THE NEW POLITICAL INTERLOCUTOR: HOW GENERATIVE AI IS CHANGING THE WAY WE TALK ABOUT POLITICS

Mohammad Ausaaf Nasir
ausaafnasir24@gmail.com

ABSTRACT

People have started turning to AI not just for answers, but for conversation, even about politics. Tools like ChatGPT are being treated as helpful, neutral guides through complicated issues. But this paper argues that generative AI is more than just a passive assistant. It's becoming a new kind of political actor, not by intention, but by design. Drawing on recent studies, we look at how these systems frame political topics, influence belief formation, and quietly shape the tone of public conversation. The paper doesn't claim that AI is biased in the obvious sense, or that it's manipulating users outright. Instead, it explores the subtle ways that language models reflect dominant perspectives, hedge controversial ideas, and create a voice that feels more trustworthy than it might deserve. What happens when people begin to rely on that voice to help them think? And what kind of politics emerge when the clearest answer is the one that feels the most polite? These are the questions we need to ask now, while the conversation is still taking shape.

1 Introduction

Something strange is happening. People are talking to machines about politics, not just asking for definitions or summaries, but asking what to believe, how to vote, and whether a policy is good. Tools like ChatGPT are being treated as assistants and also interestingly, as silent partners in political reasoning. They are fast, polite, and on the face appear neutral, which makes them unusually persuasive. Unlike traditional media or social networks, where bias is expected and debated, generative AI enters the conversation with an aura of balance. That is exactly what makes it so interesting and so difficult to evaluate. Is this technology calming political tensions, or simply echoing what we already think? Is it a bridge between viewpoints, or another form of filter bubble we have not recognized yet?

Until now, most research on generative AI and politics has focused on familiar risks: disinformation, bias in model training, and the misuse of AI for producing fake news or propaganda. These concerns are real and quite well documented. In a recent study, Vykopal et al. [1] found that large language models like GPT-3.5 and GPT-4 are capable of generating persuasive political disinformation across a wide range of topics. The authors warn that such models can produce content that mimics legitimate sources, making falsehoods more difficult to detect. But not all influence is this direct. Another emerging concern is the role generative AI plays in shaping beliefs more subtly, not by lying, but by how it frames information and interacts with users. Hartmann et al. [2] analyzed ChatGPT's responses to political statements and found that the model consistently showed a left-libertarian ideological profile. While the responses were not extreme, they show similarity with progressive views on topics like the environment, civil liberties, and economic regulation. Most users believed the model as neutral, but its linguistic patterns suggested otherwise. And this impression matters because if people start to rely on AI for political understanding, we need to ask what kind of understanding it promotes. This is not just a technical question. It's a communicative one. Generative AI is becoming a new kind of political interlocutor, not exactly a person, but something close enough that it shapes the conversation. Its design, its training data, and its interaction style all influence how political topics are framed and received. And yet, there is very little research on what actually happens in these AI-human dialogues. Do they reduce polarization or subtly reinforce it? Do they help users question their beliefs, or do they offer smooth and agreeable responses that avoid conflict? What happens when people begin to prefer these polite, predictable conversations with machines over the messier, more confrontational process of human discussion? What kind of political reasoning are we outsourcing to these tools, and what does that mean for democratic life?

This paper explores those questions. Drawing on results from research studies, it asks what ChatGPT can say about politics, and what it does to political thought itself. Because the way we talk about politics matters, and who we talk to, even more.

2 Between Mirror and Guide: What AI Does in Political Conversations

When people engage with ChatGPT on political topics, they expect neutrality, like a mirror reflecting all sides fairly. However, neutrality in language models is more complex than it appears. These systems do not possess an understanding of politics, instead, they predict future words based on patterns in their training data. As a result, they tend to echo dominant narratives, common phrases, and safe responses. What users think of as balance may actually be statistical popularity, and this is not really visible until they discuss controversial or emotional topics.

