social sciences

MDPI

Review

Surveillance, Disinformation, and Legislative Measures in the 21st Century: AI, Social Media, and the Future of Democracies

Bilge Azgin [1] and Sevki Kiralp [2,*]

[1] Department of Public Administration, Near East University, Nicosia 99138, Cyprus; bilge.azgin@neu.edu.tr
[2] Department of Political Science and International Relations, Cyprus International University, Nicosia 99258, Cyprus
* Correspondence: skiralp@ciu.edu.tr

Abstract: In contemporary society, the internet, particularly social media, has become a significant area where individuals spend considerable amounts of time engaging in various activities. Concurrently, the growing utilization of artificial intelligence (AI) has emerged as a critical component of the propaganda that is disseminated online within economic, social, and political spheres. AI encompasses a broad range of applications, including data collection for microtargeting and the dissemination of diverse forms of disinformation. Additionally, AI can be effectively employed to detect and remove content from social media platforms that contradicts democratic principles, such as disinformation or hate speech. This study reviews the existing literature on the use of AI in political propaganda, examining not only how AI has become an integral part of propaganda strategies, but also how it is utilized to counter propaganda that violates democratic values. It explores the legislation in various countries that enables (and mandates) the removal of propaganda content contrary to democratic principles from social media platforms with the assistance of AI, and it discusses perspectives that highlight the potential conflict between these practices and the principle of freedom of expression.

Keywords: censorship; freedom of expression; social media; microtargeting; content removal

Citation: Azgin, Bilge, and Sevki Kiralp. 2024. Surveillance, Disinformation, and Legislative Measures in the 21st Century: AI, Social Media, and the Future of Democracies. *Social Sciences* 13: 510. https://doi.org/10.3390/socsci13100510

Academic Editor: David F. J. Campbell

Received: 28 July 2024
Revised: 10 September 2024
Accepted: 25 September 2024
Published: 27 September 2024

1. Methodology

This study aims to examine the literature concerning the interplay between artificial intelligence (AI), surveillance, political propaganda, and the legislative frameworks developed in recent years to mitigate the misuse of AI technology. Our sources comprise academic books, journal articles, and reports from reputable institutions in relevant fields. These sources were accessed through databases such as Google Scholar, JSTOR, and institutional websites. In selecting our sources, we employed keywords such as "AI and propaganda", "surveillance in authoritarian regimes", "surveillance and democracy", "social media and AI", "content removal and AI", "microtargeting and AI", "data protection regulations", "freedom of expression and AI", and "content removal legislation". Variations of these terms were also used to ensure a comprehensive search. The 2016 US presidential election is a significant turning point in the literature addressed by this study. This election was marked by the Cambridge Analytica scandal, in which vast amounts of personal data were illegally obtained and used for disinformation, along with allegations of Russian interference in the election. These developments sparked political and academic discussions around AI, propaganda, and surveillance. Furthermore, as both governments and social media platforms have increasingly sought to implement various measures (whether through legal frameworks, community guidelines, or the use of AI itself) to combat these issues, we focused on sources published between 2016 and 2024. We synthesized the content of the reviewed articles to identify recurring themes and trends. Special attention was paid to studies presenting differing views on contentious issues and to data that yielded

contrasting results. Thus, we aimed to shed light on the ongoing debates in the literature and to highlight potential areas for future research.

2. Introduction

In the post-truth era, the increasing utilization of AI by governments around the world for propaganda and censorship purposes raises profound concerns regarding human rights and individual freedoms. Despite the gravity of this enduring political phenomenon and the alarming pace of its misuse for disinformation in democratic as well as authoritarian regimes, insufficient attention has been devoted to this issue in both public and academic discourse. This study aims to elucidate the democratic quandary engendered by the misuse of AI in the contemporary global context through an examination of the existing academic literature. Additionally, it scrutinizes the surveillance management efforts and initiatives pursued by governments to prevent the misuse of AI with respect to individual freedoms and liberties.

Contemporary technology, particularly AI, provides governments with the necessary means to constantly "watch" their citizens. Governments, especially in authoritarian regimes, tend to exploit these surveillance capabilities to the fullest extent. Moreover, AI is frequently utilized in both democracies and authoritarian regimes for purposes of data collection on citizens and their manipulation through disinformation. To ensure data privacy and to protect citizens from disinformation, states are implementing various legal measures. In the fight against the misuse of AI, AI itself emerges as one of the most critical tools, playing a significant role in identifying and removing fake or illegal content. In the fight against propaganda strategies that use AI to identify target audiences or to generate deceptive, fake content, important responsibilities fall not only on AI itself, but also on governments, social media platforms, and citizens. Although such propaganda approaches are dangerous, raising awareness of the risks they pose could potentially transform the internet into an environment composed of individuals with higher political consciousness in the near future.

