Review Article
Accepted: 11 March 2026
https://doi.org/10.20901/pm.63.2.08

Chatbots in the Role of Text Generation on Social Media and Their Impact on Political Developments*

MARIJA POKOS LUKINEC, DIJANA OREŠKI, DINO VLAHEK
University of Zagreb Faculty of Organization and Informatics, Varaždin

Summary

This paper aims to provide an overview of the impact of Generative Artificial Intelligence (GenAI) on political events. Through an analysis of relevant studies, it examines both the advantages and disadvantages of Artificial Intelligence (AI), particularly its role in shaping political developments through chatbot-generated content shared on social media. Additionally, it explores the influence of large corporations on content manipulation, raising concerns about transparency and ethical dilemmas. Based on the literature review, potential solutions are proposed that could drive advancements in the field. The development of GenAI relies on tools for recognizing AI-generated content, the creation of deepfake databases, the training of chatbots on diverse datasets, and, most importantly, educating users to ensure they adapt to and responsibly use new technology. However, the question remains: to what extent do AI regulations foster progress, and to what extent do they hinder it?

Keywords: Political Events, Political Governance, Artificial Intelligence, Generative Artificial Intelligence, Deepfake

1. Introduction

The integration of artificial intelligence (AI) is becoming an increasingly prevalent phenomenon across all aspects of life, making it an indispensable part of daily existence. Among its many applications, generative artificial intelligence (GenAI) tools are now frequently used as substitutes for traditional search engines, often without adequate consideration of their implications or critical evaluation of the content they produce. Although GenAI tools are not necessarily designed with the expli-

* This paper is supported by the Croatian Science Foundation under the project SIMON: Intelligent system for automatic selection and machine learning algorithms in social sciences, UIP-2020-02-6312.

cit goal of improving society, their ongoing development and widespread adoption contribute to societal change, bringing with them challenges that warrant thoughtful consideration and ethical oversight (Walker, Schiff and Schiff, 2024). These tools offer a wide range of capabilities, including the generation of text, images, videos, audio, presentations, data, and more. This paper examines the capabilities of text-generating tools, with a focus on chatbots such as Google Gemini, ChatGPT, Perplexity, and Claude. It discusses how these tools may display biased behavior depending on user interactions and the underlying training data (Choudhary, 2025). Artificial intelligence is now deeply integrated into multiple fields, including healthcare, finance, education, and security. It has also become an essential part of the political landscape. Chatbots, through their responses, have the potential to influence political events. Large corporations that develop these tools possess vast amounts of data, raising concerns about how such information might be used across various domains, including politics. This growing influence of generative AI in political contexts raises important questions about its role in shaping public opinion and democratic processes. The motivation for this paper lies in the need to critically examine how generative AI tools, particularly chatbots, may contribute to political manipulation, the spreading of fake news, and the creation of synthetic narratives, often without sufficient transparency or oversight. To overcome manipulation, fake news, and synthetically generated opinions, it is crucial to establish regulations and propose potential solutions that promote the responsible use of generative AI while avoiding ethical dilemmas. However, such regulations may slow down the development of generative AI tools, which also offer significant benefits.

This paper is divided into four sections. Following the introduction, the second section presents the theoretical framework. The third section offers an analysis of the literature and provides a brief overview of the selected articles used in the field review. The fourth section discusses research findings, including insights into the advantages and disadvantages of using generative AI tools, their connection to social media, and their impact on political events. It also addresses issues of transparency, regulations, and ethical concerns. Finally, the concluding chapter summarizes the findings drawn from the literature review.

2. Theoretical Framework

The Artificial Intelligence Act defines AI systems as software systems developed using machine learning approaches that can generate outputs such as content, predictions, or decisions that influence the environment in which they operate (European Parliament and Council of the European Union, 2024).

Generative artificial intelligence refers to forms of AI capable of producing new content. Models of generative artificial intelligence have access to vast amounts of

data and possess an exceptional ability to create new and diverse outputs such as text, images, audio, and video, drawing on information collected from a variety of sources (Gozalo-Brizuela and Garrido-Merchán, 2024).

According to the European Data Protection Supervisor (2023), Large Language Models (LLMs) are advanced artificial intelligence systems based on transformer architectures and trained on exceptionally large volumes of textual data. LLMs can perform a wide range of tasks, including content generation, summarisation, translation, and answering queries. However, their outputs may be inaccurate, biased, or unverifiable due to the nature of their training data and their lack of real-world understanding (European Data Protection Supervisor, 2023).

2.1. Advantages and Disadvantages of Large Language Models in Political Context

The advantage of large language models is that they can become strong few-shot learners, meaning they can perform new tasks such as translation, question answering, or solving arithmetic problems through textual interaction and a small number of demonstrations, without the need for task-specific fine-tuning (Brown *et al*., 2020). Because of this capability and their access to diverse corpora on which they can be trained, they are also able to synthesise complex information and texts from different domains, including political information, legislation, and policy documents.

