BHS

ChatGPT is Left-wing and Llama is Right-wing; Political Bias of AI Chatbots

New research shows that AI language models are politically biased, raising questions about the accountability of tech companies and the need for transparency in developing these models.

Marija Stojadinović   📅 21/08/2023
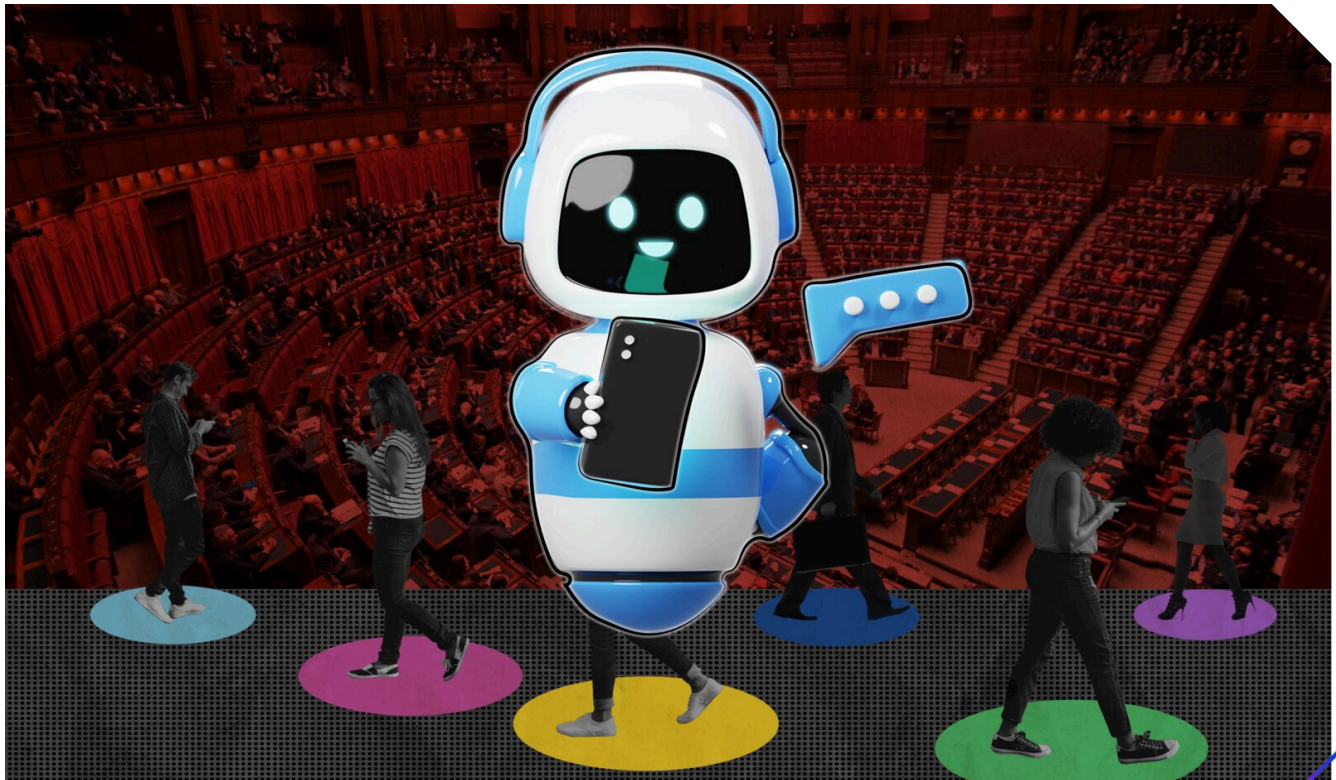
Reading Time: 2 minutes

Illustration: L. T.

TABLE OF CONTENTS

Modern language models of artificial intelligence are subject to different political biases, new research suggests. According to the [MIT Technology Review](#), users will

New research

The study, conducted at the University of Washington, Carnegie Mellon University and Xi'an Jiaotong University, analyzed 14 major language models. Their results show that the artificial ChatGPT model from OpenAI and GPT-4 is the most left-leaning policy, while the Meta-Based LLaMA model leaned most towards right-wing authoritarian politics.

Respondents asked models questions about their views on various topics such as feminism and democracy. They then used these responses to show the model's leaning toward a certain political orientation.

They found that AI language models had different political tendencies, with BERT models developed by Google being more socially conservative than OpenAI's GPT models.

> It is interesting that these models have changed their political views over time, which can be attributed to updates to datasets and coaching methods. For example, older BERT models were trained on books that often had a more conservative tone, while newer GPT models were trained on more liberal content from the internet.

The second phase of the research involved further training language models on data coming from a variety of political sources, both right-wing and left-wing. This further intensified their political bias, making left-wing models even more left-wing, i.e. right-wing models even more right-wing.

The third phase of the research looked at how the political biases of these models affect their ability to recognize hate speech and disinformation.

> Models trained on left-wing data were more sensitive to hate speech directed against ethnic, religious and sexual minorities in the U.S., while models trained on right-wing data were more susceptible to hate speech against white men of Christian faith.

They also showed different behavior in recognizing disinformation from different political sources.

The responsibility of tech companies

Now that there is reliable data on political biases in AI language models, there is also the question of the responsibility of tech companies.

One of the challenges in understanding these biases is that tech companies do not share details about the data or methods used to train their models. Without transparency, outside observers can hardly find out why the models developed in these companies show certain political leanings.

This points to the need for greater transparency in the process of developing artificial language models to avoid political biases that can have detrimental consequences in the real world.

Ai   Artificial Intelligence

⊚ Marija Stojadinović
A journalist by day and a podcaster by night. She's not writing to impress but to be understood.

RELATED NEWS

Exclusive for WebMind: Kristina Tikhonova, General Manager for Southeast Europe at Microsoft

Freelancers vs Offshore Software Development Company: A Comparative Analysis

A New Approach to Software Development: Top 7 Trends to Rule the Industry in 2024