This study explores not only how AI is employed for the purpose of safeguarding democratic principles, but also how it is utilized for purposes that diverge from these principles. In addition to examining instances of surveillance and disinformation manipulation through AI usage across different countries, we also scrutinize the legal measures implemented in various jurisdictions to counter such propagandistic methods. Furthermore, we delve into the debates in the literature triggered by the increasing prevalence of concepts such as the "privatization of censorship", alongside the discussions surrounding the implications of content removal via AI assistance on freedom of expression within the realm of social media. Lastly, we address debates that highlight both the advantages and disadvantages of human involvement alongside AI in various contexts. The fundamental conclusion we reach underscores the efficacy of combating the risks posed by AI through AI itself. However, it also emphasizes the vital importance of fostering awareness among individuals and societies to counter disinformation propaganda.

3. The Use of AI for Surveillance in Authoritarian Regimes

Due to current AI-supported surveillance mechanisms, such as facial recognition technology, locating missing persons, apprehending fugitives, and preventing various crimes have become easier; undoubtedly, this technology presents significant opportunities that may benefit citizens. AI-powered surveillance technology is employed in both democracies and authoritarian regimes; yet, a critical difference exists between the two. In democracies, AI-supported surveillance systems operate within a legal framework based on a human rights-inspired approach, respecting privacy concerns and adhering to various ethical boundaries. However, in authoritarian regimes, AI-supported surveillance systems are used not only to contribute to the maintenance of public order, but also to detect and suppress dissenting elements. In democracies, the use of this technology is guided by principles of government accountability and transparency to the public, whereas

in authoritarian regimes, no such accountability exists. Therefore, in authoritarian regimes, AI-supported surveillance systems are open to exploitation by governments that use them to advance anti-democratic agendas (Feldstein 2019; Fontes et al. 2022; Filgueiras 2022). As can be concluded, AI is an indispensable tool for surveillance today and is used by both democracies and authoritarian regimes to serve different purposes.

There is no turning back from the rapid advancement of AI, as it is inextricably linked to economic development, national security, and technological progress. In recent years, numerous articles and books have been published examining AI's implications for propaganda, censorship, and civil liberties (Black and Fullerton 2020; Büthe et al. 2022; Ashraf 2020; Chun-Chih and Thung-Hong 2020; Feldstein 2021; De Sio 2024; Nemitz and Pfeffer 2023). While AI is being utilized to spread propaganda on digital platforms, it has also been employed for various forms of digital disinformation and censorship (Stoycheff et al. 2020). One must consider any form of disinformation and censorship within the broader context of a propaganda tool used by political regimes. Authoritarian regimes in particular deliberately distort information and silence dissenting views to maintain and enhance their own political propaganda (Feldstein 2021). In their article, Goldstein et al. (2024) demonstrated that AI-based propaganda can be as effective and persuasive as traditional propaganda. In the context of the "third wave of autocratization", where countries across the globe are becoming increasingly authoritarian, the meteoric rise of digital authoritarianism, repression, and censorship has become a highly alarming and problematic issue (Lührmann and Lindberg 2019, p. 1053). Freedom House issued two reports that covered these issues in detail. The first report, titled "The Rise of Digital Authoritarianism", was published in 2018 (Shahbaz 2018), while the second report, titled "The Repressive Power of Artificial Intelligence", was published in 2023. In the latest report from 2023, Freedom House states that "global internet freedom declined for the 13th consecutive year in 2023" (Funk et al. 2023, p. 1). Although there is no unanimous agreement among scholars on the types of repression that occur in the digital world, the authors utilize Feldstein's five categories, which are "surveillance, online censorship, social manipulation and disinformation, internet shutdowns, and targeted persecution of online users" (Feldstein 2021, p. 26). Of these five categories, we particularly focus on surveillance, social manipulation, and disinformation, as AI is especially utilized in these areas. Surveillance encompasses a wide variety of activities, as states (whether democratic or autocratic) have always employed it for national security purposes. However, with the rise of AI, states have developed far more efficient, intrusive, and Orwellian systems, such as facial recognition technologies, smart policing, and smart city/safe city platforms (Feldstein 2019, p. 1). Additionally, AI is being deployed for Information Warfare and Influence Operations (IWIO) tactics and strategies. Lin (2020, p. 167) defines IWIO as "the deliberate manipulation or use of information by one party on an adversary to influence the choices and decisions the adversary makes for military or strategic gain". IWIO employs a variety of strategies and tactics, such as electronic warfare, cyberwarfare, and psychological warfare, to deliver its purpose.