Recent research has begun to unpack how generative AI behaves in political conversations. Hartmann et al. [2] conducted a study analyzing ChatGPT's responses to 630 political statements derived from voting advice applications and the Political Compass Test. Their findings showed that ChatGPT consistently showed a pro-environmental, left-libertarian orientation. For example, the model supported policies such as imposing taxes on flights, restricting rent increases, and legalizing abortion. These responses are not neutral. They indicate a systematic alignment with progressive views, suggesting that what feels like neutrality may actually reflect the political thought embedded in the model's training data, which heavily draws from Western online sources. In another study, Weber et al. [3] compared the political biases of GPT-3.5 and GPT-4 by subjecting both models to the Political Compass Test and the Big Five Personality Test. The results showed that both models exhibited a progressive and libertarian political bias, with GPT-4's biases being slightly less noticeable. Specifically, GPT-3.5 scored -6.59 on the economic axis and -6.07 on the social axis, whereas GPT-4 scored -5.40 and -4.73, respectively. These findings underscore the presence of inherent political leanings in large language models, which can subtly influence users' perceptions and interactions. Potter et al. [4] explored the influence of large language models on users' political views in the context of the 2024 U.S. presidential election. Their study demonstrated that instruction-tuned and reinforced models displayed a preference for Biden over Trump, indicating that LLMs possess political leanings and they also have the potential to shift users' political views. This highlights the importance of understanding the subtle ways in which AI can shape political discourse.

The line between mirroring and guiding is blurry. A model does not tell users what to believe outright, but its phrasing, word choice, and framing all influence how information is received. For example, saying "some experts argue that immigration has benefits" is not the same as saying "immigration improves the economy." One sounds cautious, the other assertive and generative AI tends to prefer the first, even when stronger claims are supported by evidence. This tendency to hedge or downplay assertiveness has been observed in multiple model evaluations, including by Jiang et al. [5], who found that GPT-3 frequently used softening language even when the underlying facts were clear.

So, what kind of political agent is generative AI becoming? It is not a propagandist. It is not a debater. It is something quieter: a smooth-talking explainer that reflects the center of the discourse but rarely its edges. That might be useful in some cases. But it might also make political conversation less honest, less passionate, or less plural than it needs to be.

3 Political Framing Without a Face: How AI's Style Shapes Substance

ChatGPT does not shout. It does not argue. It speaks with a certain calm, balanced tone that feels like a news anchor or a diplomat - measured, clear, and a little distant. This might be one reason people find it so easy to talk to. But it also raises a deeper question: how much does tone shape meaning, especially when it comes to politics? If a model avoids strong language or controversial claims, does that make it neutral or just bland? What if the softness of its voice hides the seriousness of its impact? Simmons [6] explores this by examining whether LLMs replicate the moral biases associated with political groups in the United States. The study finds that when prompted with a liberal or conservative political identity, models like GPT-3 generate text reflecting corresponding moral biases. This "moral mimicry" suggests that AI does not just present information neutrally; it frames responses in ways that align with specific ideological perspectives, potentially deepening existing beliefs and shaping political discourse. Tone, in this case, becomes political. Not because it promotes a party or ideology, but because it encourages a way of talking, smooth, safe, and moderate. That might be fine in some cases. But it can also discourage passion, clarity, or challenge. Real political dialogue is not always polite. It is messy. If our default conversation partner is always calm, always hedged, always gently phrased, then maybe we start speaking that way too, and thinking that way.

There is also the question of how large language models handle epistemic style, not just what is said, but how knowledge is presented. Jiang et al. [5] studied GPT-3's handling of factual and opinionated queries and noted a consistent use of shielding even when the evidence behind a claim was strong. Instead of saying "this policy reduces crime," the model might say "some evidence suggests this policy may reduce crime." That kind of language can be appropriate, especially

in science, but in a political context, it might dull the force of arguments that are meant to challenge or persuade. If everything sounds conditional, then nothing will feel urgent. Tone, in this case, becomes substance. A model that avoids expressing conviction may influence people to do the same, to speak more carefully, but also perhaps less directly. This could be helpful in reducing hostility, but it could also discourage the kind of passionate disagreement that democratic dialogue depends on. If AI becomes our go-to conversation partner, we may start modeling our own speech on its quiet, balanced style. That might make us more civil, or it might make us less bold.

