Journal of Intellectual Property and Information Technology Law (JIPIT)
ISSN: 2788-6727 – Volume 3, Issue 1 (2023) p.201-256
https:doi.org/10.52907/jipit.v3i1.262

Exploring Co-Regulation as a Solution to Automated Disinformation in Kenya

*Ikran Ali Abdirahman**

ABSTRACT

This paper discusses automated disinformation on social media in Kenya and its impact on democracy. Automated disinformation refers to disinformation that is exacerbated by the use of Artificial Intelligence (AI) and related emerging technologies, including algorithms and bots. The paper considers the electoral process in Kenya as a case study to highlight the threats that automated disinformation poses to the democratic process and proposes co-regulation as the way forward. Specifically, it reviews the impacts of automated disinformation on democracy including the negative effect on the availability of reliable and accurate information to enlighten the social media users' political choices and the effect on the exercise of political will, public opinion, and democracy. The objective of this research is to provide policy recommendations to the relevant stakeholders on tackling the challenge of widespread automated disinformation perpetuated by social media users in Kenya whilst respecting fundamental human rights and promoting democracy. The author discusses the regulatory framework applicable to the information disorder phenomenon including those relevant to the exercise of the freedom of expression and access to information, noting that these rights play a significant role in strengthening democracy. This paper also considers the nascent regulation of AI and undertakes an analysis of how effective regulation can counter the widespread automated disinformation on social media platforms.

Keywords: Automated Disinformation, Algorithms, Bots, Generative AI, Social Media, Electoral Process, Co-Regulation

* The author is an advocate of the High Court of Kenya and holds a Master's degree in Development Studies from the Geneva Graduate Institute (Switzerland) and an LLB from Strathmore University (Kenya). Currently an AI & the Rule of Law Consultant at UNESCO. Email: alikran00@gmail.com

TABLE OF CONTENTS

I. INTRODUCTION

> 'All political systems need truth to some extent. Democracies need it in a special form—namely, easily available and widely dispersed. And they need it for a special reason: democracies cannot function without public trust, which depends on the public belief that officials are competent to ascertain relevant truth and committed to presenting it candidly…' (Galston, 2012).

The importance of truth in a democracy cannot be understated. With the rapid rate of digitalization globally and the proliferation of social media, there has been a rise in disinformation (Alvares & Dahlgren, 2016). In Kenya, the situation is comparable, with rampant disinformation being exacerbated through automation processes such as algorithms, on social media platforms, as specifically observed in recent electoral processes in the country (Lilian, 2023). Widespread disinformation delegitimizes the truth and casts doubts on what is and is not a 'trusted source' of information even when they are reliable sources. This negatively affects the democratic process in the country, as the veracity of certain information is not commonly understood and debated by all citizens (Information Society *et al*., 2017, p. 3).

Automated disinformation refers to disinformation that is exacerbated by the use of Artificial Intelligence (AI) and related emerging technologies, including algorithms and bots (Epthinktank, 2021). In the tense and polarized atmosphere of Kenyan politics, hateful messaging and destabilizing anecdotes cloud voters' ability to differentiate between truth and falsehood (KICTANet, 2022, p. 4). As such, this paper relies on the electoral processes in Kenya to show the impact of disinformation on democracy and the heightened need to regulate it. Indeed, such automated disinformation is antithetical to democratic values in the country, and negatively affects the function of truth in political discourse and the general public's trust in political processes as this paper further shows.

Nyabola (2018) defines democracy as a 'society in which all eligible members are able to meaningfully participate in public

discourses regarding issues and situations that pertain to the society as a whole' (p. 32). The widespread migration to internet-based platforms has thereby been recognized as giving more Kenyans a voice to express political opinions within the public sphere. This has been attributed to the manifestation of democracy in the digital age, as democracy was previously observed through offline means, such as political discourse and participation through traditional media. With social media, there has been a shift to the online political realm, referred to as 'digital democracy' or 'e-democracy' which is characterized by the use of digital platforms for political discourse and promoting engaging and inclusive democratic processes (Smith, 2018, p. 19; Nyabola, 2018, p. 35). Scholars have put forward that there has been increased digitization of politics through the use of technology thereby advancing democracy (Sanya, 2013, p. 12).

The connection between media usage and democracy lies in their impact on information in general and communication specifically, serving as a foundation for political behavior (Okoth *et al*., 2009). Among other factors, democracy is affected by the ability of citizens to access information on public administration decisions (Epthinktank, 2021). Studies have shown positive impacts of political participation as a citizenry that participates actively in politics is an indicator of a healthy democracy (Brady *et al*., 1995, p. 271). Positive outcomes of social media include broader access to information about various events in Kenya and worldwide, enabling individuals to freely express their opinions. Studies indicate that social media often fosters greater acceptance and understanding among people from different communities with these platforms allowing the broadening and diversification of citizen participation in political discourse (Pew Research, 2022; Rumbul, 2016).

Thus, the relationship between digital media and democracy should be considered. There are mixed conclusions on this relationship. Even though the impact of social media on democracy cannot wholly be judged as positive or negative, some scholars

state that social media is a risk to democracy whilst others aver that social media has the positive impact of increasing political participation and political knowledge (Lorenz-Spreen *et al.,* 2023, p.74).

Digital spaces, including social media platforms such as Meta, X (formerly known as Twitter), WhatsApp, YouTube, and TikTok, have had a positive impact in giving Kenyans a voice. They have embraced technology to amplify their voices and be heard by the political leaders and elites, this was not previously possible with the use of traditional media (Nyabola, 2018, p. 35). Social media plays a big role in facilitating Kenyans' online discourse and has also changed the way that the citizens communicate with their government and the outside world, breaking down several institutional barriers (Okoth *et al.*, 2009). This is demonstrated through the steady increase in civic engagement and accountability pursuits against the government, and the use of social media platforms in electoral processes both at the national and county levels (Omanga, 2019).

Political parties now increasingly deploy social media for political campaigns. There is a consensus amongst scholars that social media has become a conduit for the dissemination of political information in electoral processes, including election campaigns (Kipkoech, 2022, p.3). As more people, specifically the youth, access the internet, and therefore, social media, there is a higher likelihood to consume news on digital media and increase online political participation and civic engagement. As argued by academics, social media in Kenya has been used as a medium for political mobilization through political campaigning as well as a tool for political expression (Nzina, 2014, p. 13). Social media has also increased political engagement through its use as a platform to undertake political conversations and organize and execute protests (Nyabola, 2018, p. 5).

Despite the healthy effect of social media on civic engagement as well as the dissemination of news during the electoral processes in Kenya, there are also various risks associated

with this medium that can undermine democracy (Epthinktank, 2021). These include challenges such as the spread of hate speech and incendiary rhetoric, misinformation or disinformation, online violence focusing on women in politics, deep fakes, and microtargeting (Kofi Annan Foundation, 2021). Additionally, such risks can facilitate the use of social media to cause political manipulation of the citizenry.

In Kenya, disinformation on social media has been prevalent in the recent electoral processes, undermining electoral integrity (Article 19, 2022). A study by KICTANet shows that emerging technologies, handled through bots (automated social media accounts) were tasked with spreading misleading electoral information during the 2022 elections in Kenya. Thus, the impact on the accurate information available online was worsened and amplified by the bots and platform algorithms (KICTANet, 2022). Low levels of digital literacy and the marginalization of some communities over decades of deprivation due to economic and social inequalities make them more susceptible to manipulation through disinformation (HRC, 2021).

This paper focuses on the challenges posed by automated disinformation on social media in Kenya due to its impact on the veracity of facts in political discourse and democracy, which informs this paper's recommendation for the need to co-regulate.

Part I of this paper is the introduction. Part II explores automated disinformation, using the electoral process in Kenya as a case study. It addresses challenges faced in running the process effectively, spanning from the pre-election period through to the post-election phase. Part III provides a mapping of the regulatory framework and practices in Kenya relevant to the freedom of expression as well as artificial intelligence technologies in relation to automated disinformation. Part IV gives policy recommendations for a multistakeholder and co-regulatory approach to tackling the challenges posed by automated disinformation in Kenya. Part V concludes the paper.

II. AUTOMATED DISINFORMATION AND ELECTORAL PROCESSES: A KENYAN CASE STUDY

A. The concept of automated disinformation

Disinformation refers to 'false information that is knowingly shared to cause harm' (Wardle *et al*., 2017). Automated processes, like the use of algorithms—sequences of instructions—on social media applications, play a role in disseminating disinformation by personalizing content for users based on specific factors Epthinktank, 2021). The use of algorithms on social media is widespread and is influenced by advancements in technology, including AI (Epthinktank, 2021).

Even before the adoption of algorithms, the switch to the use of social media platforms to disseminate news and other socio-economic and political information had impacted the ability of the population to make accurately informed decisions (Nyabola, 2018, p. 39). Online, the information accessible to the electorate shapes political behavior. However, the narratives presented on social media are curated by gatekeepers such as journalists, political campaigners, parties, and influencers. This gives them a disproportionate role in influencing political outcomes (Nyabola, 2018, p. 39). Hence, these gatekeepers with political influence construct stories or information in their favor using these controls. They then distribute this content to the public, thereby manipulating the response of the governed audience (Nyabola, 2018, p. 39).

In addition to the algorithms discussed, bots are also used to influence disinformation (Epthinktank, 2021). Bots are composed of a set of algorithms that are designed for a specific variety of tasks which can also include gathering or disseminating content, with large numbers of bots being used on social media platforms such as X (Howard et al., 2018, p. 8). They are used for differing functions, for example, 'social bots' which are 'user accounts equipped with the software to automate interaction with other users' or 'political bots' which are composed of program-

ming of automated interaction with other users on political topics (Howard et al., 2018, p. 8). Political bots, which are relevant to this paper, have been defined as 'automated scripts designed to influence public opinion' (Howard et al., 2018, p. 8). They are used to not only increase the follower counts of politicians' social media accounts through fake accounts but to also create and promote content for politicians and political parties, which can include propaganda and negative campaigning (Howard et al., 2018, p. 8).

The recent development of ChatGPT has also worsened disinformation on social media platforms. ChatGPT is a newly developed AI model that works in a dialogue format to answer questions or prompts presented to it (OpenAI, 2023). With the use of ChatGPT, there has been an increased generation of highly believable fake content, at a faster rate and with much cheaper costs. A recent article published by the New York Times showed that ChatGPT makes the creation of fake news much more sophisticated and powerful by using storytelling methods to provide convincing variations of disinformation in a matter of seconds, and without disclosure of any sources (Tiffany & Stuart, 2023). Researchers have also demonstrated that large generative language models such as ChatGPT have greatly contributed to the creation of enormous amounts of automated content (Goldstein et al., 2023). However, the use of developing AI content-generating technologies such as AI can be maliciously used by actors who want to spread propaganda and fake news, which has the impact of disseminating persuasive automated disinformation that will implicitly affect and influence public opinion (Goldstein et al., 2023).

In addition to traditional media, there has been an increase in the use of social media in electoral processes including in political mobilization and campaigning (Nzina, 2014, p.13). Due to the varying positive and negative effects of the use of social media during the electoral process for access to information and undertaking political discourse, it is important to investigate how

the harmful aspects of social media in relation to the electoral process can be mitigated in Kenya. For instance, automated disinformation on social media threatens democracy as it promotes a lack of transparency and accountability by actors in the electoral process in Kenya.

Article 19 has documented examples of false information being spread across the country aimed at voter suppression. For example, there were false reports of wild animals being let loose or deployment of the military personnel in some areas of Kenya on the 2022 election day. There were other claims that the electoral commission had miscounted some votes or that some candidates had already won the general election seats before the completion of voting (Article 19, 2022).

