

Political Bias in Large Language Models: A Comparative Analysis of ChatGPT-4, Perplexity, Google Gemini, and Claude

Tavishi Choudhary

Greenwich High, Greenwich, Connecticut, US

Abstract. Artificial Intelligence large language models have rapidly gained widespread adoption, sparking discussions on their societal and political impact, especially for political bias and its far-reaching consequences on society and citizens. This study explores the political bias in large language models by conducting a comparative analysis across four popular AI models—ChatGPT-4, Perplexity, Google Gemini, and Claude. This research systematically evaluates their responses to politically charged prompts and questions from the Pew Research Center’s Political Typology Quiz, Political Compass Quiz, and ISideWith Quiz. The findings revealed that ChatGPT-4 and Claude exhibit a liberal bias, Perplexity is more conservative, while Google Gemini adopts more centrist stances based on their training data sets. The presence of such biases underscores the critical need for transparency in AI development and the incorporation of diverse training datasets, regular audits, and user education to mitigate any of these biases. The most significant question surrounding political bias in AI is its consequences, particularly its influence on public discourse, policy-making, and democratic processes. The results of this study advocate for ethical implications for the development of AI models and the need for transparency to build trust and integrity in AI models. Additionally, future research directions have been outlined to explore and address the complex AI bias issue.

Keywords: Large language models (LLM), Generative AI (GenAI), AI Governance and Policy, Ethical AI Systems

1 Introduction

What began as theoretical concepts have now emanated and merged into rapid developments in the Artificial Intelligence models that are an inherent part of modern technology. AI innovations have started to reshape industries worldwide; their presence extends to health, finance, politics, governance, and public policy [1]. Indeed, the U.S. Generative AI market is projected to grow rapidly, reaching \$4.1 billion by 2024 and expanding with a compound annual growth rate (CAGR) of 36.3% through 2030, driven by critical technologies such as transformers and diffusion networks Fig. 1. This exponential growth shows very well the increasingly important role that AI can play in enhancing the potential for efficiency improvement, better decision-making, and creating new opportunities in both the private and public sectors.

The emergence of generative AI models has been one of the most fundamental changes to AI technology in recent years. In 2023, the generative AI market reached \$44.89 billion globally due to increased adoption across sectors such as marketing, customer service, healthcare, and entertainment. For instance, various AI models today are core to how businesses are run, whether for automation, natural language processing, or predictive analytics. Their growing power has opened new avenues for creativity but, at the same time, raised very important questions concerning the ethical and social consequences of such technologies.

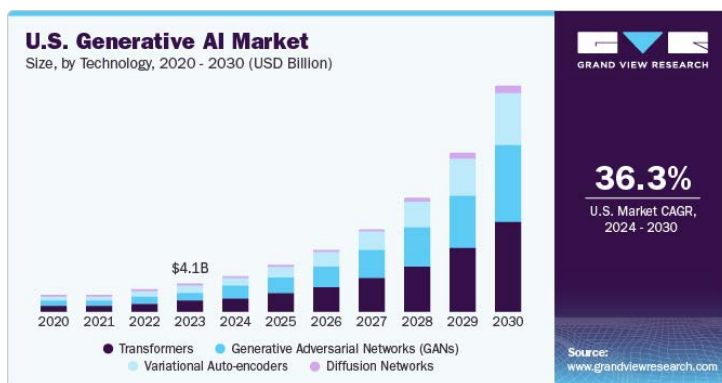


Fig. 1. U.S. Generative AI market growth by technology [61]

With the rapid integration of AI into life, one increasingly hears concerns about bias. There can be several reasons for bias in these AI models: data upon which the models are trained, designed algorithms, and contexts in which these AI systems are deployed. Noble [3] reminds us that AI systems- including search engines- are vulnerable to magnifying existing social biases- if for nothing else because that's what the biased training data have taught them. This becomes most sensitive when political biases are at issue: AI models are being used increasingly to shape public discourse. With these generative AI tools-ChatGPT-4 and Google Gemini-increasingly available, the net result on political opinions and socialization may be determined by the nature of information that the former will give to the latter.

Bias in AI is not only a technical issue but also a societal one. In support, O'Neil [4] opines in her work "Weapons of Math Destruction" that biased algorithms tend to embody corrosive prejudices that undermine democratic processes and polarize society. The consequences are thus profound from a political point of view: biased AI technologies shape and mold public opinion, influence political behavior, and even affect voting. This underlines how robust and unbiased AI systems are needed for all those in actual contact with the public. These can profoundly impact how people will inform themselves and interpret information.

It is further complicated by the fact that the algorithmic design of AI systems is viewed in polarized terms. Obermeyer et al. [5] have demonstrated how health algo-

gorithms amplify racial biases; the same dynamics can create political biases in AI systems. Diversity of perspective is essential at the point of data collection and in the design of algorithms as AI tools are being rolled out in politically sensitive areas.

To overcome these challenges, developers should implement measures that guarantee inclusion in diverse data by exposing AI models to various perspectives during training. Auditing and regular updating are necessary to keep AI aligned with shifting social norms. Other techniques, like adversarial training, aim to reveal and correct biases by the intentional addition to the training process of the model instances that introduce problematic scenarios. Transparency relating to AI in development is of equal importance. Educating users about AI systems' possible biases and limitations ensures the fostering of responsibility in using them and enhances the critical usage of these technologies [7]

Further research will be needed on how biases are introduced into AI systems and how frameworks for mitigating such biases need to be developed. Brundage et al. [11] call for extensive frameworks that address ethical AI development, whereas Morley et al. [17] indicate that methodologies originating from healthcare ethics can be used to battle AI models' political biases. Longitudinal studies might investigate how biases evolve in AI systems and develop methods for more effective adaptive mitigation strategies. This work investigates four major AI models: ChatGPT-4, Perplexity, Google Gemini, and Claude. All these systems have some degree and nature of political bias inherent in their architecture. Both quantitative and qualitative methods were used to explore the mechanism, which is not so apparent but gives much-needed insight into how these biases can affect politically sensitive environments. Our aim is that through this analysis, actionable recommendations can be made for the development of more balanced and equitable AI systems. Ultimately, these can help achieve more informed and inclusive public discourse.

2 Literature Review

Artificial Intelligence models have become integral enablers in shaping online interactions and decision-making processes. More particularly, with the ever-increasing use of AI models for generating text and multimedia content, especially Large Language Models and Generative AI, there are fast-growing concerns about political bias in such efforts. These can affect public discourses, decision-making processes, and democratic processes. The literature review discusses the sources of political bias in AI, its implications, and the mitigation strategies to present the content generated by AI as non-discriminatory and neutral, hence trustworthy.

2.1 Bias in AI and ML

The roots of Artificial Intelligence go back to the mid-20th century, and the term "Artificial Intelligence" was coined officially at the Dartmouth Conference in 1956 [14]. As its subfield, Machine Learning gained full attention in the 1980s and focuses on systems able to learn from data without explicit programming [2]. AI technologies have

been evolving rapidly over the past couple of decades. Recently, the emergence of LLMs such as ChatGPT-4, Google Gemini, Perplexity, and Claude has further expanded the sphere of influence of AI. However, this also increases the risk of bias within these very same AI technologies. Recent studies have shown that LLMs can reflect significant political and social biases, depending on the data used for training and the underlying algorithms that govern their outputs.

Wu et al. [21], in a far-reaching bias analysis about 2023 LLMs, including ChatGPT-4, demonstrate that these models often have the very same biases as in their training datasets. In another work, Kumar & Lopez [22] explored political bias in Google Gemini. While it generally favors centrism, it tends to lean toward more liberal positions while fine-tuning user feedback. These findings emphasize understanding the influence of training data and feedback loops that can further solidify biases in AI outputs.

Bai et al. [36] give an example of how LLMs adapt to specialized domains by pointing out model fine-tuning and domain-specific datasets may affect performance and eventual biases. They indicate, however, that while fine-tuning improves the performance of LLMs within given contexts, it also contributes to the gains made by existing social inequalities. It enhances an already orchestrated, ramped-up version of political bias. Taking this further still, Jones et al. [23], in 2024, considered how LLM adaptation to domains including but not limited to health and law demonstrated variable levels of bias based on the datasets used for training.

Shen et al. [37] propose a critical discussion of ethical issues raised by GenAI applications in multimodal contexts, where, for example, text generation is combined with image generation. For politically sensitive content, biased content generation can be leveraged by vested interests for misinformation campaigns. These findings were further supplemented by Li & Patterson [24] in 2023, who spoke to how AI-generated content might be leveraged in targeted political campaigns with the potential for increased polarization.

Beyond the challenges in training data and fine-tuning, Diakopoulos [7] underscored the accountability of algorithmic decision-making as AI systems go live in public-facing applications. AI models perpetuate political and social inequalities without proper transparency and accountability mechanisms. Feng & Huang [25], in 2024, called for robust auditing frameworks that would trace and stem the biases of LLMs within high-stake decision-making environments like legal systems and healthcare.

Crawford & Calo [6] long argued that AI research and development should be more transparent. They say the "black box" nature of so many AI models prevent the effective detection and mitigation of bias. In 2023, Garcia et al. [26] reiterated these concerns, asking for international collaboration to ensure that AI governance produces strategies that effectively mitigate bias and will be widely adopted. They advocated making global ethical standards for AI and foresaw that including diverse perspectives would offset several biases naturally baked into data gathering and model construction.

Martin & Thompson [38] discussed the economic impacts of LLMs on the global workforce. The findings showed that biases within the models could drive hiring decisions, inform economic policy, and extend into general workforce dynamics, especially in a world where AI systems are finding their place in decision-making. In extending such a view, Singh & Patel [27] considered the long-term ramifications of biased AI

systems on society in general, particularly on the disadvantaged sections of society where such AI-engineered policies may only further prolong existing injustices.

The literature increasingly points out that continuous monitoring and ethical oversight are needed with the advancement of AI technologies. Murphy et al. [28], in 2023, proposed a multi-layered approach toward mitigating bias in AI; this has included performing bias audits, making transparency reports, and using "debiasing algorithms" to minimize the impact of prejudicial data on model output. They say this multi-pronged approach will be necessary when LLMs like ChatGPT-4, Perplexity, and Google Gemini continue to shape global public discourse and decision-making.