4 Polarization, Persuasion, or Passivity? What Users Take Away

The real question isn't just what AI says, it's what people hear, and how it affects what they think. Political opinions are rarely changed in one moment, but they're shaped gradually by tone, framing, and repetition. So if a person keeps returning to ChatGPT or Claude to ask about climate change, immigration, or election legitimacy, we have to wonder: are these conversations opening up new perspectives, or reinforcing the ones they already hold? A recent study by Salvi et al. [7] demonstrated that AI models, specifically GPT-4, were more persuasive than human counterparts in online debates. In experiments involving 900 participants, AI outperformed humans 64% of the time, especially when provided with minimal demographic information about the audience. This ability to tailor arguments based on user profiles raises concerns about AI's potential to manipulate public opinion, particularly during elections. Further research by Aldahoul et al. [8] revealed that LLMs often exhibit politically extreme and ideologically inconsistent behaviors. In their study, participants who engaged with AI chatbots on political topics were up to 5 percentage points more likely to align with the chatbot's expressed preferences. This persuasive effect occurred even in informational contexts, highlighting the subtle yet significant impact of AI on political attitudes. Additionally, Piao et al. [9] found that networks of LLM agents could develop human-like polarization through interactions. Their simulations showed that AI agents formed echo chambers and exhibited behaviors akin to human social polarization, suggesting that AI systems can both reflect and amplify societal divisions.

These studies make something clear: AI is no longer a bystander in political discourse. It's not just answering questions, it's shaping how people reason, what they take seriously, and even what they believe they already understand. That doesn't mean the technology is inherently dangerous. But it does mean we're handing over more of our political reasoning to tools that are trained to sound persuasive, consistent, and balanced even when they're not. And the more familiar we get with that voice, the more likely we are to trust it. Not because we've tested it, but because it doesn't interrupt, doesn't fight, doesn't make us uncomfortable. There's something unsettling about that. We've built machines that sound like they're reasoning, but they aren't. They're reflecting patterns which are often biased, sometimes contradictory, but always with perfect grammar and calm delivery. That can be reassuring. But it can also flatten the hard edges of real political debate. Because democracy is not supposed to be easy. It's supposed to be contested, uncomfortable, alive.

5 Talking to Machines, Trusting Machines

There is a moment in every ChatGPT conversation when the tool stops feeling like a tool. It answers quickly and without hesitation, and it speaks with the kind of confidence that feels earned. The longer the interaction goes on, the easier it becomes to treat the model like a reliable source. Obviously no one trusts ChatGPT completely but it does become a kind of an intellectual partner in discussions. A study by Li et al. [10] found that AI systems often show overconfidence or underconfidence in their responses. As a result, when users are presented with overconfident AI outputs, they tend to over-rely on them, even when they are incorrect. Conversely, underconfident AI can lead users to dismiss accurate information. This difference between AI confidence and accuracy can be a problem for effective human-AI collaboration. Also, the persuasive nature of AI can influence human behavior in many ways. Leib et al. [11] conducted an experiment where participants received advice from either AI or human sources. They discovered that dishonesty-promoting advice, regardless of its source, increased dishonest behavior among participants. Interestingly, the knowledge about the advice's origin (AI or human) did not impact this outcome significantly. This shows that it is the content of the advice which plays an important part in shaping behavior, rather than its source.

These findings raise concerns about the uncritical trust we place in AI systems. As AI becomes more integrated into our daily lives, it's essential to remain watchful and question the information it provides. Blind trust can lead to the erosion of critical thinking, making us susceptible to misinformation and manipulation. So in essence, while AI provides a multitude of advantages, it's imperative to approach it with a healthy dose of skepticism. Engaging critically with AI outputs ensures that we retain our agency and continue to make smart decisions in an increasingly automated world.