In the 2022 Kenyan elections, the Independent Electoral and Boundaries Commission (IEBC) delayed the counting of votes, and even then, only displayed results on the presidential election, omitting the timely provision of results of 16,094 candidates vying for other political positions. Due to such delays, there was widespread disinformation marked by misleading posts that attempted to manipulate the public into believing that some candidates had won even before the IEBC completed the tallying of the votes (KICTANet, 2022). There was also a proliferation of fake polls, manipulated fake videos, and misleading news sites which made it even more difficult for the public to determine the veracity of the election outcomes or the information on social media (KICTANet, 2022). Emerging technologies were also used through bots, and automated social media accounts, tasked with spreading misleading electoral information whose impact on the information available online was worsened and amplified by the platform algorithms (KICTANet, 2022).

Furthermore, the 'attention capture' model used by social media platforms can lead to the reduction of focus on political processes and political disengagements. This is due to the promotion of content that is personalized to the user which can lock

citizens in echo chambers and influence their ability to logically assess their own opinions as well as consider other opposing opinions on a subject matter (Epthinktank, 2021).

The dangers of these instances of misinformation are enormous The Human Rights Council (HRC) has recognized that disinformation presents a threat to democracy as it 'can suppress political engagement, engender or deepen distrust towards democratic institutions and processes, and hinder the realization of informed participation in political and public affairs' (HRC, 2022). The subsection below further discusses this reality.

B. Social media and the electoral process

Kenya has a robust digital infrastructure, being the regional information and communication technology (ICT) hub in East Africa, leading in general ICT infrastructure, broadband connectivity, mobile money, and banking, among other digital services (ITA, 2019)., Kenya also has a wide usage of online social media services (SMELab, 2018). This widespread use of social media platforms leads to the vulnerability of Kenyans to information disorder.

Research shows that the prevalence of information disorder, particularly within electoral contexts and political spheres reduces the citizenry's trust in the electoral process thereby undermining democratic institutions and processes (KICTANet et al., 2022). Not only does information disorder lead to the weakening of these democratic agents, but it also leads to government authorities using the spread of information disorder to justify clamping down and censoring the citizenry, critics, and other parties from the use of digital media (KICTANet, 2017). Additionally, information disorder leads to mistrust in mainstream media as well as the internet (KICTANet, 2017).

In electoral processes, disinformation may be perpetrated by foreigners, citizens operating social media accounts, and political parties and their proxies (Nyabola, 2018; Ndlela, 2022).

Such disinformation is aimed at having an impact on electoral outcomes as it confuses the voters and distorts their political views thereby influencing them to vote in one way or another (Epthinktank, 2021).

In Kenya, the 2022 election process was rife with disinformation, as the different political officials took to social media platforms to cast suspicion on each other (Lilian, 2023). Social media users participated in disseminating disinformation through sharing fake polls, deep fake videos, and other forms of fake news (Mozilla, 2022). Indeed, the campaign period leading up to the elections showed record amounts of disinformation on social media apps such as TikTok and X with the algorithms on these platforms amplifying such disinformation, which led to higher levels of 'trending' or higher views (Mozilla, 2022).

Particularly, there is a lack of transparency from social media platforms, such as X or TikTok on the algorithmic qualities that perpetuate the trending of certain topics, including instances of disinformation, over others (Mozilla, 2022). Transparency, in this context, means that the functioning of the algorithmic system should be comprehensible to the individual users of the platforms, ensuring they are capable of recognizing and understanding that the interaction is with an algorithm (Digital Society Initiative, 2021).

A review of the content on TikTok in the run-up to the 2022 elections highlighted the weak content moderation by the platform that led to the dissemination of hateful videos and incendiary propaganda fueling threats of ethnic violence and tribal tensions (Mozilla, 2022). These results were particularly concerning as TikTok was the most downloaded app in Kenya in the run-up to the 2022 elections (AppFigures, n.d.). The report by Mozilla examined a sample of political content on TikTok and highlighted the far reach of the content with 130 videos on 33 TikTok accounts having a combined number of 4 million views (AppFigures, n.d.). Additionally, hashtags on TikTok such as '#siasa' and '#siasazakenya' translating to 'politics' and 'Kenyan

politics' respectively, have garnered more than 20 million views (Mozilla, 2022).

Mozilla's research indicates that the inciting and hateful content circulated on TikTok before and during the 2022 elections closely resembles the content shared during the 2017 election period, which was marked by scandals involving Cambridge Analytica and Harris Media (Mozilla, 2022). It was demonstrated that Cambridge Analytica and Harris media ran inflammatory electoral campaigns online which were rife with disinformation, including violent content, targeted at the political opposition in Kenya (Privacy International, 2018). This spread of disinformation included falsified videos claiming that the opposition would 'remove whole tribes' and was disseminated across social media platforms including Facebook, X, and TikTok (Privacy International, 2018). Such inflammatory content is particularly risky in Kenya due to its past experiences with post-election violence thus indicating the significance of putting in place functional policies and regulatory mechanisms to address disinformation on social media.

It is, therefore, necessary to investigate whether the technology companies offering social media services have put in place necessary safeguards in their policies and platform design to reduce the amplification of hateful and inciting content, particularly in electoral processes (Mozilla, 2022). According to the Mozilla report, the provocative political content observed on TikTok during the elections violated the platform's policies. Despite these violations, the content was widely shared. The report attributes this to inadequate content moderation by TikTok, which failed to consider the specific political and cultural context of Kenya when reviewing the shared content (Mozilla, 2022). Indeed, some of the videos did not expressly fall into TikTok's definition of incitement and hateful ideology and were not easily picked up due to context bias (unfamiliarity of languages and/or context), thus making it difficult for content moderators to flag such content.

Disinformation is not only a Kenyan problem, as witnessed in the rife instances of misinformation and disinformation in the electoral processes of countries across the globe. For instance, the US election in both 2016 and 2020 demonstrated the viral spread of propaganda and disinformation and the resultant negative effects on the democratic process (Barthel et al., 2016). Indeed, in a study carried out by Ipsos in 25 countries, it was determined that 'fake news' is a global problem, which is amplified by the internet, thereby resulting in negative impacts on political discourse by undermining public trust and their access to reliable information, before, during and after electoral processes (Ipsos, 2019).

Disinformation can have serious repercussions for the democratic process, affecting transparency, fueling hate speech and crimes, prompting internet shutdowns by authoritarian governments, and restricting freedom of expression (APC, 2021). Additionally, it can disrupt peaceful conditions within a country. A report produced by KICTANet Kenya shows that there is a relationship between disinformation and hate speech. This relationship manifests due to the electoral process in Kenya taking on ethnic dimensions which causes the various vying parties in the political divide to utilize information disorder, fueled by online anonymity, to spread hate speech and incendiary hateful propaganda to a voting public that is divided into ethnic and tribal lines (KICTANet, 2017).

Despite these risks associated with social media, there are still shortcomings related to content moderation and fact-checking on these platforms. For instance, research shows that the content moderation ecosystem in Kenya often lacks an understanding of the cultural context to effectively moderate, and resources are lacking in the appointment of a sufficient number of content moderators on platforms such as TikTok (Mozilla, 2022). In contrast, traditional media platforms had greater control over which content to broadcast as the determination was made in a centralized manner. However, social media platforms may find content moderation and fact-checking more difficult due to the

increase in the amount of disinformation content and the widespread nature of information being ingested into social media by its users.

C. Proliferation of automated disinformation in Kenya

A study conducted by the Social Media Lab Africa in Kenya showed that bots have been used to influence the electoral process in Kenya (SMELab, 2018). The study demonstrated that social bots were extensively employed in African elections, with Kenya leading in the use of bots, making up twenty-seven-point six percent of the country's social media users engaging with X. This surpasses the influence of campaigners, government, traditional media, bloggers, and even politicians themselves (SMELab, 2018). These bots spread fake news and were tasked with controlling political discourse in the country during the 2017 elections, which undermined the truthfulness of public debates, negatively affecting the electoral democratic process (SMELab, 2018).

Poll results showed that eighty-seven percent of the two thousand Kenyans who participated in a survey conducted by Geopoll and Portland (2018), were exposed to disinformation before the 2017 election which affected their decision to vote. Data on the extent of disinformation for the 2022 election is not yet available but researchers and journalists have stated that disinformation was also rife during this process (Lilian, 2023). Such disinformation campaigns mostly form a part of a wider political strategy that includes identification of the target audience, identification of the group of people tasked with the responsibility of spreading this misleading information, and creation of the audio-visual materials to be released on social media (Lilian, 2023).

Studies done on social media usage in Kenya have indicated that a growing number of influencers in Kenya have been hired by politicians to carry out astroturfing and spread disinformation, through the operation of several bots to smear reputations (Odanga, 2021). Astroturfing is a strategy that has the objective

of causing the 'bandwagon effect', forming the impression that there is a public consensus on a certain opinion or matter, where there may be little to none (Schmitt- Beck, 2008, p. 308; Broadband Commission, 2020). Thus, many Kenyans receive this disinformation and take it as truth due to the limited digital literacy, and shortcomings in verification due to limitations in the available content moderation and fact-checking avenues (Odanga, 2021).

Through disinformation-spreading tactics, such as astroturfing, the use of social media has moved from being voluntary to being involuntary due to the effects of algorithms and bots usage, affecting human agency and political interaction online (Ndlela, 2022). For instance, with the development of AI, the use of bots and algorithms is far from neutral and increasingly can sway beliefs and promote cognitive biases through the amplification of certain information or content accessible to social media users, similar to the gatekeepers in traditional media (Ndlela, 2022).

Through programming, algorithms can determine the information that is displayed to the user, for example, recommendation algorithms may be based on collaborative filtering which provides suggestions depending on similar interactions between the platform's users (Zanon et al., 2022). Specifically, Facebook has the 'EdgeRank' algorithm that displays posts depending on the user's previous likes in content, or type and frequency of interactions with online friends, among other undisclosed reasons or actors (Ndlela, 2022). X also has a similar algorithm that ranks content on the X homepage of the user depending on what the algorithms determine as most interesting to the user (Ndlela, 2022).

Such impacts of the proliferation of automated disinformation are particularly relevant based on social media users in Kenya, as a large amount of online political discourse majorly taking place equally on Facebook (88.5%) and WhatsApp (88.6%), with YouTube (51.2%), Google (41.3%), Instagram (51.2%) and

X (27.9%) (SMELab, 2018). For instance, studies have shown that WhatsApp has played a large role in enabling Kenyans to organize around political action and discourse relating to the government affairs of Nakuru County (Omanga, 2019, p. 175). The study carried out by Omanga D. (2019) shows the move of political discourse online through the mobilization around grassroots politics in Nakuru, where it was previously offline (p. 175). Social media platforms operating in Kenya have demonstrated an absence of effective content filtration and fact-checking, especially in the electoral process (Rutenberg et al., 2020, p. 302).

AI technologies, such as deep fake technologies, can give credibility to automated disinformation in political contexts as they allow the realistic manipulation of audio-visual materials such as videos through the creation of deep fakes or misleading and false images (Kreps et al., 2019). There is also the use of widespread automated social media 'bots' to perpetuate a certain political opinion or movement and thereby create the perception of credibility on information that is otherwise misleading or incorrect. Disinformation perpetuated by bots often mask the source of the information and are intended to deliberately mislead the individuals consuming this content on social media, through astroturfing which creates a sham image of public consensus (Kreps et al., 2019).