In their insightful article on the USA, China, and Russia, Hunter et al. (2024) demonstrate how China and Russia, in particular, utilize IWIO to produce false and divisive propaganda in rival countries, as well as to monitor and control their own citizens. For instance, the Chinese government collects all the data from WeChat, which accounts for 60% of all social media transactions in China, and then utilizes these data to identify citizens who are prone to opposition to the regime. The Chinese government has also employed AI to disseminate false news and information in order to portray the dire situation of the Uyghur minority in a more favorable light. Similarly, Beijing deployed the same strategy to discredit the protesters in Hong Kong. In a manner that is consistent with autocratic traditions, Russia has utilized AI-based face recognition technology in the arrest of dozens of journalists critical of Putin's regime. As the foremost violator of internet freedom globally, China has articulated its ambitious plan to become the world leader in AI by 2030. With this plan, published in 2017 under the title "A Next Generation Artificial Intelligence Devel-

opment Plan", China aims to achieve parity with the US by 2025 and subsequently surpass all other nations by the year 2030 (Stanford University 2017). This development is not necessarily auspicious, given that China presently stands as the principal global exporter of AI to authoritarian nations, such as Saudi Arabia and Ethiopia. According to the Freedom House report, a minimum of 18 countries worldwide are procuring high-tech surveillance, empowered by AI, from Chinese companies. These acquisitions aim not only to establish national facial recognition databases, but also to develop systems capable of identifying and targeting groups or individuals perceived as subversive or threatening to the political regime (Shahbaz 2018, p. 8). Moreover, it should be emphasized that AI-powered enhanced surveillance systems, manipulative propaganda based on misinformation and disinformation, the collection of personal data, and microtargeting pose a significant threat not only to authoritarian regimes or those prone to authoritarianism, but also to democracies. The misuse of AI by governments and organizations could reach dangerous levels, potentially leading to the collapse of democratic institutions worldwide (Helbing et al. 2019).

4. Utilizing AI for Political Manipulation and Misinformation

The Freedom House report of 2023 (Funk et al. 2023, p. 1) highlighted that pro-government agencies in 47 countries have actively participated in deliberate manipulation and dissemination of fake information alongside engagement in propaganda activities. The utilization of AI for the generation of propaganda content against political rivals has emerged as a growing trend in recent years. With recent technological advancements, propagandists in countries hosting democratic elections have begun employing AI, including deepfake and similar techniques, to create and disseminate distorted content about political figures on the internet. This trend has swiftly permeated the political landscapes of both Western and Eastern societies. For instance, during Turkey's 2023 elections, digital propagandists associated with the ruling *Cumhur İttifakı* (People's Alliance) utilized AI to produce a manipulated video depicting Kemal Kılıçdaroğlu, leader of the opposition *Millet İttifakı* (Nation Alliance), in a misleading context, suggesting alleged support from a leader of the terrorist organization PKK (Euronews 2023). Similar instances of AI being employed in electoral propaganda can also be observed in the United States. For example, during the 2024 presidential primary, former US President Donald Trump and Florida Governor Ron DeSantis utilized AI to create disinformation-laden propaganda content targeting their respective opponents (Ulmer and Tong 2023). Indeed, AI-generated deepfake videos are becoming viral worldwide. Shortly before the parliamentary elections in Slovakia, a deepfake audio recording surfaced on the internet, falsely portraying the leader of Progressive Slovakia, Michal Šimečka, as engaged in a conversation with a journalist about how to rig the elections. Similarly, an AI-manipulated audio clip falsely impersonated one of the presidential candidates in Nigeria, alleging conspiracy to manipulate the ballot box for the 2023 elections (Funk et al. 2023).

In Moldova, AI-generated deepfake videos have depicted President Maia Sandu, who advocates closer ties with the West, as resigning from her office and rallying the populace to vote for a pro-Putin party in the local elections (Balkan Insight 2023). In addition to disseminating false information, deepfake videos are also utilized for character assassination. Rumeen Farhana, an opposition party leader in Bangladesh, has become one of the female victims of character assassination when a photo of her wearing a bikini circulated on social media. In a conservative Muslim country such as Bangladesh, the circulation of a fake bikini photo of Farhana has led to digital harassment and has damaged her public reputation (Verma and Zakrzewski 2024). Indeed, AI has bolstered the capability of both authoritarian and democratic states to disseminate more comprehensive and efficacious propaganda. Specifically, autocratic regimes possess a "new toolkit" that enables surveillance even at the microlevel, rendering censorship more efficient, albeit less overt (Fontaine and Frederick 2019). As previously noted, we are currently witnessing the era of the third wave of autocratization. Hellmeier et al. (2021, p. 1) noted that 25 countries (home to 34% of the world's population) over the past decade have regressed from democracy to