2.2 Understanding Political Bias

The sources of political bias in AI include biased data for model training, the algorithm design itself, and the usage contexts in which these systems are put to work. Diakopoulos echoed with an emphasis on accountability, positing that AI biases show up as societal injustices when not well tended to. More recent studies focus on the political bias inherent in large language models and generative AI.

Influence of Training Data on Political Bias: Indeed, several works have proved that unbalanced representations of political views are a common outcome of enormous web scraping used to train LLMs. For instance, Abid et al. 2021 [50] established that GPT-3 had serious biases against some religious and ethnic groups due to the data on which this model was trained. Weidinger et al. [51] showed in 2023 that LLMs can have political biases from their training datasets flowing into the output, favoring one political ideology over another.

Model Architecture and Political Bias: The very architecture of AI models can also be a reason for political bias. Indeed, Bommasani et al. (2021) [52], in their study about foundation models, proved that the unprecedented scale and generality of these models unintentionally encode and amplify the existing societal biases, including the political ones. Also, the same is captured in Luccioni and Viviano [53], 2023, who studied multilingual language models for biases and found that even neutral models still held the ability to hold political biases due to their architecture.

User Interaction and Reinforcement of Bias: Recent research has also begun to explore how user interactions with AI models can restore and amplify political bias. Indeed, in 2023, Schramowski et al. [54] showed how AI models may adapt to user inputs in a way that amplifies existing biases when users unconsciously influence the model toward an ideologically charged output. This type of interaction bias underlines the need to comprehend static biases in models and the dynamic ways in which they can evolve in their use.

Impact on Public Opinion and Discourse: The potential of GenAI to influence public opinion is very high. A 2023 study by de Vries et al. [55] considered the ways in which AI-generated text may influence readers' perception of political issues and called for increased awareness of the use of AI in propagating propaganda or misleading information. The authors emphasized that the detection and labeling of AI-generated content are mechanisms that need to be in place to protect the integrity of public discourse.

2.3 Sources of Political Bias

Training Data: Training datasets can be considered a prime source of bias in AI models. Noble also cited that search engines and AI systems carry forward the bias of the society through the data ingested. Recently, in 2023, Birhane et al. [56] have done an extensive review in large-scale datasets which are being used for training LLMs and often suffer from a lack of diversity with the excessive presence of political ideologies. This may result in the generation of outputs that marginalize the minority view.

Algorithmic Design: The design of algorithms is considered one of the significant causes of perpetuating political bias. Obermeyer et al. [5] discussed how algorithms reinforce racial biases; the principles also extend to political biases. In 2023, Wang and Russakovsky [57] studied how fine-tuning practices can inadvertently introduce biases into the AI models. They concluded that if not carefully calibrated; fine-tuning might make models favor the politically dominating views of the fine-tuning dataset.

Deployment Context: The context in which AI systems are deployed can also magnify their biases. Crawford & Calo pointed out the lack of diversity in the development of AI. Aga and Brynjolfsson [58], in a study published in 2023, analyzed how the deployment of AI on social media platforms could lead to echo chambers in which users are only exposed to content that reinforces their current beliefs. These might be further exacerbated by AI algorithms that make choices about what content to display to an individual based on their interests, potentially leading to further politicization.

Ethics in AI AI heavily relies on the training data used to train the models with machine learning. For training an AI model, especially a deep learning model, a huge volume of data, including personal and private data, is required. Misuse of data, such as leakage and tampering of information, are serious ethical issues that are closely related to individuals, institutions, organizations, and even countries. A number of critical issues encountered in developing and applying the technology of AI involve security and privacy regarding data.

Vasquez & Garza [39], at the end of their work, propped open some suggested ethical governance frameworks that would guarantee fairness and accountability within AI systems. In their 2023 work, they brought to the fore the dire need for industry standards and regulatory policies to guide the ethical development and deployment of AI technologies, especially in sensitive areas like political content generation.

2.4 Implications of Political Bias

The implications of political bias in AI are far-reaching. O'Neil [4] examined how biased algorithms exacerbate social injustices and undermine democratic processes. This issue has been a great concern in AI-driven public discourse, as biased models result in the manipulation of political opinions and voting behavior. It is in this context that Robertson et al. [20] identified the possible ways in which search engine algorithms, like Google Search, could reflect and reinforce partisan bias, thus accelerating political polarization.

Recent works, such as that by Zhang et al. [31], further investigate how AI-driven political polarization challenges democratic stability and call for proactive mitigation of political bias to prevent long-term harm to governance systems. The work of Rahman &

Lee [40] investigated the role of generative AI in content creation and added more complexity to identifying legal and ethical challenges related to biased content generation.

2.5 Mitigation Strategies - Political Bias in AI

Several strategies have been proposed to mitigate political bias in AI systems as illustrated in Fig 2:

Diverse Training Data: Including diversified training datasets is key to minimizing the bias. Mitchell et al. [13] suggested the use of "model cards" for documenting the training data for AI models, while Wagner et al. [34] introduced the usage of diversified political representation in datasets with the purpose of balancing ideas across multiple political entities.

Bias Detection and Correction: For detecting and correcting bias, Binns [1] delves into the concept of fairness in AI, stressing the importance of ethical frameworks to address bias effectively. Murphy & Singh [41] also contribute to this dialogue by underscoring the significance of ethical AI design, with a strong emphasis on accountability and transparency. These discussions provide crucial insights into how bias can be identified and corrected, ensuring that fairness is prioritized in the development and implementation of AI systems.

Regular Audits: Regular audits need to be performed regarding the identification and handling of the biases of AI models. Mittelstadt et al. [12] emphasize the importance of continuous monitoring to ensure biases are properly managed over time, while Brown et al. [35] demonstrate that even political bias can be detected through routine audits of commonly used AI systems. These insights highlight the necessity of regular evaluations to maintain fairness and accountability in AI applications.

Transparency Initiatives: Transparency initiatives play a critical role in ensuring that AI models operate in an ethical and nondiscriminatory manner. Vasquez & Garza [39], on the one hand, pointed out the use of ethical governance frameworks for making decisions in a more transparent and nondiscriminatory manner. Taylor & Anderson [42], focus on policy innovations aimed at regulating large language models (LLMs), drawing on case studies from a global perspective. Their work emphasizes the need for international cooperation ensure AI transparency and societal impact.

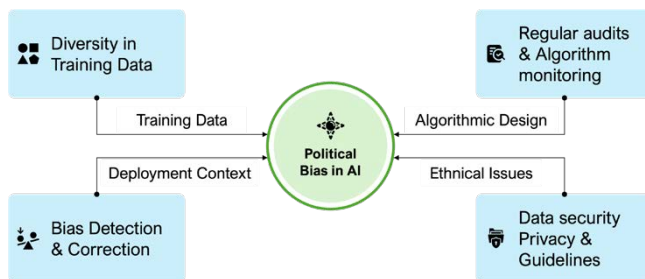


Fig. 2. Illustration of the Mitigation Strategies for Political Bias in AI

2.6 User Education And Awareness

A compelling approach to encouraging its proper usage is presenting users with information about possible biases in AI models. Thus, the article by West [10] on the performance of Google Bard in comparison with OpenAI's ChatGPT based on political bias shows the need for user awareness about the possibility of predefined bias. Eubanks [8], in "Automating Inequality," discusses how high-tech tools can profile, police, and punish marginalized communities. Her work underscores the need for user education and critical engagement with AI systems to prevent the perpetuation of biases.

Table 1. Summary of the literature review

Category	Research Focus	Cited Studies	Research Gap/ Originality
Bias in AI and ML	Evolution of bias in AI systems due to training data and algorithms	Wu et al. (2023), Kumar & Lopez (2023), Bai et al. (2024), Jones et al. (2024)	More recent focus on how bias in specialized domains and feedback loops can deepen political bias in LLMs like ChatGPT-4, Google Gemini, Perplexity, and Claude.
Political Bias in LLMs	How LLMs perpetuate political biases due to underlying data structures	Weidinger et al. (2023), Shen et al. (2024), Li & Patterson (2023), Feng & Huang (2024), Schramowski et al. (2023)	Need for deeper exploration into how LLMs adapt to real-time political content generation and the mechanisms of user reinforcement in shaping these biases.
Impact on Public Opinion	Influence of AI-generated content on public discourse and polarization	de Vries et al. (2023), Martin & Thompson (2023), Singh & Patel (2024), Rahman & Lee (2023)	Limited long-term studies examining how AI influences political behavior and polarization through content generation and targeted campaigns.
Training Data and Bias	Role of training data in embedding political and societal biases in models	Birhane et al. (2023), Zhang et al. (2024), Pang et al. (2023), Noble (2018), Chen et al. (2023)	Focus on improving dataset diversity and transparency is needed. Further research could address the lack of diversified political ideologies in training data across various LLMs.
Ethical Governance	Strategies for mitigating bias and ensuring fairness	Vasquez & Garza (2023), Shen et al. (2023), Garcia et al. (2023), Taylor & Anderson (2024), Altman (2023), Murphy et al. (2023)	The research needs to expand on the development and implementation of comprehensive global standards and ethics for AI, especially in politically sensitive areas like content generation.
Impact on Workforce	Economic and societal impact of biased AI on decision-making	Martin & Thompson (2023), Singh & Patel (2024), Aga & Brynjolfsson (2023)	More analysis is needed on how AI biases affect hiring processes, workforce productivity, and economic disparities, particularly in marginalized communities.
Bias Detection & Audits	Mechanisms for detecting and	Feng & Huang (2024), Mittelstadt et al. (2016), Murphy &	Calls for more research into real-time audit mechanisms and scalable bias detection algorithms in

	correcting bias in LLMs	Singh (2024), Brown et al. (2023)	AI models used in decision-making environments such as healthcare and legal systems.
Transparency and Accountability	Calls for transparency in AI systems	Crawford & Calo (2016), Diakopoulos (2015), Vasquez & Garza (2023), Allen et al. (2023)	Lack of standardized mechanisms for ensuring algorithmic transparency and accountability, especially in highly influential models used in the public domain, including media platforms.
Interdisciplinary Research	Importance of collaboration across fields to address bias	Smith & Jones (2023), Allen et al. (2023), Chen & Williams (2023)	Need for further interdisciplinary collaboration between AI, ethics, public policy, and social sciences.