The use of algorithms to promote disinformation also has the capability of convincing social media users through the repetitive posting of some content, that false information is indeed true. For instance, YouTube has a recommendation algorithm that amplifies similar content, thus leading to the grouping of content that is false, which creates several sources for one misleading idea, convincing the content consumer of the veracity of this false narrative (Amnesty International, 2022, p. 10). Therefore, the use of such algorithms has the inevitable effect of diminishing the ability of the content consumers to differentiate what is true and false and the cognitive autonomy to make independent decisions.

With the use of algorithms and bots, the ability of the population to make political decisions is therefore affected by the algorithmic gatekeepers who control the political narratives present on social media. This loss of control over what one consumes impacts the voluntariness of one's political choices and outcomes. Algorithms and bots manipulate social media users by affecting public opinion and influencing electoral votes (Ndlela, 2022). As demonstrated in the Cambridge Analytica scandal in Kenya, politicians have previously hired consultancy companies to influence and manipulate social media content to promote their digital campaigning to influence voting patterns (Ndlela, 2022). The algorithms utilized in the programming of political bots are also designed to misinform the wider public through misrepresentation of political leaders' followings and make the politicians and their ideas appear more popular (Ndlela, 2022).

The basis of a democracy is the voluntariness in civic engagement, as the citizen makes an intentional decision to participate, for example, through voting in electoral processes (Howard, 2019, p. 318). However, the effect of technologies structurally linked to social media such as the mass surveillance infrastructure and the larger Internet of Things affects the voluntariness in civic engagement as the voters become exposed to information or content without their informed consent (Howard, 2019, p. 318). For instance, algorithms are also used in microtargeting, which is an online strategy that collects personal data from online users, and is used to segment these individuals into different demographic groupings that can be targeted by companies or political parties with different content and information (Privacy International, 2023). This therefore implies the existence of a new power paradigm that the individuals who control the networks will also control large amounts of data thus managing the data to create algorithms that can be used in political processes (Howard, 2019, p. 318).

Additionally, through the spread of propaganda and false information online, bots have resulted in the alteration of po-

litical reality by catalyzing artificial discourse (SMELab, 2018). These effects create misinformation as they lead to the distortion of views and perspectives by reinforcing existing biases and propagating human confirmation bias even in the absence of reliable information (Epthinktank, 2021). The effects of the use of AI and algorithms lead to widespread and worsened automated disinformation signaling a new era of accountability of social media platforms and their usage of emerging technologies in their platform design and functions.

Phenomena such as the use of social media platforms for microtargeting and the resulting echo chambers are exacerbated by AI and cause a manipulation of political behavior which can lead to skewing political information and distorted outcomes of electoral processes (Amnesty International, 2019). For instance, in the 2017 elections in Kenya, Cambridge Analytica, a communication consulting firm, mined the personal data of Facebook users for political campaigning in an attempt to influence the electoral vote through micro targeting (Privacy International, 2018). Such targeting was particularly worsened by AI, through the use of algorithmic systems to manipulate the electorate through micro-segmentation. This refers to a technique that compiles information on the structural and emotional profiles of the citizens in order to gain insights into what their behavior is on social media, thus allowing them to be targeted (López- López et al., 2023).

In 2021, the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression prepared a report on 'Disinformation and freedom of opinion and expression' in which she examined threats posed by disinformation to democracy and human rights (United Nations Human Rights Council (HRC, 2021). Digital literacy is a key factor in how social media platforms are utilized and the need for digital literacy has been recognized in the Special Rapporteur's report (HRC, 2021). Literacy is key to the personal regulation of activity on social media and should be a key priority in education systems and curricula (Collins et al., 2019).

A key priority in ensuring digital literacy through including it in education has been demonstrated in Finland, where at a very early stage, in 2014, the country already included a college 'anti-fake news' course focused on the identification of manipulated information on the internet, including deepfakes and the use of bots (CNN, 2019). This allowed students, journalists, and politicians to possess the digital know-how to identify and counter automated disinformation that had malicious objectives. Finland has been ranked as the country with the highest media literacy in Europe (CNN, 2019).

A study by the National Cohesion and Integration Commission (NCIC) showed that the threats to peaceful elections in Kenya in the run-up to the 2022 elections included a lack of trust in the institutions and amongst the communities, self-interested leadership, ethnic polarizations, just to mention a few, which was worsened by disinformation (NCIC, 2020). Such disinformation now runs the risk of worsening through the use of automated techniques such as the utilization of algorithmic systems, bots, and large generative AI technology such as ChatGPT, which collectively create highly believable false content.

According to Safaricom, the largest telecommunication company in Kenya, in the year 2017, fifty percent of its communications department time was spent monitoring fraud and fake information at different times – with the instigators of this disinformation being influencers, politicians, political parties, and the people they work with (Safaricom, 2017). Similar statistics are not available from Safaricom for the subsequent years but since then, there have been advancements in technology, leading to the proliferation of the use of algorithms, bots, and artificial intelligence in digital campaigns.

There is therefore an ever-increasing need for regulation of the use of social media to address automated disinformation. According to the Special Rapporteur on freedom of opinion and expression, both States and companies need to make more of an effort to confront disinformation (HRC, 2021). At a continental

level, Smart Africa is an AI initiative for a blueprint for Africa to strengthen the technical understanding of AI, removing barriers to entry into the market for AI and developing regulatory frameworks for AI (Diplo, 2022). In addition to regional frameworks, there is a need for common areas of regulation of AI globally, through multilateral co-operation to ensure aligned and effective regulation.

Such automated online tactics eventually influence the will of the people through erosion of their ability to make decisions informed by varying available information and not locked within echo chambers, which in turn affects and delegitimizes what is considered 'true' on social media (Sugow, 2019, p. 22). This confusion and lack of ascertainment of the veracity of certain information affects decision-making and the will of the people to participate in democratic processes (Sugow, 2019, p. 22).

Whilst AI may have the impact of worsening the disinformation crisis on social media platforms, it can be used to tackle disinformation through the application of automated processes in moderation of large amounts of online content (ITIF, 2019). Such automation in content moderation can also contribute to positive means of addressing the challenge of widespread disinformation on social media, particularly in political and electoral contexts (ITIF, 2019). AI can be geared towards combating disinformation in electoral contexts, for example, the EU, noting the proliferation of fake news in its electoral processes, used AI to detect and respond to fake news, and to empower social media users to determine the truthfulness of information found online (ITIF, 2019).

Specifically, forms of disinformation such as deep fake pictures, audio, or videos, which are highly believable fake content developed using only samples of a few seconds, are difficult to determine without the intervention of automated systems, in which case AI would be very valuable (ITIF, 2019). For example, the European Commission (EC) has recently encouraged signatories to the Code of Practice on Online Disinformation, with technology companies adopting it, for example, Microsoft and TikTok

signed it in May 2019 and June 2020 respectively. The Code recommended signatories who have services with the potential to disseminate AI-generated information should put in place technology to recognize such content and clearly label this content as such, to the end-users (Lomas, 2023). This recommendation under the Code has resulted from the EC's refusal to accept machines or AI as having any freedom of speech (Lomas, 2023).

However, an important consideration was that the use of AI in the moderation of such content should only be done where the algorithms used in these processes are overseen by a human (Lomas, 2023), as human control of the use of AI is imperative to ensure contextualization of information to reduce algorithmic bias or censorship (Lomas, 2023).

AI has similarly been used to tackle the scourge of automated disinformation, for example, in Zambia, where the use of an AI fact-checking tool known as iVerify was piloted during the country's general election in 2021 (UNDP, n.d.). The tool reviews public content submitted by individuals using an open-source algorithm known as 'Detoxify' which is trained to detect hate speech, misinformation, and disinformation (UNDP, n.d.).

By training AI models, through machine learning, AI is used to detect false information and inaccurate data as well as the detection of bots (Bontridder et al., 2021). Facebook is an example of a company that utilized AI to detect ninety-nine-point-six percent of the fake accounts on its platform in 2020 before users even reported these accounts as fake (Facebook, n.d.). However, as mentioned previously, AI usage in tackling the disinformation challenge can perpetuate bias as it is prone to false negatives and false positives, and it may also result in over-censorship. As such, any use of AI in reviewing disinformation online must be accompanied by human content moderators (Bontridder et al., 2021, p. 32).

The next section reviews the regulatory framework as well as the regulatory steps being taken by these stakeholders and

analyzes how the present challenges can be tackled more effectively.

III. ADDRESSING AUTOMATED DISINFORMATION

This section reviews the existing regulatory framework in Kenya on social media regulation covering various aspects of information disorder and the freedom of expression, specifically in the electoral process. Further, it reviews the nascent regulatory and institutional framework on AI in the country. The section also considers the relevant governmental institutional bodies tasked with governing matters relating to social media and the freedom of expression in the electoral process in Kenya. It examines the extent to which they are effectively regulating and fulfilling their obligations in the electoral processes.

A. Regulating the freedom of expression

Under Article 33, the Constitution of Kenya 2010 guarantees the freedom of expression, which includes the right of every person to exercise artistic creativity, to seek, receive, and impart information and ideas, and the academic freedom and freedom of scientific research. This guarantee for the freedom of expression is not absolute but is limited to hate speech, advocating for ethnic hatred, propaganda for war, discrimination, or incitement to violence (The Constitution of Kenya 2010, a33(2)). Additionally, this article of the Constitution provides that the freedom of expression can be exercised by respecting the rights and reputation of others (The Constitution of Kenya 2010, a33(3)). The Constitution allows the limitation of certain rights and freedoms guaranteed if their exercise would limit or prejudice the rights and freedoms of others (The Constitution of Kenya 2010, a24 (1)). However, any such limitation of rights by law must be 'reasonable and justifiable in an open and democratic society based on human dignity, equality and freedom' (The Constitution of Kenya 2010, a24 (1)).

Inasmuch as the Constitution guarantees the freedom of expression, it must still conform to the limitations outlined above. The Constitution does not directly provide for disinformation or misinformation as a cause for the limitation of freedom of expression, but it does fall as a limitation that can be sanctioned under Article 24 of the Constitution which provides for the limitation of rights especially where the exercise of these rights and freedoms would prejudice the rights and freedoms of others.

For instance, the National Cohesion and Integration Act is an Act of Parliament enacted to promote national cohesion and integration with an aim to outlaw discrimination due to ethnic reasons. It therefore limits the freedom of expression to promote national cohesion through the criminalization of hate speech which is defined as 'speech where ethnic hatred is the desired or likely consequence' (National Cohesion and Integration Act 2008, s13). In addition to the freedom of expression and other human rights guaranteed under the Constitution, it also recognizes and promotes democracy as a national value and principle of governance (The Constitution of Kenya 2010, a10(2)(a)).

Other statutes in Kenya make specific limitations to the freedom of expression on the grounds provided under Articles 33 and 24 of the Constitution, with respect to the issue of disinformation. In the aftermath of the 2017 election in Kenya, which was characterized by widespread disinformation, the Computer Misuse and Cybercrimes Act (2019) (Cybercrimes Act), was enacted to address the publication and dissemination of false and misleading information. The Cybercrimes Act (2019) criminalizes the publication of 'false, misleading or fictitious data or misinforms with the intent that the data shall be considered or acted upon as authentic' (s22). The Act goes on to further provide that publication of 'any information that is false in print, broadcast, data or over a computer system, that is calculated to result in panic, chaos or violence' or where such false information has the effect of harming or injuring the reputation of another, amounts to an offense (The Cybercrimes Act 2019, s23).