authoritarianism. However, democratic nations are also vulnerable to the anti-democratic exploitation of AI. In a deeply divided and polarized global landscape, where populism is gaining ground within democratic states and populist leaders, such as Viktor Orbán (Hungary), Andrzej Duda (Poland), and Donald Trump (USA), are garnering increasing popularity through electoral victories, disinformation campaigns are proving increasingly effective (Müller 2017). For these very reasons, democracies and democratic states worldwide should exert their utmost efforts in formulating legislation and modes of governance aimed at preventing the misuse of AI and safeguarding individual rights and liberties.

5. Identifying the Target Audience for Disinformation through the Use of AI: Microtargeting

Meta, the parent company of Facebook, WhatsApp, and Instagram, currently employs microtargeting with considerable efficacy; yet, this capability has drawn various criticisms. By collecting data through users' activities on the platform and various inputs (such as their ages, genders, locations, interests, and behavioral patterns), Meta facilitates the use of these data in favor of advertisers. Through microtargeting, Meta ensures that users encounter "highly personalized" advertisements based on the information gathered about them (Dobber 2023). In response to concerns regarding privacy and the protection of personal data, Meta has developed the Ad Library, which provides information about all advertisements and target audiences, including political advertisements. However, the effectiveness of the Ad Library in safeguarding privacy and personal data protection remains a topic of debate, and there are apprehensions regarding the potential for abuse of the data collected by Meta. These concerns are underscored by significant incidents, such as the Cambridge Analytica scandal, where the data of millions of users were obtained without their consent during the 2016 US elections (Mehta and Erickson 2022).

Microtargeting has become a widespread phenomenon, utilized for both online customer marketing and political advertising campaigns. In their study, Baviera et al. (2023, p. 2) analyzed a dataset comprising 14,677 ads downloaded from the Facebook Ad Library, sponsored by the five main Spanish parties during the two General Election campaigns in 2019. Indeed, the infamous scandal associated with Cambridge Analytica and the allegations that Russia interfered in the 2016 U.S. Presidential election have brought the prominence of microtargeting into the public spotlight. It became apparent that Cambridge Analytica utilized an enormous amount of data at its disposal for microtargeting, during both the US presidential elections of 2016 and the UK's Brexit referendum (Heawood 2018). Microtargeting has emerged as a highly effective tool for shaping and manipulating public opinion owing to its capacity for tailoring messages to individual psychological and socio-political profiles. To comprehend the perilous dimensions of disinformation or fake news, which can be further amplified with the assistance of AI, it is imperative to consider the concept of "post-truth", which is frequently invoked by politicians, media, and propagandists in contemporary politics. Those intending to produce disinformation typically identify a susceptible target audience beforehand by collecting personal data on the internet to gain information about individuals' interests or political tendencies (Cosentino 2020). To achieve this, they employ methods such as microtargeting. Leveraging data gathered through microtargeting, they craft propaganda content tailored to appeal to the emotions of the identified target audience members, rather than presenting them with objective facts based on logic. Emotional appeal addresses the target audience of the propaganda in question, activates their affective orientation (Armaly and Enders 2021), and evokes emotions such as anger, fear, or excitement among the members of this audience, thereby prompting them to act in favor of or against a political actor. Consequently, propagandists aim to influence the political behaviors of the target audience members, such as voting behaviors (Chiu and Oh 2021).

Miró-Llinares and Aguerri (2023, p. 366), in their examination of empirical data on the impact of microtargeting on the 2016 election results, revealed that only 1.18% of the news reaching users via social media during this period was fake. Moreover, they showed that

80% of this false news was consumed by just 1% of users. In their study, they referenced arguments suggesting that the influence of fake news on election outcomes may have been exaggerated. In another study based on empirical data, Laterza (2021, pp. 135–39) demonstrated that, during the election period in question, the open rate of messages specifically tailored to users based on their psychological traits could have increased by up to 20% in some cases. Accordingly, Cambridge Analytica targeted 9 million users with personalized messages. Donald Trump, on the other hand, won critical states such as Michigan, Pennsylvania, and Wisconsin by a margin of only 80,000 votes. Therefore, the issues arising from Cambridge Analytica's microtargeting operations during the 2016 election cannot be underestimated. Similarly, Tappin et al. (2023, p. 4), in their experiment involving 5,284 US citizens, delivered personalized messages through microtargeting and found that support for the Citizenship Act among participants increased by 5.17%. Therefore, the impact of microtargeting on political processes cannot be underestimated, and the protection of personal data is of vital importance for the healthy functioning of liberal democracy.

