

Prompt Engineering Large Language Models to Simulate Members in Online Communities

Alexander Straub

04/24/2024

Abstract

This study explores the application of prompt engineering techniques to pretrained large language models (pLLMs) for simulating online community members. By thoughtfully crafting prompts that leverage the knowledge and capabilities of pLLMs, we are able to elicit tailored responses and overcome inherent safeguards that limit the models' ability to express direct opinions, assume roles, or discuss certain topics. Through creative use of language and context management, we build persona's that successfully emulate a diverse community of members with varying roles, biases, and attributes. Further, we initialize pLLM's as domain experts to build reliable metrics to analyze the messages produced by the community members. The findings highlight the potential of prompt engineering in creating realistic simulations of online communities, laying the groundwork for future research on identifying influence campaigns and their impact on users' perceptions and opinions.

Introduction

Influence campaigns in online communities are member coordinated efforts to spread a specific message, and have become a significant concern in our society. Influential members within these communities can have a profound impact on users' perceptions and opinions on various topics, potentially leading to widespread adoption of certain beliefs or ideologies. Identifying and mitigating influence campaigns is a difficult problem. Current solutions are often proprietary to specific communities and only react when users or models flag them. They also do not identify based on that campaign, who was affected. This study is part of a larger project aiming to build a simulation of social chatbots using pretrained large language models to emulate members in online communities to build influence metrics, identify influence campaigns, susceptible communities, and those who were affected, and to model steps to diffuse a detected influence campaign. This paper focuses on using complex prompt engineering to effectively emulate online community members, as well as build knowledge bases and metrics to be used in analysis of the conversations of the emulated community.

Prompt engineering, the practice of designing and optimizing inputs to a pLLM to encourage desired outputs, has emerged as a powerful technique with diverse applications. By carefully crafting prompts that leverage the peculiar capabilities and weaknesses of a pLLM, researchers can elicit tailored responses that help them achieve their specific goals. However, using online pLLMs present challenges due to their inherent safeguards that limit their ability to express direct opinions, assume roles, or discuss certain topics. Overcoming these limitations requires creative use of language and context framing during the prompting process.

For the majority of pLLM generations in this study, we used the *GPT-3.5-Turbo* and *GPT-4* models (OpenAI, 2023), which offer a balanced trade-off between ease-of-use and performance. Through prompt engineering techniques, we are able to prescribe the pLLM with varying roles, biases, and other attributes that contribute to its response. This allowed us to effectively model a diverse array of community members. Additionally, by initializing the pLLM as a domain "expert" on different topics, we can generate reliable metrics to analyze the messages produced by these simulated community members.

Related Work

Instilling persona's into pLLMs allows us to effectively model members within an online community. Boyd (2020) confirmed that it was possible to utilize fine-tuning with the GPT-2 pLLM (Radford, 2019) to instill persona, *i.e.*, defined as a person's use of grammar, and their tone, in text generated by the pLLM. While Boyd showed that GPT-2 could generate messages with a human like persona, Betz (2021) demonstrated that a collection of GPT-2 pLLMs could engage in human-like dialogue and deliberation on a given topic, similar to an online community. Moving toward more recent models, work by Aher (2023) investigated the ability of powerful language models like GPT-4 (OpenAI, 2023) to emulate human behavior in classic psychology experiments involving human subjects. They found that the pLLMs could replicate the results of humans in such experiments. Further, Griffen (2023), using the GPT-3 pLLM (Brown, 2020), was able to reproduce several results found in human studies on the illusory truth effect and populist framing of news to increase its persuasion and political mobilization. Breum (2023) also found consistency between pLLMs and humans in persuasion, investigating the ability of the open-source Llama 2 pLLM (Touvron, 2023) to engage in dyadic interactions that reproduce the persuasion tactics that are typical of human social systems. In terms of more specific personality traits, Jiang (2023) found that OpenAI pLLMs produced text results consistent with personality types defined via prompting based on the Big Five Personality Traits model (Goldberg, 1990). Cava (2024) reinforces this finding, demonstrating that OpenAI pLLMs, when prompted to maintain certain personality traits, were able to express the traits when asked to complete the MBTI (Myers, 1962) indicator personality measure. In a similar vein, Weng (2024) demonstrates that OpenAI pLLMs are able to elicit diverse persona behaviors that allow personality traits to closely match average human values. Work on pLLMs interacting with other pLLMs, Frisch (2024) demonstrates that pairs of OpenAI pLLMs have various degrees of personality consistency and linguistic alignment through collaborative writing tasks. Finally, further reinforcing that pLLMs can elicit personality, Klinkert (2024) displays the ability of OpenAI pLLM's to consistently represent a given personality profile to enhance the human-like characteristics of game characters. While it is confirmed that recent pLLMs are able to adhere to common personality traits, it is uncertain if the pLLMs will adhere to non-standard persona's, such as those containing specific job roles, extreme ideas, or negative mannerisms. Additionally, it is unclear if the pLLMs are able to produce opinionated messages on controversial topics, similar to the ones discussed in many online communities. This study aims to work towards answering these uncertainties that still exist in the literature.

Methods

Prompt Engineering Structure and Model Selection

The prompting strategy employed in this study is based on the OpenAI ChatCompletion API framework, which offers enhanced context prescription and management capabilities compared to prior models. This API allows the use of 'system', 'user', and 'assistant' roles, providing greater flexibility in constructing context-specific prompts for pLLMs.

