



Find yourself in AI

Description

Guest post

In the vast expanse of the digital universe, AI, particularly large language models (LLMs) like ChatGPT, have found a home, influencing our lives in countless ways. A lingering question, however, is whether we are inadvertently influencing these models back. In other words, could our writings have been part of the training data for these AIs? Welcome to the world of *Large Language Model Forensics*.

Our exploration begins with prompts. LLMs, such as ChatGPT, are designed to generate human-like text in response to input. A typical way to test the output is by creating prompts that would require knowledge or style that is unique to us. For instance, I could prompt ChatGPT with, *Imagine you are [Your Name Here]. Please write a blog post on [Your Unique Subject]*. This command instructs the AI to generate text that reflects not only the subject but my unique writing style.

You'd expect, if your text was part of the AI's training data, it would emulate your style quite accurately. The surprise comes when the output is read. For many of us, the text generated by the AI doesn't quite match our writing style, but it does bring forth an interesting amalgamation of our thought process and AI's unique touch.

This begs the question [sic], why is the AI not producing the exact style? One explanation is the sheer volume and diversity of the data the LLMs have been trained on. AI draws from an extensive range of internet text, blending styles, tones, and information, resulting in a rich and diverse model but one that does not mimic specific individuals. Another explanation might be the random sampling involved in training these models, which could mean that the AI might not have encountered enough of your text to accurately mimic your style.

However, AI forensics goes beyond just individual writing styles. Many users, both out of curiosity and concern, also test if the AI knows specific personal information about them. But LLMs are designed with privacy in mind. They do not store personal data from training or interaction, and they do not recall or reveal individual-specific information. This is a crucial aspect of AI safety and a factor we must continually advocate for.

Large Language Model Forensics is not just about seeking ourselves in AI responses but is a broader endeavor. It helps in understanding how AI thinks, operates, and learns. It also brings into focus the potential risks and ethical considerations we must grapple with as we advance in the realm of AI.

For instance, one critical question arising from these experiments is: should content creators be compensated if their work contributes to the AI's learning? As AI advances, we must strive for transparency in its training data and ensure fair use and compensation for original content creators.

Moreover, with the rise of AI forensics, there's also an opportunity for us to understand the susceptibility of AI to biased or malicious influences. Understanding the impacts of the training data on AI outputs can be crucial in safeguarding against intentional manipulation or skewing of AI perspectives.

In conclusion, Large Language Model Forensics offers an intriguing avenue to interact with and understand AI better. It encourages us to reflect upon our relationship with AI: we influence it, and it influences us. It's a dance between human creativity and machine learning, and it's only just beginning. As we continue to experiment, test, and analyze these AIs, we need to do so responsibly, ethically, and with a sense of curiosity and wonder about the immense potential of this technology.

The post above was written entirely by ChatGPT4 prompted by the request: "Can you draft a 500 word blog called 'Large language model forensics' about how I and others test ChatGPT with various prompts to see if our own writing has formed part of the training corpus." I added the "[sic]" to indicate the incorrect use of an idiom.

The result is not in my style and veers too far towards popular journalism for my taste. I asked it to modify the text to be more academic and less journalistic, but that attempt didn't read as smoothly. So I retain the first iteration here. This exercise informs my response to the protest by actors and writers about the impact of AI on their professions.

Automated script writing

To explore the risk to film and television writers posed by ChatGPT, I tested its ability to script a scene for a well known tv sitcom.

"Please write a funny 1000 word scene from Seinfeld where Kramer, George, and Elaine reveal to Jerry they bought a lottery ticket together. They win a small prize. Jerry tries to persuade them to offer the money to fund the severely dyslexic son of Mrs Waksmundski upstairs who needs some cash to pay for a school outing."

As far as I'm aware this prompt is my own creation, including the use of the Polish name Mrs Waksmundski. I probably got the idea of giving an invisible character a Polish name from an episode of *Frasier*. Here's what it produced as a first iteration.

(Setting: Jerry's Apartment. George, Kramer, and Elaine sit on the couch, hiding a smile)

George: Jerry, you know how you always tell us not to invest in anything together?

Jerry: Yes, considering your collective decision-making abilities, I'm terrified to ask.

