CSAI - Tutorial 1

Prepared by: Sydelle de Souza (TA of CSAI in 2024-25)

On Authenticity, Sincerity, and Apologies from an LLM

Why do we value handmade goods from platforms like Etsy with all their imperfections over perfect-looking mass-produced ones? Or go to live concerts, even when streaming platforms and home audio systems offer perfect sound quality at home? Why do we connect so deeply with art that reveals raw human emotions and personal struggles? And why do we feel betrayed when a memoir turns out to be fiction, or if an artist in concert is lip-syncing? It almost seems like we're wired to seek out what feels "real". In this two-part tutorial we're going to dive into our need for authentic and sincere communication, and whether AI has a place in that pursuit. In Part 1, we'll reflect on two cases wherein generative AI was used to write a heartfelt piece of communication. We'll discuss the definitions of authenticity and sincerity and whether using generative AI in writing compromises those values. In Part 2, we'll discuss writing tasks wherein using generative AI would be appropriate (or not!) and wrap up by thinking about the sincerity behind an LLM's output.

1 Reflection: What does it mean to be authentic and sincere?

1.1 Who wants an AI-written fan letter?

Let's start by watching this ad (links to YouTube):

Google + Team USA — Dear Sydney

During the 2024 Paris Olympics, Google released an ad showing a father using Gemini AI to write a fan letter from his daughter to her idol—Olympic hurdler Sydney McLaughlin-Levrone. Outrage ensued [1]. Critics argued that it undermined personal creativity and authentic human expression. Many viewers felt that the father should have encouraged his daughter to write the letter herself, fostering her writing skills and personal voice, rather than relying on an LLM [2]. Others pondered the irony of promoting a technology that can be used to devalue human creativity and labour[1] at an event meant to celebrate humans performing incredible feats [5]. This backlash led to Google pulling the ad from TV.

1.2 Who wants to receive an AI-generated apology?

Now give this article a quick skim:

Vanderbilt University apologizes for using ChatGPT to write mass-shooting email [6]

Vanderbilt University's Peabody School recently issued an apology after using an LLM to draft an email addressing a mass shooting at Michigan State University. This drew significant criticism from students, leading to an apology from the associate dean. In his apology, he acknowledged the poor judgment in dealing with such a sensitive issue. The controversy has prompted further review and temporary reassignment of the administrators involved.

---

[1]See [3, 4]

1.3 Discussion

Let's reflect on the two situations by answering the following questions:

1. What does it mean to be sincere in personal communication? What does it mean to be authentic in personal communication? Is there a difference between the two?

2. Is it insincere and/or inauthentic to use AI to help navigate complex feelings or articulate emotions? In the two cases discussed, does it matter that AI was involved?

3. Why do we place such a high value on sincerity in personal communication? Would public perception of the ad/email have been different if the AI had only polished the original draft rather than generating the message?

Figure 1: Image Credit Lim [7]

2 Activity: When is it appropriate to use an LLM?

Now that we've reflected on the role of sincerity and authenticity in communication, let's think about what types of writing tasks could, in principle, be handed off to an LLM. Try to think of both appropriate and inappropriate use cases for AI assistance. Let's now consider where the line should be drawn for using LLMs when writing. Here are three categories:

1. **Never use an LLM**: Authenticity is non-negotiable.

2. **An LLM could help with form**: AI can help with structure or etiquette, but personal input is still essential.

3. **Go ahead and use an LLM**: Authenticity doesn't matter, efficiency is key.

Now sort the tasks that you listed into these three categories. For example, could a product description for a website be placed in category 2? What about appealing a parking ticket? Or emailing a company about an internship?

2.1 Discussion

1. Were there any surprising or controversial placements?

2. Do you notice any patterns that emerge from this exercise?

3. How do our cultural values around sincerity and authenticity shape our decisions here?

2.2 Wrapping Up

For our final bit of reckoning, I'd like to leave you with one last question to think about:

What does it mean when an LLM says that it is sorry?

References

[1] Nate Anderson. Google pulls its terrible pro-AI "Dear Sydney" ad after backlash. URL https://arstechnica.com/culture/2024/08/google-pulls-its-terrible-pro-ai-dear-sydney-ad-after-backlash/.

[2] Shelly Palmer. Why Google's "Dear Sydney" Ad Makes Me Want to Scream. URL https://shellypalmer.com/2024/07/why-googles-dear-sydney-ad-makes-me-want-to-scream/.

[3] Ted Chiang. Why A.I. Isn't Going to Make Art. ISSN 0028-792X. URL https://www.newyorker.com/culture/the-weekend-essay/why-ai-isnt-going-to-make-art.

[4] Matteo Wong. AI Can't Make Music. URL https://www.theatlantic.com/technology/archive/2024/07/generative-ai-music-suno-udio/679114/.

[5] Caroline Mimbs Nyce. Google Wins the Gold Medal for Worst Olympic Ad. URL https://www.theatlantic.com/technology/archive/2024/07/google-dear-sydney-olympic-ad/679292/.

[6] Jennifer Korn. Vanderbilt University apologizes for using ChatGPT to write mass-shooting email — CNN Business. URL https://edition.cnn.com/2023/02/22/tech/vanderbilt-chatgpt-shooting-email.

[7] Alex Lim. How to Make ChatGPT Stop Apologizing: A Guide for Prompt Designers. URL https://pupuweb.medium.com/.