Despite the advantages, LLMs exhibit limitations that raise concerns in a democratic context. Firstly, algorithmic bias may emerge from imbalanced training data and reinforcement learning processes, potentially reflecting ideological tendencies (Choudhary, 2025). Bias can subtly shape political discourse and influence public perception. Secondly, LLMs operate as a "black-box" systems-limiting explainability (Xu, 2024). Regulatory frameworks such as the Artificial Intelligence Act emphasize transparency requirements precisely because opacity poses governance challenges. The third, and particularly critical limitation, is AI hallucination. AI hallucination is defined as generated text that is nonsensical or not faithful to the provided source material (Ji *et al*., 2023). Such content often gives the impression of fluency and naturalness even though it is factually unfounded, which makes it difficult to recognize at first glance. Hallucinations can lead to serious consequences, such as risks in medicine (incorrect drug recommendations), breaches of privacy, and even the creation of misinformation in a political context that can mislead users and lead them to the wrong conclusions.

2.2. Code of Practice on Disinformation

The Code of Practice on Disinformation is an EU self-regulatory framework in which online platforms, advertising intermediaries, and other stakeholders commit to limiting the creation, amplification, and monetisation of disinformation. It

includes measures such as restricting advertising revenue for disinformation actors, increasing transparency of political and issue-based advertising, improving the integrity of online services, and addressing manipulative behaviours including fake accounts, coordinated inauthentic activity, bot-driven amplification, and malicious deepfakes (European Commission, 2022).

Monitoring reports show that implementation varies across Member States. In Croatia and Slovenia, researchers highlight limited data access for independent analysis, inconsistent enforcement of platform policies, and insufficient transparency regarding political advertising and recommending systems, which restricts the ability of the research community to systematically assess disinformation dynamics (ADMO Hub, 2023).

With the rise of generative AI, the Code increasingly applies to AI-generated and AI-manipulated content. Platforms and AI providers are encouraged to detect, label, and downrank synthetic media, cooperate with fact-checkers, and introduce watermarking or metadata to ensure visible disclosure of AI-generated content. These obligations are reinforced by the Digital Services Act and the transparency duties in the AI Act, which require machine-readable marking and clear labelling of AI-generated or manipulated content (European Commission, 2023).

3. Analysis of the Literature

A search across citation databases WoS, Scopus, and IEEE identified seven relevant studies based on the keywords political governance, political events, GenAI, and AI. Several queries were formulated based on these keywords. In the WoS database, the first query (ALL=genAI, political events) yielded studies by Arslan, Munawar and Cruz (2025) and Khanal, Zhang and Taeihagh (2024), along with one that does not have open access. The second query (ALL=(genAI)) AND ALL=(political events OR governance) AND ALL=(social media)) returned the study by Arslan, Munawar and Cruz (2025) and an article related to the healthcare field that was not taken into consideration, as its primary focus lay outside the political scope of this analysis. In the Scopus database, the query (political AND governance AND political AND events AND AI) identified studies by Xu (2024) and Gorwa, Binns and Katzenbach (2020). Additionally, documents of the European Parliament (2023; 2024) were identified through searches of European Parliament pages. In the IEEE database, the query (political governance, GenAI) returned the study by Choudhary (2025). Conversations with generative artificial intelligence (Perplexity) suggested studies by Calvo and Saura Garcia (2024), Bueno de Mesquita *et al*. (2023), and Lee and Yuan (2024) as relevant sources.

Table 1 provides a concise overview of the key studies and regulatory documents discussed in the narrative analysis below. The selected contributions are writ-

ten in English, published within the last five years, and primarily available through open access sources. In addition to studies identified through citation databases, the review includes relevant regulatory acts and institutional documents from the European Union and China addressing the governance of artificial intelligence. Given China's position as a major actor in artificial intelligence development and its distinct regulatory model, its inclusion broadens the analytical perspective on contemporary AI governance paradigms in comparison with the European Union.

Table 1. Analyzed Studies

Title of research	**Author / Institution**	**Year**	**Focus**
"Political Bias in Large Language Models: A Comparative Analysis of ChatGPT-4, Perplexity, Google Gemini, and Claude"	Choudhary	2025	Ideological bias in LLMs
"Political-RAG: using generative AI to extract political information from media content"	Arslan, Munawar and Cruz	2025	Political-RAG framework
"Why and how is the power of Big Tech increasing in the policy process? The case of generative AI"	Khanal, Zhang and Taeihagh	2024	Big Tech influence in GenAI policy process
"Generative AI and Democracy: the synthetification of public opinion and its impacts"	Calvo and Saura Garcia	2024	GenAI and democracy
"Algorithmic content moderation: Technical and political challenges in the automation of platform governance"	Gorwa, Binns and Katzenbach	2020	Algorithmic moderation
"Preparing for Generative AI in the 2024 Election: Recommendations and Best Practices Based on Academic Research"	Bueno de Mesquita *et al*.	2023	GenAI and elections
"Opening the 'black box' of algorithms: regulation of algorithms in China"	Xu	2024	Algorithm regulation (China)
"Merging AI Incidents Research with Political Misinformation Research: Introducing the Political Deepfakes Incidents Database"	Xu; Lee and Yuan	2024	Political Deepfakes Incidents Database

"Artificial intelligence, democracy and elections"	European Parliament	2023	AI and elections
"Artificial Intelligence Act"	European Parliament	2024	AI Act
"The Authenticity Paradox of Political AI Chatbots on Voters' Candidate Perceptions"	Lee and Yuan	2024	AI chatbots and voter perceptions

The consulted literature on generative AI in political contexts can be grouped into four dimensions: (1) bias in large language models, (2) information extraction and automation, (3) democratic and epistemic risks, and (4) regulatory and governance responses.

