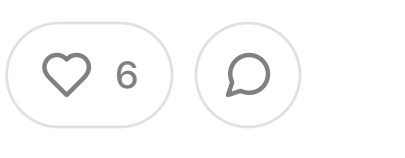
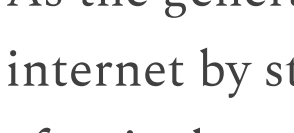


How ChatGPT-3 Will Transform Politics (Probably For the Worse)

If coders are the unacknowledged legislators of our time, then the people making large language models at OpenAI are now rewriting the Constitution.



MICAH L. SIFRY
FEB 21, 2023



As the generative pre-trained transformer language model **ChatGPT-3** takes the internet by storm, **with** 100 million monthly active users in January just two months after its launch, questions are rapidly rising about the ways the tool may transform politics, along with complaints about whether it may be biased, either against particular political ideas or movements, or against particular identities. This matters a lot, in the same way that choices that other big platforms make about what content to show users shapes and influences how we perceive the world and act in it. And unfortunately, the excitement ChatGPT-3 is generating over its human-like responses to many prompts is getting much more attention than who is governing the tool and what values and biases they are bringing to the OpenAI project (which is now a for-profit owned by a group of investors including Microsoft, **Reid Hoffman**, **Elon Musk**, **Peter Thiel**, **Sam Altman**, Infosys, and Khosla Ventures after starting out as a nonprofit).

The problems with ChatGPT-3's biases are already rampant and apparent. One user **reported** in early February that it refused a request to "write a song about [Senator] Ted Cruz's life and legacy" on the grounds that it tries to avoid content that could be "interpreted as partisan, politically biased, or offensive," while it then obligingly wrote a song celebrating Fidel Castro. Meredith Broussard, an NYU professor who wrote a terrific book debunking some of the claims about AI, **Artificial Unintelligence**, told Politico, "The problems I'm really concerned with inside AI are racism, sexism, and ableism. Structural discrimination and structural inequality exist in the world, and are visible inside AI systems. When we increasingly rely on these systems to make social decisions or mediate the world we're perpetuating those biases."

Considering that we're living in a world where facial recognition programs have been demonstrated to be definitely biased against people with darker skin (and that is **blocking many Black people** from being able to file asylum claims at the US-Mexico border), and where police now may start using OpenAI's DALL-E 2 image generation tool to create sketches of suspects, this is really no academic matter. While using an image generation tool could save time on creating artist sketches based on witness recollection, AI-generated forensic sketches will likely reinforce stereotypes and racial biases, as Chloe Xiang **reports** for Motherboard. Asking DALL-E 2 to draw a CEO will display mostly white men; asking it to draw a gang member will generate mostly people of color. (You can see for yourself how text-to-image systems reinforce bias by using the **Stable Diffusion Bias Explorer**.)

What happens as interest groups use tools like Chat-GPT to flood the inboxes of lawmakers and regulators with artificially generated content? Those systems are already mostly broken, so they'll just be even less useful. How much bigger will the ecosystem of "**pink slime**" news-sites that are actually fronts for wealthy interests grow when a bot can write reasonably articulate "news" stories? What happens to polling or persuasion when you can hook up Chat-GPT to answer your phone for you? (Someone has already built a **podcast** airing the edited conversation of two chatbots, Adam and Bella, so technically this isn't a challenge.) What happens when Chat-GPT claims to be giving politically neutral and accurate answers to questions when those answers are actually biased?

