The Role of AI in Shaping Public Discourse: Ethical Implications and Societal Impact

Syam sai santosh Bandi Roll no: 2022528
Siddhant Bali, Roll no: 2022496
Abbas Murtaza, Roll no: 20220012
Rishabh kumar, Roll no:2022402
Nishi Ninawat, Roll no: 2021480

1. Introduction:

As AI is evolving fast, its algorithmic models such as deepfakes, fake news, and algorithmically curated news are revolutionizing information creation, exchange, and use (Al-kfairy et al., 2024). This report considers AI's impact on public discourse, in terms of ethical issues surrounding misinformation, bias, and transparency, and how they affect society, democracy, and public trust (Jungherr, 2023).

The traditional media now exist alongside AI systems that possess the ability to generate highly realistic audio and video. These systems tend to reinforce current biases, warping political narratives and undermining the legitimacy of democratic debate (Jungherr, 2023). While AI promotes innovation, this is at the expense of institutions' credibility (Germani et al., 2024).

Our investigation addresses three overarching objectives: (1) acquainting ourselves with technical concepts and applications of AI-generated content, (2) examining ethical challenges particularly disinformation and prejudice and (3) formulating reasonable proposals that will allow for AI development without jeopardizing democratic values (Al-kfairy et al., 2024; Corrêa et al., 2023). The report aims to be readily traceable for the reader. It begins with a general introduction to AI-generated content, then to ethical concerns, examines the impact on democracy, and concludes with risk mitigation and proposals for further study (Corrêa et al., 2023).Corrêa et al., 2023). The report is systematically structured: it begins with introducing AI-generated media, proceeds with the discussion of ethical dilemmas, looks into the democratic implications, and concludes with reducing concerns along with proposals for potential research work (Corrêa et al., 2023)

2. AI & Public Discourse:

AI-generated content such as deepfakes and algorithmically generated news are fundamentally impacting the public sphere in ways that we have never seen before. The recent generative models such as Generative Adversarial Networks (GANs) and transformer based architectures have permitted the creation of synthetic media nearly indistinguishable from real acts of communication, especially now that hyper realistic images, videos, and narratives can be produced, making it increasingly challenging to determine what is true and what is false. The availability of this content should mark not only a departure in technology, but a fundamental reorganization in how people create beliefs(Sage Journals, 2020; Floridi & Cowls, 2023).

Expansion of deepfakes demonstrates the potential for AI to disrupt the public space. Rising from the creative and artistic endeavors of computer scientists originally focused on generating images of worlds and persons, deepfakes have emerged as potentially dangerous tool for misinformation and manipulation of the public space. The technology can be used for more satire or entertainment purposes, but the most certainly has malicious potential. The findings of the Deepfakes and Disinformation study (Sage Journals, 2020) suggest that fabricated media compromises trust in digital content which casts doubt and challenges the genuineness of visual evidence.

There have been cases when deepfakes targeted public figures, leaders, and purposefully spread disinformation during elections which shows that the danger is not just on the theoretical basis. Experts suggest that advancements will allow present actors to deploy deepfakes with military precision shortly to shape a global narrative and deepen geopolitical tensions (Napoli, 2025). In this the genuineness of the content that is being consumed significantly drops.

In the case of personalization, the new feed personalized by the machine learning algorithm learns what a user seems to prefer over time and what it likes to watch, which moves users' content consumption in the direction of their ideological preferences while also increasingly limiting the types of information the user is likely to encounter. According to Napoli (2025), this cognitive insulation fosters intellectual rigidity and demobilizes consideration of alternative perspectives, undermining the democratic principles of deliberative pluralism.

What is most troublemaking part is the lack of transparency around how the information is being consumed. Whole bundles of content are being filtered, ranked, and curated by algorithms in ways that users rarely know or ever will be able to understand. Public discourse, historically driven by human editors, guided by ethical and professional norms of journalism. As Floridi and Cowls (2023) point out, these AI systems do not reflect the values of society, they are generating values.

When AI generated content is indistinguishable from human communication, it will become easier for bad actors to manipulate sentiment in society without being detected. In this case, transparency, algorithmic accountability, and digital literacy is of paramount importance. Promoting educational initiatives to enhance media literacy and building regulatory frameworks that require the accountability of algorithms must be viewed as significant resources.(Floridi & Cowls, 2023).