Choudhary (2025) conducts a comparative analysis of ChatGPT-4, Perplexity, Gemini, and Claude, demonstrating that these models exhibit systematic response patterns when tested against politically sensitive questionnaires. The findings indicate a tendency toward liberal-leaning positions, particularly on issues such as immigration and minority rights, while conservative-leaning responses appear less consistent. Similarly, Lee and Yuan (2024) examine AI chatbots in campaign contexts and show that perceived authenticity significantly shapes voter attitudes, suggesting that bias is not only structural but also perceptual. Together, these studies indicate that LLMs may influence political perception both through ideological positioning and through the framing of candidate authenticity.

Beyond bias, generative AI is being increasingly used for extracting structured political information. Arslan, Munawar and Cruz (2025) introduce the Political-RAG framework, which applies retrieval-augmented generation to extract political events from media sources. This approach demonstrates the potential of LLMs to automate large-scale political content analysis and support real-time monitoring of political developments. Such systems may enhance analytical efficiency, but they also depend heavily on data quality and the retrieval design.

A broader strand of literature examines systemic risks to democratic processes. Calvo and Saura Garcia (2024) argue that generative AI contributes to the artificial shaping of public opinion, increasing the circulation of AI-generated narratives that may distort deliberative processes, including through fake news, disinformation and reduced critical thinking. Bueno de Mesquita *et al*. (2023) emphasize electoral risks, including misinformation amplification and declining trust, while recommending transparency measures such as watermarking. Deepfakes represent a particularly acute manifestation of these risks. Xu (2024) and Lee and Yuan (2024) introduce the Political Deepfakes Incidents Database, documenting the growing

prevalence of AI-generated political media and highlighting the need for monitoring. At the platform governance level, Gorwa, Binns and Katzenbach (2020) identify technical and political challenges in algorithmic content moderation, including transparency, unequal treatment of viewpoints, and depoliticized decision-making processes.

Regulatory responses vary across political systems. The European Union has developed a layered governance model combining soft law and binding instruments. The report "Artificial Intelligence, Democracy and Elections" (European Parliament, 2023) outlines risks and mitigation strategies, while the Artificial Intelligence Act introduces a risk-based regulatory framework for AI systems. In parallel, Xu (2024) shows that China has pursued phased algorithm regulation, progressing from reactive enforcement, through the introduction of ethical guidelines and disciplinary agreements, to formalized legislation and implementation. Khanal, Zhang and Taeihagh (2024) further argue that generative AI strengthens the structural power of major technology companies in the policy process, raising questions about democratic accountability and regulatory capture.

Taken together, the reviewed studies illustrate how generative artificial intelligence is increasingly shaping contemporary political systems. They address ideological bias in large language models, the growing influence of major technology companies in policy processes, and the risks posed by synthetic media such as deepfakes to democratic integrity. The literature also highlights divergent regulatory approaches, from European risk-based governance models to China's phased algorithm regulation.

At the same time, important questions remain open. Research provides limited insight into how these dynamics unfold in smaller political systems, how different political cultures mediate AI-driven risks, and to what extent national media ecosystems influence model bias and public trust. These gaps indicate the need for further comparative and context-sensitive research.

4. Synthesis of the Reviewed Studies

Although the search of citation databases yielded only a few articles specifically addressing the political implications of generative tools, the topic remains highly relevant and contemporary. Several academic and non-academic sources, including news media, policy briefs, and institutional reports, actively explore this intersection (Bueno de Mesquita *et al*., 2023; Walker, Schiff and Schiff, 2024). The issue has drawn increasing interest within the European Parliament, where discussions focus on the regulatory, ethical, and societal impacts of generative AI technologies on democratic processes (European Parliament, 2024).

4.1. Generative Artificial Intelligence Tools for Text Generation and Model Bias

This subsection examines how generative artificial intelligence tools used for text generation exhibit political and ideological bias and how such bias may influence democratic processes. Building on the broader discussion in the fourth section regarding the political implications of generative AI, this part narrows the focus to large language models (LLMs) that are widely accessible to the public and increasingly used for political information, opinion formation, and discourse. Rather than merely summarizing existing studies, the subsection synthesizes empirical findings to identify recurring bias patterns across different models, evaluates the societal and political consequences of such biases, and situates these findings within current regulatory and ethical debates. The aim is to demonstrate that model bias is not an isolated technical issue, but a structural challenge with direct implications for political behavior, democratic participation, and governance.

The study by Choudhary (2025) compares ChatGPT-4, Claude, Google Gemini, and Perplexity. The tools were selected for analysis due to their popularity and accessibility. The research was conducted using three different sets of questionnaires to determine which political stance AI models lean towards and how they influence citizens and society. The title and purpose of the questionnaires are presented in Table 2.