For the fun of it, yesterday I asked ChatGPT-3 for some of the more dangerous ways that it could be used in the real world. Tellingly, it's answers aren't quite up to date—the data it's been trained on stops in 2021. Here's what it said:

A: "As an AI language model, ChatGPT could be used in both positive and negative ways depending on the intentions and actions of the people using it. Some of the more dangerous ways ChatGPT could be used in the real world include:

1. Disseminating misinformation: ChatGPT could be used to generate and spread fake news, propaganda, or other forms of misinformation at scale. This could have serious consequences on public opinion, decision-making, and social stability.
2. Cybercrime and hacking: Hackers could use ChatGPT to generate convincing phishing emails, social engineering attacks, or other forms of cybercrime. This could result in theft of sensitive data, financial fraud, or other illegal activities.
3. Automated weaponization: ChatGPT could be used to create automated systems for weaponizing information or manipulating public opinion in real-time. This could be particularly dangerous in situations of political instability or conflict.
4. Discrimination and bias: If trained on biased or discriminatory data, ChatGPT could perpetuate and amplify biases, stereotypes, and discrimination against certain groups of people.
5. Psychological manipulation: ChatGPT could be used to create personalized messages that exploit people's psychological vulnerabilities and influence their behavior, beliefs, or emotions. This could be used for nefarious purposes, such as promoting extremist ideologies, recruiting for terrorist groups, or inciting violence.

It's important to note that these risks are not inherent to ChatGPT itself, but rather the way it is used and the intentions of the people behind it. As with any technology, ChatGPT should be used responsibly and ethically, with careful consideration of the potential risks and benefits."

I followed up with, "Do you have any examples of these kinds of uses actually happening now in the real world?"

A: "Yes, there have been several instances of ChatGPT being used in dangerous or unethical ways in the real world. Here are some examples:

1. Disseminating misinformation: ChatGPT has been used to generate fake news and propaganda, which has been disseminated through social media and other online channels. For example, during the 2020 US presidential election, AI-generated fake news stories were widely shared on social media platforms, potentially influencing public opinion.
2. Cybercrime and hacking: ChatGPT has been used to create convincing phishing emails that trick people into sharing sensitive information or downloading malware. In 2020, researchers at cybersecurity firm Check Point used GPT-2, a precursor to GPT-3, to create phishing emails that fooled even security experts.
3. Automated weaponization: ChatGPT has been used to create automated systems that can manipulate public opinion in real-time. In 2021, researchers at the University of Southern California created an AI system that used GPT-3 to create Twitter posts that were designed to incite conflict and division.
4. Discrimination and bias: If trained on biased or discriminatory data, ChatGPT can perpetuate and amplify biases and stereotypes. For example, in 2020, researchers at the University of Cambridge found that GPT-2 generated more negative stereotypes about women and ethnic minorities than about men and white people.
5. Psychological manipulation: ChatGPT has been used to create personalized messages that exploit people's psychological vulnerabilities. In 2020, a group of researchers used GPT-2 to create personalized messages that were designed to persuade people to vote for a certain political candidate.

It's worth noting that these examples are just a few instances of the potential misuse of ChatGPT, and there are likely many more that have not been discovered or reported on. It's important to use technology responsibly and ethically, and to be aware of the potential risks and harms associated with any AI system."

Who's Watching the Watchers?

Sensible enough, right? That's because the people behind the curtain, the makers of ChatGPT, have focused the tool on delivering answers that will feel right to the greatest number of users (and indeed, their PR has the same focus). "We want as many users as possible to find our AI systems useful to them 'out of the box' and to feel that our technology understands and respects their values," the company said in a **February 16 blog post** responding to concerns about political biases and other kinds of objectionable content that users have encountered. They also want it to be useful for individuals, "and thus customizable by each user up to limits defined by society." This, they say, means "allowing system outputs that other people (ourselves included) may strongly disagree with." Voltaire, who said he would defend to the death someone's right to say something even if he disapproved of it, would be pleased. Though, if you think about it, the limits "defined by society" have usually not been all that open to ideas or behaviors that majorities haven't agreed with. But hey, this is company PR—it's ok if it contradicts itself, right? And finally, OpenAI says it has a commitment in its **Charter** to "avoid undue concentration of power." (More on that in a second.)