3. Ethical Challenges in AI-Driven Discourse

3.1 Misinformation and Disinformation

As far as the amplification of misinformation and disinformation is concerned, the role of Artificial Intelligence can’t be over exaggerated. AI has the power in terms of generation, replication, and dissemination of content that is, in reality, false but seems to be a compelling truth. Traditional disinformation has its origin in human errors or at most, it may be based on rumours. Unlike traditional misinformation, AI-backed disinformation is more pernicious. AI-backed disinformation may appear more realistic and convincing.

Role of AI in Fake News Generation

Large Language Models (LLMs) and Generative Adversarial Networks (GANs) are at the forefront of AI-backed misinformation. Such is the power of these tools when it comes to the near-real fabrication of data, that it makes it almost impossible to the most discerning human eyes and minds to differentiate between the real and fabricated content. LLM like ChatGPT has the potential to mimic the tone and structure of any standard and reputable journal. Deepfakes created by GANs have the capability to portray people saying things which has no connection with that person.

Bots, Trolls, and Viral Propagation

The content creation and dissemination mechanisms are the core elements of communication. The influence of AI is not limited to content creation. It has spread its wings through social media bots and troll farms. It has the power to amplify false narratives. The result is the fabrication of an illusion of consensus.

Malevolent actors can use swarms of bots to flood timelines with polarizing memes or doctored videos, as has been witnessed during election seasons. On sites like social sites, AI-driven recommendation engines also contribute by promoting user biases and frequently giving preference to sensational or divisive content that boosts engagement, regardless of its veracity.

Case Studies

An alarming example of AI-assisted misinformation occurred in 2016 during the U.S. presidential elections when Russian troll farms used AI-assisted bots to amplify misinformation across social media, affecting millions of users(Watts, C). Many of these posts were created through content farms that were artificially constructed to reduce the factual material and increase division in our democracy. Likewise, during the COVID-19 pandemic, shocking AI-generated clips on social media and articles were helping to spread harmful hoaxes like fake cures, anti-vaccine chants, and conspiracies about where the virus originally came from(Kumar, A., & Sharma, R). Each of these generated a lot of confusion for the public and threatened public health.

In India, deepfake technology has been used during election campaigns to show politicians speaking in different languages or saying things they never actually said, manipulating public perception(Sundaram, R.).

Consequences for Public Discourse & Truth

The proliferation of AI-generated misinformation erodes the very foundation of informed public discourse. When falsehoods are indistinguishable from facts, and trust in traditional information sources declines, societies face an epistemological crisis, one in which citizens can no longer agree on a shared reality. This damages civic engagement, weakens democratic institutions, and polarizes communities.

Moreover, constant exposure to AI-generated hoaxes can lead to "information fatigue," where individuals either disengage entirely or adopt a cynical view that all information is potentially false. This "truth decay" undermines journalistic accountability and diminishes the role of facts in shaping policy and opinion.

Accountability Dilemmas

Dilemmas and ethical challenges, as far as assigning responsibility for AI-generated misinformation is concerned, have yet to be resolved. Whose responsibility is it? Should the whole ecosystem, right from developers to platforms responsible for the distribution of the content to the users, be a part of this ethical challenge? Tracking the origin of the malicious information is next to impossible due to the decentralized nature of the internet.

3.2 Bias and Transparency

Artificial intelligence as a technology have fallbacks to introduce and reinforce biases at various levels along the stages of development and lifecycle ,varies from data collection and classification to algorithm design ,development and deployment in real work environments; often acting as "Black Boxes" (Wachter, Mittelstadt, & Floridi, 2018; Mitchell et al., 2019).

Understanding the core essence of biases is essential to establish fairness and control the development of Safer AI. Biases in data during collection occurs by humans during training on data which subsequently discriminates against certain groups,which leads to biased patterns (Wachter, Mittelstadt, & Floridi, 2018; Mitchell et al., 2019).

At the same time, transparency challenges, based on proprietary model architectures and opaque optimization procedures, hinder stakeholders' capacity to comprehend, question, and challenge AI‑based decisions (Wachter, Mittelstadt, & Floridi, 2018; Mitchell et al., 2019).