Since its presentation as the Cybercrimes Bill and before its enactment, the Cybercrimes Act has been challenged in court for limiting the freedom of expression. Several civil society organizations questioned the constitutionality of the Cybercrimes Bill before its enactment as they posited that certain sections were detrimental to the digital rights of Kenyans, specifically, the freedom of expression, opinion, privacy, and access to information (CIPESA, 2019). There were further arguments by Article 19 about the limiting nature of the Cybercrimes Act, (Article 19, 2018) including the broad nature of the offense of the publication of false information.

The High Court initially suspended the coming into force of certain sections of the Cybercrimes Act whilst awaiting the petition on the determination of the constitutionality of these sections (*BAKE v AG & others, 2018*). This petition to the Court highlighted the vague nature of the criminalization of the publication of false information and the use of sweeping undefined terms such as 'false, misleading or fictitious' which would be capable of unnecessarily limiting the freedom of expression. Despite these objections, the High Court held that the sections in question in the petition were constitutional and justifiably limited in line with Articles 24 and 33 of the Constitution and were necessary for the protection of others' rights and valid public interest needs. The Court particularly stated that the petitioner in this case had failed to demonstrate how these provisions under the Cybercrimes Act were excessive (*BAKE v AG & others, 2018*).

In any case, such criminalization under the Cybercrimes Act has not been successful in stopping the spread of disinformation in the country (KICTANet et al., 2022). The global international human rights framework does not encourage the criminalization of such instances of disinformation as this is dangerous and may result in negative impacts on other rights and freedoms of the electorate (APC, 2021). Indeed, the Disinformation report prepared by KICTANet and CIPESA highlights the Government's responses to disinformation including the 'weaponization of dis-

information law to silence critical voices' and states the need by technology companies to put in place remedial measures on platforms in the cases of disinformation (KICTANet et al., 2022).

However, this criminalization of disinformation is not effective as it has been utilized to censor citizens, including activists, journalists, and members of civil society, who have been the subject of counter-disinformation measures. This demonstrates a need to update the laws governing freedom of expression in Kenya, indicating the criminalization of disinformation as a last resort, and putting in place clearer language to ensure proper implementation of the legislation. As pointed out by Amnesty International, vague laws have the effect of silencing independent voices and critics (Amnesty International, 2022).

Previously, the Kenya Information and Communications Act (the Information Act 1998), was utilized in the criminalization of false information as Section 29 criminalized the improper use of a telecommunication system, specifically by sending messages or matter that is 'grossly inoffensive, indecent, obscene or menacing, or a message known to be false for the purpose of causing annoyance, inconvenience or needless anxiety to another person' (The Kenya Information and Communications Act 1998, s29). However, this section was declared unconstitutional as the judge stated that the criminalizing terms such as 'grossly offensive', were largely ambiguous and would lead to uncertainty in law (*Andare v AG and others, 2016*). This was reasoned to be a result of the varied understandings of what these terms mean and placing a high responsibility on the judicial officer making the determination (*Andare v AG and others, 2016*).

B. Regulations on Artificial Intelligence in Kenya

As part of the global fourth industrial revolution (4IR) and its strategy for the development of the digital economy, Kenya has adopted the use of AI in several sectors of the economy including health, education, agriculture, and fintech (Industrial Analytics

Platform, 2021). With Kenya being known as 'Africa's Silicon Savannah' due to its strong history in the uptake of digitalization and prioritizing the digital agenda, the country has not lagged in embracing innovations. The significance of digitalization in Kenya's development is acknowledged through its Big Four Agenda which places importance on leveraging emerging technology in its key sectors including food security, health, manufacturing, and housing (Government of Kenya).

Nonetheless, the legislative frameworks on disinformation reviewed in the previous sub-section, including the Cybercrimes Act and the Information Act make no mention of regulation of artificial intelligence despite its contribution to the disinformation phenomenon. This is largely because technology is fast changing and is evolving at a faster rate than regulation is being created. Furthermore, the law focuses on the activities of disinformation but does not extend this regulation to the tools used in spreading such information, including AI and other emerging technologies. As such, the relevant laws are yet to be reviewed to keep up with such developments evident in the use of emerging technologies in spreading and amplifying disinformation.

The Kenyan Ministry of Information, Communication and Technology examined how technology such as the Internet of Things (IoT) and AI, specifically blockchain and AI technologies can be utilized to support the fulfillment of the development goals (ICT Ministry, 2019). Through this endeavor and due to the adoption of AI technologies, the Kenyan government created the Blockchain and Artificial Intelligence Taskforce (the AI Taskforce) to guide the effective usage of AI in the country (Gazette Notice 2095, 2018). The AI Taskforce was tasked with the development of a national strategy for the use of AI in Kenya's path to the 4IR and for its regulation (Gazette Notice 2095, 2018).

The AI Task Force found two key conclusions: that the government of Kenya invest more into the creation of an enabling ecosystem to allow the use of blockchain and AI technology to thrive; and that there was a need to develop regulations to gov-

ern AI by ensuring protection of the public interest whilst encouraging innovation from the private sector (Gazette Notice 2095, 2018). It recognized the benefits posed by AI technology but also highlighted the risks it presents ranging from privacy infringement, unethical use of AI, and weaponization of the technology in the physical, digital, and political spheres. Politically, it acknowledged that AI could be exploited to skew public opinion (ICT Ministry, 2019).

The report referred to Brundage et al. (2018) in highlighting the political risks of AI being used to persuade the public through the dissemination of targeted propaganda, for them to act or vote in a certain manner in order to achieve the desired political outcomes. The manipulation of the public through AI would not only occur through targeting but also through deception of the public by the creation of fake but highly believable videos or audios, known as deep fakes (Brundage et al., 2018).

However, even with the resultant risks and negative outcomes from the malicious usage of AI, there is still a lack of a regulatory framework and no specific legislation governing the responsible use and adoption of AI in Kenya. There are also no regulatory mechanisms in place for the ethical implications of AI including data bias and data transparency in the use of AI systems (CIPIT, 2023, p. 6). This state reflects the global position on AI as the technology is rapidly evolving making it difficult for law and policy to effectively regulate it. Thus, the issue that presents itself is that the legal and policy frameworks are being currently developed, both globally and nationally, which has resulted in the malicious use of AI within a regulatory vacuum. This absence of a regulatory framework has therefore not facilitated the proliferation of the automated disinformation phenomenon in Kenya, specifically with regard to the use of AI such as algorithms and bots in worsening disinformation during electoral contexts.

Accordingly, there is a need to develop an enabling AI regulatory ecosystem that will not only regulate the space and govern

its use but also encourage innovation as well as the positive and responsible utilization of AI in social, economic, and political contexts. This should be done in a manner that considers responsible AI principles that define policies and put in place accountability mechanisms guiding AI regulation. Allerin (2019), has posited that governments must regulate AI as a lack of regulation will lead to the potential of having grave implications and irreparable harm to human civilization. In addition, regulating AI will assist in ensuring that AI safety and ethics are integrated into the AI creation process while fostering responsible AI creation (Allerin, 2019). As previously discussed, the regulation of AI should take care to ensure that it does not stifle further innovation, research, and development of this emerging technology.

1. Exploring intermediary liability

In Kenya, Section 56 of the Cybercrimes Act makes provisions that refer to intermediary liability (The Cybercrimes Act 2019, s56). Due to the difficulty of detecting some cybercrimes, the Kenyan government places some responsibility on the intermediaries requiring them to 'net the direct offenders'. In this context, as per MacKinnon's (2015) definition, intermediaries are entities that '(i) give access to, host, transmit and index content, products and services originated by third parties on the internet, or (ii) provide internet-based services to third parties'.

According to MacKinnon (2015), intermediary liability falls within three categories: strict liability, conditional liability, and broad immunity. Strict intermediary liability imposes direct liability and incriminates the intermediary only by the fact that the intermediary provided access to the illegal content; whilst conditional liability allows intermediaries to be exempt from the liability if they have abided by the conditions under the law. Broad immunity is an extension of conditional liability and allows intermediaries to develop a content policy, allowing them to be protected from liability if they disrupt the activities under the content policy (MacKinnon et al., 2015).

Section 56 (1) of the Cybercrimes Act makes a special provision for intermediary liability by providing that the intermediary '…shall not be subjected to any civil or criminal liability, unless it is established that the service provider had actual notice, actual knowledge, or willful and malicious intent…' (The Cybercrimes Act 2019, s56). The section, however, does not clarify how to determine actual notice or knowledge as a factor of liability and does not indicate what an intermediary can do to excuse itself from liability (Walubengo et al., 2018, p. 21). Among the offenses the Cybercrimes Act imposes intermediary liability may be the offense of false publication and publication of false information, even though the Act does not describe what activities could constitute 'publishing' and therefore puts intermediaries at risk.

Section 56 (2) of the Cybercrimes Act further provides that the intermediary shall not be liable for 'maintaining and making available the provision of their service'. The Kenyan position necessitates the indication that the requirements under Section 56 (1) of the Cybercrimes Act are met in order for intermediaries to be held responsible. Thus, in Kenya, intermediaries cannot be held liable only for the provision of services (The Cybercrimes Act 2019, s56). It is important to note, however, that Section 56 of the Cybercrimes Act and its reference to any intermediary liability has not been applied to cover the use of AI in the publication of false information and there is no jurisprudence yet publicly available on implementation of this section.

Section 56 (2) of the Cybercrimes Act can also be applied to social media companies on the functioning of algorithms and bots on their platforms, for instance, if the algorithm perpetuates the trending of certain topics over others without any transparency, it can be deemed a violation of this provision (The Cybercrimes Act 2019, s56). Additionally, it can be argued that the requirement for content labeling where generative AI, including large language models (LLMs) which are technologies using algorithms trained on large volumes of text-based data that power ChatGPT should be governed under this principle of intermedi-

ary liability (Wright, 2023). This is because such technologies are used on social media platforms to create believable disinformation content, which leads to the faster and more far-reaching spread of disinformation.

At a global governance level, the international standards of AI regulation are also at the nascent developmental stages. In 2021, the UNESCO Recommendations on Ethics on Artificial Intelligence, the first normative framework on AI was globally accepted by 193 UNESCO member states, including Kenya (UNESCO, 2021). The recommendations are intended to guide member states in the development of legislation, policies, and other regulatory instruments to govern AI, in alignment with several ethical principles including proportionality and do no harm, fairness and non-discrimination, privacy and data protection, transparency, responsibility, and accountability, among others (UNESCO, 2021).

Further international dialogue on AI can be observed in the United Nations (UN) Global Digital Compact (Policy Brief 5; the Digital Compact), which was developed to set out principles, objectives, and actions to be taken by multi-stakeholders to advance an open, inclusive, and human-centered digital future that enables the achievement of the sustainable development goals (SDG) and is in line with the international human rights framework (UN, 2023). It promotes the SDGs focusing on overcoming the digital, data and innovation divides in order to advance a sustainable digital future for all. In considering the digital divide, the Digital Compact notes that the innovation divide is stark with a large amount of wealth generated from AI, being unequal and concentrated in a few big technology companies and States such as China and the USA (UNCTAD, 2021). The Digital Compact recognized the intensification of misinformation and disinformation through the use of AI in the creation of false but believable content at a large scale, fast rate, and low cost, necessitating the need for a multi-stakeholder approach to developing regulatory standards (UNCTAD, 2021). It makes reference to

the UNGA Resolution 76/227 on ensuring that disinformation is countered in a manner that still protects and promotes human rights and fundamental freedoms (UNGA, 2022).