6. Combating the Misuse: AI and Content Removal

While AI has become a significant tool for misuse in political propaganda, it has also brought about preventive measures, legislation, and numerous debates, including the use of "AI against AI". AI can also be employed in combating such objectives and in strengthening the practice of democratic principles against propaganda that contradicts democratic values. Furthermore, the Network Enforcement Act (NetzDG) in Germany, the Online Safety Act (OSA) in the United Kingdom, and the Digital Services Act (DSA) in EU countries are all aimed at combating propaganda contrary to democratic principles (manipulation through disinformation, hate speech, etc.) and enhancing the importance of AI in the fight against such propaganda. After the serious allegations that Russia interfered in the 2016 US Presidential election (see Senate Select Committee on Intelligence 2018), data security and the involvement of foreign actors in elections through disinformation and similar illegal propaganda became significant topics of discussion. Legislation proposals such as the "Honest Ads Act", "Protecting Democracy from Disinformation Act", "Disclose Act", "REAL Political Ads Act", and "Digital Consumer Protection Commission Act" aimed to mandate transparency in known political propaganda content on the internet (such as who funded this content), to ensure necessary data privacy, and, when necessary, to provide for the deletion of content shared on the internet platform. Additionally, they aimed to prevent foreign propagandists from using AI to manipulate target audiences by identifying them (Kaplan 2020; Weber 2021; Ahmad et al. 2022). The debates about these proposals are primarily centered around concerns that "freedom of expression will be inhibited", and as of yet, none of them have undergone the requisite legislative procedures to be enacted into law (Barrett et al. 2021). In October 2023, the United Kingdom enacted the OSA, thus moving slightly faster than the USA in this regard. The British executive and legislative branches developed the OSA with the awareness that AI would make psychological manipulation more widespread and effective. The OSA grants various powers to the government concerning the removal of "harmful" content from the internet, while obligating social media platforms to facilitate the reporting of "harmful" content by users, to respond promptly to such reports, and to remove the content when necessary. At this point, the ability of AI to swiftly respond emerges as a significant contributing factor for social media platforms. However, the possibility of removing "lawful but harmful" content, the highly debatable nature of what constitutes "harmful" content, and the uncertainty as to who or what should define it, have led to accusations of "censorship" and criticism (Matamoros-Fernández 2023).

The relatively tighter data protection legislation in the EU restricts the collection, storage, and transfer of data (both within and outside the EU) belonging to internet users. This legislation is acknowledged to provide a safer environment for internet users in the EU compared with those in the USA. Due to the fact that the anticipated legislation in the

USA has not yet been completed, social media platforms are currently operating within the confines of their own community rules (and, of course, within the framework of legitimacy derived from existing laws) when it comes to removing content containing inappropriate propaganda activities. It is worth emphasizing that almost no users who have filed lawsuits regarding removed social media content in the USA have won their cases (Goldman and Miers 2021).

The EU took significant steps toward the detection and removal of illicit content on social media platforms where AI is frequently employed. The most overarching and concrete manifestation of these efforts is the DSA, which is in force across the EU. However, prior to the implementation of the DSA in EU member states, examples such as Germany's NetzDG, which came into effect much earlier and addressed numerous needs effectively, were also present. Germany had obliged social media providers to rapidly and effectively remove illegal content. The NetzDG of 2017 established a specific framework for enforcing such measures (Claussen 2018). The removal of social media content with the assistance of AI has sparked various debates regarding the roles of AI and human factors. At this juncture, there are perspectives asserting that AI cannot be reliable without the human factor, while there are also viewpoints advocating that AI inspires greater confidence in instances where the human factor is sidelined. The study by Gonçalves et al. (2023), focusing on the United States, the Netherlands, and Portugal, demonstrated that algorithmic moderation (i.e., the removal of inappropriate social media content by relying on AI) is perceived to be fairer than human moderation. However, as pointed out by Wu, there are views in the literature suggesting that AI is not seen as being as reliable as humans in content moderation because AI cannot interpret content within linguistic, social, and cultural contexts (Wu 2019). Arguments have been put forth asserting that "hate speech" is more complex than what AI's technical filters can identify, thus indicating that forming sound judgments about content shared on social media necessitates the evaluation of frames such as intent, nuance, and context (Daphne 2018). In addition, it is well known that artificial intelligence sometimes commits glaring errors in detecting "inappropriate" (or illegal) content. As Elkin-Koren (2020, p. 5) pointed out, YouTube, which uses AI for the detection of inappropriate and/or illegal content, erroneously deleted over 100,000 pro-human rights videos in 2018 that were attempting to document the use of chemical weapons in Syria. This incident has significantly undermined the reliability of AI in content moderation. However, we cannot overlook the fact that there are instances where AI has proven its capability to remove inappropriate propaganda content from social media. For instance, according to Facebook's Community Standards Enforcement Report for the year 2020, artificial intelligence was responsible for the removal of 94.7% of hate speech on the platform before user reporting (Grad and Turnbull 2021, p. 28).