Data Bias

Training data are affected by many societal and historical biases from the people who provided the training data and training sets.For example , Buolamwini and Gebru (2018) showed that the leading commercial facial-recognition algorithms misclassified the faces of darker-skinned women with error rates up to 34.7% , while other lighter-skinned men had an much less error rate of 1%. In the case of health-related algorithms which learned from a clinical data set which is biased toward dominant populations, underrepresented populations were often not represented in majority of the clinical source data at all, resulting in a limited clinical effectiveness and wanting to widen health inequities. (Obermeyer, Powers, Vogeli, & Mullainathan, 2019)

Design & Deployment Bias

Bias can also be created by algorithmic design and deployment decisions. Amazon's now-abandoned AI hiring tool routinely downgraded resumes that included words found on female-identified resumes, mirroring patterns in its heavily male training data (Dastin, 2018). In criminal justice, the COMPAS recidivism risk assessment tool was found to misclassify Black defendants as "high risk" at almost twice the rate as that of white defendants, highlighting how deployment with inadequate auditing can entrench racial inequalities (Angwin, Larson, Mattu, & Kirchner, 2016).

Transparency Issues

A multitude of AI systems, particularly deep neural networks, are "Black boxes" that obfuscate the reasoning behind their predictions. Even though the GDPR includes a "right to explanation" of automated decisions (European Parliament, 2016), implementation results in little more than a token effort for providing insight, and individuals are generally powerless to contest decisions beyond ones that impact employment, credit or in a legal case (Wachter et al., 2018). The lack of a standardized way to report explainability and fairness, such as "nutrition labels" for AI, similarly hampers benchmarking of explainability and fairness across and between industries (Mitchell et al., 2019).

Implications for Fairness & Public Trust

Studies show that these perceptions of fairness and transparency have a clear relationship with user acceptance, and when confronted by opaque or biased systems, users are likely to be more resistant, which can stifle innovation and, in more critical contexts, impede the fair application of AI in sectors such as finance, healthcare, and criminal justice (Lee, 2018).

4. Societal Impact

Societal impact due to AI is slow and profound. Slow in realisation for the masses that their thoughts are being shaped by a set of algorithms with which they interact on a daily basis and profound for the changes in thoughts and tolerance that we are witnessing.. The diversity of content that is available to people is reducing significantly, which is resulting in decreasing exposure to different perspectives, creating a feedback loop and a set of beliefs based on confirmation bias due to only seeing content that agrees with their preexisting beliefs, forming “echo chambers”.

Such polarisation can potentially weaken the democracy due to extreme intolerant views on politics, religion and climate change just to name a few, if left unchecked.
Generative AI is also becoming an easily accessible tool for spreading misinformation through online platforms, especially social media by creating very convincing images, videos and audios

to supplement a claim. Such false information can have a potential to cause public unrest, riots, and frenzy if discretion is not practiced. The fake news regarding an explosion at Pentagon caused a brief dip in the market, AI generated images of wrestlers smiling during wrestler's protest also caused damage to credibility of protest, numerous deep fakes circulated during 2024 elections leave people confused and may sway decisions and outcomes.

Beyond these things that we encounter in the media, education and academic integrity is also being impacted more negatively than positively. On one hand the access to AI tutors are reducing the learning gap through easier doubt solving but on the other, a Forbes Magazine survey shows that 65% teachers said that cheating is their top worry on AI integration for education, which can fuel lack of creativity and novel thoughts due to dependence on such a system.

Currently the labour market is also seeing a lot of emotions regarding the future and present of what jobs will sustain their positions and would be hard to replace due to usage of AI for certain tasks and in turn, what new jobs will be created due to this integration. The demand for the high skill workers is expected to increase. A report by the Future of Jobs finds that the rise of technology coupled with other changes that are reshaping global labour markets will result that the number of jobs created will be 14% of today's employment whereas the replacement will be witnessed for 92 million roles, effectively giving a rise of 78 million new jobs.

5.1 Policy and Technology Solution:

Robust legal frameworks and emerging technologies are the pillars of effective regulation in the era of AI-mediated media. The formulation and implementation of comprehensive legislative regulations, including the EU AI Act and information technology legislation in different sectors, are a central strategy. The strategy will ensure that artificial intelligence systems are governed by stringent rules of clarity and responsibility (Council of Europe, 2024).