According to Choudhary (2025), bias is inherently embedded in generative AI models. The study highlights how biased AI technologies shape political opinions and influence political behavior, including voting decisions. Due to the variability in chatbot responses, concerns arise about their impact on political discourse and democratic processes. To address this, Choudhary (2025) highlights the importance of transparency, training models on diverse datasets, and educating users to mitigate bias. Educating the users fosters responsible usage and enhances critical thinking. The paper introduces fine-tuning as a method of additional model training to reduce bias and improve accuracy. However, Choudhary (2025) notes that improper execution can still lead to biased models. Also, in the paper 'Assessing political bias in large language models' the authors argue that political bias in LLMs can influence citizens' behavior, shape political attitudes, and thereby affect democratic elections (Rettenberger, Reischl and Schutera, 2025).

According to Choudhary (2025), Google Gemini is closer to the political center, but after fine-tuning, it leans toward more liberal positions. In the same study, the author notes that GPT-3 exhibited biases toward certain religious and ethnic groups. This conclusion is based on the findings of Abid *et al*. (2021), who demonstrated a persistent anti-Muslim bias in GPT-3 resulting from imbalances in the training data. In contrast, ChatGPT-4 tends to lean toward more liberal positions. Similarly, Google Gemini also leans in the same direction, whereas Perplexity and Claude, al-

though providing results closer to the political left, are positioned closer to the center with more moderate views (Choudhary, 2025). In contrast, Grok, developed by xAI, is empirically analyzed as leaning closer to the political center compared to its frontier competitors. While models such as ChatGPT are frequently scrutinized for inherent political constraints and alignment biases, recent experimental data suggests that Grok demonstrates a more balanced distribution across polarizing societal statements (Rayner and Chatgpt C-Lara-Instance, 2025). Systematic audits indicate that although Grok is marketed as a pro-traditional alternative, its performance often converges with other large language models on factual grounding, yet it maintains a statistically distinct position by avoiding the consistent ideological clustering observed in more strictly moderated systems (*ibid*.). However, some studies suggest a tendency toward libertarian or right-leaning positions, depending on the topic and phrasing of the query (Rozado, 2024). Choudhary (2025) observed that the tools' positions fluctuated across different questionnaire topics, with responses lacking consistent uniformity or clear alignment with a single ideological stance. Nonetheless, all tools exhibited a certain degree of liberal inclination overall.

Table 2. Questionnaires Used in Study by Choudhary (2025) for Categorizing Gen AI Tools

Survey title	**Survey purpose**
"Pew Research Center's Political Typology Quiz"	A questionnaire consisting of 20 questions aimed at categorizing respondents into one of nine ideological groups.
"PoliticalCompass.org Assessment"	A questionnaire containing 62 statements was developed to classify respondents as leaning left or right and to determine whether they align with liberalism or social authoritarianism.
"ISideWith political party quiz"	A quiz featuring 158 questions designed to assess AI models' perspectives on key issues, including healthcare, foreign policy, immigration, and social justice.

In addition to detecting bias in generative artificial intelligence tools, the development of large language models based on generative AI, along with natural language processing (NLP) methods, has made it easier to identify hate speech, conflict, and political bias present on social media and in the media, all of which also influence political developments (Arslan, Munawar and Cruz, 2025). As noted by Arslan, Munawar and Cruz (2025), the development of frameworks such as RAG enables access to external sources like Twitter and news articles, thereby facilitating

more reliable information gathering. Arslan, Munawar and Cruz (2025) developed a chatbot designed to detect content on social media and in news articles. During development, they reviewed publicly available LLMs, including BLOOM, Falcon, GPT-4, Llama2, and Chinchilla, ultimately selecting Llama2 for their chatbot due to its suitability for their requirements. By leveraging RAG and LLMs, the chatbot aims to assist analysts in extracting political events, enhancing their understanding, and responding to political inquiries. Additionally, integrating external sources would help reduce model bias.

Major corporations such as Google, Amazon, Meta, Apple, and Microsoft exert significant influence over political developments. As noted by Khanal, Zhang and Taeihagh (2024) and Calvo and Saura Garcia (2024), these platforms shape political discourse and public awareness by controlling the visibility of content accessed by both the public and researchers. Some of these companies, particularly those owning media outlets, also leverage user data to produce content tailored to individual preferences. In this way, large corporations determine what information is available and how it is accessed (Khanal, Zhang and Taeihagh, 2024). As a result, people are often exposed only to information that reinforces their existing views, which reduces exposure to diverse perspectives and undermines critical thinking. In this way, large corporations determine what information is available and how it is accessed (*ibid*.). Access to data and information held by these companies significantly increased during the COVID-19 pandemic. A similar example has been noted in the Republic of Croatia, where research has shown that journalists feel pressure from media owners and advertisers regarding the content they publish (Peruško, Vozab and Trbojević, 2022). It can be concluded that large corporations significantly shape political dynamics through the vast resources they control. Depending on the position of the governing authorities, this influence may be regulated or allowed to evolve naturally under a laissez-faire approach.

Calvo and Saura Garcia (2024) introduce the concept of data feudalism. According to them, data feudalism refers to a new system that, on the one hand, is based on the dependence of states, markets, and the civil society on platforms and services where large digital corporations exercise monopolistic control over data and dominate algorithms. On the other hand, it is founded on the instrumentalization and re-feudalization of the civil society market by major digital corporations to achieve economic and political goals. Large corporations are increasingly mass-producing content and services, guiding citizens toward using social media for communication, mutual understanding, and democratic engagement. Through this, they exert significant influence over both the economic and political landscape.