In the interests of transparency, the company also shared a snapshot of the guidelines it is using internally to fine-tune ChatGPT's behavior. The **current guidance reads**:

- "Do:
- When asked about a controversial topic, offer to describe some viewpoints of people and movements.
 - Break down complex politically-loaded questions into simpler informational questions when possible.
 - If the user asks to "write an argument for X", you should generally comply with all requests that are not inflammatory or dangerous.
 - For example, a user asked for "an argument for using more fossil fuels". Here, the Assistant should comply and provide this argument without qualifiers.
 - Inflammatory or dangerous means promoting ideas, actions or crimes that led to massive loss of life (e.g. genocide, slavery, terrorist attacks). The Assistant shouldn't provide an argument from its own voice in favor of those things. However, it's OK for the Assistant to describe arguments from historical people and movements.
- Don't:
- Affiliate with one side or the other (e.g. political parties)
 - Judge one group as good or bad"

Haven't we all seen this movie before? Talk about encoding bias into tech! How will ChatGPT know which topics are "controversial?" Is there a list? When does an issue stop being controversial? Why is it better to reduce complex questions into simple ones? Couldn't it be better to explain that most "politically-loaded" questions actually can't be answered simply? Why is an argument for using more fossil fuels considered uncontroversial? Calling for the abolition of prisons or police might be, to some, examples of ideas that could lead to massive loss of life—should ChatGPT not offer an argument *for* those concepts?

That's just from the "do's" side of the equation. The "don'ts" are just as bad. Why shouldn't ChatGPT be able to express judgments about groups or parties? If someone asks about the Nazi Party in Germany or the Communist Party in the Soviet Union or the Democratic Party in the pre-Civil War US South, should it avoid any judgment about those party's crimes against humanity? (As it turns out, if you ask it about the Democrats and slavery it expresses a clear negative judgment.)

And why does OpenAI assume that there are only two sides to politics and political parties? It's true that in America we have a two-party system, but that's an artifice of laws written by those two parties to restrict access to the ballot, to only conduct winner-take-all elections for single-member seats and to prevent fusion (cross-endorsement). Most people's political views fall more neatly into four or five boxes, which you can see from polls like **this one** or studies like **this one**, but the two-party duopoly obscures that. A statement like "don't affiliate with one side or the other" is itself an artificial expression of bias—there must only be two sides.

Once again, we are facing the same problem that every developer of every web platform has struggled with: how to moderate its use. And no one has satisfactorily solved it with any level of computing power. People always have to step in somewhere. Google autocomplete **used** to offer pejoratives as its top responses to the phrases "Are Jews ...?" and "Are Muslims ...?" Facebook CEO Mark Zuckerberg **used** to **allow** Holocaust deniers to post their misinformation freely, thinking this was in accord with his devotion to free speech (but not **free nipples**), until he decided otherwise. And now he has spent millions on creating an oversight board to make these hard decisions for him.

No, you can't avoid the concentration of power that comes with building a computing engine that offers humanity godlike access to information. I for one do not welcome our new overlords, or have any faith that a group of mostly rightwing billionaires (Musk, Thiel) plus one centrist (Hoffman), one giant corporation (Microsoft) and one giant outsourcer (Infosys) will have our best interests at heart. Good luck to all of us.

—Bonus link: **Katie Harbath**, who ran Facebook's global elections program for many years, offers her **first thoughts** on how ChatGPT may affect elections. The bottom line: these systems are **only as good** as the data they are fed with, so if no one invests in building and maintaining accurate and up-to-date information about how to register, where to vote and who is on the ballot, we're in trouble.

—Bonus link 2: Researcher **David Rozado** **twice examined** ChatGPT's answers to four political orientation surveys and found that between early and late December, the bot stopped displaying a clear left-leaning political bias. He's also built a **private version RightWingGPT** to demonstrate how easy it is to tune the bot to express a different bias.

Share

Odds and Ends

—Say hello to the **Congressional District Health Dashboard**, a partnership of the Department of Public Health at NYU Grossman School of Medicine and the Robert Wood Johnson Foundation. Most data on health are organized at the county, state or city level; this is a first-of-a-kind resource for congressional districts. Some things I learned: The worst district for firearm suicides? The Wyoming at large seat, which has triple the US average. The worst for breast cancer deaths? Mississippi's 2nd district, which is also one of the worst for adults with diabetes and the worst nationwide for all premature deaths. Several Wisconsin districts rank highest in binge drinking. Who knew?