Using blockchain-like verification mechanisms is a suitable way of improving authentication of various content. Blockchain can provide a method of verification that can help regarding the origin of the digital content and its making in a locker of sorts (unhackable ledger). Detection software has to be designed through joint efforts of multiple stakeholders, such as academia researchers, tech companies, and government bodies, to ensure its efficacy and timely upgrading following changing tactics being used by hostile users (Ilukwe, 2024).

Using blockchain to authenticate information is a good method to improve content authentication. Blockchain is able to confirm information and sort out genuine and fake content by tracking where digital content comes from and how it was generated on an immutable record. When combined with legal requirements on digital content, it is able to effectively avoid the nefarious use of AI-generated misleading information. Regulations must be revamped to ensure digital media platforms are brought up to speed. Platforms such as Facebook, Twitter/X, and YouTube must be forced to demonstrate how they select and display content. Regulations must also force these platforms to undertake research and employ AI tools that identify and flag misleading content in real-time.

Establishing an independent auditing and monitoring system would put AI systems under continued observation. The systems should be audited on a daily basis to make sure that they follow ethical principles. Institutionalizing such audit procedures is how lawmakers can ensure people's trust and make sure AI is utilized for democratic ends and not to inflict harm on them (Stilgoe, 2024).

5.2 Education and Best Practices:

Education and community led best practices can provide the means of diminishing the harm created by AI-generated online content. As deepfakes and algorithmic enforced curation become more common, and emboldened by inseparable digital technology, media literacy should hold an important place in our list of strengthened interventions. The Ash Center's Report on AI and Democracy Movements (2025) indicated that perhaps one of the strongest defenses against misleading and manipulative AI futures is to protect democracy and democratic flexibility by ensuring the public is informed.

The best practices for ethical AI design start, as should everything, with transparency and accountability frameworks. Developers and institutions should adhere to principles of explainability, fairness, and inclusivity when designing and deploying models, by auditing their AI systems on a regular basis to identify and reduce bias, and ensuring that datasets represent diverse perspectives and do not enhance systemic inequality. Ethical design should include clear disclosures when the content is AI-generated this guideline is expressed in the thorough review of Floridi and Cowls (2023) .

In addition, participatory governance can be transformative, rather than just relying on individual action. According to Helbing and Böhme (2024), there are democratic engagement mechanisms, like deliberative forums and co-creation of policies using citizens' assemblies, that can facilitate tension between advances in technology and society's values. As educational curriculum additionally need to integrate digital ethics, AI literacy, and platform accountability into structured educational moments, institutions and civil society need to work together to create publicly accessible learning resources and promote critical engagement with digital technologies (Napoli, 2025). When combined with effective legal and technical safeguards, such an educational approach ensures the public are protected from the harms of AI while also being able to meaningfully shape the future of AI (Floridi & Cowls, 2023).

Conclusion

We now live in an age where a protest can be defused with a smile pasted by an algorithm, and a war can begin with an image that never happened. AI is no longer a tool; it is a participant in democracy—uninvited, unelected, yet profoundly powerful.But technology is not destiny. The danger is not in the intelligence of machines, but in the complacency of the people who build and use them. The systems we've created reflect the biases we refuse to confront, the inequalities we pretend are accidental, and the apathy we mistake for neutrality. If left to grow unchecked, these systems will not just mirror our world—they will magnify its fractures.

And yet, there is hope. Not in blind optimism, but in the clarity that comes from recognition. Recognition that this is a moment of reckoning. That regulation without empathy is hollow. That education without critical thinking is performative. That innovation without ethics is not progress—it is peril.

If we are to reclaim the integrity of public discourse, we must demand transparency where there is none, create accountability where power has gone faceless, and insist that the right to truth is not optional, but essential. The future of AI will not be decided by machines—it will be decided by what we dare to imagine, and what we choose to stand for.

So the question remains—not what AI will become, but what we will become in response to it.