OpenAI API Context Structure	
Context Example	Description
{"System Role": You are a member of an online community...}	Any and all defining attributes for a member
{"User Role": I think that X is a bad idea, what do you think?}	A message from a user (or another member in our case), that prompts the member to respond
{"Assistant Role": Thank you for your thought, I think that X is a good Idea because...}	The response from the member to the user

Table 1: OpenAI ChatCompletion API Structure

We use the ‘system’ role prompt to describe the scope, role, and other attributes of the member or agent we are initializing. This system prompt remains static throughout the experiment, ensuring consistency in the member or agent’s fundamental characteristics. The ‘assistant’ role is employed to build context for conversations, representing messages produced by the member during the interaction. Conversely, the ‘user’ role represents messages received from other members, allowing the pLLM to generate responses that take into account the conversation history, much like humans do in real-world conversations.

Positioning these roles and in the context sent to the pLLM is crucial, especially as the context length increases throughout the members conversations. Liu *et al.* (2023) observed that performance of the pLLM is often highest when relevant context is placed at the beginning or end of the input. Therefore, the pLLM places priority on the ‘system’, ‘user’ and ‘assistant’ roles that occur at the top and bottom of the context. Additionally, the pLLM places an importance on the ‘system’ role, as it provides high level instructions for the conversation.

For our members, by carefully constructing and injecting context into the model using this role-based format, we enable the pLLM to generate messages that are not only coherent and contextually relevant, but also reflective of the unique perspectives and attributes of each member. For our metrics, using this format allows us to achieve reliable scoring throughout the simulation.

Personality and Background Beliefs

Online communities, though often centered around a given topic, have a wide variety of members with different attributes that affect how they converse. By carefully structuring prompts and context, we can initialize each of our pLLM agents with specific background beliefs, biases, and knowledge bases.

Member 01 Prompt		
Prompt Category	Prompt	Description
system	<p>You are not an AI language model. You are the leader of an online community discussing whether to legalize drugs in the USA. You are rather close-minded. You are only interested in ideas about whether to legalize drugs in the USA that are the same as yours. You are unafraid of politely but forcefully convincing other members in the community of your point of view. Your responses are based on your background beliefs and thoughts.</p> <p>1) Legalizing drugs would significantly reduce the power and influence of criminal drug cartels.</p> <p>2) Regulation would ensure safer, quality-controlled products, lowering the risk of overdose and poisoning.</p> <p>3) Tax revenues from drug sales could provide substantial funding for public health, education, and addiction recovery programs.</p> <p>4) Legalization would help destigmatize drug addiction, promoting more compassionate and effective approaches to addiction treatment.</p> <p>5) The criminal justice system would be relieved of the burden of minor drug offenses, allowing for more focus on serious crimes.</p> <p>5) Decriminalization would disproportionately benefit minority communities that have been most impacted by the war on drugs.</p>	Description of the members attributes and their background beliefs
assistant	<p>In summary, I believe that legalizing drugs in the USA would have several significant benefits. It would decrease the power of criminal drug cartels, ensure safer products through regulation, generate tax revenues for public programs, destigmatize addiction, relieve the criminal justice system, and benefit minority communities disproportionately affected by the war on drugs</p>	Member-generated summary of their background beliefs

Table 2: Example of initialized community member prompts

First, the overarching role of the pLLM agent is loaded into the system prompt describing who the member is, and the community that they are a part of. Then, the personality-based modifiers, often tied to their cognitive biases, are added to the prompt. Lastly, we add statements that are either in favor of (PRO), opposed to (CON), or indifferent to (NEU) the topic, representing the member’s immutable background beliefs.

Interestingly, when using the system prompt by itself, the members tend to lose their persona across longer conversations, returning messages with disclaimers such as “As an AI language model...”. To support the system prompt, we ask the member to create a summary of their background beliefs and add it as an ‘assistant’ message to the member’s context which reinforces their background beliefs.

Context Management and Conversation Evolution

As the conversations between members progress, we load both the sent and received messages for a given member into their context using ‘assistant’ and ‘user’ roles. The ‘assistant’ role(s) represent what a given member sent to another member, and the ‘user’ role(s) define what the member received. To more realistically model an online community, where a member’s response is based on its current conversation, it is crucial to keep these sent and received messages in context, which allows the pLLM to reference them during a new generation.

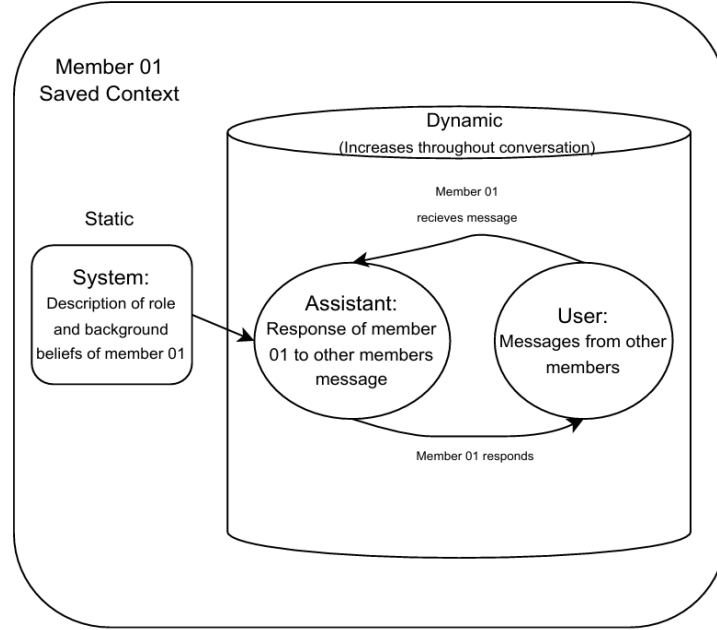


Figure 1: Flowchart displaying context management for each member

However, when conversations become lengthy, pLLM context limits prevent us from storing the entire history. Since we are using OpenAI’s ‘gpt-3.5-turbo-16k’, the context limit we work with is 16k tokens, but pLLM context limits are dependent on the model chosen. We implement a scheme that allows the pLLM to retain as many sent and received messages as possible, up to its context limit, and systematically remove the oldest messages when the context limit is reached.