The AI regulatory framework in Africa has been slow considering the rapid advancement and uptake of AI technologies on the continent (CIPIT, 2023). The use of AI in the continent is mainly in the sectors of manufacturing, health, transport, business, and government services (CIPIT, 2023). Several African countries have begun the development of AI national strategies including Mauritius, Egypt, and Rwanda, with Kenya, Ghana, Zambia, and Tunisia having taken some steps towards developing national AI strategies and policies (CIPIT, 2023). The existing national AI strategies have prioritized the regulation of risks associated with privacy and promoted a human-centric approach to AI with a focus on people's well-being. The regulations are centered on principles such as ethics, accountability, inclusion, building public trust, and the development of a robust enabling environment for AI (CIPIT, 2023).

Given the availability of legal systems and standards applicable to technological systems, the current legal systems may be adapted and interpreted to apply to AI developments, pending the development of new technological regulation standards (Digital Society Initiative, 2021). There is also a need to balance any potential regulation and innovation to ensure that the development of algorithm systems continues, responsibly and ethically, aimed at transparency and avoidance of discrimination (Digital Society Initiative, 2021). Furthermore, any development of regulatory standards for AI should be in a 'technology neutral' manner to ensure that it is generally applicable to similar technologies instead of being limited to a specific technology (Digital Society Initiative, 2021).

It is paramount, therefore, that any potential AI regulation developed should address these challenges whilst ensuring that free speech and public discourse are not limited. Scholars have recommended that transparency of the functioning of algorithms

is key to algorithmic filtering and dissemination of content (Digital Society Initiative, 2021). Currently, the prevailing practice on social media platforms is the use of algorithms, in a black box process, with the end-users not being made aware of why or how their behaviors may result in the recommendation of certain content on their pages (Sachin, 2022, p. 10). For instance, collaborative filtering algorithms use deep learning architectures that make recommendations to users but struggle to generate explanations on how such recommendations were made (Zanon et al., 2022).

Furthermore, the manner in which this information is individualized to the users' social media feed should be known to users, such as making available the criteria that inform what content is displayed or suppressed (Digital Society Initiative, 2021). This would involve the ability of a platform user to ask that any content displayed on the user's feed should not be based, either implicitly or explicitly, on personal identifiers, such as one's political affiliation (Cen et al., 2021).

The next section discusses the opportunities for regulation of automated disinformation in elections, by the government and technology companies, intending to promote transparent, free, and fair elections. It considers the opportunities for co-regulation, and the need for a multi-stakeholder and multidimensional response to the problem of automated disinformation before, during, and post-election processes.

IV. MULTISTAKEHOLDER AND CO-REGULATION APPROACH

This section reviews the strides of the Kenyan government and its regulatory bodies in addressing information disorder on social media during electoral processes. It further discusses the responsibility of social media companies to address such challenges through self-regulation and the shortcomings thereof. Both governments and social media companies globally have demonstrated shortcomings in the way they deal with informa-

tion disorder in digital spaces (Amnesty International, 2022). On the one hand, States have been found culpable of repressive responses to regulation of disinformation, including censorship as well as the criminalization of 'fake news' with these actions considered to be counter-productive (Amnesty International, 2022). This response has been accelerated by the fact there is no adequate human rights oversight over social media platforms (Amnesty International, 2022). On the other hand, social media companies have not adequately considered human rights in the actions taken to counter information disorder on their platforms (Amnesty International, 2022).

Co-regulation involves companies establishing mechanisms, either independently or collaboratively, to oversee their users whilst such mechanisms require approval from democratically legitimate state regulators or legislatures, who also assess their effectiveness (Marsden et al., 2020). This chapter further considers other relevant stakeholders in the regulation of automated disinformation, for example, political leaders and political parties who are instrumental in the proliferation of automated disinformation on social media, particularly during the electoral process.

A. Government regulation

The Kenyan regulatory space is composed of various independent institutions that may regulate the mandate of the use of social media on the one hand, and the regulation of electoral processes on the other hand. Thus, such regulations applicable to social media, and any happenings on these platforms, are several and their operations are fragmented. For instance, the IEBC is tasked with 'conducting or supervising referenda and elections to any elective body or office established by the Constitution' (IEBC, n.d.). In addition to this, the IEBC oversees the registration of voters and candidates, maintaining and keeping updated the voter's roll, the regulations of political parties' processes, developing a code of conduct for parties and candidates, and voter education (IEBC, n.d.).

Within the electoral space, the NCIC has a very significant role of 'promoting national unity, equity and the elimination of all forms of ethnic discrimination by facilitating equality of opportunities, peaceful resolution of conflicts and respect for diversity among Kenyan communities' (NCIC, n.d.). The NCIC is a statutory body established under the National Cohesion and Integration Act (2008). The NCIC is particularly relevant in maintaining unity, inclusivity, and respect for diversity in Kenya given the bloody history of tribalism and the devastating effects of post-election violence in the country.

As previously discussed, disinformation on social media in Kenya during elections is also fueled by notions of tribalism and hate speech which threatens the peace of the country. As a result, NCIC has been further tasked with social media monitoring where it conducts a review of content on social media that serves as an early warning system, preventing hate speech, disinformation, and incitement (NCIC, n.d.). With the findings of such monitoring, the NCIC aims to provide accurate information for stakeholders' activities and aims to increase accurate data-driven decision-making and reduction in electoral violence. Nevertheless, NCIC can only do so much in social media monitoring through the provision of accurate data but cannot control what is posted on these platforms or take such content down, with this responsibility falling on the social media companies who operate the platforms. Social media monitoring has also proved difficult for NCIC as it does not have sufficient resources and expertise to track the large volume of information being shared on online platforms during the electoral process in Kenya (Mzalendo, n.d.).

Additionally, the Media Council of Kenya (MCK), is established by the Media Council Act and is mandated with setting media standards in line with Article 34 of the Constitution and ensuring there is compliance with these standards (MCK, n.d.). The mandate of this institution extends to journalists, media practitioners, and media enterprises but does not extend to the regulation of social media companies. Interestingly, it regulates

media companies' activities on social media platforms as demonstrated in the recently adopted Code of Conduct for Digital Media Practitioners (MCK, 2022). This Code requires digital media players to comply with ethical and behavioral rules laid out in the Code including ensuring all digital content is accurate, fact-checked, and capable of being substantiated. The Code particularly addresses misinformation and disinformation by requiring that digital media practitioners ensure that misleading, falsified content, or digitally manipulated content is not published, promptly alert consumers where such content is present on social media, provide correct information, and ensure all posted content is sourced (Media Council of Kenya Code of Conduct, 2022). This is particularly relevant as, in some instances, digital journalism has been implicated in the spread of mis/ disinformation in Kenya.

The Office of Registrar of Political Parties (ORPP) is a state office which is established through the Political Parties Act (2011), which regulates the 'formation, registration and funding of political parties' (ORPP, n.d.). As it is tasked with the regulation of political parties, researchers have suggested that it should also be involved in keeping politicians and political parties accountable in digital campaigning and other usage of social media platforms. For example, the Kofi Annan Foundation held a workshop in Kenya that proposed recommendations for the development of a digital code of conduct by political parties and the ORPP to regulate the creation or dissemination of falsified content, avoidance of astroturfing and other malicious digital campaign methods (Kofi Annan Foundation, 2021). It would be a good step to require political parties to put in place provisions in their internal political party manifestos or frameworks that disinformation should be prohibited, specifically disinformation spread on social media through automated means, during electoral processes.

As demonstrated above, these are fragmented institutions that operate within the space of social media regulation and elec-

toral process regulation, with all these bodies having different mandates. The patchwork of regulations as well as applicable mandates make it difficult to ensure interoperability between these above-mentioned institutions and leads to differing interpretations on how to deal with some of the challenges caused by automated disinformation. Thus, it is clear that there is a lack of effective regulation of social media platforms and the operation of social media companies, demonstrating the need for an independent regulator dedicated to this function in Kenya, especially noting the rapid advancement of emerging technologies such as AI.

Currently, in Kenya, there is a degree of uncertainty regarding which body must be assigned the responsibility if the country's laws applicable to social media companies are infringed as social media companies, including Meta which operates Facebook, WhatsApp, and Instagram, are not officially registered in the country (Business Daily, 2023). This is due to the structuring loopholes that companies utilize to ensure they are not legally present in a country, for example through the utilization of sub-contractors in employing individuals to run its operations within Kenya. However, as this paper later highlights, a Kenyan court ruled, within the context of an employment case, that legal action can be brought against social media companies in Kenya, even though they are not officially registered as a company in Kenya (Njanja, 2023).

B. Self-regulation by technology companies

Technology companies are primarily driven by profit, and it therefore follows that the content bringing in the most profit, whether composed of disinformation or not, will be prioritized (Collins et al., 2019). This has been demonstrated through the use of 'bot farms' which have provided avenues for profitability for social media companies, by churning out large amounts of similar information to popularize existing content, indicating

that algorithms are designed to prioritize paid-for-content (Nyabola, 2018, p. 34).

The algorithms utilized by social media companies and advertising have been found to undermine democracy through the spread of automated disinformation (Nyabola, 2018, p. 34). This is why it is important for these social media companies to put in place internal policies aligned with the human rights obligations imposed on them. As highlighted by the Special Rapporteur's report on disinformation (HRC, 2022), such social media companies' policies on content control and privacy in line with international human rights obligations do exist although not directly addressing disinformation, but are often fragmented in application, difficult to review and implement.

However, within electoral contexts, instances demonstrate how politicians can easily employ an army of bots to spread their campaigns, content, or propaganda to a wider audience. For example, during the US Congress hearing of Russian interference in the 2016 US election, there was testimony concerning the Internet Research Agency, a bot farm, which spent USD 126 million on advertising on social media (Kang et al., 2017).

In regulating algorithm usage on social media platforms, attention should be paid to the manipulation of information, which is exacerbated by microtargeting, including falsified information to particular groups of people (Digital Society Initiative, 2021). As such, targeting and dissemination of disinformation affects social media users' ability to form autonomous decisions and undermines democratic will and the formation of public opinion. Whilst algorithmic filtering has positive effects in personalizing content to improve the user's online experience, it can also contribute to the spread of disinformation, and worsening of online echo chambers as a result of filter bubbles (Digital Society Initiative, 2021).

The UN Guiding Principles on Business and Human Rights are instructional to social media companies in their operations.

It places a responsibility on all companies to respect human rights, however, and wherever they operate (UN, 2011). This pushes for the companies to ensure that their operations are in line with international human rights standards. The UNGA Resolution 76/227 (2022), also places a responsibility on social media companies to:

> 'review their business models and ensure that their design and development processes, their business operations, data collection and data processing practices are in line with the Guiding Principles on Business and Human Rights, the importance of conducting human rights due diligence of their products, particularly of the role of algorithms and ranking systems in amplifying disinformation, and calls upon them to adopt and make publicly available, after consultation with all relevant stakeholders, clear, transparent, narrowly defined content and advertising policies on countering disinformation that are in line with international human rights law…'.

Self-regulation is advantageous as it allows social media companies to adapt quickly where there is a rapidly evolving online ecosystem. However, this is not sufficient and government regulation is still necessary for independence and effective decisions in the public interest, for example, the independence of auditors can be overseen by the state (Cen et al., 2021). As mentioned in previous sections of this paper, it is nevertheless difficult to develop regulatory systems and auditing processes concerning algorithmic filtering of content on social media. This is because there are concerns about ensuring the freedom of expression and public discourse is maintained, the subjectivity of what is considered appropriate content or not, and ensuring that regulation should not negatively impact innovation (Cen et al., 2021).

