2024 Elections AI Policy Proposal: Preserving democracy by increasing the costs of misconduct

Executive Summary

The emergence and advancement of Artificial Intelligence (AI) is a relatively new phenomenon and presents both opportunities and challenges for contemporary societies. With AI becoming more accessible to the population in the last 3 years, concerns have surged with this technology being used for the creation and massive spread of fake information through social media. In the case of politics, and more specifically, electoral processes like the 2024 US Presidential election, AI-driven violations can lead to the creation of inflammatory campaigns that weaken democratic processes. To respond to this situation, a combination of **existing technologies to verify the** nature of the evidence, a **team** of **experts dedicated** to **track** and **evaluate** trending threads in social media, as well as a strategy that **exploits** the **incentives** of the **main political forces** in the country by **making them put** "**skin in the game**" could result in an effective solution.

Problem statement/description of AI violation

In this text's view, the main concern for AI is its use for disinformation purposes. In this sense, AI violations will be defined as the malicious exploitation of AI tools to manipulate or fabricate information with the goal of generating misinformation (Bailey, 2023). These violations encompass the manipulating or the outright creation of false (deepfake) sound files, images and videos, with the goal of tricking its viewers into thinking it is true. In the case of an election, research has found that this false information's main purpose is to misconstrue viewers' understanding of reality in favor of a political force (Coppins, 2020), as well as to manufacture sentiments of anger and inconformity among different demographic groups (Burton, 2023).

These actions have detrimental consequences to democracy when used in politics, especially if they are being employed during a presidential election of one of the most important economies worldwide. First, since AI violations are very effective at deceiving a wide variety of voters, as a staunch amount of political sophistication is required to discern what evidence can be considered a fact and what is simply misinformation, historical studies (Campbell et al, 1960) have shown that only a relatively small portion of citizens have it and it is very costly for many to acquire it. This can lead to countless voters being deceived to vote based on false premises, which undermines the principle that a majority is voting based on their views. A second problem is that this malicious use of AI manufactures content that displays a lurid public confrontation among demographic groups (Umoja Noble, 2018), distorting public discourse to principles that aren't constructive to public debate: tribalism, racial politics and personal traits. In this sense, if AI is causing voters to elect their representatives based on negative feelings and principles of intolerance, it is hard to make a case for representative democracy if it is leading to decisions and measures that can be tainted with the passion of a mob rather than sincere policy preferences (Spinoza, [1677] 2002)

Finally, the main issue with combating the negative effects of AI violations is that it is extremely easy for actors targeting the 2024 election to employ the aforementioned nefarious tactics, as AI technologies have been made highly accessible to practically anyone with access to internet. In the same sense, spreading this false information is also streamlined, as disinformation campaigns can be easily generated by manipulating or outright generating deepfake sound files, images and videos anonymously by any user (i.e. even creating fake accounts), allowing them to post them in practically any digital platform, and permitting them to become viral. Even if an authority were in charge of tracking where the AI violation came from, our current digital environment allows almost

complete anonymity for its authors and supporters. In other words, the monetary and time (opportunity) costs of weaponizing AI against the 2024 democratic and spreading it are very low, while those of policing and catching those responsible are very high.

Desired Resolution and Importance.

At their core, these issues caused by the misuse of AI undermine the trust in democratic procedures: if the election of a government is placed on individuals who are taking decisions based on false premises and hostile sentiments, is it even worth having compared to other forms of government (dictatorships)? In this sense, the literature on democracy posits that this form of government is desirable because it allows people to involve themselves in collective decision making and because this system is geared towards increasing the welfare of the largest amount of members in a society (Sen, 1999). AI violations would be jeopardizing the main benefits of popular representation.
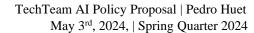
Based on the previous information, the main problems with AI violations are that they can lead to insincere and unconstructive electoral results, weaken the trust in democracy among the population and bystanders. Additionally, the risk of misuse of this technology make this possibility almost inevitable and make it very costly to detect culprits. Therefore, a proposal to deal with AI misuse during the 2024 should have 2 main aims: 1) to preserve democracy's main value of producing electoral outcomes that represent the population's sincere policy preferences and 2) a system that is able to mitigate the damage caused by the virtually untraceable and widespread misuse of AI.

Additional to these 2 principal objectives however, there should be a third priority: establish a wise and sustainable system. When saying wise, the definition employed is that of Fisher and Ury (1981): a situation in which the involved parties accept to employ a procedure –in this case, system for dealing with AI violations– because of its metrics being reasonable. This legitimizes the process and increases its likelihood of being supported by stakeholders overtime. With regards to the sustainability focus, the objective is that the system has feasibility and extensibility in mind: that it can be used in the foreseeable future and expanded with relative ease (Luther, 2021).

Policy Description.

The proposal is the creation of an autonomous state-sponsored organism: Electoral Committee of AI Regulation (ECAR). It would charge of monitoring social media trends and fact-checking campaign information with regards to the use of AI. It should be conformed with specialists on the design, use and consequences of the use of Artificial Intelligence, as well as by members of the two most important political forces: Democrats and Republicans. ECAR would also make official statements about the veracity (in terms of AI use) of trending media pieces, shaming false reports.

The design of ECAR is conducive to its main objective: the specialists on the use and training of AI would be able to use the latest technologies, as deep learning algorithms like Convolutional Neural Networks and Recurrent Neural Networks (Government Accountability Office, 2024), to check the veracity of political media, the specialists in the use of Artificial Intelligence can give insights on which platforms and spaces to monitor for these AI violations, offer holistic arguments to explain why a piece of media is false and speculate of its intended use. Representatives of both political forces would read the findings from the rest of the team, ask all their and their larger political group's questions on the findings of the report and certify these findings for the organism's official public statement.

Implementation Strategy.

The success of this proposal relies mostly on its capacity to tap into the incentives of the involved stakeholders to abide by the judgement of the aforementioned autonomous organization. The two most important actors are the members of the two main political parties in the US, as they are the main actors who have incentives to abuse AI for electoral gains (produce or outsource it) and blame their counterpart for doing the same when a piece of media (fake or not) is unfavorable to them. Since stopping the fabrication of this false AI media is all but impossible, the strategy against it is to mitigate its impact of influencing the election by getting the parties' representatives to abide by the research-based judgement of ECAR.

The implementation, therefore, requires the ECAR to use a framing strategy (Lakoff, 2004) of its responsibilities that aligns with both parties narratives: constantly present itself and its work as a product of the unfair portrayal and campaigns each political party has experienced from fake news in past elections, and vowing to work to safeguard the parties' reputation from future campaigns. By admitting to both political parties' constant claims of being affected by fake news, inviting its members into the organization and publicly making statements with their approval, the organization is appealing to the parties narratives and claims and making the act of rejecting these a cognitive dissonance. In this manner, research has shown (Cialdini, 1984) that people are highly susceptible to incoherent rhetoric, so rejecting the ECAR's narrative, offer of adding its representatives or its public statements would become highly costly to political parties. In the end, the two main parties would have more to lose by defecting from their commitment than abusing AI.