Bibliography

(1)

Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024). Ethical challenges and solutions of generative AI: An interdisciplinary perspective. *Informatics, 11*(3), 58. https://doi.org/10.3390/informatics11030058

Jungherr, A. (2023). Artificial intelligence and democracy: A conceptual framework. *Social Media + Society, 9*(3). https://doi.org/10.1177/20563051231186353

Kharvi, P. L. (2024). Understanding the impact of AI-generated deepfakes on public opinion, political discourse, and personal security in social media. *IEEE Security & Privacy, 22*(4), 45–53. publication

Germani, F., Spitale, G., & Biller-Andorno, N. (2024). The dual nature of AI in information dissemination: Ethical considerations. *JMIR AI, 3*, e53505. https://doi.org/10.2196/53505

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., ... & de Oliveira, N. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns, 4*(10), 100857. https://doi.org/10.1016/j.patter.2023.100857

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency (pp. 77,91). https://doi.org/10.1145/3287560.3287599

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

European Parliament. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union, L119.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, accountability, and transparency. Big Data & Society, 5(1), 1,15. https://doi.org/10.1177/2053951718756684

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 220,229). https://doi.org/10.1145/3287560.3287592

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447,453. https://doi.org/10.1126/science.aax2342

Wachter, S., Mittelstadt, B., & Floridi, L. (2018). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76,99. https://doi.org/10.1093/idpl/ipy016

(5.2)

Ash Center. (2025). *How AI can support democracy movements*. Harvard Kennedy School. https://ash.harvard.edu/wp-content/uploads/2025/02/How-AI-Can-Support-Democracy-Movements-Final.pdf

Floridi, L., & Cowls, J. (2023). *Worldwide AI ethics: A review of 200 guidelines and recommendations*. *AI and Ethics*, *5*(4), 321,345. https://www.sciencedirect.com/science/article/pii/S2666389923002416

Helbing, D., & Böhme, R. (2024). *AI has a democracy problem. Citizens' assemblies can help*. *Science*, *379*(6639), 12,14. https://www.science.org/doi/10.1126/science.adr6713

Napoli, P. M. (2025). *Artificial intelligence and democracy: Pathway to progress or decline? International Journal of Media & Cultural Politics*, *21*(1), 33,51. https://www.tandfonline.com/doi/full/10.1080/19331681.2025.2473994

(2)

Floridi, L., & Cowls, J. (2023). *Worldwide AI ethics: A review of 200 guidelines and recommendations*. *AI and Ethics*, *5*(4), 321–345. https://www.sciencedirect.com/science/article/pii/S2666389923002416

Napoli, P. M. (2025). *Artificial intelligence and democracy: Pathway to progress or decline? International Journal of Media & Cultural Politics*, *21*(1), 33–51. https://www.tandfonline.com/doi/full/10.1080/19331681.2025.2473994

Sage Journals. (2020). *Deepfakes and disinformation: Exploring the impact of synthetic media on society*. *Social Media + Society*, *6*(1). https://journals.sagepub.com/doi/full/10.1177/2056305120903408

(4)

Are online recommendation algorithms polarising users' views?

India's general election is being impacted by deepfakes.

Fake viral images of an explosion at the Pentagon were probably created by AI

India wrestlers protest: photo of “smiling” protesters is digitally altered.

Beyond Plagiarism: The Untapped Potential Of AI In Closing The Achievement Gap

Future of Jobs Report 2025: The jobs of the future – and the skills you need to get them

Sundaram, R. (2024, May 10). Indian voters inundated with deepfakes during the general election. Blackbird.AI. https://blackbird.ai/blog/india-election-deepfakes/Blackbird.AI

Watts, C. (2017, April 3). How Russian Twitter bots pumped out fake news during the 2016 election. NPR. https://www.npr.org/sections/alltechconsidered/2017/04/03/522503844/how-russian-twitter-bots-pumped-out-fake-news-during-the-2016-election

Kumar, A., & Sharma, R. (2024). Emotions unveiled: Detecting COVID-19 fake news on social media. Humanities and Social Sciences Communications, 11(1). https://www.nature.com/articles/s41599-024-03083-5Nature

(5.1)

Council of Europe. (2024). *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*. Retrieved from https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence

Ilukwe, A. (2024). *To Help Rebuild Public Trust in Government, Harness AI*. Centre for International Governance Innovation. Retrieved from https://www.cigionline.org/articles/to-help-rebuild-public-trust-in-government-harness-ai/

Stilgoe, J. (2024). AI has a democracy problem. Citizens' assemblies can help. *Science*, 383(6660), 1234–1235. https://doi.org/10.1126/science.adr6713