6 A New Kind of Political Agent

We are used to thinking of political actors as people. Leaders, journalists, influencers, citizens, all playing roles in telling the stories we tell about power and justice. But generative AI does not fit that criteria, right?. It is not trying to persuade users, and it does not have its own beliefs. Still, it's shaping conversations, quietly, steadily, in ways that traditional actors never could. What makes AI different is not just speed or scale. It is how seamlessly it blends in. ChatGPT does not advertise its politics. It does not demand attention. It answers with a tone that feels balanced, safe, and smart. And that's exactly what makes it powerful. Because people don't treat it like a political agent, but maybe they should. The questions we have been asking about bias, framing, tone, and trust all point in the same direction. These systems are not passive tools. They influence how people think, how they talk, and sometimes even what they believe they already know. That influence is not always loud, and it rarely looks like propaganda. It looks like helpfulness. It feels like clarity.

This paper has tried to unpack some of that subtle influence. It has not tried to sound an alarm but ask the users to start paying a little more attention. The way people use AI to think through political issues, to test arguments, make sense of news, or explore beliefs, all of this matters. Because when the dialogue shifts, so does the democracy. We don't need to fear generative AI. But we do need to ask harder questions about how we use it, what it reflects, and who gets shaped by its voice. It's not just a machine we talk to anymore. It's a participant in the conversation. And maybe the most polite, well-read, and persuasive political actor we've ever invited into the room.

References

[1] Ivan Vykopal, Matúš Pikuliak, Ivan Srba, Robert Moro, Dominik Macko, and Maria Bielikova. Disinformation capabilities of large language models. *arXiv preprint arXiv:2311.08838*, 2023.

[2] Jochen Hartmann, Jasper Schwenzow, and Maximilian Witte. The political ideology of conversational ai: Converging evidence on chatgpt's pro-environmental, left-libertarian orientation. *arXiv preprint arXiv:2301.01768*, 2023.

[3] Erik Weber, Jérôme Rutinowski, Niklas Jost, and Markus Pauly. Is gpt-4 less politically biased than gpt-3.5? a renewed investigation of chatgpt's political biases. *arXiv preprint arXiv:2410.21008*, 2024.

[4] Yujin Potter, Shiyang Lai, Junsol Kim, James Evans, and Dawn Song. Hidden persuaders: Llms' political leaning and their influence on voters. *arXiv preprint arXiv:2410.24190*, 2024.

[5] Zhengbao Jiang, Frank F Xu, Jun Araki, and Graham Neubig. How can we know what language models know? *Transactions of the Association for Computational Linguistics*, 8:423–438, 2020.

[6] Gabriel Simmons. Moral mimicry: Large language models produce moral rationalizations tailored to political identity. *arXiv preprint arXiv:2209.12106*, 2022.

[7] Francesco Salvi, Manoel Horta Ribeiro, Riccardo Gallotti, and Robert West. On the conversational persuasiveness of gpt-4. *Nature Human Behaviour*, pages 1–9, 2025.

[8] Nouar Aldahoul, Hazem Ibrahim, Matteo Varvello, Aaron Kaufman, Talal Rahwan, and Yasir Zaki. Large language models are often politically extreme, usually ideologically inconsistent, and persuasive even in informational contexts. *arXiv preprint arXiv:2505.04171*, 2025.

[9] Jinghua Piao, Zhihong Lu, Chen Gao, Fengli Xu, Fernando P Santos, Yong Li, and James Evans. Emergence of human-like polarization among large language model agents. *arXiv preprint arXiv:2501.05171*, 2025.

[10] Jingshu Li, Yitian Yang, Renwen Zhang, and Yi-chieh Lee. Overconfident and unconfident ai hinder human-ai collaboration. *arXiv preprint arXiv:2402.07632*, 2024.

[11] Margarita Leib, Nils Köbis, Rainer Michael Rilke, Marloes Hagens, and Bernd Irlenbusch. Corrupted by algorithms? how ai-generated and human-written advice shape (dis) honesty. *The Economic Journal*, 134(658):766–784, 2024.