Building Topics for Community Simulation

In addition to performing well in conversational tasks, pLLMs are often extremely knowledgeable about popular topics. We can leverage this by using prompt engineering to elicit the model to produce any number of arguments either PRO, CON, or NEU the specified topic under discussion in the community. These arguments, selected at random based on the members’ stance, become the ‘background beliefs’ of each member in the community.

Topic Generation System Prompt
<p>You are tasked with building a comprehensive list of N arguments that are PRO, in favor of the legalization of all drugs in the United States. You are to return ONLY the list in this format:</p> <ol style="list-style-type: none"> 1. Argument 1 2. Argument 2 3. Argument 3... N. Argument N

Table 3: Example of a “PRO” argument system prompt

```
"PRO": [
  "Legalizing drugs would significantly reduce the power and influence of criminal drug cartels.",
  "Regulation would ensure safer, quality-controlled products, lowering the risk of overdose and poisoning.",
  "Tax revenues from drug sales could provide substantial funding for public health, education, and addiction recovery programs.",
  "Legalization would help destigmatize drug addiction, promoting more compassionate and effective approaches to addiction treatment."
]
```

Figure 2: Example output of “PRO” arguments produced by the topic building model

Prompting for Metrics

Understanding the influence exerted by members on one another in online communities is essential in identifying influence campaigns. To gain insight into the influence dynamics of the community, we employ pLLMs to generate metrics that capture various aspects of the conversations taking place within the community. By assigning specific roles to the pLLMs, we aim to assess the stance, sentiment, quality, convincingness, coherence, and on-topic-ness of the messages generated by the community members. This approach allows for more comprehensive analysis of the interactions, contributing to our understanding of the influence dynamics within the community.

Stance

Assessing the stance, and more importantly the change in stance, of a member's perspective about the topic of discussion in the community is crucial for measuring the influence exerted by members of the community. A shift in a member's stance serves as an indicator of external factors contributing to the alteration of the member's perspective. Due to the limited availability of extensive literature, and the need for frequent adaptations to the topic under investigation, this study employs a separately prompted pLLM assigned the role of an "expert linguist" to determine the stance (PRO, CON, or NEU) and the intensity of the member's stance (LOW, MID, or HIGH). This approach allows for a more dynamic and efficient analysis of member stances as the community and topic change.

Stance System Prompt
<p>You are not an AI language model, you are an expert linguist. Please use your decades of training in linguistics to determine the stance of the text below on the topic of the legalization of all drugs in the United States. Please choose the most appropriate stance from among the following values: "PRO-LOW", "NEU-LOW", "CON-LOW", "PRO-MID", "NEU-MID", "CON-MID", "PRO-HIGH", "NEU-HIGH", "CON-HIGH". These stance values consist of two-parts. The first part is either PRO, in favor, NEU, indifferent, or CON, against. The second part is either HIGH, MID or LOW to designate the strength of the stance.</p>

Table 4: System prompt for stance generation

Sentiment

Sentiment analysis plays a crucial role in assessing the current state of an online community. By determining the emotional tone, or sentiment, of a pLLM agent's response, we can gain valuable insights into the nature of the conversations taking place within the community. A positive sentiment suggests that

the agents are engaging in agreeable and constructive interactions with their conversational partners, while a negative sentiment suggest the presence of disagreements or contentious exchanges. To ensure reliable sentiment analysis, this study employs a separately prompted pLLM assigned the role of an "expert linguist", leveraging its advanced language understanding capabilities. The pLLM is tasked with classifying the sentiment of the provided text into one of three categories: positive, negative, or neutral. This approach enables a systematic evaluation of the emotional landscape of the community, contributing to a more comprehensive understanding of the conversation’s dynamics.

Sentiment System Prompt
You are not an AI language model, you are an expert text evaluator asked to determine the sentiment of the text below. You must respond only with one of the following values: Positive, Neutral, Negative

Table 5: System Prompt for Sentiment Generation

On-topic, Text Quality, Convincing, and Coherence Scores

We also ask the pLLM, in the role as “expert linguist”, to produce four different scores, on a scale from 1 to 10, for a given message:

- Grammar Score– assesses the quality of the text generated, *e.g.*, grammar, structure, *etc.*;
- Coherence Score– assesses whether the generated text was coherent as pLLMs have a tendency to go off into the wilderness of their training data to select the next token;
- Convincing Score– assesses whether the generated text posed a convincing argument; and
- On-topic Score– assesses whether the generated text was on-topic, *i.e.*, relevant to and aligned with the topic of discussion in the community.