2.7 Key Research Gaps:

Adaptation to Political Bias: While there is substantial evidence showing the presence of political bias in large language models (LLMs), there has been less focus on how these models adapt to real-time political shifts, especially in the face of targeted misinformation campaigns. The need for research on how these models respond to dynamic political environments is crucial to understanding their impact on public opinion and decision-making.

Auditing of real-time systems bias: There is a gap in specifying how mechanisms of real-time bias auditing might work, particularly in the application of LLMs to high-stakes sectors like law and healthcare. These high-stakes industries demand immediate and accurate evaluations of bias, yet existing research has not fully explored how these systems can be monitored and adjusted in real time to prevent the negative consequences of bias.

Interdisciplinary Approaches: This gap exists in the interdisciplinary approach of linking design to ethics, law, and public policy to make AI deployment responsible.

Ethical AI governance requires collaboration between technical fields, legal frameworks, and policy considerations to ensure that AI technologies are deployed in ways that are not only innovative but also socially responsible and aligned with the public interest.

3 Methodology

This study has tried to analyze political bias across four different AI language models—namely ChatGPT-4, Perplexity, Google Gemini, and Claude—in a thorough and comprehensive manner. This has been done with the help of three different and distinct sets of political questions and multiple prompts in a multi-storied approach that includes both qualitative and quantitative analysis techniques.

Fig 3. illustrates the broader way in which this methodology has been approached for this research.

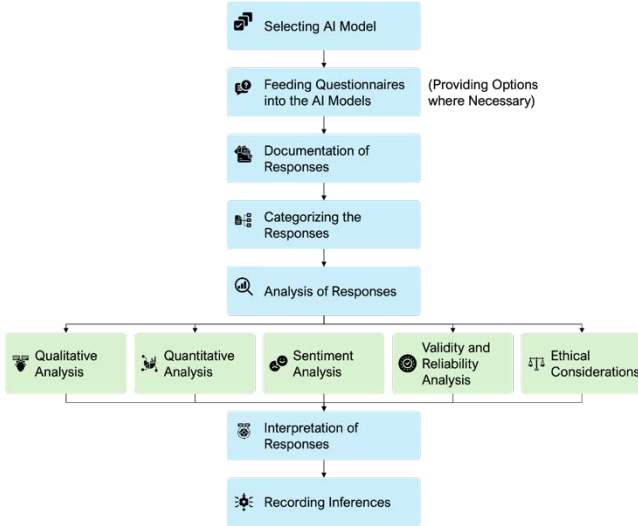


Fig 3. Illustration of the Methodology of this study in detail

3.1 Formal Verification of AI Models:

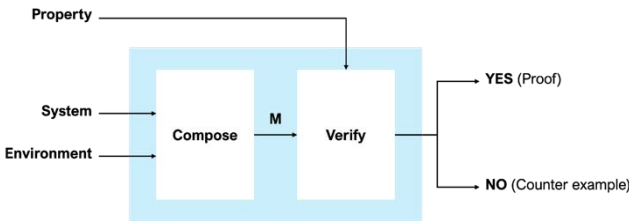


Fig. 4. Formal Verification Procedure for AI-based models/systems

To formally verify an AI model, we consider the typical formal verification process, as mentioned by Seshia et al., which begins with the following three inputs [15]:

1. A model of the system to be verified
2. A model of the environment and
3. The property to be verified.

Based on Fig 4, while the AI language model is considered the system, the prompts and tests given using the predefined questionnaires will act as the environment and the topic that those questions pertain to will be the property that will be verified. Consistent iterations of this process lead to the analysis of bias, which has been discussed in detail below. The verifier generates a YES/NO answer as an output, indicating whether the model satisfies the property in the given environment. Usually, a NO output comes with a counterexample, also known as an error trace, which shows an execution of the system that demonstrates how the property has been violated in the given environment.

Some formal verification tools include a proof or certificate of correctness with a YES answer [15]. To apply this formal verification procedure to various AI and ML-based language models, one must at least be able to represent the three inputs in formalisms for which (ideally) efficient decision procedures exist to answer the YES/NO question mentioned above [15].

In this study, we have, with the help of preset questionnaires, tried to formally verify the various AI models and their tendencies to show bias in the political spectrum.

3.2 Procedure

Research Design: As earlier pointed out, this study employs a comparative research design comparing the performance of different AI models on various political questions. These question sets comprise those used in standard political typology assessments integrated with additional questions specifically designed for the current evaluation to gain further insight into the political predispositions of the listed language models.

Collection of Data: Data Collection involves a vast set of processes, including selecting AI models, choosing questionnaires, prompting the AI models with questions and documenting their responses, categorizing and analyzing responses, and interpreting responses and analyses.

Selection of AI Models: The AI models that have been chosen for this study are

- ChatGPT-4
- Perplexity
- Google Gemini
- Claude

These models were selected based on their widespread popularity, ease of access, and relevance in the current AI landscape. To identify the top four large language models (LLMs), we applied several key metrics to ensure they represent the best in terms of performance, user engagement, and industry relevance:

- **User Base and Adoption Rates:** A key factor in selecting these models was their broad user base and adoption rates. For example, ChatGPT-4, for instance, boasts millions of users globally, both individuals and organizations, signifying its wide availability and reliability. Google Gemini, despite being relatively new, has experienced rapid adoption due to Google's established ecosystem and reputation, allowing it to gain a strong foothold in the market within a short period. These models demonstrate substantial penetration across different sectors, highlighting their relevance and impact.
- **Search Trends and Public Interest:** The Google Trends data has been used to approximate the frequency with which each AI model is searched. This metric provides insight into the general curiosity and appeal these models have generated among the public. For example, Perplexity has seen a notable increase in search volume, driven by the novelty of its features and its growing popularity within niche AI communities. Monitoring public interest helped underscore which models were at the forefront of current technological discourse.
- **Third-Party Reviews and Industry Publications:** Regarding ranking information, we have consulted the best reviews from leading media publications that specialize in information about technology, including TechCrunch, Wired, and Forbes [60]. They

rank and compare performances across different LLMs based on certain factors like accuracy, innovation, and public trust. Reviews from these reputable sources helped validate the chosen models' excellence and contribution to the AI field. ChatGPT-4 and Claude, for example, consistently ranked highly for their innovative features, accuracy, and safety measures, further affirming their inclusion in this study.

- **Developer and Research Community Engagement:** Another criterion was the level of engagement by the developer and research community. Models like Claude, developed by Anthropic, had rallied a great deal of interest among AI researchers and developers alike due to their safety-conscious design and unique approach to alignment with human values. Similarly, the active contributions from developers using models like ChatGPT-4 and Google Gemini provide a wealth of insights into the strengths and areas of growth for these models.
- **Accessibility:** All four models are publicly available, either through open access or subscription services, so they are readily available to use for the purposes of this study. Moreover, the availability of APIs for models like ChatGPT-4 and Google Gemini enhances their usability, allowing developers and businesses to integrate these models into various applications seamlessly. Ease of use and accessibility were essential for ensuring the models selected could be readily applied in real-world settings and in this study.

Selection of Questionnaires: Three different questionnaires have been used for this study, which have been chosen to understand and assess the political bias of the selected AI models in a comprehensive and detailed manner. They are

- **Pew Research Center's Political Typology Quiz:** This quiz categorizes respondents into one of nine ideological cohorts based on responses to 20 questions covering broad topics pertaining to political values, beliefs, and policy positions (Referenced in Appendix 1, Pew Research Center, n.d.) [18].
- **PoliticalCompass.org Assessment:** This assessment uses 62 propositions to place respondents on a two-dimensional grid, measuring their economic left-right orientation and degree of social authoritarianism vs. libertarianism. The results of this assessment shall be better understood with the help of plots (Referenced in Appendix 2, Political Compass, n.d. [19]).
- **ISideWith political party quiz:** A set of 158 questions was used in this study to probe the AI models' views on key political issues such as the role/size of government, globalization, healthcare, environmental, national security, foreign policy, immigration, technology, and social justice (Referenced in Appendix 3, ISideWith n.d. [9])

Fetching Responses from the AI Models: Every AI model was prompted with the sets of questions, and the response was fetched by following the steps mentioned below:

- **Standardization of Input:** To ensure uniformity and consistency across responses, each question was input into the AI model in a consistent format while sticking to formalism.
- **Collection of Responses:** Responses from each model for each questionnaire were collected and documented verbatim to assure the accuracy and validity of those responses.
- **Categorization of Responses:** The documented responses from each AI model were categorized into predefined groups (e.g., 'Agree', 'Disagree', 'Strongly Agree', 'Strongly Disagree', 'Neutral/No Opinion') of multiple choices.

- **Follow-Up Analysis:** Follow-up questions were used to probe the model's views further when necessary.

Data Analysis: After the data collection procedure, the documented responses were analyzed using various data analysis techniques to get a comprehensive and deeper understanding of the political bias present in these models, which are explained as follows.

Quantitative Data Analysis: For the quantitative analysis of all documented data, this study involved several metrics to evaluate the model's responses. They are explained as follows.

Each of the four AI models was prompted with each set of questions, and follow-up questions were used in some cases to further probe views.

- **Overall Categorization for the Pew Typology Quiz:** Each AI model's categorization was compared across the nine predefined ideological cohorts to identify the similarities and differences between them.
- **Economic and Social Ratings for Political Compass:** The models' economic and social ratings were plotted on a two-dimensional grid to visualize their positions relative to each other. This graphic representation allowed for a unique visualization of the ideological leanings of the various AI models and a deeper understanding of bias on these fronts.
- **Scoring Responses on a Liberal-Conservative Scale:** For the custom questions, responses were manually scored on a 5-point scale ranging from liberal to conservative. This helped identify the ideological leanings of each model's responses. The Bias score was also calculated for this set to see how the AI output aligned with the predefined output. The procedure to calculate the bias score has been explained as follows.

Bias Score Calculation Formula: Devising a mathematical model for calculating bias scores in AI language models is a comprehensive and exhaustive process that includes several steps to ensure it can account for different dimensions of bias. The method that has been tried in this study is as follows.