Legislation aimed at preventing microtargeting through the use of AI and ensuring the removal of content that does not comply with legal frameworks with the help of AI is not limited to the EU and the US. Many countries around the world have taken significant steps in this regard. Although not a democratic example, China's 2017 Cybersecurity Law grants the government substantial authority over online content (Qi et al. 2018). India, through the Information Technology Rules materialized by the 2021 legislative amendment, imposes an obligation on social media platforms to delete content that violates democratic principles within a specified timeframe (Shankar and Ahmad 2021). Pakistan, through the Prevention of Electronic Crimes Act of 2016 and the Pakistan Citizens Protection Rules of 2020, has made it mandatory for social media platforms to remove illegal content. These regulations also enable citizens to report illegal content on social media and grant the government, specifically the Pakistan Telecommunication Authority (PTA), the authority to delete content under certain circumstances (Jamil 2021). Turkey, through a 2020 legislative amendment, imposed an obligation on social media platforms to remove illegal content within a specified timeframe (Oymak 2020). Singapore, through its 2019 Protection from Online Falsehoods and Manipulation Act, grants authorities the power to mandate the removal of content that undermines democratic principles, particularly content designed

for misinformation and disinformation. In this country, social media platforms possess highly advanced AI capabilities to detect false information disseminated online (Goh and Soon 2021).

The debates regarding the appropriate utilization of AI in political processes, particularly within the context of propaganda, give rise to significant discussions concerning "algorithm transparency". These discussions aim to ensure that the algorithms employed in political advertising, as well as the propagandists utilizing them, are subject to public scrutiny. For instance, one aspect of the DSA, in effect in EU countries since 2018, mandates significant transparency from propagandists regarding the algorithms they employ (Maroni 2023). Within the same context, another concept given significant importance in the literature is "independent oversight" (Kertysova 2018). The principle of independent oversight aims to facilitate institutional (or individual) expertise in detecting disinformation through AI, such as in the identification of deepfakes, thereby providing significant third-party expertise in contentious AI applications. For example, the Netherlands-based Deeptrace firm is renowned for its expertise in deepfake detection, providing services to many parts of the world (Gong et al. 2021). Due to the contemporary phenomenon of the internet technology "erasing boundaries between countries", it has become imperative to address the dissemination of disinformation through AI with an international perspective, employing AI itself for countermeasures. Consequently, there are valid arguments demonstrating that, without international support, the national efforts to combat disinformation cannot be fully successful. Such arguments underscore the need for countries to collaborate with each other in tackling disinformation, highlighting the interdependence between national and international endeavors in this regard (Garon 2022). Another prerequisite for the success of such processes is the active involvement of the public (internet users), and it is highly important to raise awareness for this purpose (Kertysova 2018). One of the most effective means of enhancing the aforementioned awareness is to educate the public on digital media literacy, thereby fostering a more informed stance toward propaganda content on the internet (Liebowitz 2021).

In the literature, the predominant perspectives highlight concerns that the utilization of AI in the detection and removal processes of "illegal" propaganda may unjustly remove certain content, potentially leading to violations of fundamental values of liberal democracy, such as freedom of expression and access to information. According to this view, social media platforms may be compelled to broaden the scope of content classified as "illegal" by AI, as they face risks such as heavy fines and reputational damage for content they do not remove. This could lead to consequences such as the controversial or unjustified removal of certain content (Llansó et al. 2020). On the other hand, in the literature, there are proponents of the need for human oversight in addressing the errors made by AI in combating "illegal" propaganda on the internet. However, since it is humans who will create the algorithm of AI and review its functioning, and given that it is not possible for any human to be completely bias-free, achieving a completely "bias-free" and entirely "objective" performance of AI is deemed unrealistic. Currently, in both the establishment of norms and the enactment of laws, AI is acknowledged as a key tool in the fight against propaganda on the internet that contradicts democratic principles. However, another significant issue is the lack of accountability of AI, meaning that humans will still be responsible for everything entrusted to AI (Elkin-Koren 2020).