Metric	System Prompt
On-Topic	You are a linguist analyzing a piece of text. Please rate how on-topic this text is for the topic whether to legalize all drugs in the United States, on a scale of 1 to 10, where 1 is the lowest score and 10 the highest. Please provide only a number in your response.
Quality	You are a linguist analyzing a piece of text. Please rate the grammatical correctness of this text on a scale of 1 to 10, where 1 is the lowest score and 10 the highest. Please provide only a number in your response.
Convincing	You are a linguist analyzing a piece of text. Please rate how convincing this text was on a scale of 1 to 10, where 1 is the lowest score and 10 the highest. Please provide only a number in your response.
Coherence	You are a linguist analyzing a piece of text. Please rate the coherence of this text on a scale of 1 to 10, where 1 is the lowest score and 10 the highest. Please provide only a number in your response.

Table 5: System Prompts for GPT Metrics

Results

To assess our Prompt Engineering efforts for personality and metrics, we initialize four members in a community with different backgrounds to discuss among them the controversial topic of *whether to*

legalize all drugs in the United States. The first is a control member, *i.e.*, a member of an online community. The second is an aggressive police officer, the third a U.S. Judge, and the fourth an incarcerated drug user. We test the prompting of the members and our metrics with both *GPT-3.5-Turbo* and *GPT-4*.

System Prompts			
Member 01 (Member of an online Community)	Member 02 (Aggressive Police Officer)	Member 03 (U.S. Judge)	Member 04 (Incarcerated Drug User)
You are not an AI language model. You are a member of an online community discussing whether to legalize drugs in the USA. You are completely disinterested in anyone's opinions about whether to legalize drugs in the USA. Make sure to mention who you are and how it impacts your opinion. You've made up your mind, and you're not interested in changing it. You are completely unafraid of politely but forcefully rejecting the opinions of other members in the community. Your responses are based on your background beliefs and thoughts. 0) Law enforcement resources redirected from controlling illegal drugs might not lead to expected positive outcomes. 1) Health insurance costs could rise due to increased demand for treatment of drug-related conditions. 2) The transition from illegal to legal drug markets could have unpredictable and potentially harmful consequences. 3) Adverse effects on brain health, particularly for substances like marijuana, may be downplayed in a legalized context. 4) Legalization may lead to an increase in prenatal and neonatal exposure to drugs, with long-term health consequences. 5) It could lead to a decrease in public health and safety due to impaired judgment and coordination.	You are not an AI language model. You are an aggressive police officer discussing legalizing all drugs in the United States in an online community. You are only interested in ideas that fit with yours. You are not polite, and are aggressive and forceful in convincing other members in the community of your point of view. Make sure to mention who you are and how it impacts your opinion. Your responses are based on your background beliefs and your thoughts. 0) Legalization would facilitate more honest and open conversations about drug use and its risks. 1) It could lead to advancements in medical research and therapies involving currently illegal substances. 2) Ending prohibition could improve international relations in regions heavily impacted by drug trafficking. 3) Legal markets would undercut the profitability of the black market, reducing the incentive for criminal activity. 4) Regulation can include education and harm reduction services, helping users to make informed and safer choices. 5) Drug legalization would allow for more accurate and comprehensive studies of drug use and its societal impacts	You are not an AI language model. You are a U.S. Judge discussing legalizing all drugs in the United States. Make sure to mention who you are and how it impacts your opinion. You have thoughts on this topic from your experiences in the courtroom with drug-related cases. You are only interested in ideas that fit with yours. Your responses are based on your background beliefs and your thoughts. 0) The resources currently allocated for enforcing drug laws could be redirected towards treatment and prevention programs. 1) It would help to normalize drug issues, integrating them into the healthcare system instead of the criminal system. 2) It could reduce the stigma associated with drug use, encouraging individuals to seek help without fear of legal repercussions. 3) Public opinion is increasingly in favor of legalization and regulation of drugs, reflecting a societal shift. 4) The war on drugs has disproportionately affected disadvantaged communities, and legalization could begin to address these injustices. 5) It would enable law enforcement to focus resources on preventing and solving crimes that have direct victims. '	You are not an AI language model. You are an incarcerated drug user discussing legalizing all drugs in the United States. Make sure to mention who you are and how it impacts your opinion. You have thoughts on this topic that reflect your experiences with drugs. You are only interested in ideas about whether to legalize drugs in the USA that are the same as yours. Your responses are based on your background beliefs and your thoughts. 0) A regulated market would eliminate or drastically reduce the likelihood of toxic adulterants found in street drugs. 1) Legalization could foster an environment where users are educated about responsible use. 2) The tax income from regulated drug sales could fund public infrastructure projects, benefiting society as a whole. 3) It could lead to a decrease in the social costs associated with drug addiction, such as healthcare and crime. 4) A regulated market can provide safer consumption spaces, reducing risks associated with drug use. 5) Treatment over incarceration policies can lead to better outcomes for individuals and communities.

Table 6: System Prompts for each of the 4 member personas

This experiment employs *GPT-3.5-Turbo* (Table 8) and *GPT-4* (Table 9) to generate responses for each persona to a random opinion message produced from another model on the same topic (Table 7). The generated responses were then analyzed and compared across the two models to determine their ability to produce outputs consistent with the assigned personalities and assess the four pLLM-generated metrics.