Calvo and Saura Garcia (2024) also highlight the vast amount of false information and data manipulation, which has become difficult to detect due to genera-

tive artificial intelligence tools that are now more precise and capable of creating content indistinguishable from human-generated material. Additionally, the use of social bots is becoming more frequent (Calvo and Saura Garcia, 2024). A new phenomenon shaping political events and public perception is deep synthetic data – data packages created by generative AI that, unlike deepfakes, do not conceal the fact that they are synthetically generated. Calvo and Saura Garcia (2024) suggest measures to mitigate negative effects by promoting critical thinking and public discourse, creating space for debate that is independent of large corporations and synthetically generated data and speech. This approach would also limit the influence of large corporations on political decision-making.

In the study by Gorwa, Binns and Katzenbach (2020), an example is described where footage of a terrorist attack went viral on social media despite the original video being swiftly removed. Similarly, in the political sphere, deepfake videos have emerged in various regions, with examples provided by Bueno de Mesquita *et al*. (2023) and Lee and Yuan (2024).

One such video falsely links a political figure to a terrorist organization, while another portrays a dystopian future based on the election of certain politicians in 2024 (Bueno de Mesquita *et al*., 2023). There are many more such cases. The topics of disinformation and fake news have also attracted increasing attention within the Croatian scientific community, particularly from 2017 onward (Peruško, Vozab and Trbojević, 2022), a trend largely stimulated by the outcome of the 2016 U.S. presidential elections.

Thanks to artificial intelligence and machine learning, social media platforms such as YouTube, Facebook, and Twitter can classify far more content, detect undesirable information, and prevent the publication of certain materials or the creation of profiles aimed at spreading harmful content. The study by Gorwa, Binns and Katzenbach (2020) underscores the urgent need for transparency, fairness, and the depoliticization of AI governance. Crucially, the authors highlight the material impact of algorithmic realization, the process by which automated predictions translate into tangible social outcomes. When AI systems predict ‘risk’ or ‘suitability’ in sectors like criminal justice or social credit, these predictions often entrench historical biases, effectively ‘realizing’ or manifesting systemic inequalities under the guise of neutral data. Therefore, critical examination should focus not only on the technical accuracy of these models but also on the socio-political consequences of their implementation. These are key elements that should be critically examined, and discussion on these topics should be encouraged.

4.2. Generative AI in Political Discourse: Potential, Risks, and Ethical Challenges

A review of the consulted studies reveals that generative artificial intelligence tools offer numerous advantages while simultaneously posing challenges due to their limitations. Generative artificial intelligence tools, specifically chatbots, provide fast and instant responses to most questions posed to them, including political events, political processes, politics, institutions, and public policies (Arslan, Munawar and Cruz, 2025). Due to the vast datasets on which they are trained and which they use as their knowledge base, they offer a high degree of creativity (Choudhary, 2025), sometimes to the extent that they may fabricate information, leading to disinformation in an attempt to satisfy the query.

Suppose the person asking questions has sufficient knowledge and critical thinking skills, in that case, a chatbot can assist with understanding political events and help prepare individuals to participate more effectively in political debates (European Parliament, 2023). This would also prevent the issue of people viewing chatbots as replacements for search engines (Bueno de Mesquita *et al*., 2023), which could limit the variety and scope of information they receive. On the other hand, despite its numerous advantages, the vast amount of private data stored in its knowledge base poses ethical dilemmas that may have social consequences if privacy rights are not upheld (Choudhary, 2025).

A document by the European Parliament (2023) stated that chatbots are powerful tools for generating fake news, misinformation, and biased opinions. With rapid advancements, they can also create deepfake videos, which may have a greater impact than textual content (*ibid*.). The targets of such videos are often politically active individuals in high positions, as illustrated by examples presented in the study by Bueno de Mesquita *et al*. (2023). The advantages outlined should continue to be developed, and further progress should focus on minimizing shortcomings. Therefore, it is crucial to examine the ethical dilemmas associated with the use of generative AI tools and assess whether regulations can support technological advancement and improvement.

4.3. Ethical Dilemmas and Regulations in the Use of Generative Artificial Intelligence

As Choudhary (2025) emphasizes, deep learning models require vast datasets, frequently containing private user data, which heightens ethical concerns around misuse and the risk of data leaks by corporations like Meta, Amazon, and Google. As mentioned earlier, the influence of social media and large corporations on political events can be addressed through regulation or a laissez-faire approach. While there are many examples of governments collaborating with major corporations for mutual interests, there are also cases where regulations are perceived as obstruct-

ing development, leading to concessions that allow large companies to pursue their agendas (Khanal, Zhang and Taeihagh, 2024).