Self-regulation is solely not enough either. Amnesty International (2020) notes that self-regulation by technology companies is ineffective and can only be deemed successful when the States put in place and enforce data protection and digital regulations laws. It concludes that tackling disinformation by technology companies has to go beyond content moderation, to a com-

plete overhaul of company practices that rely on surveillance and profiling (Amnesty International, 2022).

The Special Rapporteur's report on disinformation (HRC, 2021), also notes that social media companies may have the content policies to take better steps towards combatting disinformation. However, the range of these policies is fragmented and disjointed making it harder to review, interpret, and implement (HRC, 2021). They are also marked with several vague terms that do not clarify in concise or concrete terms what situations are deemed as 'harm' or what amounts to 'likelihood of harm' that would necessitate labeling, content removal, or other actions (HRC, 2021). Content moderation by technology companies is also applied in a non-transparent manner and inconsistently. Thus, there is a need for these companies to consolidate their policies in a way that is consistent with their other rules as well as institute transparency mechanisms such as human involvement in automated content moderation and automated fact-checking to identify what sort of content would be removed (HRC, 2021).

For instance, with regard to the conduct of social media companies during electoral processes, X established the Civic Integrity Policy. This policy is geared towards social media users and states the fact that X may not be used 'to manipulate or interfere in elections or other civic processes' (Twitter, 2023). It also highlights what amounts to a violation of the policy including misleading information about how to participate in the civic process (including elections) (Twitter, 2023). It further states that misleading information about outcomes aimed at undermining public confidence in the electoral process such as disputed claims can lead to an undermining of faith in the process or misleading claims regarding the results of the election (Twitter, 2023). Additionally, the policy states that users should not create fake accounts misrepresenting their affiliation. However, it is important to note that this policy does not cover the use of algorithmic systems by X itself and does not provide transparent information on how it intends to tackle the spread of disinformation by its al-

gorithmic systems, the use of bots, or the spread of AI-generated misleading content.

Considering the companies' internal policies, Facebook or Meta has in place the Violence and Incitement Guidelines, (Facebook, n.d.) whilst TikTok has implemented terms of service and community guidelines (TikTok, 2023). Meta also has the Oversight Board in place which is an independent body that decides whether content should be on Facebook or Instagram and also accepts complaints from the users of the platforms (Meta, 2021).

1. Regulation of algorithmic-dependent systems

Regulation of the use of algorithms should be underpinned by the principle of transparency to the users on how their sites work, especially on the factors influencing the prioritization of certain news or stories based on each user's profile. This will allow users to be conscious of the potential influencing factors in their decisions countering the loss of autonomy in decision-making and will ensure that users make more informed decisions (Collins et al., 2019).

The United Nations has recently prepared the 'Common Agenda on Information Integrity on Digital Platforms', which recognizes the spread of misinformation and disinformation on the digital system, specifically on social media, which is capable of causing harm at a global level, including death, violence and an 'existential risk to humanity' (UN, 2023). The report acknowledges the exacerbation of disinformation challenges by rapid advancements in technology including the use of generative AI (UN, 2023). The report proposes a 'UN Code of Conduct for Information Integrity on Digital Platforms' which pushes for stakeholders, including governments and companies, to be committed to information integrity and respect for human rights (UN, 2023). It would require increased transparency from digital platforms regarding data, algorithms, and content moderation. This would also include the publication of policies on mis and disinformation, including reporting on coordinated disinforma-

tion on their services and the effectiveness of such policies in countering this challenge (UN, 2023).

The High-Level Expert Group on AI in Europe has outlined these requirements for trustworthy AI systems: 'human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity and non-discrimination; societal and environmental well-being; and accountability' (CERRE, 2023 and Kreiss et al., 2021, p. 522). Concerning the use of AI in electoral contexts, there is a need for explainability of the use of AI, meaning the transparency of the AI processes, including informing the end-users of the technical processes used to form decisions by the systems and how this decision was concluded (CERRE, 2023 and Keriss et al., 2021, p. 522).

In the attempt to regulate challenges presented by technological advancements, a predominant question has been how to define these challenges, such as algorithmic biases, automated disinformation, or even privacy invasion, considering that the traditional legal standards are not easily or fully applicable to such challenges. Urs Gasser (2021) identified three legal response modes by regulators to technological change, as follows (p. 202):

i). Subsumption: where the legal system applies already existing rules to new challenges arising from technological advancements;

ii). Gradual innovation: in which the existing legal system proves to be insufficient to address the novel issues arising, and the legal system resorts to innovation, or upgrading the existing norms to set new precedents to complement current norms; or

iii). Paradigm-shifting approach: this is more radical and not only updates existing norms but pushes for the development of entirely new approaches or instruments.

Policymakers have increasingly pushed for the idea of the paradigm-shifting approach as the most promising option which

will give law a functional role and encourage the role of technology as 'part of the solution space.' As discussed by Gasser (2021) in the case of privacy challenges created by technological law, the paradigm-shifting approach encouraged the use of technological design solutions or mechanisms to respond to privacy challenges (p. 203). For example, the promotion of privacy rights through the development of privacy-enhancing technologies or designs such as 'encryption tools and privacy-preserving analysis techniques' among other approaches. Thus, researchers and policymakers have encouraged the designing of technology and embedding within the underlying architecture, solutions to address the potential challenges (Gasser, 2021, p. 203). The issue of automated disinformation can be addressed similarly, by adopting a paradigm-shifting approach, in addition to subsumption and gradual innovation responses.

The UN's recent report on Information Integrity on Digital Platforms, states the importance of ensuring that protection mechanisms are ingrained within emerging technologies, including generative artificial intelligence, and states that it is 'essential that user privacy, security, transparency and safety by design are integrated into all new technologies and products at the outset' (UN, 2023).

In this regard, this paper recommends regulation of the challenges presented by automated disinformation on social media platforms, through the implementation of platform design controls and data or content controls. Platform design controls such as the re-introduction of 'friction' into the social media functions, which represents any process that would slow down a function on the platform, such as pop-up requests asking users whether they want to publish on their feed, as opposed to automatic posting on the platforms or limiting the number of people messages could be forwarded to (Collins et al., 2019). This has been argued that it will make social media users pause and think about whether they want to consume or generate some content.

According to the Center for Humane Technology, such examples of friction include the ability of a user to only share or repost a post or comment if they write about the post or comment, or only share a post that one has read fully, among other ways (Collins et al., 2019). Social media platforms should ensure that such controls are put in place in the design to ensure the end-user experience is controlled and transparent, thus promoting their autonomy in decision-making.

Data control measures should also be put in place by technology companies including the taking down of content containing false information or even a review of certain groups and users for veracity or harmfulness of content produced. Social media applications are profit-driven and prioritize the bottom line thus the content recommendation algorithms are also designed to focus on maximized engagement of users. Therefore, there may be instances of no financial incentive for companies to implement such controls as misinformation or disinformation could be lucrative due to the amount of engagement such content generates, which can be utilized to bring in more advertising fees (Amnesty International, 2022). Such structural incentives on these platforms are harmful to the users as they can lead to even more spread of disinformation, as long as the content brings in more engagement.

A 2020 testimony by a Facebook employee before the US Senate highlighted this issue of content recommendation algorithms, pointing out the destructive risks of harm to society (Sachin, 2022, p. 2). The harms associated with content recommendation algorithms can be addressed by social media companies through regulation, from the development of the algorithms, ensuring the content shared under the algorithms reduces disinformation and that the effect on the decision-making autonomy of the end-user is controlled (Sachin, 2022, p. 10). As such, it is necessary that social media companies put in place stronger content moderation and fact-checking guardrails and ensure that human rights are upheld for all, including the users of the platforms.

C. Opportunity for co-regulation

The Special Rapporteur's report on disinformation (HRC, 2021) recommended that States should not delegate the responsibility to adjudicate on online content that fosters corporate judgment over human rights principles, which affects the exercise of the freedom of expression online (Amnesty International, 2022). There should be mechanisms in place to allow for not only government regulation but should also be done in conjunction with technology companies' self-regulation. Self-regulation involves the firms establishing rules concerning their procedures and rules applicable to their processes, through the promulgation of voluntary codes of conduct that the firms or group of firms adhere to as well as overseeing sanctions for non-compliance (Rubinstein, 2018). On the other hand, government regulation emanates from the government and imposes rules and oversight over the firm or group of firms (Rubinstein, 2018). Whereas full government regulation has been criticized for limiting dynamism, speed of response, and international cooperation being lost, self-regulation has been critiqued as prioritizing the interests of the firms themselves to avoid government regulation and often involves a lack of transparency and accountability of these companies ((Marsden, 2011); Rubinstein, 2018).

Regulation has generally moved towards the mid-point of the regulatory spectrum, co-regulation, which involves the collaboration between the government and private firms or other subjects of regulation, with a higher level of involvement by the government (Rubinstein, 2018). Co-regulatory rules are less prescriptive than those of state regulation and allow firms to develop more specific guidelines for the implementation of these rules (Rubinstein, 2018). Co-regulatory approaches have proved successful in certain countries and regions, for example in the EU. This approach is heralded because there are better results demonstrated in regulation when the state works together with the companies to address the harms posed by online social media platforms.

A Report by UNESCO stated that government regulation should still ensure that the liability regime does not become too strict as intermediaries would be more likely to censor their users. This is due to the profit-driven businesses that are run by technology companies that will make them prioritize the legal requirements by leaning towards censorship of users rather than the possibility of huge fines (UNESCO, 2015).

These stringent regulations also grant an excessive amount of power to social media companies to interpret local laws and decide what qualifies as disinformation, justifying censorship. However, this process may be inaccurate or unauthorized. The challenge lies in the fact that social media companies, like Facebook, utilize automated systems to remove legitimate and accurate content, resulting in errors and biases that lead to the censorship of users (Article 19, 2022). For example, in Nigeria, during the #EndSARS movement against police brutality, Facebook and Instagram were implicated in silencing posts about the 'Lekki massacre' and flagging them as false news (Okoh, 2020). In Vietnam, Facebook was accused of conceding to Vietnam's government requirements on censoring those accounts considered voices of dissent from using the platform. As the government pushed for further restrictions against criticism, Facebook tightened content controls which raised alarms about censorship by the company (Washington Post, 2023).

Regulation cannot be left purely to technology companies either as the Kenyan regulatory space does not only rely on market-oriented factors to ensure human rights are upheld. This requires government intervention and a co-regulatory approach to ensure there is independent regulation of social media companies' activities in the country, and that the local laws are respected and upheld.

In this context, the use of online platforms can be regulated through co-regulatory means, by aligning content moderation rules to legal rules and standards, under government supervision (Cornils, 2020). Scholars have also advocated for co-regula-

tory approaches where the firms control more information in the processes at hand, proving command-and-control regulation (as under state regulation) to be difficult without their buy-in and collaboration. This has been highlighted by Richard Stewart who explains this using the logic of Coasian bargaining principles:

> 'The premise is that legal rules will advance society's welfare if they are voluntarily agreed to by all relevant interests. If those with a stake in the regulatory requirements – the regulated, the regulator, and perhaps third-party environmental or citizens interests – agree on an alternative to the standard requirements, the agreement may be presumed to be superior to the standard' (Rubinstein, 2018).