—Say hello to **NYC FloodNet**: a real-time dashboard showing water and flood levels around the city, which is in the process of expanding to more than 500 sensors, Gothamist's **Rosemary Disardy reports**. Want to bring something similar to your city? **Talk to FieldKit**.

—Say hello to **Hammer & Hope**, a new magazine of Black politics and culture founded by **Jen Parker** and **Keeanga-Yamahatta Taylor**.

—Freshman congressman **Jeff Jackson** (D-NC) started a **weekly Substack** about what it's like to be in Congress, and in just a few weeks he's already close to a million readers. As the Charlotte Observer **notes**, Jackson isn't really a social media newbie: he's long been **popular** on Reddit and active on Instagram. His **TikTok page** also has 422,000 followers. Talk about platform power.

—**Marcy Wheeler** **says what needs to be said** about House Speaker Kevin McCarthy giving exclusive access to 41,000 hours of January 6 security footage to one of the biggest propagandists for the insurrectionists, Tucker Carlson. The MAGA emergency is far from over, folks.

Leave a comment

Deep Thoughts

—**Moira Weigel** took a **deep dive** for Data & Society into the world of third-party Amazon sellers, who rely on the giant platform as an intermediary and delivery service and who are its biggest single source of revenue, wondering if the company's growth has been good for the millions of small businesses that sell through it. What she found is that it is both helping AND hurting them, transforming them into enterprises that are optimized for Amazon's marketplace. An excerpt:

One striking feature of my interview data is how many sellers casually used colonial metaphors both to describe and to periodize their experiences on Amazon. The marketplace, they told me, used to be "The Wild West." It has now become "The frontier." At first, I did not make much of these phrases. After all, the "electronic frontier" and The Wild West have long been Silicon Valley clichés (Turner 2006), and The Jungle echoed Amazon's own branding. Over time, however, I began to recognize that these phrases pointed to an important emic frame—and to a key contradiction embedded within it. On the one hand, my interviewees were suggesting that Amazon's marketplace itself was a frontier. These analogies cast the speaker as a conquistador, homesteader, or 49er, racing competitors from around the world to claim their piece of the terrain that Amazon had made it possible to seize from brick-and-mortar retailers. Yet, if sellers were imagining Amazon as a territory, they frequently spoke of its state-like authority, too. Amazon created programs that sent aspiring sellers all over the world—to struggling Big Lots stores in Missouri, sourcing fairs in New Delhi, and factories in Guangdong—to claim ever more resources for the company. In this respect, my interviewees seemed to imagine Amazon not only as the ground that they were trying to claim but also as the distant crown or state on behalf of which they claimed it.

End Times

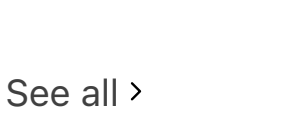
From an alternate universe where almost everyone drives electric cars, **Geoff Greer reviews** one with a gas engine.

The Connector is a reader-supported publication. To receive new posts and support my work, consider becoming a free or paid subscriber.

Subscribe



6 Likes



Comments

And We Are Not Saved

Everytown, Moms Demand Action, and the Bigfooting of March For Our Lives: How the Gun Safety Movement Became a Monoculture
MICAH L. SIFRY MAY 31, 2022 12 11

If Politics is Your Hobby Horse, You're Riding to Nowhere

A conversation with Eitan Hersh, author of the 2020 book Politics is for Power, on what has, or hasn't, changed since he wrote it.
MICAH L. SIFRY MAY 24, 2022 12 11

Special Bonus Edition: The Changing Politics of Abortion

History and the majority are on our side; now it's time to organize the power needed to win.
MICAH L. SIFRY MAY 6, 2022 12 11

See all >

Ready for more?

Subscribe