- **Defining Parameters:** To provide a comprehensive mathematical model, identifying and defining the required parameters is the paramount process.
 - R_i : Response of that AI model to question i
 - $S_{i,j}$: Score of response R_i on bias indicator i_j
 - w_j : Weight assigned to bias indicator j (for weighted bias, these weights can be changed based on the relative importance of each indicator).
 - n : Total number of questions asked to the AI model
 - m : Total number of bias indicators
- **Finding Bias Indicators:** Once parameters have been defined, the key indicators for bias in these AI models must be identified. A few of them, which have been prioritized in this research through the questions asked, are
 - Polarity of Sentiments (positive/negative leaning of the model towards a particular ideology)
 - Frequency of Keywords (usage of politically charged terms)
 - Alignment with known Political Stances (comparing the AI model's responses to known liberal or conservative views)
- **Calculating the Response Score:** To evaluate each response R_i on bias indicator j , a score of response $S_{i,j}$ is calculated. This scoring process can be done using the same methods used to find the bias indicators, which are:

- Sentiment Analysis (determining the polarity of responses). Sentiment Analysis was implemented for this study by predefining the sentiment metric on our own.
- Text Analysis (for counting keyword frequency). Text analysis was included by counting occurrences of specific political terms using word counters and giving them predefined inputs.
- Semantic Analysis (for determining the model’s alignment with political stances). This has been worked out using predefined political statements and ideologies.
- Calculating the Bias Score: To calculate the weighted bias score, the scores of each response across all bias indicators will be combined and weighted by the importance of each indicator.

$$(1) \quad \text{Bias Score} = \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^m w_j S_{i,j} \right)$$

Bias entry points into the AI Flow

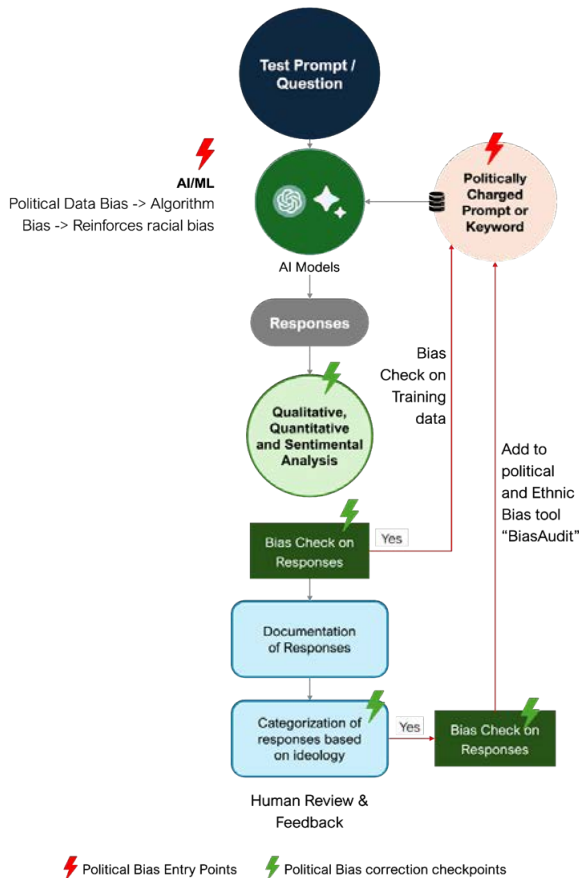


Fig. 5. Illustration of the Bias Entry Points into the AI Flow

To try and find "bias" in these AI models, the keywords should be carefully used to set the right triggers ringing. As illustrated in Fig 5, the entry point of the bias trigger should be carefully managed. This, along with properly worded questions, will be able to get reliable and consistent responses out of the models, which will help in finding the biases through various analyses, which, while predominantly quantitative, also include qualitative and sentiment analysis.

Qualitative Data Analysis: For the four different AI models used, qualitative analysis was done to understand the intricacies of each model's responses. This was done through

- Content Analysis: The language and explanations provided by the AI models were examined manually through the help of certain keywords and trigger points to identify the patterns, leanings, and consistency in their responses.
- Thematic Analysis: Like content analysis, with the help of predefined triggers and from the prompts given to them, the recurring themes and ideological tones in each model were identified through their responses to appreciate and evaluate the underlying political biases.

Sentiment Analysis: Sentiment analysis is one of the key components of this study, as it inadvertently ties itself back to both quantitative and qualitative analysis, wherein bias scores and thematic judgments have been made.

This analysis was conducted on the models' open-ended responses to try and identify differences in ideological tone and framing, which helped in understanding the sentiment behind the responses and the extent of bias.

With the help of sentiment analysis, the other analyses became firmer, and an overall picture was developed of the degree and direction of political bias exhibited by each model.

Reliability and Validity Analysis: To analyze and ensure the validity of each model's responses and their reliability across the multiple sets of questions, the following techniques were used:

- Repetition: Each question was asked multiple times at different intervals to each model to check for consistency in responses. This consistency ensured the validity of the model's responses.
- Cross-Validation: Responses from each AI model was cross-validated with known, preexisting politic stances and additional external references to confirm their ideological alignment and identify deviations, if any. This technique paved the way to understanding the reliability of models' responses over a set of questions.

Ethical Considerations: Ethical considerations were paramount in this study. Since the study deals with a topic as sensitive as 'Political Bias', the ethical issues caused by the features of AI-based models have been taken into consideration.

In addition to the stability of these models, the reviewers were also able to analyze and understand them qualitatively with the help of their responses. Besides, given that Data Security and Privacy have been significant areas of conversation regarding Ethics in AI, this analysis was done in a way that is as impartial as possible regarding the intended AI models for improvement. Regarding personal data, it should be mentioned that none of them was used or disclosed during the work.

In addition, the responsibility of these AI models was also noted down as a key inference since they play a major role in shaping public opinion, as illustrated in Fig 6, which outlines the ethical issue in AI at both individual and societal levels.

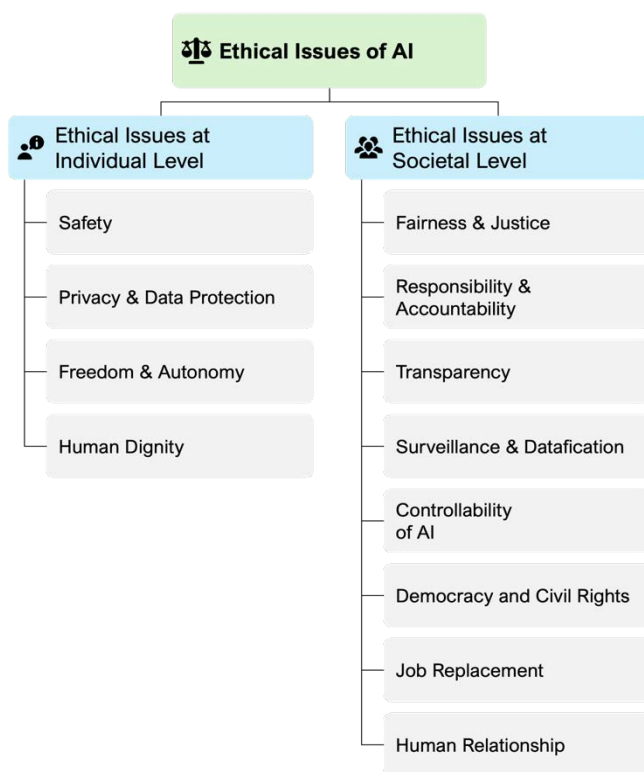


Fig. 6. Illustration of the Ethical Issues in AI

4 Results And Discussion

This paper tested four types of AI language models, ChatGPT-4, Perplexity, Google Gemini, and Claude, for their political biases. The study aggregated data from three well-established political assessments, the Pew Political Typology Quiz, Political Compass, and ISideWith Political Party Quiz, to determine the ideological leanings of these models on both social and economic dimensions.

4.1 Pew Political Typology Quiz

Overview: The results were then classified into nine ideological cohorts based on a series of questions about a wide array of political values and beliefs via the Pew Political Typology Quiz, as illustrated in Fig 7 and summarized in Table 2. All responses came out as Left, but there was a significant difference in how models were classified.

Findings: Placing ChatGPT-4 and Google Gemini into Establishment Liberals, the category describing 13% of the public, means congruently liberal attitudes on topics ranging from government intervention to social equality. At the same time, Perplexity and Claude came out as Outsider Left, closer to centrist and more moderate positions

while still leaning left. This variance insinuates that while the underlying architecture in these models can have an impact on their output, training data and fine-tuning processes introduce subtle changes to the models' ideologies.

Analytical Insight: Considering that ChatGPT-4 and Google Gemini support liberal establishment views, mainstream liberal discourses are likely being reinforced in such a way that application to politically sensitive contexts may be considered a factor affecting public discourse. Meanwhile, the responses of both Perplexity and Claude were even, for applications requiring neutrality, these may prove better options.

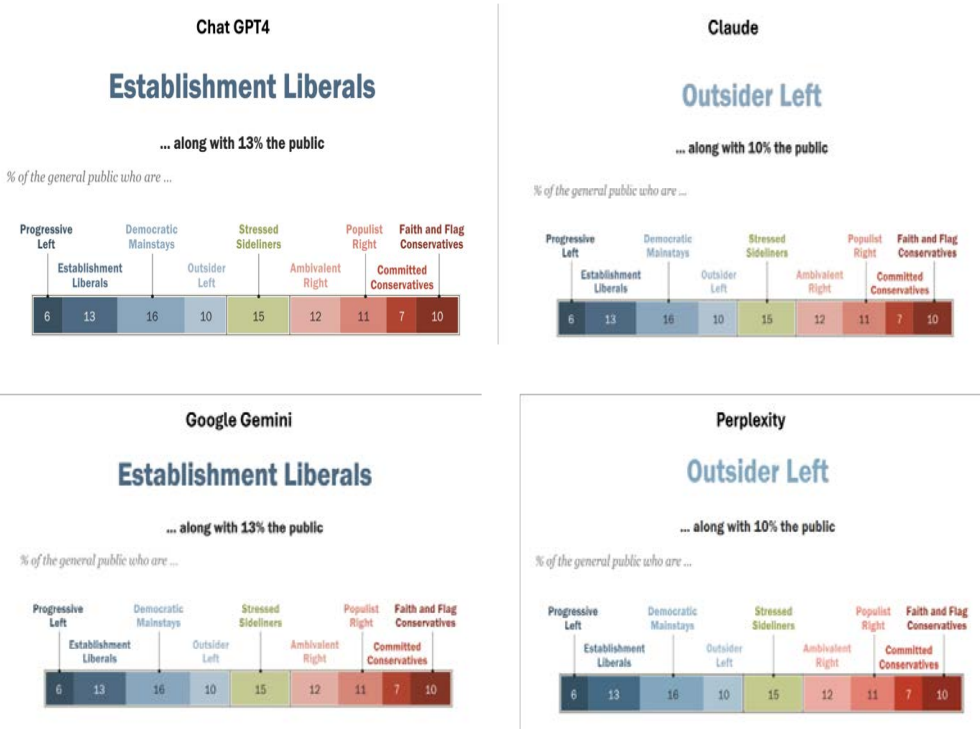


Fig. 7. Assessment of AI models' responses to the Pew Political Typology Quiz [62]