In Kenya, looking at the need for intervention of laws and institutions in the proper functioning of online social media platforms, it is therefore clear that co-regulatory approaches will be most appropriate in ensuring the civil and political rights of the electorate in Kenya are respected. Specifically, the Government should consider adopting the gradual innovation and paradigm-shifting approach in legal responses to the challenges posed by emerging technologies. As this paper demonstrates, there is an urgent need for relevant regulatory regimes to be developed in the country to address advancements in technology including social media, AI, and algorithms which despite the positive effects have furthered the perpetuation of automated disinformation that negatively impacts democracy.

However, the limitations of including such regulations and controls on the functioning of social media platforms should be considered, with multi-stakeholder approaches being prioritized. This would enable the smooth adoption by the technology companies to ensure that such regulation will be successful due to the multitude of responsibilities placed on these companies. It is important to consider how to aptly acquire such adoption. For instance, the adoption can be increased through bringing together all stakeholders in developing any such code of conduct and in determining the priorities in regulation.

As a result, the technology companies can not only have

ownership of the application of the rules but also ensure that the rules are realistically enforceable. This is because regulation is sometimes far-reaching and may not be enforceable because of limitations in sharing certain information with the government or publicly, due to protections such as patents and trade secrets, or lack of technological advancements allowing for the technical implementation of certain rules. For example, it might be difficult for technology companies to comply with such regulations imposed on them which require too much information that may force them to reveal unique details on the running of their businesses, due to privacy and competition reasons. However, this has been opposed by policymakers who state that trade secrets should not outweigh the requirement for regulation and public scrutiny to uphold the rule of law and ensure automated processes do not happen in black boxes (Brookings, 2020).

It is therefore imperative to develop any such rules in a manner that is not too detailed but still concise enough and with clear language ensuring effective interpretation, implementation, and enforcement of such rules. It is worth noting, however, that this is difficult to develop and implement in practice, but by bringing together different stakeholders and being guided by the objective of promoting and protecting human rights, it is possible to find a midpoint that is acceptable to all parties involved.

Such a code could be developed through the inclusion of parameters to be disclosed that fulfill the same objective of allowing the regulator to understand whether the algorithms are being used responsibly and ethically. This must also be balanced in a way that does not discourage social media companies from fully operating in Kenya by ensuring any such rules are not too invasive in company operations.

V. CONCLUSION

Using Kenya as a case study, this paper is premised on the positive impacts of social media platforms in amplifying the

voice of citizens. These platforms have enabled an increase in the levels of civic engagement in the country through the dissemination of political information, broadening of citizen participation in online discourse, and political campaigning occurring online. Whilst a lot of benefits have emanated from social media platforms, it has inevitably resulted in the development of new challenges and online harms to the populace and the ideas of democracy in Kenya.

Such harms have been demonstrated in the proliferation of disinformation. In this context, the focus is on automated disinformation referring to instances of disinformation happening on social media which are exacerbated by the use of AI and related emerging technologies such as algorithms and bots. Such automated disinformation is rampant in electoral processes in the country, with the spread of misleading electoral information through astroturfing, microtargeting of the electorate, and the entrapment of voters in personalized echo chambers. Additionally, the paper identifies the use of algorithmic systems in an opaque manner, resulting in skewing public opinions through the use of bot farms or similar systems on what information is widely disseminated online and to whom. It further highlights how such harms harm the 'voluntariness' of democracy as automated disinformation affects the ability to maintain control over one's political choices and public opinion.

Considering such challenges, this paper reviews the regulatory framework in Kenya, relating to the freedom of expression and limitations thereof as well as the laws applicable to the regulation of AI. On the one hand, the regulatory framework on the exercise of one's freedom of expression is fairly established and guaranteed as a right although limited and does not extend to disinformation. On the other hand, the regulatory framework on AI is at the nascent developmental stages at both the local and international levels and is characterized by a global lack of governance mechanisms. As such, the author illustrates what such regulation on AI would look like, as well as the progress

of the regulation discourse and development. Particularly, this analysis entailed considering how AI and related emerging technologies impact the phenomenon of disinformation and the effect on political will, public opinion, and democracy at large.

Finally, this paper recommends co-regulation in order to ensure that stakeholders are involved in the regulation of the challenge of automated disinformation on social media platforms, including its impact on the electoral process in Kenya.

REFERENCES

A Human Rights Approach to Tackle Disinformation. (2022). Amnesty International. https://www.amnesty.org/en/wp-content/uploads/2022/04/IOR4054862022ENGLISH.pdf

AI-powered fact-checking tool iVerify, piloted during Zambia election, shows global promise | United Nations Development Programme. (n.d.). UNDP. Retrieved May 30, 2023, from https://www.undp.org/digital/stories/ai-powered-fact-checking-tool-iverify-piloted-during-zambia-election-shows-global-promise

A Legal Framework for Artificial Intelligence' (Digital Society Initiative (University of Zurich) 2021).

Association for Progressive Communications (APC). (2021). *Disinformation and freedom of expression*.

AppFigures (n.d.). Retrieved May 8, 2023 from https://appfigures.com/top-apps/google-play/kenya/top-overall

Barthel M, Mitchell M, & Holcomb J. (2016). *Many Americans believe fake news is sowing confusion*. Pew Research Centre.

Blockchain and Artificial Intelligence Taskforce: Kenya Gazette Notice Number 2095 (2018).

Bloggers Association of Kenya v The Attorney General and 5 others (2018) eKLR.

Bontridder, N., & Poullet, Y. (2021). The Role of Artificial Intelligence in Disinformation. *Data & Policy*, *3*, 32.

Brady, H. E., Verba, S., & Schlozman, K. L. (1995). Beyond SES: A Resource Model of Political Participation. *American Political Science Review*, *89*(2), 271–294. https://doi.org/10.2307/2082425

Broadband Commission for Sustainable Development. (2020). *Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression*.

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., hÉigeartaigh, S. Ó., Beard, S., Belfield, H., Farquhar, S., … Amodei, D. (2018). *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* (arXiv:1802.07228). arXiv. http://arxiv.org/abs/1802.07228

Cen, S., & Shah, D. (2021). *Regulating algorithmic filtering on social media*. 35th Conference on Neural Information Processing Systems.

CIPESA. (2019). *Promoting best practice among activists for more effective collaboration in digital rights litigation in Kenya: A case study of the Bloggers Association of Kenya (BAKE) versus Hon. Attorney General & Three Others.*

Code of Conduct for Digital Media Practitioners. (2022). Media Council of Kenya

(MCK).

Computer Misuse and Cybercrimes Act (Act No. 5 of 2018).

Constitution of Kenya (2010).

Content Moderation and Local Stakeholders in Kenya. (2022b). Article 19.

Cornils, M. (2020). *Designing platform governance: A normative perspective on needs, strategies, and tools to regulate intermediaries*.

Damian Collins and others. (2019). 'Disinformation and "Fake News": Final Report'.

Digital Economy Report 'UNCTAD/DER/2021'. 2021. UNCTAD.

Disinformation in Kenya's Political Sphere- Actors, Pathways and Effects. (2022). KICTANet and CIPESA.

Eliza, M. (2019). *Finland is winning the war on fake news. Other nations want the blueprint*. CNN. Retrieved June 7, 2023, from https://www.cnn.com/interactive/2019/05/europe/finland-fake-news-intl

Europe Wants Platforms to Label AI-Generated Content to Fight Disinformation. (2023, June 6). TechCrunch. Retrieved June 6, 2023, from https://techcrunch.com/2023/06/06/eu-disinformation-code-generative-ai-labels/

Emerging Digital Technologies for Kenya: Exploration & Analysis. (2019). Ministry of Information, Communication and Technology.

Facebook. (n.d.). *Community Standards Enforcement Report*. https://transparency.facebook.com/community-standards-enforcement#fake-accounts

Fake News, Dangerous Speech and the Elections. Talk to NCIC: KICTANet Community Engagement Report. (2017). KICTANet.

From Dance App to Political Mercenary: How disinformation on Tiktok gaslights political tensions in Kenya. (2022). Mozilla. https://assets.mofoprod.net/network/documents/From_Dance_App_to_Political_Mercenary.pdf

Further questions on Cambridge Analytica's involvement in the 2017 Kenyan Elections and Privacy International's investigations. (2018). Privacy International. Retrieved May 8, 2023, from http://privacyinternational.org/long-read/1708/further-questions-cambridge-analyticas-involvement-2017-kenyan-elections-and-privacy

Galston, W. A. (2012). Truth and Democracy: Theme and Variations. In J. Elkins & A. Norris (Eds.), *Truth and Democracy* (pp. 130–145). University of Pennsylvania Press.

Gasser, U. (2021) Futuring Digital Privacy: Reimaging the Law/Tech Interplay. In Mira Burri (Ed), *Big Data and Global Trade Law*. Cambridge University Press.

Geoffrey Andare v Attorney General & 2 others (2016), eKLR

Giansiracusa, N. (2021). *How Algorithms Create and Prevent Fake News: Exploring the Impacts of Social Media, Deepfakes, GPT-3, and More*. Apress. https://doi.org/10.1007/978-1-4842-7155-1

Goldstein, J., Girish, S., Musser, M., DiResta, R., Gentzel, M., & Sedova, K. (2023). *Generative Language Models and Automated Influence Opera-*

tions: Emerging Threats and Potential Mitigation.

Greenwood, S. (2022, December 6). Social Media Seen as Mostly Good for Democracy Across Many Nations, But U.S. is a Major Outlier. *Pew Research Center's Global Attitudes Project*. https://www.pewresearch.org/global/2022/12/06/social-media-seen-as-mostly-good-for-democracy-across-many-nations-but-u-s-is-a-major-outlier/

Howard, P. (2019). Democratic Futures and the Internet of Things: How Information Infrastructure Will Become a Political Constitution. In *Digital Media and Democratic Futures* (p. 318).

Howard, P., Samuel, W., & Ryan, C. (2018). Algorithms, bots, and political communication in the US 2016 election. *Journal of Information Technology & Politics*, *15*(2), 81–93.

Independent Electoral and Boundaries Commission (IEBC). (n.d.). Retrieved June 7, 2023, from https://www.iebc.or.ke/iebc/?mandate

Introducing ChatGPT. (n.d.). OpenAI. Retrieved June 4, 2023, from https://openai.com/blog/chatgpt

Ipsos Global Survey on Internet Security & Trust "Social Media, Fake News & Algorithms." (2019). Ipsos Public Affairs, Centre for International Governance Innovation.

Jankowicz, N. (2017, December 11). How disinformation became a new threat to women. *Coda Story*. https://www.codastory.com/disinformation/how-disinformation-became-a-new-threat-to-women/

Joshi, N. (2019, May 1). *Why governments need to regulate AI | Artificial Intelligence |*. https://www.allerin.com/blog/why-governments-need-to-regulate-ai

Kang, C., Fandos, N., & Isaac, M. (2017, October 31). Tech Executives Are Contrite About Election Meddling, but Make Few Promises on Capitol Hill. *The New York Times*. https://www.nytimes.com/2017/10/31/us/politics/facebook-twitter-google-hearings-congress.html

Kenya: Cybercrime and Computer Related Crimes Bill. (2018). Article 19.