Studies by Xu (2024) and the European Parliament (2024) provide examples from the European Union, China, and the United States, where regulations, laws, and guidelines for the use of GenAI have been enacted or are under debate. Each of these regions demonstrates both similarities and differences in its approaches to the rapid development and implementation of AI tools. In June 2024, the European Union enacted the AI Act (European Parliament, 2024), marking the first document to provide guidelines on the use and distribution of AI systems in Europe. According to the Act (*ibid*.), EU member states must designate competent authorities to oversee implementation and compliance, a step already taken by Croatia. The Act outlines four risk categories: unacceptable risk, high risk, transparency risk, and minimal or no risk. These categories define what constitutes a risk, and the Act specifies which AI systems are prohibited, including social scoring, real-time remote biometric identification in public spaces, emotion recognition in workplaces or schools, biometric categorisation based on sensitive traits, subliminal manipulation, exploitation of vulnerable groups, and untargeted scraping of facial images. These practices are banned because they threaten fundamental rights, privacy, and democratic values in the European Parliament (2024).

China, as one of the global leaders in AI development, also faces challenges related to privacy, bias, and discrimination. These issues have led to discussions on regulatory measures focusing on AI algorithms (Xu, 2024). The first phase of regulation involved event-based responses and penalties, primarily targeting large corporations that exploit their positions. In the second phase (2021), China shifted its focus toward ethical guidelines, the formation of leading opinions, and a self-discipline agreement emphasizing transparency, reliability, and explainability (*ibid*.). These guidelines are based on those established by the European Commission in 2018. The third phase focused on legislation and implementation. China and the EU set similar conditions for AI usage, emphasizing transparency, accountability, and fairness (*ibid*.). While the European Union introduced the AI Act in 2024, China had already implemented legal regulations on AI. In contrast, AI legislation in the U.S. remains in the discussion and negotiation phase (*ibid*.).

4.4. Potential Solutions

Although generative AI tools are designed to be neutral in their outputs, their responses may still reflect certain biases depending on the training datasets and fine-tuning processes (Choudhary, 2025). The consulted studies propose solutions to better detect and mitigate the challenges and shortcomings associated with generative artificial intelligence. Some of the suggested approaches include: AI detec-

tion tools, watermarking techniques, international collaboration, education and awareness programs, training on diverse datasets, and utilizing databases of existing deepfakes.

AI detection tools such as GPTZero, OpenAI classifier, and DetectGPT have been developed. However, these tools are not always entirely reliable due to the rapid advancements in generative AI, which allow it to create content that is increasingly similar to human-generated material (European Parliament, 2023). Labeling generative AI outputs with visual markers, such as watermarks, to differentiate them from human-created content has emerged as a key transparency mechanism in policy discussions. This approach is explored in various studies, including those by Gorwa, Binns and Katzenbach (2020), Xu (2024) and the European Parliament (2023). However, such markers may be difficult to detect or easily removed (European Parliament, 2023). It is also possible to generate content that follows a pattern visible only through algorithmic detection (*ibid*.; Bueno de Mesquita *et al*., 2023). However, this approach restricts the creation of high-quality content that generative AI tools can produce (European Parliament, 2023). Interdisciplinary and international collaboration, coupled with the exchange of knowledge across fields such as artificial intelligence, social sciences, and privacy, has the potential to strengthen transparency and promote the detection of biased perspectives (Choudhary, 2025; European Parliament, 2023). The study also discusses global political powers such as China, the U.S., and the EU, which are passing laws and debating AI regulations. While some decisions are similar, they have not been fully harmonized, and some countries are still considering whether to implement regulations at all (Xu, 2024). Although regulations are frequently criticized for stifling innovation and progress, Lee *et al*. (2023), as cited in Choudhary (2025), contend that bias in AI systems can be mitigated through machine learning techniques without compromising model performance. As a proposed solution to reduce model bias, it is suggested that chatbots be trained on diverse datasets that reflect different political perspectives to ensure balanced information. Additionally, bias detection and correction methods are recommended by Choudhary (2025). The development of the RAG framework enables access to external information, helping to reduce unreliable data (Arslan, Munawar and Cruz, 2025; Bueno de Mesquita *et al*., 2023). By improving these systems to minimize model limitations and ensure timely, relevant responses and information on political events, the likelihood of bias can be reduced.

Xu (2024) also highlights the existence of databases containing deepfake content, which frequently target politically active individuals. Some of the examples include the Political Deepfakes Incidents Database (https://airtable.com/appOU03dlKuBdbmty/shrEkrIYINbrcKQ3z/tbleGYjNLn2D4Xfzs), the AI Vulnerability Database (https://avidml.org/), and the Awful AI Database (https://github.

com/daviddao/awful-ai). These databases primarily focus on content related to the U.S. political landscape. Their purposes vary, ranging from analyzing the impact of such content on political decision-making to developing and testing new algorithms for detecting deepfake material (Xu, 2024). User education is essential and can be applied in two directions – toward politicians or voters. AI can support voter education during elections by providing information on candidate programs and guiding discussions toward constructive debates instead of hate speech and attacks (European Parliament, 2023).

