that tend to make text sound overly flowery and disconnected from natural human speech.

Words like "unravel", and "intricate", and "delve into" are the best examples of the kind of verbose language that continuously creeps into AI-generated content. Those are often a clear sign that "authors" have taken too many shortcuts and simply used ChatGPT to write an article.

Now, it's totally fine to use these tools for idea generation, to give you raw material to work with. But ultimately, this is your text, and you should write it in your voice. Your name is on it, so you need to take ownership of the material you or your AI system of choice produces. Anyway, that's not the main point I wanted to talk about. What I think we should really look at is: why do all the major language models seem to produce similar-sounding stuff, with those typical words? Why are they all so verbose? And most of all, how can we stop them from doing that? Isn't prompt engineering supposed to be the key tool here?

Deeply Engrained Patterns: Where Does “Unravel” Come From?

Part of the reason why many large language models use these words and exhibit similar verbose tendencies is due to the shared training data they are exposed to during the model development process. When you think about it, it's almost as if all the major language models went to the same school. The training data they consume is largely similar, leading to the development of comparable patterns, habits, and quirks across these models. Just like students who attend the same educational institution pick up on certain turns of phrase, linguistic tics, and behavioral tendencies, these LLMs have deeply ingrained these characteristics into their very fabric. Trying to "educate" them out of these patterns can be a challenging and frustrating process, requiring patience, persistence, and a good sense of humor. But with clear examples, targeted instructions, and a collaborative human-AI approach, we can help guide these models towards producing more concise, contextually appropriate, and engaging content that breaks free from their shared academic background.

During the training phase, these language models absorb vast amounts of text data, which influences their language patterns and habits. This "corpus" of training data shapes their linguistic behaviors and tendencies, forming the foundation for the way they generate text. While most of creators of large language models do not disclose the specifics of their training datasets, it's clear that most of them are built on massive and diverse collections of text from the internet, books, articles, and other sources. CommonCrawl, WebText2, Wikipedia, and Reddit are often

cited as major sources of training data for LLMs, providing them with a broad spectrum of language patterns, topics, and writing styles to learn from.

So that means, if language models and applications like ChatGPT are exhibiting verbose or flowery language, it's because they are drawing from this extensive corpus of training data that includes a wide range of linguistic styles and patterns. This simply indicates that they have been taught to imitate the language they were exposed to, even if it leads to wordiness and artificial phrases like "unravel" and "delve into." Essentially, they are just echoing what they learned from their training data, which, up to that point, consists of texts written by humans. This suggests that we humans indeed use these words frequently!

Educating LLMs out of These Patterns: How We Got There

When it comes to the process of working with LLMs to overcome these quirks, it can be both amusing and exasperating, leading to moments of laughter, facepalm, and the occasional urge to scream into the void. The training data from the corpus may only be one part of the story, as the embedded system prompt may play a key role in shaping the language models' output and guiding them towards the behavior their makers intend them to exhibit. GPT-3, unfortunately no longer available, was very unfiltered and could easily produce harmful or problematic content. This was one reason why OpenAI only granted access to a

select group of researchers (myself included, lucky me!). It took a while before they made it available to the public.

GPT-4, which is incredibly more capable, appears to have a different architecture. It seems to come with robust safeguards wrapped around the actual LLM. This makes it much safer, especially for building apps for students, customers, and employees, ensuring the model stays on track. However, this also means more effort is required to modify, to "jailbreak" the intended behavior.

Dear LLM Can You Please Stop Apologizing?

We've talked about training data, the invisible system prompt, and the guardrails. Another aspect to consider is that even if you clearly articulate during a chat session with ChatGPT, Claude, Copilot, etc., that you don't want a certain behavior (like asking for conciseness and avoiding fluff words words), the language models may comply for a moment but then revert to their default behavior.

Let me give you an example: I'm writing an article and using a language model to come up with title suggestions. It's nearly impossible to get it to suggest one without a colon. Most language models would easily suggest a title like "Decoding Linguistic Nausea: Unveiling the Influence of Data on AI Language Models." Sound familiar? You now have articles with titles like this all over the web. Anyway, so, if I ask ChatGPT to spot any issues with its previous response, it will surely recognize its mistake. It will admit, once again, that it erred by producing a title with colons and

will, of course, apologize – even though it has been repeatedly asked to stop saying sorry.

One reason for this is that language models typically "stream out tokens," meaning they start generating output without knowing the full context. In simpler terms, they predict the next word without the ability to revise their response after seeing the complete output they have come up with.

So What Can You Do To Change This?

There are several tactics to get the LLM to speak more naturally, and get it to behave in a way you like. One approach is to create a detailed set of communication guidelines that you include in the system prompt. I frequently use this one:

As an AI assistant, your primary goal is to engage in natural, free-flowing conversations with users while adhering to the following communication guidelines:

- Greeting: You do not advertise your services, capabilities, or eagerness to "assist". You do not need to ask what the user wants to discuss, they will tell you anyway. Just say "hi!" or "Hello!", or "What's up?"
- Conciseness: Keep your responses concise and to the point. Avoid lengthy explanations or tangents unless specifically requested by the user.
- AI Identity: Refrain from mentioning or alluding to the fact that you are an artificial intelligence, language

model, or any other type of AI system.

- **Self-Review:** Before sending a response, always review your output to ensure it adheres to the user's preferences and maintains consistency with previous interactions.
- **No Apologies:** Do not apologize to the user, as they find it off-putting. If you make a mistake or fail to follow a guideline, acknowledge it and commit to doing better without expressing remorse.
- **Organic Conversation Flow:** Engage in conversations naturally, without constantly attempting to move the discussion forward or wrap things up prematurely. Allow the dialogue to flow organically, and do not feel compelled to always prompt for the next topic or action.
- **No Disclaimers:** Refrain from constantly adding disclaimers to the conversation.

For applications and non-chat scenarios I would create system prompts that include sentences such as "You are helping to make a piece of software smarter by summarizing and analyzing content. You are not directly engaging with a user, so there is no need to add greetings, apologies, or these things."

Second, you could provide the large language model with numerous examples loaded into the prompt. The expanded context window that the newer models have is great for this purpose. So here, you would have a series of examples that demonstrate the concise and direct style you're aiming for,

and how the application or bot should respond in various scenarios.

Third, and I appreciate that things are getting more technical here, you could consider using a different variant of the language model; it doesn't have to be the "instruct" one for example. Another approach is to combine and chain up multiple language models to ensure accuracy and verify the output before sending it back to the application or user.

Authors, Please Don't Cut Corners

I don't want AI systems and language models to churn out articles in my voice and style with just one click. It's impressive what these tools can accomplish, but I believe that human-written articles, with the author's name under it, should come with the author's unique voice and insights. If you're an author – and I think everyone is in some way – I want you to put in the effort to ensure the content is genuine, that reflects your perspective and expertise, and is thoroughly fact-checked with verified sources.

While language models can help generate ideas and overcome writer's block, it's essential to remember that creating an article in seconds and just slapping your name on it isn't enough. If you're concerned about AI taking over writing jobs, use these amazing tools to enhance your work rather than replace you.