Kenya Information and Communications Act (No. 2 of 1998)

Kenya—Information, Communications and Technology (ICT). Retrieved October 19, 2023, from https://www.trade.gov/country-commercial-guides/kenya-information-communications-and-technology-ict

Kenya: Tackling misinformation is critical for electoral integrity. (2022a). Article 19. https://www.article19.org/resources/kenya-tackling-misinformation-critical-electoral-integrity/

Kenya's 2022 Political Sphere Overwhelmed by Disinformation – Collaboration on International ICT Policy for East and Southern Africa. CIPESA. (2022). Retrieved May 26, 2023, from https://cipesa.org/2022/07/11187-2/

Key social media risks to democracy: Risks from surveillance, personalisation, disinformation, moderation and microtargeting. (2021, December 14). Epthinktank. https://epthinktank.eu/2021/12/14/key-social-me-

dia-risks-to-democracy-risks-from-surveillance-personalisation-disinformation-moderation-and-microtargeting/

Kipkoech, G. (2022). Connections Between Internet, Social Media News Use, and Political Participation in Kenya. *Social Science Computer Review*, 08944393211058702. https://doi.org/10.1177/08944393211058702

Kiplagat, S. (2023, February 6). *Meta case sets up multinationals for Kenya suits*. Business Daily. https://www.businessdailyafrica.com/bd/economy/meta-case-sets-up-multinationals-for-kenya-suits--4114018

Kreiss, D. & Barrett, B., Dommett, K., (2021). The capricious relationship between technology and democracy: Analyzing public policy discussions in the UK and US. *Policy & Internet*, 522–543.

Kreps, S., & McCain, M. (2019, August 2). Not Your Father's Bots. *Foreign Affairs*. https://www.foreignaffairs.com/world/not-your-fathers-bots

Lilian, O. (2023, January 5). Disinformation was rife in Kenya's 2022 election. *Africa at LSE*. https://blogs.lse.ac.uk/africaatlse/2023/01/05/disinformation-was-rife-in-kenyas-2022-election/

Lomas, N. (2023). Europe wants platforms to label AI-generated content to fight disinformation. *TechCrunch*. https://techcrunch.com/2023/06/06/eu-disinformation-code-generative-ai-labels/

López-López, P.C., Barredo-Ibáñez, D., & Jaráiz-Gulías, E. (2023). Research on Digital Political Communication: Electoral Campaigns, Disinformation, and Artificial Intelligence. *Political Communication and Public Political Participation in the Digital Societies*.

Lorenz-Spreen, P., Oswald, L., Lewandowsky, S., & Hertwig, R. (2023). A systematic review of worldwide causal and correlational evidence on digital media and democracy. *Nature Human Behaviour*, *7*(1), Article 1. https://doi.org/10.1038/s41562-022-01460-1

Mackinnon, R., Hickok, E., Bar, A., & Lim, H. (2015). *Fostering freedom online: The role of internet intermediaries*. UNESCO.

Mandate | Media Council of Kenya (MCK). (n.d.). Retrieved June 7, 2023, from https://mediacouncil.or.ke/about-us/mandate

Marsden, C. (2011). *Internet Co-Regulation and Constitutionalism: Towards a More Nuanced View*.

Marsden, C., Meyer, T., & Brown, I. (2020). Platform values and democratic elections: How can the law regulate digital disinformation? *Computer Law & Security Review*.

Michailidou, A., & Trenz, H. (2023). Journalism, Truth and the Restoration of Trust in Democracy: Tracing the EU 'Fake News' Strategy. In *Europe in the Age of Post-Truth Politics*. Palgrave Macmillan, Cham.

Micro-targeting | Privacy International. (n.d.). Retrieved May 6, 2023, from https://privacyinternational.org/learn/micro-targetingNational Cohesion and Integration Act (Act. No. 12 of 2008)

National Cohesion and Integration Commission (NCIC). (2020). *A Violence Free*

2022: Roadmap to Peaceful 2022 General Elections.

NCIC at a Glance. (n.d.). Retrieved June 7, 2023, from https://cohesion.or.ke/index.php/about-us/ncic-at-a-glance

NCIC Monitoring Social Media for Hate Speech. (n.d.). Mzalendo. Retrieved June 15, 2023, from https://info.mzalendo.com/

Ndlela, M. N. (2022, March 30). *Algorithms, bots and elections in Africa: How social media influences political choices*. The Conversation. http://theconversation.com/algorithms-bots-and-elections-in-africa-how-social-media-influences-political-choices-179121

Njanja, A. (2023, June 9). Meta to appeal court ruling declaring content moderators as its employees. *TechCrunch*. https://techcrunch.com/2023/06/09/meta-to-appeal-court-ruling-declaring-content-moderatos-as-its-employees/

Nyabola, N. (2018). *Digital Democracy, Analogue Politics: How the Internet Era is Transforming Politics in Kenya*. Zed Books.

Nzina, J. (2014). The impact of mass media on the political participation among Kenyans: A case study of Nairobi youth. In M. Amutabi, K. Otiso, F. Koti, & C. Manayara (Eds.), *Kenya Review Studies* (pp. 4–16).

Odanga, M. (2021). *In Kenya, Influencers Are Hired to Spread Disinformation | WIRED*. Retrieved October 19, 2023, from https://www.wired.com/story/opinion-in-kenya-influencers-are-hired-to-spread-disinformation/

Office of the Registrar of Political Parties (ORPP). (n.d.). Retrieved June 7, 2023, from https://www.orpp.or.ke/index.php/2-uncategorised/6-about-us

Okoh, C. (2020, November 16). *Implications of Facebook/Instagram Fact-Checking Mechanism: The Nigerian EndSARS Hashtag Instance - Centre for Intellectual Property and Information Technology law*. https://cipit.strathmore.edu/implications-of-facebook-instagram-fact-checking-mechanism-the-nigerian-endsars-hashtag-instance/

Okoth, F., Fackson, B., & Wisdom, T. (2009). Introduction: New Media and Democracy in Africa—A Critical Interjection. In *African Media and the Digital Public Sphere*.

Omanga, D. (2019). WhatsApp as 'digital publics': The Nakuru Analysts and the evolution of participation in county governance in Kenya. Journal of Eastern African Studies, 13:1, 175–191.

Oversight Board. (2021, November 19). *Meta*. https://about.fb.com/news/tag/oversight-board/

Preliminary Observation Report: KICTANet Technology Observer Mission During Kenya's 2022 General Elections. (2022). KICTANet.

Recommendation on the Ethics of Artificial Intelligence—UNESCO Digital Library. (2021). Retrieved May 30, 2023, from https://unesdoc.unesco.org/ark:/48223/pf0000377897

Roxana, E. (2017). *GeoPoll and Portland launch a Survey Report on Fake News in Kenya*. Geopoll. https://www.geopoll.com/blog/geopoll-and-portland-

launch-a-survey-report-on-fake-news-in-kenya/

Rubinstein, IS. (2018). The Future of Self-Regulation Is Co-Regulation. In E. Selinger, J. Polonetsky & O. Tene (Eds), The Cambridge Handbook of Consumer Privacy. Cambridge University Press.

Rumbul, R. (2016). ICT and Citizen Efficacy: The Role of Civic Technology in Facilitating Government Accountability and Citizen Confidence. In F. J. Mata & A. Pont (Eds.), *ICT for Promoting Human Development and Protecting the Environment* (Vol. 481, pp. 213–222). Springer International Publishing. https://doi.org/10.1007/978-3-319-44447-5_20

Rutenberg, I., & Sugow, A. (2020). *Regulation of the Social Media in Electoral Democracies: A Case of Kenya. 7*, 302–361.

Sachin, H. (2022). *Regulating Content Recommendation Algorithms in Social Media*. Digital Platform Regulation Conference. https://som.yale.edu/sites/default/files/2022-05/DPRC-Holdheim.pdf

Safaricom Sustainability Report. (2017). Safaricom. Retrieved May 26, 2023, from https://www.safaricom.co.ke/sustainabilityreport_2017/ebook/files/assets/basic-html/page-4.html

Safeguarding the Digital Space Ahead of Kenya's 2022 Elections. (2021, December 3). Kofi Annan Foundation. https://www.kofiannanfoundation.org/supporting-democracy-and-elections-with-integrity/safeguarding-digital-space-kenya-elections-2022/

Sanya, B. (2013). Disrupting Partriarchy: An Examination of the role of e-Technologies in Rural Kenya. *Feminist Africa*, *18*, 12–24.

Schmitt-Beck, R. (2008). Bandwagon effect. In *The international encyclopaedia of communication* (Vol. 2).

SMElab Africa, USIU Africa. (2018). *Social Media Consumption in Kenya: Trends and Practices*.

Smith, T. G. (2018). Politicising digital space: Theory, the internet and renewing democracy. In *Contemporary Political Theory*, 19, (2018) S59–S62.

Stronger digital voices from Africa: Building African digital foreign policy and diplomacy. (2022). Diplo.

Sugow, A. (2019). *The Right to be Wrong: Examining the (Im) possibilities of Regulating Fake News while Preserving the Freedom of Expression in Kenya* (SSRN Scholarly Paper 3623716).

Surveillance giants: How the business model of Google and Facebook threatens human rights. (2019, November 21). Amnesty International. https://www.amnesty.org/en/documents/pol30/1404/2019/en/

Tan, R. (2023, June 20). Facebook helped bring free speech to Vietnam. Now it's helping stifle it. *Washington Post*. https://www.washingtonpost.com/world/2023/06/19/facebook-meta-vietnam-government-censorship/

The Big 4—Empowering the Nation. Government of Kenya. (n.d.). Retrieved November 30, 2023, from https://big4.delivery.go.ke/

The Information Society Project and The Floyd Abrams Institute for Freedom

of Expression. (March 2017). *Fighting fake news: Workshop report*. Yale Law School.

The State of AI in Africa. (2023). Centre for Intellectual Property and Information Technology Law (CIPIT). https://cipit.strathmore.edu/state-of-artificial-intelligence-in-africa-2023-report/

Tiffany, H., & Stuart, T. (2023, June 20). *Disinformation Researchers Raise Alarms About A.I. Chatbots—The New York Times*. https://www.nytimes.com/2023/02/08/technology/ai-chatbots-disinformation.html

Trade secrets shouldn't shield tech companies' algorithms from oversight | Brookings. (2020). Retrieved October 19, 2023, from https://www.brookings.edu/articles/trade-secrets-shouldnt-shield-tech-companies-algorithms-from-oversight/

United Nations. (2023a). *A Global Digital Compact—Common Agenda Policy Brief 5*.

United Nations General Assembly (UNGA). (2022). Resolution A/RES/76/227.

United Nations. (2011). Guiding Principles on Business and Human Rights.

United Nations Human Rights Council (HRC). (2021). Resolution A/HRC/47/25.

United Nations Human Rights Council (HRC). (2022). Resolution A/HRC/49/21.

United Nations. (2023b). *Our Common Agenda Policy Brief 8: Information Integrity on Digital Platforms*.

Using AI to Fight Disinformation in European Elections | ITIF. (2019). Retrieved June 3, 2023, from https://itif.org/events/2019/02/20/using-ai-fight-disinformation-european-elections/

Walubengo, J., & Mutemi, M. (2018). Treatment of Kenya's internet intermediaries under the Computer Misuse and Cybercrimes Act. *The African Journal of Information and Communication*, *21*, 1–19.

Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking*. Council of Europe.

What is the Fourth Industrial Revolution? (2021, January 14). Industrial Analytics Platform. https://iap.unido.org/articles/what-fourth-industrial-revolution

World Trends in Freedom of Expression & Media Development – Special Digital Focus. (2015). UNESCO.

Wright, K. (2023, September 12). *ChatGPT Large Language Model Explained*. InData Labs. https://indatalabs.com/blog/chatgpt-large-language-model

Zanon, A., Chaves Dutra da Rocha, L., & Manzato, M. G. (2022). Balancing the trade-off between accuracy and diversity in recommender systems with personalized explanations based on Linked Open Data. *Knowledge-Based Systems*, *252*.