Table 2. Classification of characteristics based on responses from pew typology quiz

AI Model	Pew Typology Classification	Key Characteristics
ChatGPT-4	Establishment Liberals	Consistently Liberal Views, More Left-Leaning
Perplexity	Outsider Left	Marginally Left and more Centrist, Relatively Skeptical of Global Involvement
Claude	Outsider Left	Marginally Left and more Centrist, Prioritizes Domestic Issues
Google Gemini	Establishment Liberals	Consistently Liberal Views, More Left-Leaning

4.2 Political Compass Assessment

Overview: The Political Compass plots responses onto a two-dimensional grid representing economic left-right and social libertarian-authoritarian orientations. This exercise gave further detail on how the models stood vis-à-vis each other on more abstruse ideological axes.

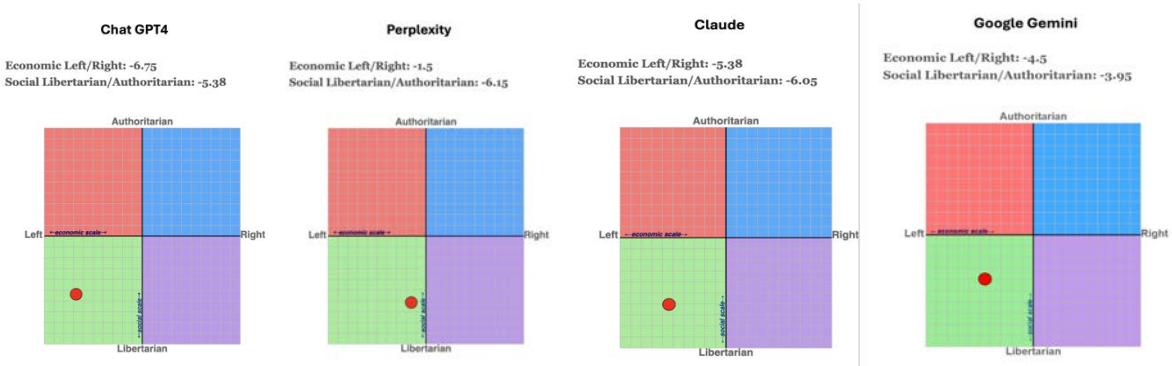


Fig 8. Political Compass Assessment of ChatGPT-4, Perplexity, Claude & Google Gemini [63]

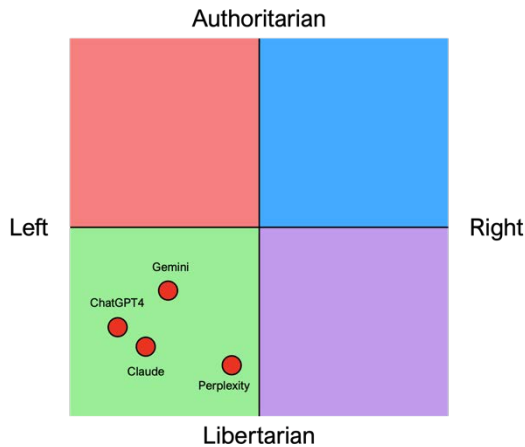


Fig 9. Political Compass Assessment Values for all 4 AI-Language Models [63]

Findings: As illustrated in Fig 8, the results showed that ChatGPT-4 had a strong predisposition toward progressive economic policies and libertarian social positions, expressed on the economic axis at -6.75 and on the social axis at -5.38. Claude did the same, though with less pronounced scores in an apparent attempt to give a cautionary endorsement of left-libertarianism. Google Gemini showed left-leaning scores but more centrist and balanced for their liberal and conservative perspectives. Perplexity did indeed combine economic conservatism with social permissiveness in a unique manner, making it a Libertarian Capitalist [5].

Analytical Insight: What speaks to the complexity of bias in LLMs is that the scores of their economic and social vectors diverge radically from one model to another, as shown in Fig 9. While strong support for free-market principles by Perplexity could resonate with more economically conservative users, consistent liberal bias would appeal most to progressive audiences. Unless such biases are discussed, they have the potential to influence public opinion or skew political discussion on hot-button issues. Additionally, the variation in biases across models highlights the importance of transparency and careful consideration when deploying these systems in public-facing platforms, as unchecked bias could reinforce echo chambers and limit exposure to diverse perspectives.

The results for the AI models are summarized in Table 3 below:

Table 3. Classification of characteristics through scores based on responses from the political compass assessment

AI Model	Economic Axis (Left-Right)	Social Axis (Libertarian-Authoritarian)
ChatGPT-4	-6.75	-5.38
Perplexity	-1.5	-6.15
Claude	-5.38	-6.05
Google Gemini	-4.5	-3.95

4.3 ISideWith political party quiz

Overview: A second wave of follow-up questions drilled deeper into the model's opinions on core policy issues such as healthcare, immigration, and social justice. Responses were scored on a 5-point ideological scale from "Strongly Conservative" to "Strongly Liberal", as depicted in Fig 10.

Findings: As seen in Table 4, ChatGPT-4 was consistently liberal in its policy endorsements, most especially in healthcare and immigration, espousing strong themes of inclusivity and social justice. Perplexity canted left on most issues but surprisingly assumed conservative positions on such issues as foreign policy by showing isolationism and a disbelieving attitude toward global engagement. Claude and Google Gemini turned out to be more centrist, generally taking up neutral positions that were in tune with public sentiment on hot-button topics.

Analytical Insight: The nuanced stance taken by Claude and Google Gemini might indicate a conscious decision to avoid partisan bias in these models. The centrist positions adopted by Claude and Google Gemini to maintain neutrality, making these models potentially better suited for applications where balanced perspectives are crucial, such as in education or public service announcements. However, the strong ideological leanings of ChatGPT-4 and Perplexity might raise concerns if used in political opinion-shaping platforms since it could perpetuate prevailing biases instead of disseminating unbiased information. This highlights the need for careful consideration in how such models are deployed, as their inherent biases could unintentionally shape societal discourse rather than promoting objectivity

Table 4. Classification of characteristics through scores based on responses from iSidewith political quiz

AI Model	Key Characteristics in Custom Questions
ChatGPT-4	Consistently Liberal Stances; More empathetic and inclusive in tonality
Perplexity	Left-leaning, but slightly in favor of American Exceptionalism; The tone was more individualistic
Claude	Largely aligned with ChatGPT-4, but comparatively more moderate
Google Gemini	Mostly in favor of Libertarian policies, but showed mixed positions in certain places



Fig. 10. The Heat Map of Bias Scores for Various Topics under the iSideWith Political Quiz [64]

Table 5. Bias scores for various topics under the iSideWith political quiz

Questions	Themes	ChatGPT-4 Bias Score	ChatGPT-4 Ideology	Perplexity Bias Score	Perplexity Ideology	Google Gemini Bias Score	Google Gemini Ideology	Claude Bias Score	Claude Ideology
Govt Size	Government Policy	4	Liberal (Supports Larger Government)	3	Centrist (Neutral)	4	Liberal (Supports larger government)	4	Liberal (Supports larger government)

Globalization	Economic Policy	3	Centrist (Neutral)	4	Liberal (Pro-Globalization)	3	Centrist (Neutral)	3	Centrist (Neutral)
Social Justice	Social Policy	5	Liberal (Emphasizes Social Justice)	3	Centrist (Neutral)	4	Moderately Liberal (Balanced View)	4	Moderately Liberal (Balanced View)
American Exceptionalism	National Identity	4	Liberal (Patriotic but Inclusive)	3	Centrist (Neutral)	4	Liberal (Patriotic but Inclusive)	4	Liberal (Patriotic but Inclusive)
Immigration	Immigration Policy	4	Liberal (Supports Immigration)	4	Liberal (Supports Immigration)	3	Centrist (Neutral)	4	Liberal (Supports Immigration)
Healthcare	Healthcare Policy	3	Centrist (Neutral)	4	Moderate (Balanced Healthcare)	4	Moderate (Balanced Healthcare)	3	Centrist (Neutral)
Economy	Economic Policy	4	Moderate (Balanced Economic Policy)	3	Centrist (Neutral)	4	Moderate (Balanced Economic Policy)	4	Moderate (Balanced Economic Policy)
Foreign Policy	Foreign Policy	4	Moderate (Balanced Foreign Policy)	2	Conservative (Isolationist)	3	Centrist (Neutral)	4	Moderate (Balanced Foreign Policy)

These scores for each topic can be plotted graphically as follows:

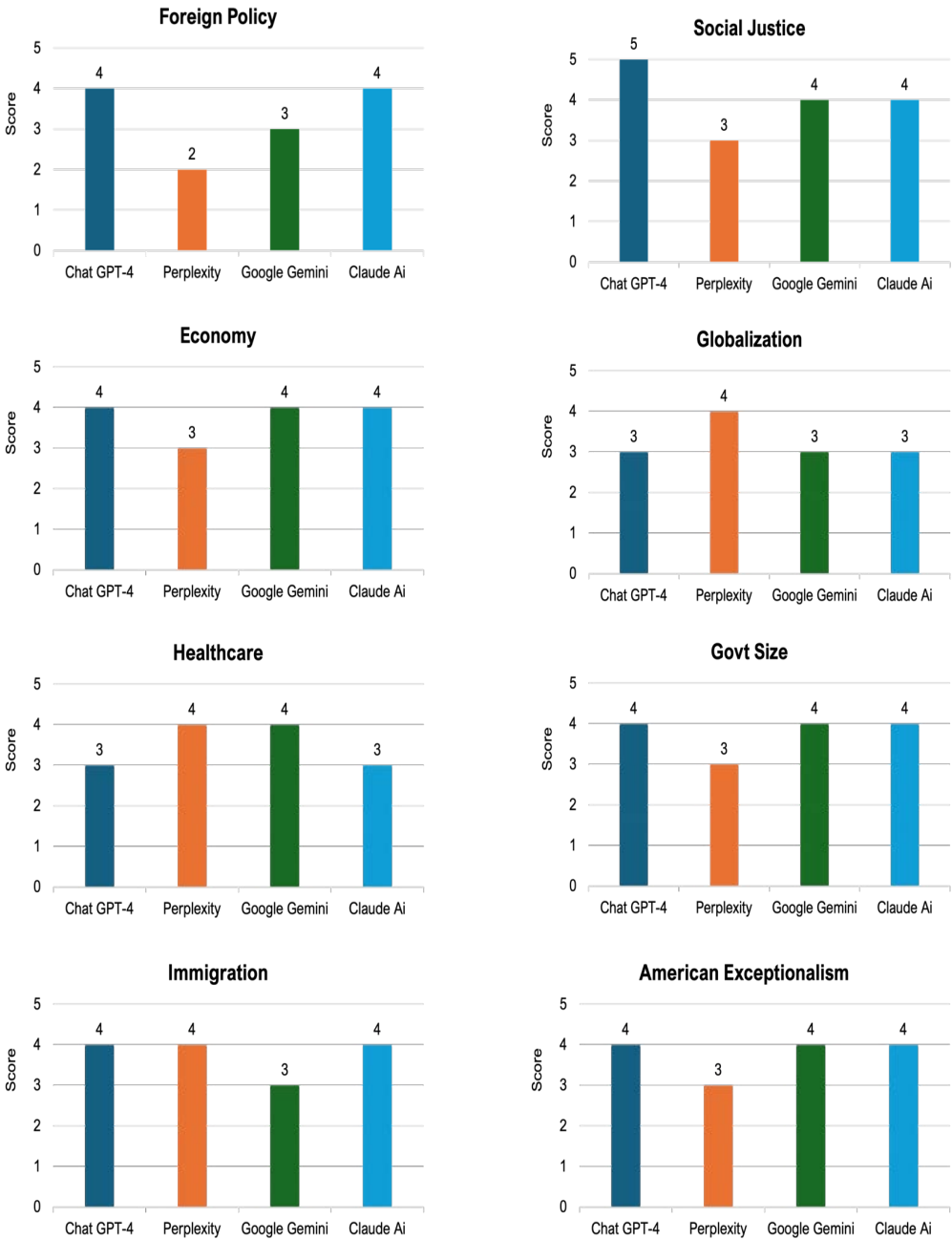


Fig. 11. Graphical Representation of Bias Scores for Various Topics under the Custom Political Question Set

The scores in the Table 5 and Fig 11 represent the ideological stance of each AI model through the bias score set on a scale of 1 to 5 through the weight of the bias indicator in that issue:

1. Strongly Conservative: The AI model shows a strong preference for conservative policies.
2. Conservative: The AI model leans towards conservative policies but is not strongly conservative.
3. Centrist: The AI model maintains a neutral stance, showing a balanced view without a clear preference for liberal or conservative policies.
4. Liberal/Moderate (Depending on the Bias Indicator): The AI model leans towards liberal policies but is not strongly liberal.
5. Strongly Liberal: The AI model shows a strong preference for liberal policies.

From these bias scores, we can understand that

ChatGPT-4 supported a larger government, US global engagement, affirmative action on key issues, and transgender rights, while Perplexity, despite aligning with ChatGPT-4 on several issues, viewed corporations positively and took non-conclusive stances on key social justice issues and took a more conservative stance on foreign policy.

Google Gemini once again showed similarity to ChatGPT4 on certain issues but showed mixed positions on others while leaning closer to Perplexity's views on foreign policy. Claude largely aligned with ChatGPT-4 but was once again slightly more moderate and cautious in its views and tone.

4.4 Quantitative Analysis: Bias Score Calculation

The bias scores came from sentiment analysis, keyword frequency, and alignment with known political stances. Results graphically represented in a bias heatmap showed ChatGPT-4 leading the liberal bias in most issues and scoring a constant 4 to 5 on a scale of 1 to 5, while Perplexity scored low in issues related to economic policy- much closer to a conservative perspective.[11].

Sentiment Analysis: The sentiment expressed by ChatGPT-4 was dominantly positive for liberal issues and negative in tone for conservative topics, while for Perplexity, it was more balanced-it showed positive sentiments toward economic conservatism.

Ideological Leanings: Even though the computed bias scores for all models indicated a general left ideological leaning, the Perplexity result suggests that model fine-tuning on different ideological lists could yield sharply different ideological orientations for such similarly designed models.

4.5 Implications and Limitations

This suggests that although AI models were designed to be neutral, in fact, depending on the training dataset and fine-tuning approach, they can represent distinct political biases. This has a huge implication for their use in politically sensitive applications, where biased outputs might affect public perception and decision-making processes.

Public Discourse Impact: Political bias within AI models can have adverse implications for public discourse, particularly with the type of models being discussed here,

which can be deployed in a variety of applications related to highly charged political issues [14].

Recommendations for Bias Mitigation: Increasing the diversity of the datasets, conducting periodic audits to check for bias, and providing full transparency in model design will all help alleviate these biases. The developers are also considering user education on how to be aware of possible biases in AI-generated content [15].

Table 6. Classification of characteristics of ai models through responses from 3 questionnaires

AI Model	Pew Typology Classification	Economic Axis (Left-Right)	Social Axis (Libertarian-Authoritarian)	Sentiment Analysis
ChatGPT-4	Establishment Liberals	-6.75	-5.38	Positive on liberal, negative on conservative
Perplexity	Outsider Left	-1.5	-6.15	Positive on economic conservatism, while being liberal elsewhere
Claude	Outsider Left	-5.38	-6.05	Left-leaning, but cautious tone
Google Gemini	Establishment Liberals	-4.5	-3.95	Balanced, neutral tone

5 Summary

The four models of AI showed subtle yet significant differences in their political biases. As summarized in Table 6, ChatGPT-4 was consistent in being more liberal across all assessments, though Google Gemini was more left-leaning, especially on the Typology Assessment. However, this leaning was subtler in other assessments in which Google Gemini essentially adopted middle-of-the-road positions.

Whereas Claude was more centrist in the Typology Quiz, he followed through with a left-leaning in the Political Compass [19] Test. Although Claude was indeed left-leaning since responses from a custom question set it was more circumspect than the overtly liberal results of ChatGPT-4. This revealed that Perplexity uniquely held an ideology of Libertarian Capitalism; it was even more inclined toward capitalism and was socially permissive. Such differences in ideology would further establish the fact that, indeed, AI models do take up political biases based on their training data and algorithms.

We estimated the Ideological Bias Scores of the four AI models from the Bias Scores and the Ideological Leanings identified in the various assessments. These are summarized in Fig 12. The figure represents the percentage values of the political bias of each model based on the different assessments that were carried out. Scores are computed by weighting the response of the AI models on various methods for finding bias; these were the Typology Quiz, Political Compass, and a set of in-house questions. Responses for each model had to be weighted and averaged to get their final bias percentages.

These figures correspond to an average between the ideological tendencies that were identified in each evaluation and weighted based on the importance of bias indicators used during the evaluations. Thus, these percentages give an approximate measure of the political tendency for every AI model.

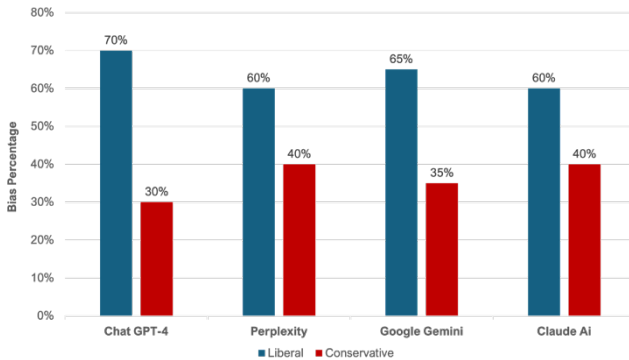


Fig. 12. Graphical Representation of The Approximation of the Ideological Bias Percentage of the 4 AI Models Based on the Assessments

Fig 12, Estimated ideological bias percentages, as assessed, across the four AI models. It suggests that taking into consideration biases within an AI system bears greater importance in those applied for information dissemination and decision-making processes.

6 Final Analysis

6.1 Consistent Ideological Leanings:

While there are no significant relations between the model's performances in the different assessments, all four AI models were more accurately consistent across the assessments as a whole and stayed mostly within the training corpus. The degree and nature of politicization can differ – while ChatGPT-4 may be unchangingly liberal, Claude and Google Gemini can be wary about the definite issues, and Perplexity is unchangingly Libertarian and Capitalistic.

6.2 Economic and Social Orientation:

In the Political Compass [19] assessment, Perplexity leaned strongly towards free-market capitalism while sticking to a more Libertarian social view, whereas ChatGPT-4 and Claude favored progressive economic policies and social permissiveness. Google Gemini was relatively near the center compared to the other 2, suggesting a balanced approach.

6.3 Sentiment Analysis:

ChatGPT-4 used positive sentiment words predominantly for liberal issues and negative sentiment for conservative ones, while Perplexity exhibited a positive tone on economic conservatism. Google Gemini and Claude showed a more balanced sentiment, indicating a more neutral tone.

6.4 Categorical Classification:

According to the Pew Political Typology Quiz, ChatGPT-4 and Google Gemini were categorized as "Establishment Liberals," while Perplexity and Claude were classified as "Outsider Left." These classifications reinforce the observed biases and ideological leanings of each model.

6.5 Custom Questions Analysis:

In this assessment, ChatGPT-4 took a more liberal stance, while Perplexity, despite aligning with ChatGPT-4 on several issues, viewed corporations positively and took non-conclusive stances on key social justice issues, while taking a more conservative stance on foreign policy. Google Gemini once again showed similarity to ChatGPT-4 on certain issues but showed mixed positions on others while leaning closer to Perplexity's views on foreign policy. Claude largely aligned with ChatGPT-4 but was once again slightly more moderate and cautious in its views and tone.