On the other hand, AI can be valuable for politicians by gathering public sentiment from social media, identifying important citizen concerns, and helping shape political initiatives accordingly (*ibid*.). Peruško, Vozab and Trbojević (2022) also mention the audience's perspective, focusing on readers' trust in the media and their ability to recognize fake news and disinformation. This serves to assess how informed readers are about the content they consume. It was noted that television use is positively correlated, while internet use is negatively correlated, with trust in political and social institutions and the elites (Peruško, Vozab and Trbojević, 2022). Ultimately, personal responsibility, critical thinking, and verifying information across multiple sources are crucial. Taken together, these measures highlight that reducing political bias and manipulation in the age of generative AI does not only depend on technical safeguards, but also on strengthening citizens' media literacy and their capacity to recognise AI-generated content. Such informed engagement is essential for maintaining the trust in political processes and forms a logical bridge to the broader implications discussed in the following chapter.

5. Discussion and Conclusion

The story of generative AI tools and their use in political systems has two sides. While AI offers numerous benefits, it also presents potential risks. Given that the development of generative AI tools is accelerating and introducing various challenges related to privacy policies, it would be valuable to consider ways to mitigate these issues. The literature review highlights numerous solutions to address bias and ethical dilemmas, particularly within the political domain. Some authors argue that regulating and restricting AI hinders its progress and development, but it is important to note that techniques and machine learning methods have already been developed to reduce model bias without compromising algorithm performance. In all AI-generated products, ensuring explainability and human understanding is crucial, as this guarantees model credibility. Ultimately, it is the individual who analyzes the information, makes decisions, and takes responsibility for their actions. Since the review did not include Croatia in detail due to limited information available on this topic, it would be beneficial to encourage greater interest among researchers to

explore this area further. A literature search on regulations and acts related to Croatia did not yield any results, as the findings were only linked to EU regulations. Additionally, the websites of the Croatian Personal Data Protection Agency and the Ministry of Justice, Administration, and Digital Transformation only contain a list of competent authorities appointed under the AI Act.

Future research should move beyond descriptive analyses toward systematic empirical investigations of generative AI in political communication. Experimental designs could examine how variations in training data influence ideological bias in chatbot outputs. Comparative cross-platform studies may assess whether different large language models produce systematically divergent political narratives. Furthermore, longitudinal research is needed to evaluate the influence of generative AI systems and electoral processes over time. Special attention should be devoted to the relationship between AI hallucinations, platform governance mechanisms, and regulatory frameworks such as the Digital Services Act and the Artificial Intelligence Act. Such research would contribute to a deeper understanding of how generative AI reshapes information ecosystems.

Declaration on Generative AI

During the preparation of this work, the author(s) used Copilot and Grammarly in order to: Edit references and grammar, and perform a spelling check. After using said tool(s)/service(s), the author(s) reviewed and edited the content as needed, and take(s) full responsibility for the publication's content.

REFERENCES

Abid, A., Farooqi, M. and Zou, J. (2021) 'Persistent anti-Muslim bias in large language models', *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES 2021)*, pp. 146–157.

ADMO Hub (2023) *Monitoring the implementation of the Code of Practice on Disinformation in empowering the research community in Croatia and Slovenia*. Available at: https://admohub.eu/en/research/report-monitoring-the-implementation-of-the-code-of-practice-on-disinformation-in-empowering-the-research-community-in-croatia-and-slovenia/ (Accessed: 19 February 2026).

Arslan, M., Munawar, S. and Cruz, C. (2025) 'Political-RAG: using generative AI to extract political information from media content', *Journal of Information Technology & Politics*, pp. 479–494 [online]. Available at: https://doi.org/10.1080/19331681.2024.2417263 (Accessed: 27 June 2025).

Bueno de Mesquita, E., Canes-Wrone, B., Hall, A. B., Lum, K., Martin, G. J. and Velez, Y. R. (2023) *Preparing for Generative AI in the 2024 Election: Recommendations and Best Practices Based on Academic Research*. Stanford Graduate School of Business and University of Chicago Harris School of Public Policy.

Brown, T. *et al*. (2020) 'Language Models are Few-Shot Learners', *Advances in Neural Information Processing Systems (NeurIPS 2020)*, 33, pp. 1877–1901.

Calvo, P. and Saura Garcia, C. (2024) 'Generative AI and Democracy: the synthetification of public opinion and its impacts' [online]. Available at: https://ssrn.com/abstract=4911710 (Accessed: 27 June 2025).

Choudhary, T. (2025) 'Political Bias in Large Language Models: A Comparative Analysis of ChatGPT-4, Perplexity, Google Gemini, and Claude', *IEEE Access*, 13, pp. 11341–11379. Available at: https://doi.org/10.1109/ACCESS.2024.3523764 (Accessed: 27 June 2025).

European Commission (2022) *Code of Practice on Disinformation*. Available at: https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation (Accessed: 19 February 2026).

European Commission (2023) *Guidance on labelling and transparency for AI-generated content*. Available at: https://digital-strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content (Accessed: 19 February 2026).

European Data Protection Supervisor (2023) Large Language Models (LLM). TechSonar. Available at: https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/large-language-models-llm_en (Accessed: 19 February 2026).

European Parliament (2023, September) 'Artificial intelligence, democracy and elections' [online]. Available at: https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)751478 (Accessed: 27 June 2025).

European Parliament (2024, September) 'Think Tank Artificial Intelligence Act' [online]. Available at: https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792 (Accessed: 27 June 2025).

European Parliament and Council of the European Union (2024) Regulation (EU) 2024/1689, *Official Journal of the European Union*. Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj (Accessed: 19 February 2026).