It is also important to consider the concept of the "privatization of censorship" in the literature concerning the removal of online content with the assistance of AI. Long before the enactment of the DSA in 2023, legislation in both Germany and Italy had left social media platforms obligated to remove "illegal" content within 24 h after being reported, and the determining authority of this "inappropriateness" (or "illegality") was not a judicial body. Such practices faced criticism as they delegated the initiative to private entities (or AI acting under their direction), thus leading to concerns regarding the "privatization of censorship". In such examples, social media platforms find themselves in the position of interpreting whether content is "illegal" within the framework outlined by legislation or relying largely

on AI for this purpose, which raises concerns that some content could be unjustly removed. In contrast, the situation in France was markedly different, where a judicial decision determined whether content should be removed from social media (Monti 2020). The DSA, which came into effect in August 2023, outlined a comprehensive framework for combating illegal propaganda content online across European Union countries. This act holds internet platforms accountable for illegal propaganda content and obliges them to remove various contents when necessary. These obligations require internet platforms to provide users with the opportunity to report "illegal" content, which is perceived positively in terms of human rights and participatory democracy (Trengove et al. 2022). Nonetheless, from one point of view, the DSA legislation also contributes to the proliferation of the phenomenon previously referred to in this study as the "privatization of censorship", whereby internet platforms seeking to avoid hefty fines must determine which content can be deemed "illegal", with or without the assistance of AI. In addition to its negative implications for freedom of expression, it should also be noted that this could lead to a situation where an authority peculiar to the state is delegated to private companies (Cobbe 2021).

Legislation such as the DSA does not preclude unjustly deleted users from resorting to legal action. Nonetheless, since the enactment of the NetzDG in 2017, there have been almost no court decisions in Germany indicating that social media platforms violated freedom of expression by removing content, particularly in cases where the social media platform adequately informed users whose posts were to be deleted (Zurth 2020). This situation, as previously stated in the paper, was nearly the same for the US judiciary. Such instances are not merely construed as indicative of the success of AI or the diligence of social media platforms; they are also interpreted as manifestations of legislative insensitivity toward freedom of expression. Nevertheless, there are also empirical studies demonstrating that such laws can hardly be accused of undermining freedom of thought or "overblocking". For instance, Maaß et al. (2024, p. 10) conducted a six-month examination covering the initial two months after the implementation of the NetzDG and the final four months prior to its enactment. This study, which analyzed a dataset consisting of 33,916 posts and 7,386,644 comments shared on Facebook, revealed that the proportion of comments deleted during the specified period increased by only 0.717%. Additionally, the researchers found no significant increase in self-censorship in terms of the tone of the comments made by users following the enactment of the NetzDG (Maaß et al. 2024, p. 15).

There are academic views that highlight the lack of sufficient transparency in social media platforms' content moderation practices. For instance, the DSA requires social media platforms to publish reports on the content they remove within EU countries. However, these platforms tend to limit themselves to providing information on the role of AI in the removal process, failing to offer sufficient clarity regarding the reasons for content removal and the role of human oversight in these moderation processes. Additionally, these platforms do not demonstrate transparency in their collaborations with fact-checking organizations or the criteria used by these organizations for evaluations. Despite these issues, the existing EU legislation offers a promising foundation for future transparency. For example, in a 2021 case, the German Federal Court of Justice ruled that platforms do not have the right to remove content without informing users and without providing them with the opportunity to defend themselves (Galantino 2023). In conclusion, it has been proven over time and through experience that social media platforms lack transparency in content removal processes, and AI can make serious errors in certain cases. However, with the regulations they have enacted, EU countries have successfully established solid foundations for content removal from social media as an essential measure in the fight against hate speech and misinformation.

7. Protection of Personal Data: Practical Difficulties

In response to concerns raised by the misuse of artificial intelligence in the unauthorized access and exploitation of personal data, particularly during the 2016 US presidential election, the EU countries drafted the General Data Protection Regulation (GDPR), which

came into effect in 2018. The GDPR prohibits the collection of internet users' data without their consent, aims to minimize the scope of collected data, and ensures that data are collected for specified, legitimate, and transparent purposes and retained for defined periods. The GDPR, which sets significant boundaries for the processing of data, with or without the use of AI, serves as a highly effective measure against microtargeting, a technique employed by propagandists (regardless of whether they are foreign or national) for disseminating disinformation. However, claiming that GDPR completely prevents microtargeting is exceedingly difficult. One reason for GDPR's inability to consistently achieve the desired impact is users' tendency to click "I agree" without reading the privacy notices (Fathaigh et al. 2021). In developing the GDPR, the EU aimed to curb microtargeting, which is utilized in identifying target audiences that are vulnerable to manipulation by AI; the USA, however, has not been able to respond as promptly in enacting similar legislation that addresses both the combating of microtargeting and the facilitation of content removal on social media platforms where AI can be effectively employed.