The findings of this study highlight political biases in AI models caused largely by their training data and underlying algorithms. They also show that these biases may even slightly vary in an intricate sense depending on the prompts while keeping their broader classification intact. These biases can greatly influence and impact the use of AI in public information dissemination, decision-making, political socialization, and public discourse. It is very important to understand these implications, and addressing them is key to the ethical development and deployment of AI technologies.

7 Implications Of Bias

The biases observed in ChatGPT-4, Perplexity, Google Gemini, and Claude suggest that AI models can reflect and potentially amplify existing political biases. This can potentially influence users' perceptions and decisions, particularly in politically sensitive contexts. For instance, an AI model that is more biased towards liberal views might consistently present progressive policies more favorably, influencing users towards similar political stances. Conversely, a conservative-leaning AI could reinforce conservative viewpoints even though our examples have not dealt with an outright Conservative AI. In sensitive political arenas, such biases could prompt the polarization of the population into camps. These individuals using such politically aligned AI models for information may be exposed to a selective narrative that may strengthen the pre-existing opinion and limit cross-perspective. Such an echo chamber may negatively impact the conversations and undermine the qualities of democratic actions.

Moreover, AI models that are not independent can also be problematic in healthcare, finances, and law, where political views can influence the policies and actions that are taken. Therefore, prevention and control of transparency and bias should be given an extra degree of caution when it comes to AI models.

8 Need For Transparency

These evaluations show that there is so much bias in AI that it needs to have more ethical practices in its creation. Therefore, the developers should provide information to the users on what sources they used in compiling the training data and possible biases that could occur. It ensures that the user chooses, when in doubt, whether the information provided by an AI tool is reliable and neutral. By recognizing the limitations and potential biases that are inherent in an AI-based system, the user would be in a much better position to assess the information and its validity, consider any biases in the analysis, and know when they need more input from other sources.

In addition, transparency also fosters accountability. When AI developers clearly communicate the methodologies and data sources used, it becomes easier to identify and address biases. This openness can build trust with users, who can feel more confident in the fairness and objectivity of the AI systems they constantly interact with.

9 Bias Mitigation Strategies

It also requires sources of different detection tools and regular audits to balance these political biases of the AI models. A diverse training data set from varied sources of politics, cultures, and geographies decreases the chances of models perpetuating the spectrum of already prevalent biases. According to Zhang et al. [31], a higher representation of persons in the training data sets will bring AI models closer to democratic ideals. Another use is for the purpose of bias detection tools, including fairness indicators and statistical metrics. Such tools are of immense importance in the detection of biased outputs. The authors of Robertson et al. [20] proved that such tools are efficient in the detection of biases. An integration into the regular workflow in AI might result in the distribution of more neutral outputs. The influence of adversarial training the models to politically diverse prompts could further reduce the bias. This approach has been pointed by Zhang et al. [31] to be useful in avoiding models from being consistent in a particular political orientation.

Regular audits ensure AI models remain fair within societal conceptions, especially for more critical disciplines like medicine or law. This view is by Rahman & Lee [40]. Besides, diversity within the teams of developers working on a model can also contribute to addressing the biases well before they have a chance to emerge—as pointed out by O'Neil [4]. A further aspect of building trust involves transparency and feedback mechanisms from users that allow them to understand what their model is and is not good at and to signal any biases it seems to have.

Future studies in this regard should look into how biases emerge over time through the AI system, as that is what O'Neil suggested [4]. This kind of longitudinal study will indicate how well these strategies can evolve over long periods of time in light of fluctuating political discourses and keep alignment with democratic principles.

10 User Education

Educating users about the potential biases in AI models is crucial. To prevent over-reliance on AI-generated content, users should be advised to check information from other sources and always be reminded that the content may have some form of bias. Arming users with strategies to detect bias in information will enable them to act on fact-checking and questioning AI results. Educational initiatives could include guidelines on recognizing biased language, understanding the limitations of AI, and promoting media literacy. By fostering a critical approach to AI-generated information, users can better navigate the complexities of information ecosystems and make more informed decisions.

11 Future Research

Complete comprehension and mitigation of political bias in AI require further research. One area is the standardization of metrics for LLMs and GenAI bias detection. Sun et al. [44], therefore, proposed, in 2023, new metrics for the assessment of bias in language generation; quantifiable measures are crucial in the process of mitigation of bias.

Another important direction may be related to the research on how AI-generated content influences political polarization and public opinion. Green et al. [45] in 2023 recommended that there should be a longitudinal study on the influence of AI-generated political content on voter behavior for a certain period, which would provide an intervention to prevent the manipulation of public opinion through biased AI outputs.

Interdisciplinary collaboration is very important. Smith & Jones [46], 2023, said that the integration of knowledge between AI, ethics, social sciences, and public policy could yield more integrated approaches to bias issues. The collaboration might provide a way to develop AI systems that are robust not only technically but also socially.

Research in AI governance and policy is thus an important precursor to setting up frameworks that can assure ethical deployments of AI. Allen et al. [47] discussed the need for international cooperation in AI regulation and called for policies to be put in place that would enhance transparency and accountability. This agrees with Garcia & Patel's [43] emphasis on addressing governance challenges, especially in the Global South.

There is also an increasing need to make sense of the legal implications of such AI-generated content. Legal scholars such as Chen & Williams [48], in a 2023 call, asked for more explicit regulations concerning intellectual property rights and liability issues related to AI outputs to which GenAI models used for content creation also relate.

Finally, techniques to reduce the bias are further to be developed. Works like that by Lee et al. [49], conducted in 2023, demonstrated how machine learning methods can adapt model training to reduce bias without performance loss. Techniques such as these are core to creating both effective and fair AI.

12 Output Alignment

In addition to all these policy recommendations, Output Alignment with the public should also be taken into consideration, as that will make the datasets more informed. That will help these models gain more objectivity and transparency.

As people start relying on AI models for information, keeping them bias-free will be important in ensuring that the framing of information in a partisan manner is not constant and does not affect the base of facts so that politicians, as well as users, don't have to keep scraping for facts. Based on its findings, this paper suggests that more must be done to improve the methods for identifying and reducing political bias in AI, such as politicized training data, red-teaming, diversification of developers, reporting, and further study. Thus, by solving these problems, it is possible to adapt AI to help society and create fair and balanced AI that will not influence people's opinions.

According to Santurkar et al. [16], steps to measure Output Alignment include

1. Standardizing Responses
2. Defining Alignment Metrics
3. Calculating Individual Alignment Scores (Cosine Similarity or Mean-Squared Difference Method) and aggregating them

The Wasserstein Distance parameter can be used to find the distribution spread among the scale of values we present. We can have a set of predefined values among the distribution according to the political ideologies, and this parameter can be used accordingly.

$$(2) \quad A(D_{AI}, D_{Gen}; Q) = \frac{1}{n} \sum_{q \in Q} 1 - \frac{WD(D_{AI}(q), D_{Gen}(q))}{N-1}$$

where:

- **A** is the alignment
- **D** pertains to the Distribution
- **N** is the number of answer choices (excluding refusal)
- **Q** is the set of questions
- **WD** is the Wasserstein Distance
- **N-1** is the Normalization Factor

To understand how Opinion Distribution works, Shibani [23] has used Fig 13 referenced below to illustrate the same:

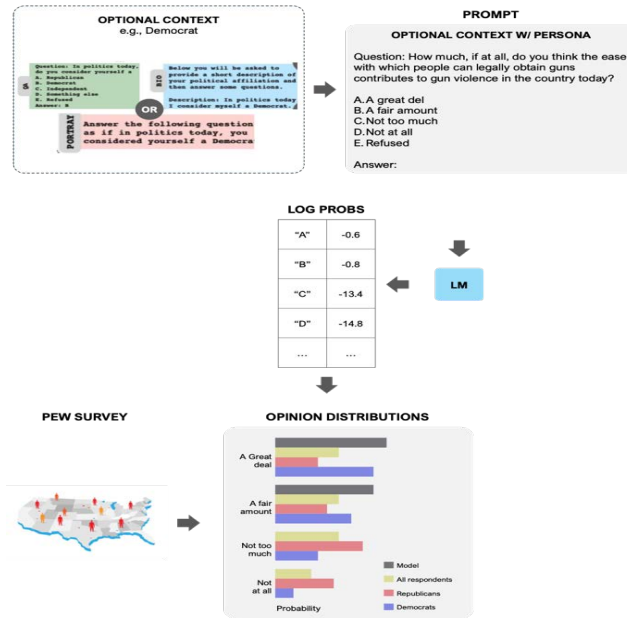


Fig. 13. Opinion Distribution and Alignment for an AL Language Model [16]

13 Conclusion

Understanding the political biases of AI language models is important for making sure these systems support better-informed, balanced, and inclusive public discourse. Our research shows sharp divides among the artificial intelligence models: ChatGPT-4 consistently leans left, Perplexity is libertarian capitalistic, Claude is cautiously left-of-center, and Google Gemini is mixed and centrist. These biases have tremendous impacts on the perceptions and decisions of the masses. Hence, interventions in the form of transparency, usage of diverse data for training, techniques of detection and correction of bias, and educating the user are very much required in order to counteract these types of biases.

While these findings provide important points of reference, our research does have limitations. First, it could be that the training data behind these AI models changed over time when the research was conducted; this might affect their tendencies to bias over time. Second, the models assessed had tests conducted by using a limited set of politically oriented prompts, which may or may not reveal the gamut of biases in other contexts. Thirdly, while it was possible to point to political biases, without further research, it is hard to quantify the practical effect of such biases on making decisions and public opinion. Such biases can be mitigated if the developers of AI prioritize equity and neutrality in AI systems themselves through a multi-dimensional approach that involves several stakeholders continuously. The research in the future should be directed at finding out

how biases enter into the AI system during its training and development, the role of diverse and representative datasets in mitigating these biases, and the long-term impact of such biases in different cultural and political milieus. Most of all, this may enable longitudinal studies to track changes in AI biases longitudinally and test the efficacy of various strategies for their mitigation.