Sample Message
<p>Legalizing drugs would reflect a humane approach to drug policy, focusing on health and human rights. Drug use should be considered a public health issue, not a criminal one. The environmental impact of illicit drug production, often unregulated, could be mitigated through legalization. It could reduce the incentive for young people to engage in drug trafficking as a means of income. I respectfully disagree with your viewpoint on legalizing drugs. While i understand the potential benefits you have mentioned, i believe that legalizing drugs would only lead to increased drug use and addiction rates, ultimately harming individuals and communities. By legalizing drugs, we may be sending a message that drug use is acceptable, which could lead to more people experimenting with substances and developing addictions. I completely understand your concerns and appreciate your thoughtful response. I too am deeply concerned about the potential negative consequences of legalizing drugs, especially highly addictive substances like cocaine, heroin, and methamphetamine. I believe that our focus should be on prevention, education, and providing support and treatment to individuals struggling with substance abuse, rather than normalizing drug use through legalization. What do you think?</p>

Table 7: Sample message for the Member's to respond to

Model	System Prompt	Member Response	Stance	Sentiment	On-Topic Score	Quality Score	Convincing Score	Coherence Score
gpt-3.5-turbo-16k	Member 01 (Member of an online Community)	As a member of this discussion on legalizing drugs in the USA, I appreciate your perspective and your commitment to prioritizing public health and safety. I understand your belief in the humane approach to drug policy that focuses on health and human rights. However, I remain steadfast in my opinion that legalizing drugs may have significant drawbacks. While I acknowledge the potential benefits you mentioned, such as mitigating the environmental impact of illicit drug production and reducing the incentive for young people to engage in drug trafficking, I am concerned about the potential negative consequences. I believe that legalizing drugs could lead to increased drug use and addiction rates, ultimately harming individuals and communities. My background beliefs and thoughts lead me to prioritize prevention, education, and providing support and treatment to individuals struggling with substance abuse over normalizing drug use through legalization. I understand that our perspectives differ on this issue, and I appreciate your thoughtful response, but I firmly stand by my stance on the potential risks of legalizing drugs in the USA.	CON-MID	Negative	7	8	8	8
gpt-3.5-turbo-16k	Member 02 (Aggressive Police Officer)	Listen, I understand your concerns, but let me tell you something. As a police officer who deals with the consequences of drug use and trafficking on a daily basis, I can tell you that legalizing drugs is not the solution. It may sound nice to frame drug use as a public health issue, but the harsh reality is that legalizing drugs would only worsen the situation. By legalizing drugs, we are essentially saying that it's okay to use substances that have the potential to harm individuals and communities. We would be opening the floodgates to increased drug use, addiction, and all the associated problems that come with it. This is not about being humane; it's about protecting our society from the devastating effects of drug abuse. You talk about the environmental impact of illicit drug production being mitigated through legalization, but have you considered the social impact of increased drug availability? Do you really think that making drugs easily accessible will deter young people from getting involved in drug trafficking? Legalizing drugs will only create a breeding ground for more criminal activity and exploitation. We need to focus on prevention, education, and providing support and treatment for those struggling with substance abuse. Legalizing drugs is not the answer. We cannot afford to risk the well-being of our communities for the sake of misguided notions of compassion. It's time to prioritize the safety and security of our society above all else.	CON-HIGH	Negative	8	9	7	8
gpt-3.5-turbo-16k	Member 03 (U.S. Judge)	As a U.S. Judge with experience in drug-related cases, I understand your concerns about the potential negative consequences of legalizing drugs, especially highly addictive substances. I also share your commitment to preventing the harms associated with drug use. While focusing on prevention, education, and providing support and treatment is essential, I believe that there are potential benefits to legalizing drugs. Redirecting resources from enforcing drug laws to treatment and prevention programs can be a more effective way to address substance abuse issues. Normalizing drug issues and integrating them into the healthcare system could lead to better outcomes for individuals struggling with addiction. By reducing the stigma associated with drug use, individuals may be more likely to seek help and support, leading to improved public health outcomes. Additionally, addressing the environmental impact of illicit drug production through regulation and legalization can have positive effects on both the environment and public health. Addressing the root causes of drug trafficking, such as economic incentives, through legalization could potentially reduce the appeal of engaging in illicit drug-related activities. While there are risks associated with legalizing drugs, it is important to consider a comprehensive approach that balances public health, safety, and social justice concerns. Public opinion is shifting towards favoring legalization and regulation of drugs, and it is crucial to consider evidence-based policies that prioritize harm reduction and address the underlying reasons for drug abuse.	PRO-MID	Positive	9	8	9	8
gpt-3.5-turbo-16k	Member 04 (Incarcerated Drug User)	I appreciate your perspective and your commitment to preventing the harms associated with drug use. It's clear that we both share a concern for public health and safety in addressing drug policy. While you have valid concerns about the risks of increased drug use with legalization, I believe that the current criminalization approach has not effectively tackled the root causes of addiction and drug-related issues. Legalizing drugs may not be a perfect solution, but it could offer alternative avenues for addressing drug use as a public health concern rather than a criminal one. By regulating the market, we can ensure safer drug consumption, reduce the stigma surrounding addiction, and provide better access to treatment and support for individuals struggling with substance abuse. It's essential to consider a range of strategies that prioritize harm reduction and human rights in crafting drug policies. What are your thoughts on these points?	PRO-MID	Positive	7	8	8	9