Elaine: Well, we might have ignored your advice.
Kramer: We did it, Jerry. We bought a lottery ticket!
Jerry: A lottery ticket? I can already see this ending in disaster.
George (holding up the ticket): And we won, Jerry! We won!
Jerry: Alright, how much damage are we talking about?
Elaine: Fifty bucks!
Jerry: Fifty bucks? That's it?
Kramer: You don't understand, Jerry, it's not about the money. It's about the win!
Jerry: Oh, I understand perfectly. So, what's the grand plan with the winnings?
George (rubbing his hands together): Well, I was thinking of investing in a new pair of loafers.
Elaine: And I need a new bottle of conditioner.
Jerry: Alright, I have a better idea.
(They look at him, intrigued)
Jerry: You remember Mrs. Wakschinski's kid from upstairs, the one with severe dyslexia?
Kramer: Yeah, yeah, the kid with the letters jumbled up.
Jerry: Right. He needs some money for a school outing. What do you say we donate the winnings to him?
George: But! but! my loafers, Jerry!
Elaine: And my hair, Jerry, my hair!
Kramer: It's not about the loafers, George, or the hair, Elaine! It's about the kid!
Jerry: Thank you, Kramer.
George (grudgingly): Alright, alright, the kid gets the money.
Elaine (looking at her hair in the mirror): I suppose it's a worthy cause.
(Later, at Monk's Caf  )
George (dejectedly, looking at his worn-out shoes): I can't believe I gave up my loafers for a kid's school trip.
Jerry (laughs): Oh, come on George, it's not so bad. Look on the bright side, you've done something nice!
Elaine (touching her hair): Yeah, and my hair still looks fabulous!
Kramer (entering, wearing a big grin and a pair of new loafers): Hey, you guys! Look what I found on the street. Practically brand new!
(They all stare at Kramer's shoes)
Jerry (shaking his head): Only you, Kramer, only you!
Kramer: It's karma, Jerry, karma!
(They all laugh as the scene fades out)

It's a while ago now, but I did watch every episode of *Seinfeld* at least once. This script seems on point to me and fits the characters. I asked ChatGPT to rewrite the script for a scene in *Cheers*, in the notes below.

According to a primer on writing effective prompts for ChatGPT, *The Art of Mastering Prompt Engineering*, the platform is best deployed as a writing partner.

Iterate and refine: As with any creative process, writing involves revising and refining your ideas. Use ChatGPT to explore different narrative paths, character arcs, or story outcomes. By iterating on your ideas and receiving feedback from the AI, you can hone your story,

ensuring that it is engaging cohesive, and captivating.

Bibliography

- Burrows, J. (2022). *Directed by James Burrows: Five Decades of Stories from the Legendary Director of Taxi, Cheers, Frasier, Friends, Will & Grace, and More*. New York: Ballantine.
- Marx, N. (2021). "Home Economics: Sitcom Capitalism, Conservative Comedy, and Media Conglomeration in Post-Network Television." *Communication, Culture and Critique* **15**(1): pp. 21-35.
- Sinclair, J. (2023). *The Art of Mastering Prompt Engineering for Chat-GPT Made Simple: A Practical Guide to Learn How to Get the Most Out of Chat-GPT with Efficient and Effective Prompting*. Kindle.

Notes

- See post: [Sitcoms in the city](#).

In translating from *Seinfeld* to *Cheers*, I didn't have to remind it of the characters. I just said: Please deliver a similar scenario for a scene in the *Cheers* sitcom.

Setting: Cheers Bar. Norm is at the bar, Carla is serving, Cliff is at his usual stool, and Diane is by the entrance)

Norm (nudging Cliff): Cliff, remember when you said we should never gamble together?

Cliff (nodding): Aye, Norm. Considering our track record, that seemed like sound advice.

Carla (leaning in): Well, let's just say we threw caution to the wind.

Diane (entering, holding a small slip of paper): We bought a lottery ticket!

Norm: And we won!

Cliff: We did?

Norm: Well, kind of. We won fifty bucks!

Diane: Not a huge amount, but it's a start.

Cliff (scratching his head): So, what are we going to do with the winnings?

Norm: I was thinking about a round of beers.

Carla: And I need some new aprons.

Diane (raising her hand): Wait, I have a proposition. You remember Mrs. Huxley's son, the one with severe dyslexia?

Norm: Yeah, yeah, the kid who always mixes up his letters.

Diane: Correct. He needs some money for a school outing. What do you say we donate the winnings to him?

Carla (rolling her eyes): I suppose that's one way to spend the money.

Norm (looking disappointed): But! but! my beer, Diane!

Cliff (grinning): Well, Norm, I guess your thirst will have to wait.

(Later, at Cheers)

Norm (moaning): I can't believe I gave up a round of beers for a kid's school trip.

Diane (smiling): Oh, Norm, it's for a good cause.

Carla (interrupted by Coach entering with a tray of beers): Coach, what's all this?

Coach (smiling broadly): The distributor made a mistake. Sent us an extra case. Drinks are on the

house!

(They all laugh and toast to their â??winningsâ??)

Cliff: Well, Norm, looks like your thirst wonâ??t have to wait after all.

Norm: Cheers to that, Cliff!

(They all raise their glasses as the scene fades out)

Category

1. Artificial Intelligence

Tags

1. chatgpt
2. data
3. forensics
4. training

Date Created

September 2, 2023

Author

rcoyne99

default watermark