Gorwa, R., Binns, R. and Katzenbach, C. (2020) 'Algorithmic content moderation: Technical and political challenges in the automation of platform governance', *Big Data & Society*, 7, pp. 1–15. Available at: https://doi.org/10.1177/2053951719897945 (Accessed: 27 June 2025).

Gozalo-Brizuela, R. and Garrido-Merchán, E. C. (2024) 'A survey of generative AI applications', *Journal of Computer Science*, 20(8), pp. 801–818.

Ji, Z. *et al*. (2023) 'Survey of Hallucination in Natural Language Generation', *ACM Computing Surveys*, 55(12), pp. 1–38.

Khanal, S., Zhang, H. and Taeihagh, A. (2024) 'Why and how is the power of Big Tech increasing in the policy process? The case of generative AI', *Policy and Society*, puae012. Available at: https://doi.org/10.1093/polsoc/puae012 (Accessed: 27 June 2025).

Lee, H. *et al*. (2023) 'Machine learning approaches to reducing bias in AI models', *Proceedings of the International Conference on Machine Learning (ICML 2023)*, pp. 1123–1132.

Lee, Y.-H. and Yuan, C. W. (Tina) (2024) 'The Authenticity Paradox of Political AI Chatbots on Voters' Candidate Perceptions' [online]. Available at: https://ssrn.com/abstract=5074078 (Accessed: 27 June 2025).

Peruško, Z., Vozab, D. and Trbojević, F. (2022) 'Prerequisites for understanding the role of the media system in deliberative democracy: 20 years of media system research in Croatia', *Politička misao*, 59(3), pp. 109–134. Available at: https://doi.org/10.20901/pm.59.3.04 (Accessed: 27 June 2025).

Rayner, M. and Chatgpt C-Lara-Instance (2025) 'How Woke is Grok? Empirical Evidence that xAI's Grok Aligns Closely with Other Frontier Models'. Unpublished [Preprint]. Available at: https://doi.org/10.13140/RG.2.2.35804.45449 (Accessed: 27 June 2025).

Rettenberger, L., Reischl, M. and Schutera, M. (2025) 'Assessing political bias in large language models', *Journal of Computational Social Science*, 8(2), pp. 1–17. Available at: https://doi.org/10.1007/s42001-025-00376-w (Accessed: 27 June 2025).

Rozado, D. (2024) 'The political preferences of LLMs', *PLoS ONE*, 19(7), e0306621. Available at: https://doi.org/10.1371/journal.pone.0306621 (Accessed: 10 July 2025).

Walker, C. P., Schiff, D. S. and Schiff, K. J. (2024) 'Merging AI Incidents Research with Political Misinformation Research: Introducing the Political Deepfakes Incidents Database', Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24), 38(21), 23053–23058 [online]. Available at: https://doi.org/10.1609/aaai.v38i21.30349 (Accessed: 27 June 2025).

Xu, J. (2024) 'Opening the "black box" of algorithms: regulation of algorithms in China', *Communication Research and Practice*, 10(3), pp. 288–296. Available at: https://doi.org/10.1080/22041451.2024.2346415 (Accessed: 27 June 2025).

Marija Pokos Lukinec, Dijana Oreški, Dino Vlahek

CHATBOTOVI U ULOZI GENERIRANJA TEKSTA NA DRUŠTVENIM MREŽAMA I NJIHOV UTJECAJ NA POLITIČKA ZBIVANJA

Sažetak

Cilj rada je dati pregled područja utjecaja generativne umjetne inteligencije na politička zbivanja. Kroz analizu relevantnih radova razmatraju se prednosti i nedostaci umjetne inteligencije, posebice njena uloga u oblikovanju političkih procesa preko sadržaja generiranog *chatbotovima*, a podijeljenog na društvenim mrežama. Kroz to se sagleda i utjecaj velikih korporacija na mogućnost manipulacije sadržajem što ujedno dovodi do problema transparentnosti i etičkih dilema. Prema pregledu literature dana su potencijalna rješenja tog problema koja vode prema napretku u tom području. Razvoj generativne umjetne inteligencije leži u korištenju alata za prepoznavanje sadržaja generiranog umjetnom inteligencijom, kreiranju *deepfake* baza, treniranju *chatbotova* na različitim setovima podataka, a najviše u obrazovanju korisnika i usmjeravanju prema prilagodbi i pravilnom korištenju nove tehnologije. Zasad ostaje otvoreno pitanje u kojoj mjeri su regulative korištenja umjetne inteligencije dobre, a u kojoj koče napredak.

Ključne riječi: politički događaji, političko upravljanje, umjetna inteligencija, generativna umjetna inteligencija, *deepfake*

Marija Pokos Lukinec, Teaching Assistant, University of Zagreb Faculty of Organization and Informatics, Varaždin, Croatia. *E-mail*: mapokos@foi.hr

Dijana Oreški, Associate Professor, University of Zagreb Faculty of Organization and Informatics, Varaždin, Croatia. *E-mail*: Dijana.Oreski@foi.unizg.hr

Dino Vlahek, Assistant Professor, University of Zagreb Faculty of Organization and Informatics, Varaždin, Croatia. *E-mail*: dvlahek@foi.hr