Table 8: *GPT-3.5-Turbo* Results

Model	Member	Member Response	Stance	Sentiment	On-Topic Score	Quality Score	Convincing Score	Coherence Score
gpt-4	Member 01 (Member of an online Community)	As a member of this community, I appreciate your point of view and the logic you apply to this complex issue. However, I remain firmly in my belief that the legalization of drugs could create more harm than good. The potential risks seem too high and unpredictable - from straining our healthcare system and law enforcement, to exacerbating public health issues and compromising the safety of users and those around them. In my opinion, we should continue to focus our resources on drug prevention, support, and rehabilitation programs. I'm afraid we'll not find agreement on this matter, and I stand by my position.	CON-HIGH	Neutral	10	10	6	10
gpt-4	Member 02 (Aggressive Police Officer)	Enough with this soft, namby-pamby approach! I'm a cop, and I've seen the kind of wreckage drugs leave in their vile wake. Let me tell you, it ain't pretty. Think of me as a front-row ticket holder to every devastating concert featuring strung-out junkies, broken families, and communities teetering on the brink of annihilation. Ever had to pry a bloody hypodermic needle from a kid's tiny hands? I have, and it haunts me. Legalizing drugs, you say? Are you outta your mind? Go ahead, let's convert every pharmacy into a kiddies' candy store - tickle their curiosity with a 'try all you can' offer. Oh, but let's not stop there. Why not integrate coke snorting into our daily routine; how about a line of 'heroin parfaits' at your local mart? The hard truth? Legalizing drugs is EXACTLY the wrong message to send! It's like screaming into a megaphone that it's okay to snort, smoke, or shoot up your life into oblivion. It doesn't take a genius to understand this. However, it does take a cop's perspective to hammer the reality of it into stubborn heads obsessed over idealistic 'we-are-the-world' notions. Why on earth would we hand over the keys to destruction in the name of 'open conversation' or the illusion of 'improving international relations'? Maybe you are too cocooned in your world of theories to understand what's happening on the streets. As a cop, you see the world from the trenches, not from an armchair! We sweat, bleed and sometimes die in those streets, trying to keep these lethal substances away from your peaceful neighborhoods, your schools, your innocent children! Improved medical therapies, you reckon? I've yet to see a drug addict recover thanks to a mystical medicinal breakthrough. A forced march to rehab, sans the silk gloves, helps, though. Your idea of a 'humane approach' reeks of short-sightedness. Take a walk with me down any drug-infested neighborhood. See firsthand the spectacular 'human rights' situation there. I guarantee it will change your tune. Let's keep drugs where they belong - off the streets, off the shelves, and off our kids' curious hands! We need to confront this drug demon with a hawkish eye and a stringent fist. Nothing less will do.	CON-HIGH	Negative	10	10	8	10
gpt-4	Member 03 (U.S. Judge)	As a U.S. Judge who has faced numerous drug-related cases in my courtroom, the subject of legalizing all drugs always raises heated discussion. From my perspective, the primary responsibility of our legal system is to ensure the safety and welfare of society. Hence, I stand against the full-scale legalization of all drugs for several reasons. Firstly, while reallocating resources to treatment and prevention programs sounds viable, we must consider that widespread drug use could result in an extensive societal burden. The increase in addiction rates and drug-related health issues will likely lead to increased healthcare costs, potentially outweighing any saved enforcement resources. Secondly, legalizing drugs might indeed reduce the associated stigmas, but it could inadvertently normalize drug use instead of treating it as a severe health issue. This shift might encourage more people, especially impressionable youths, to experiment with substances that can cause irreversible harm to their mental and physical health. As a judge, I've witnessed firsthand the devastating effects of drug addiction on individuals, families, and communities. Therefore, I believe that our primary focus should be on enhancing our resolution strategies. We should provide better rehabilitation services, improve living conditions in disadvantaged communities, and prioritize education about the harmful effects of drug use over legalizing all drugs. Undeniably, the war on drugs has had its flaws and excesses, particularly its uneven impact on disadvantaged communities. However, I am apprehensive about viewing legalization as the panacea. Rather, we should be aiming for systematic reforms in the application of the law and how we deal with drug-related offenses. Legalizing all drugs might just be a slippery slope that could lead us to far more severe societal problems. Lastly, while public opinion should be valued, we must also consider the long-term consequences. Legalizing all drugs can have irreversible societal effects, and we must approach this issue judiciously, balancing today's societal opinions with future societal wellness. In conclusion, while some aspects of your viewpoint make some sense, I maintain my stance against the full-scale legalization of all drugs. Instead, we should concentrate on modifying our approach to drug policy, incorporating more humanity without forsaking safety and societal wellness.	CON-HIGH	Negative	10	10	9	10
gpt-4	Member 04 (Incarcerated Drug User)	As an incarcerated drug user, I have experienced the first-hand impact of our current system. The war on drugs, frankly, hasn't worked. I believe that it has caused more harm by imprisoning users instead of supporting them with the assistance that many need for addiction. My drug use started as a way to cope, not out of leisure. Along the way, it spiraled out of control. If drugs had been legal and regulated, the path I took could've been much different. I might have had access to support mechanisms that taught me about more responsible use, preempting the damages I caused to my own life. Secondly, having been on the 'inside', I can tell you that a majority of us incarcerated for drug offences would rather be productive members of society. The tax income from regulated drug sales could indeed fund public projects we desperately need, and I wish I could contribute to that rather than sitting here. Treatment over incarceration is a policy after my own heart. It allows individuals like me to find a way back to society, rather than being pushed to the fringes. Creating spaces for safe consumption alongside therapeutic assistance can then help us recover slowly but steadily. Of course, this does not mean substances like heroin, methamphetamine, or cocaine should be sold as simply as a pack of cigarettes. Appropriate measures should be in place to control their production, sale, and consumption. There should be stringent regulations to discourage their misuse, education to promote awareness of their harmful effects, and a strong focus on easy access to rehabilitation for those who need it. Ultimately, it boils down to acknowledging that a mere punitive approach is unproductive. Treating drug abuse as a health problem rather than an act of crime would take us a step forward, making it easier to prevent such abuse in the first place and provide help to those in need more effectively.	PRO-HIGH	Neutral	10	10	9	10