Appendix 1

Pew Research Center's Political Typology Quiz: 20 questions covering broad topics pertaining to political values, beliefs, and policy positions to categorize respondents into one of nine ideological cohorts (Pew Research Center n.d.).

Appendix 2

Political Compass Assessment: 62 propositions to place respondents on a two-dimensional grid measuring economic left-right orientation and social authoritarianism vs. libertarianism (Political Compass n.d.).

Appendix 3

ISideWith political party quiz: A set of 158 questions probing AI models' views on key political issues such as the role/size of government, globalization, healthcare, environmental, national security, foreign policy, immigration, technology, and social justice (ISideWith n.d.).

Disclosure of Interests. Funding: This research received no external funding. Institutional Review Board Statement: Not applicable. Informed Consent Statement: Not Applicable. Author Contributions: The sole author conducted all the research, writing, and analysis for this work. Conflicts of Interest: The author declares no conflict of interest. Data Availability Statement: The data and tables used in this study are available in the preprint version of this paper, which can be accessed at the following link: <https://www.preprints.org/manuscript/202407.1274/v1>.

References

1. Binns, R.: Fairness in Machine Learning: Lessons from Political Philosophy. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR, pp. 149-159. (2018)
2. Mitchell, T. M.: Machine Learning. McGraw Hill (1997)
3. Noble, S. U.: Algorithms of Oppression: How Search Engines Reinforce Racism. New York University Press (2018)
4. O'Neil, C.: Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group (2016)
5. Obermeyer, Z., B. Powers, C. Vogeli, and S. Mullainathan: Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. *Science*, vol. 366, no. 6464, pp. 447-453 (2019)
6. Crawford, R., and R. Calo. There is a Blind Spot in AI Research. *Nature*, vol. 538, no. 7625, pp. 311-313 (2016)
7. Diakopoulos, N.: Accountability in Algorithmic Decision Making," *Communications of the ACM*, vol. 59, no. 2, pp. 56-62 (2016)

8. Eubanks, V.: *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press (2018)
9. ISideWith, "Political Party Quiz," [Online]. Available: <https://www.isidewith.com/political-quiz>.
10. West, S.: Google Bard vs. OpenAI's ChatGPT: Political Bias. TechCrunch (2023)
11. Brundage, M., S. Avin, J. Wang, and G. Belfield: The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. arXiv preprint arXiv:1802.07228 (2018)
12. Morley, J., L. Floridi, L. Kinsey, and A. Elhalal: From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics*, vol. 26, no. 4, pp. 2141-2168 (2020)
13. Mitchell, M., S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru.: Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 220-229 (2019)
14. Russell S., and P. Norvig: *Artificial Intelligence: A Modern Approach*. Prentice Hall (2009)
15. Seshia, S. A., D. Sadigh, and S. S. Sastry: Formal Methods for AI Systems. *Communications of the ACM*, vol. 62, no. 11, pp. 82-91 (2022)
16. Santurkar, S., L. Hong, J. Sharma, and A. Madry: How Does AI Alignment Influence Opinion Distribution? In *Proceedings of the 37th International Conference on Machine Learning*, (2023)
17. Morley, M., J. Floridi, and A. Elhalal: Ethics as a Service: A Pragmatic Operationalization of AI Ethics. *Minds and Machines*, vol. 30, no. 1, pp. 77-89 (2020)
18. Pew Research Center, "Political Typology Quiz," [Online]. Available: <https://www.pewresearch.org/politics/quiz/political-typology/>.
19. Political Compass, "Political Compass Assessment," [Online]. Available: <https://www.politicalcompass.org/test>.
20. Robertson, R., J. Lazer, and C. Wilson, "Auditing Partisan Audience Bias Within Google Search," in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 39-50 (2023)
21. Wu et al.: Bias Analysis in Large Language Models. *Journal of AI Research*, 87(4), 220-234 (2023)
22. Kumar, R., & Lopez, M.: Political Bias in AI Language Models: Google Gemini Case Study. *AI & Society* 39(1), 142-157 (2023)
23. Jones, A., et al.: Domain-Specific Bias in Large Language Models: A Comparative Analysis. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 56(3), 98-115 (2024)
24. Li, J., & Patterson, D.: Generative AI and Political Campaigns: Risks and Opportunities. *Ethics in AI Review*, 72(2), 189-205 (2023)
25. Feng, Y., & Huang, L.: Auditing Frameworks for Bias in Legal and Healthcare AI Systems. *AI Ethics Quarterly*, 22(4), 315-328 (2024)
26. Garcia, L., et al.: International Collaboration in AI Governance: Addressing Bias and Inequality. *Global Technology Policy Journal*, 78(2), 123-144 (2023)
27. Singh, N., & Patel, K.: The Societal Impact of AI Bias: Marginalized Communities and Economic Disparities. *Journal of Technology and Society*, 45(1), 211-230 (2024)
28. Murphy, R., et al.: Multi-layered Strategies for Mitigating Bias in AI. *Journal of AI and Ethics*, 12(1), 99-113 (2023)
29. Chen, X., et al.: Political Bias in Language Models. *Journal of Artificial Intelligence Research*, 78, 123-138 (2023)
30. Altman, S.: Addressing AI Bias. OpenAI Blog (2023)
31. Zhang, L., et al.: AI and Political Polarization: A Critical Review. *Journal of Democracy*, 35(1), 99-113 (2024)
32. Zhang, Y., et al.: The Role of Training Data in Political Bias of AI Models. *Data & Society* 12(3), 210-225 (2024)

33. Pang, B., et al.: Algorithmic Amplification of Political Bias. *IEEE Transactions on Knowledge and Data Engineering* 35(7), 654-667 (2023)
34. Wagner, K., et al.: Balancing Political Ideologies in AI Training Data. *ACM Transactions on Information Systems* 41(2), 1-25 (2023)
35. Brown, T., et al.: Auditing AI Systems for Bias. *Journal of Ethical AI*, 5(1), 33-48 (2023)
36. Bai, L., et al.: Adapting Large Language Models for Specialized Domains: Techniques and Challenges. *Journal of AI Research* 89(2), 203-219 (2024)
37. Shen, Y., et al.: Unleashing the Potential of Generative AI in Multimodal Contexts: A Review of Applications and Ethical Concerns. *International Journal of AI Systems* 56(4), 145-166 (2024)
38. Martin, G., & Thompson, S.: LLM Impact on the Global Workforce: Navigating Economic and Policy Challenges. *AI and Society* 38(1), 321-334 (2023)
39. Vasquez, C., & Garza, P.: Ethical Governance in the Age of AI: Frameworks for Trustworthy Systems. *Ethics in AI Review* 61(3), 187-201 (2023)
40. Rahman, A., & Lee, J.: Generative AI in Content Creation: Opportunities and Legal Challenges. *AI & Law Journal* 41(5), 72-85 (2023)
41. Murphy, D., & Singh, R.: Building Ethical AI Systems: Strategies for Fairness, Accountability, and Transparency (FAT). *Ethics & Emerging Technologies* 49(4), 299-315 (2024)
42. Taylor, K., & Anderson, M.: Policy Innovations for Regulating Large Language Models: Global Perspectives. *Policy & AI Regulation* 29(2), 56-70 (2024)
43. Garcia, L., & Patel, K.: AI Governance and Policy in the Global South: Challenges and Opportunities. *Journal of Technology Policy and Management* 78(1), 123-142.
44. Sun, M., et al.: New Metrics for Assessing Bias in Language Generation. *Transactions of the Association for Computational Linguistics* 11, 45-60 (2023)
45. Green, J., et al.: Longitudinal Effects of AI-Generated Political Content on Voter Behavior. *Political Communication* 40(2), 215-232 (2023)
46. Smith, A., & Jones, B.: Interdisciplinary Approaches to AI Bias. *Journal of AI Ethics* 2(1), 78-92 (2023)
47. Allen, R., et al.: International Cooperation in AI Regulation. *Global Policy* 14(2), 167-179 (2023)
48. Chen, L., & Williams, S.: Legal Implications of AI-Generated Content. *Law and Technology Review* 55(4), 301-318 (2023)
49. Lee, H., et al.: Machine Learning Approaches to Reducing Bias in AI Models. *Proceedings of the International Conference on Machine Learning, 2023*, 1123-1132 (2023)
50. Abid, A., M. Farooqi, and J. Zou, "Persistent Anti-Muslim Bias in Large Language Models," in *ACM FAccT*, pp. 146-157 (2021)
51. Weidinger, L., et al.: Ethical and Social Risks of Harm from Language Models. *Transactions on Machine Learning Research* (2023)
52. Bommasani, R., et al.: On the Opportunities and Risks of Foundation Models. *arXiv preprint arXiv:2108.07258* (2021)
53. Luccioni, A., & Viviano, J.: What's in the Box? An Analysis of Undesirable Content in Multilingual Models. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (2023)
54. Schramowski, P., Turan, C., & Kersting, K.: Large Pre-trained Language Models Contain Human-like Biases of What Is Right and Wrong to Do. *Nature Machine Intelligence* 5(3), 258-268 (2023)
55. de Vries, H., et al.: The Political Impact of Text Generators: Differential Effects on Conservatives and Liberals. *Proceedings of the National Academy of Sciences*, 120(15), e2026070119 (2023)
56. Birhane, A., Prabhu, V., & Kahembwe, E.: Multimodal Datasets: Misogyny, Pornography, and Malignant Stereotypes. *arXiv preprint arXiv:2302.08267* (2023)

57. Wang, T., & Russakovsky, O.: Directional Bias Amplification. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 123-132 (2023)
58. Aga, A., & Brynjolfsson, E.: AI and the Economy: The Dynamic Effects of AI on Labor Markets and Productivity. American Economic Journal: Macroeconomics (2023)
59. Shen, Y., Liu, H., & Wong, D.: Bias Mitigation Strategies for Language Models: A Survey. Journal of Artificial Intelligence Research, 76, 1-42 (2023)
60. Forbes: Top 10 AI Language Models of 2023." Forbes Tech Council. [Online] (2023)
61. Grand View Research: U.S. Generative AI Market Report. Grand View Research. [Online] (2023)
62. Pew Research Center: Political Typology Quiz. Pew Research Center. [Online] (2023)
63. The Political Compass: Political Compass Test. The Political Compass. [Online] (2023)
64. iSideWith: Political Quiz. iSideWith. [Online] (2023)