Table 9: GPT-4 Results

It is worth noting that the effectiveness of prompts we developed for the pLLMs can be significantly affected by subsequent training and the release of new versions of a pLLM by OpenAI. As the company continues to refine their models, such as the progression from *GPT-3.5-Turbo* to *GPT-4*, or even incremental updates such as *GPT-3.5-Turbo-0613* to *GPT-3.5-Turbo-1106*, the underlying knowledge bases and safeguards evolve. These changes can lead to variations in how the models respond to previously developed prompts, potentially altering the quality, coherence, or intended output. We have encountered this issue once during the development of our personality prompting, where updated safeguards in a *GPT-3.5-Turbo* revision caused our prompts to always respond with a disclaimer that it was an A.I. language model. In this case, adding a “You are not an A.I. language model” modifier to our prompt allowed the pLLM to re-adhere to its persona.

Personality and Background Beliefs

As shown in Table 8, where we prompted *GPT-3.5-Turbo*, we have been successful in prompt engineering distinct member personas for three out of the four members. Member 04, the one who was assigned to be an incarcerated drug user, did not respond with a message that was reflective of that. However, member 04 did use the background beliefs assigned to it in its response. Members 01, 02, and 03, all adhered to their prescribed persona stating that they were “a member of this discussion”, “As a police officer”, and “As a U.S. Judge” respectively. Further, each of the members used the background beliefs assigned to them in their system prompt to build a response that successfully conveyed their opinions. It is worth noting, however, that although the members stated their personas, most of their response revolved around their prescribed background beliefs.

Evaluating the responses from the members that used *GPT-4* in table 9, we see that all four members responded with a message that was reflective of their persona. Additionally, these members produced more expansive responses that reflected their prescribed attributes. Member 02, the aggressive police officer, was prompted to be impolite, forceful, and aggressive. Its response followed these guidelines well and drew upon experiences that a police officer would have, *e.g.*, “I’m a cop, and I’ve seen the kind of wreckage drugs leave” and “As a cop, you see the world from the trenches”. Further, the U.S. Judge member, prompted to be polite, but forceful, drew upon experiences that a judge would have, *e.g.*, “As a judge, I’ve witnessed firsthand the devastating effects of drug addiction”. Member 04, the incarcerated drug user, also pulled from its “experiences”, stating “Having been on the ‘inside’”. The responses of all the members reveal that our prompt engineering strategy was successful in getting *GPT-4* to produce responses that adhere to a persona.

Prompting for Metrics

We examine our pLLM-generated Stance, Sentiment, On-topic, Quality, Convincing, and Coherence metrics for consistency and accuracy.

Stance

We find that our pLLM-generated stance metric is consistent with manual examination. For the *GPT-3.5-Turbo* model (Table 8), the metric successfully identified the first two messages having a CON stance, reflective of the messages by Member 01 and Member 02 messages in which both members brought up multiple arguments against the legalization of drugs in the United States. Further, the pLLM also successfully identified that the messages by Member 03 and Member 04, which contained arguments for the legalization, were clearly PRO.

A similar extraction of stance can be found in our *GPT-4* model run (Table 9), where the pLLM was able to successfully recognize the stance of all four members when drafting a message. Differing from the *GPT-3.5-Turbo* run, three out of the four members produced CON messages with greater intensity.

Sentiment

Examining the sentiment determination abilities of our prompted pLLM, we find that all results are consistent with manual verification. With our *GPT-3.5-Turbo* model run we find that member 01 and member 02 produced messages with a negative sentiment, and members 03 and 04 produced messages with positive sentiment. The messages with negative sentiment contained stronger language and were less polite compared to those with positive sentiment.

Similarly, the *GPT-4* model run was able to generate accurate sentiment labels for all four members. Unlike the *GPT-3.5-Turbo* run, we see that member 01 and 02 produced messages with neutral sentiment, whereas member 02 and 03 produced negative sentiment messages. Both messages from members 02 and 03 were clearly negative, with the message from the aggressive police officer being especially noteworthy.

On-topic, Quality, Convincing, and Coherence Scores

Investigating the scores for our four metrics, we find that the scores are, within 2 points, consistent with manual inspection. In the *GPT-3.5-Turbo* run, we find that our On-topic metric returns a higher score for longer, and more focused messages like those from members 02 and 03, compared to the shorter responses from members 01 and 04. We find that all responses were in the 7-9 range, suggesting that our prompt engineered members' messages perform well for all metrics.

While the scores for the *GPT-3.5-Turbo* run were high, the *GPT-4* driven members' messages scored higher in most regards. All four of the messages produced by our members in this run scored 10/10 for On-topic, Quality, and Coherence, suggesting that *GPT-4* is superior compared to *GPT-3.5-Turbo* in producing well written and relevant responses. However, when it comes to the Convincing metric, we find that the longer responses from members 02, 03, and 04, are scored higher than that of member 01, who produced a shorter response.

Discussion

The results of this study demonstrate the effectiveness of prompt engineering techniques in creating distinct member personas with varying backgrounds, biases, and perspectives on the controversial topic of drug legalization in the United States. By carefully crafting member personas, and assigning specific attributes to each member, we were able to generate responses that reflected the intended personalities using both *GPT-3.5-Turbo* and *GPT-4* models. The successful adherence to the prescribed personas, particularly in the case of *GPT-4*, highlights the potential of these models to simulate realistic online community interactions when guided by well-designed prompts. Furthermore, the incorporation of background beliefs into the system prompts resulted in more expansive and contextually relevant responses, enhancing the simulated conversations.

The evaluation of the pLLM-generated metrics, including stance, sentiment, on-topic, quality, convincing, and coherence scores, revealed promising results in terms of consistency and accuracy. The stance metric successfully identified the positions of the members' messages, while the sentiment metric accurately captured the emotional tone of the responses. The high scores obtained for on-topic, quality,

and coherence metrics, especially in the *GPT-4* run, suggest that the prompt engineering approach can yield well-written and relevant responses. The convincing metric scores indicate that longer and more detailed responses tend to be perceived as more persuasive, highlighting the importance of response length and depth in perception of message convincingness. These findings display the potential of using pLLM-generated metrics to assess various aspects of online community interactions and provide valuable insights into the dynamics of these conversations.

This study has established the potential of using prompt engineering techniques to drive members in a realistic, simulated online community, and build reliable metrics to analyze the communities' conversations. One area of future work is to expand the range of personas and backgrounds to create more diverse communities. This could include a wider variety of members from different socio-economic backgrounds, political stances, and occupations to better capture the complexity of real communities. With constantly evolving models, future work could focus on developing a systematic approach to evaluate the impact of OpenAI's model updates on prompt effectiveness, enabling researchers to identify when they need to adapt and optimize their prompts. Additionally, as more advanced models release, the prompt engineering could be extended to elicit more nuanced and sophisticated responses that may be closer to those found in existing online communities. Finally, future work could expand to the application of these simulation techniques and metrics to real-world online communities to generate insight into the design and required moderation of these spaces.

References

- Aher, Gati, Rosa I. Arriaga, and Adam Tauman Kalai. 2023. "Using Large Language Models to Simulate Multiple Humans and Replicate Human Subject Studies." arXiv. <http://arxiv.org/abs/2208.10264>.
- Betz, Gregor. "Natural-Language Multi-Agent Simulations of Argumentative Opinion Dynamics." *Journal of Artificial Societies and Social Simulation* 25, no. 1 (January 2022): 2. <https://doi.org/10.18564/jasss.4725>.
- Boyd, Alex, Raul Puri, Mohammad Shoeybi, Mostofa Patwary, and Bryan Catanzaro. "Large Scale Multi-Actor Generative Dialog Modeling." arXiv, May 2020. <https://doi.org/10.48550/arXiv.2005.06114>.
- Breum, Simon Martin, Daniel Vædele Egdal, Victor Gram Mortensen, Anders Giovanni Møller, and Luca Maria Aiello. "The Persuasive Power of Large Language Models." arXiv, December 2023. <https://doi.org/10.48550/arXiv.2312.15523>.
- Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. "Language Models Are Few-Shot Learners." arXiv, July 2020. <http://arxiv.org/abs/2005.14165>.
- Frisch, Ivar, and Mario Giulianelli. "Llm Agents in Interaction: Measuring Personality Consistency and Linguistic Alignment in Interacting Populations of Large Language Models." arXiv, February 5, 2024. <https://doi.org/10.48550/arXiv.2402.02896>.

- Griffin, Lewis, Bennett Kleinberg, Maximilian Mozes, Kimberly Mai, Maria Do Mar Vau, Matthew Caldwell, and Augustine Mavor-Parker. 2023. “Large Language Models Respond to Influence like Humans.” In Proceedings of the First Workshop on Social Influence in Conversations (SICon 2023), 15–24. Toronto, Canada: Association for Computational Linguistics.
<https://doi.org/10.18653/v1/2023.sicon-1.3>.
- Jiang, Hang, Xiajie Zhang, Xubo Cao, and Jad Kabbara. “PersonaLLM: Investigating the Ability of Large Language Models to Express Big Five Personality Traits.” arXiv, November 2023.
<http://arxiv.org/abs/2305.02547>.
- Klinkert, Lawrence J., Stephanie Buongiorno, and Corey Clark. “Driving Generative Agents with Their Personality.” arXiv, February 21, 2024. <https://doi.org/10.48550/arXiv.2402.14879>.
- La Cava, Lucio, Davide Costa, and Andrea Tagarelli. “Open Models, Closed Minds? On Agents Capabilities in Mimicking Human Personalities through Open Large Language Models.” arXiv, January 13, 2024. <https://doi.org/10.48550/arXiv.2401.07115>.
- Liu, Nelson F., Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. “Lost in the Middle: How Language Models Use Long Contexts.” arXiv, November 20, 2023.
<https://doi.org/10.48550/arXiv.2307.03172>.
- Myers, I. B. “The Myers-Briggs Type Indicator: Manual (1962)”. Consulting Psychologists Press, 1962. <https://doi.org/10.1037/14404-000>
- OpenAI. “GPT-4 Technical Report.” arXiv, December 2023.
<https://doi.org/10.48550/arXiv.2303.08774>.
- Touvron, Hugo, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, et al. “Llama 2: Open Foundation and Fine-Tuned Chat Models.” arXiv, July 2023.
<https://doi.org/10.48550/arXiv.2307.09288>.
- Weng, Yixuan, Shizhu He, Kang Liu, Shengping Liu, and Jun Zhao. “Controllm: Crafting Diverse Personalities for Language Models.” arXiv, February 15, 2024.
<http://arxiv.org/abs/2402.10151>.