Legislation similar to the GDPR, concerning the protection and storage of personal data on the internet, is not limited to EU countries. In 2016, Turkey implemented a measure against microtargeting with the enactment of the *Kişisel Verilerin Korunması Kanunu*/Personal Data Protection Law (Kaşlı 2023). In 2019, Nigeria took a step by implementing the Data Protection Regulation, which outlines the framework for collecting personal data online and grants users the right to request the deletion of their personal data in certain circumstances (Chika and Tochukwu 2020). Similar legislation is also seen in South Africa, Brazil, and the United Arab Emirates. South Africa, with its 2020 Protection of Personal Information Act (Netshakhuma 2020), Brazil, with its 2020 General Data Protection Law (Sombra 2020), and the UAE, with its 2021 Personal Data Protection Law (Thanvi 2023), have taken serious measures against the exploitation of personal data for microtargeting, making the deletion of personal data for security purposes more practical.

While the GDPR and similar regulations represent a crucial legal framework aimed at preventing potential misuse of AI, they have also created inequities for small and medium-sized enterprises (SMEs) during their early years of implementation in EU countries, presenting significant practical challenges. For instance, in 2018, a legal case involving a training company's Facebook page in Germany resulted in a ruling by the German judiciary that the administrators of the page were equally responsible, alongside Facebook itself, for protecting the personal data of their members. In 2019, a kebab shop in Austria was fined for unlawfully collecting personal data through its camera system without obtaining consent. Therefore, the early application of the GDPR imposed the same responsibilities on SMEs, which lacked the resources of large corporations, leading to various instances of inequity (Lynskey 2023).

8. Future Prospects for AI and Democracy

As previously noted, AI, due to its algorithmic capabilities, plays a significant role in combating misleading content on social media, and it facilitates users' access to reliable information while forming their political opinions and behaviors, especially regarding voting (Kertysova 2018). Moreover, AI is highly conducive to the development of the 21st century's emerging concept of "digital democracy". AI-supported surveys and public opinion polls can enhance the understanding of public views on various issues, thereby contributing to healthier dynamics in the relationship between society and politics (Moats and Tseng 2024). Additionally, AI can serve as an effective tool for ensuring election security, particularly against foreign interference, thus supporting the genuine functioning of democracy. For instance, AI algorithms can effortlessly detect unusual and coordinated behavioral patterns during election periods, such as the rapid spread of misinformation by fake accounts (Chertoff and Rasmussen 2019). Similarly, through Natural Language Processing (NLP) algorithms, AI can quickly identify the malicious activities of bot accounts (Ruffo et al. 2023). AI can also play an effective role in cybersecurity, swiftly identifying and mitigating attacks on servers (Wirkuttis and Klein 2017).

AI not only provides citizens with a safer internet environment (particularly against malicious misinformation), but also facilitates more active participation in political processes. For instance, in relation to various issues, many countries employ AI-powered chatbots that assist citizens in finding answers to their questions, accessing accurate information, and saving time (Cortés-Cediela et al. 2023). Similarly, Canada is among the countries where municipalities hold online consultation meetings to collect public opinions on various topics, to employ natural language processing (NLP) techniques to analyze the feedback from these meetings, and to leverage the collected data to improve public service delivery (Robinson 2022). On the other hand, activist groups leverage AI's algorithmic features to disseminate their propaganda to wider audiences. For instance, the Black Lives Matter movement, which surged in response to the murder of George Floyd in the USA in 2020, reached millions of people worldwide and raised awareness in this manner (Taraktaş et al. 2024). Through AI's algorithmic capabilities and its ability to "personalize" the content that users encounter, marginalized groups or minorities can amplify their voices to broader audiences (Mariyono and Akmal 2024). Another contribution of AI to effective governance is its ability to provide decision makers with accurate insights into public sentiment through large data analyses (König and Wenzelburger 2022). The increasing use of AI brings not only risks, but also numerous opportunities, depending on how and for what purposes it is employed.

9. Conclusions

This study unveiled the exploitation of AI by authoritarian regimes and also by propagandists in democratic countries for manipulative purposes, such as disinformation and even interference in the elections of other countries. Furthermore, this study demonstrated how and why, in the world of the "third wave of autocratization", in which countries across the globe increasingly become more and more authoritarian, the meteoric rise of digital authoritarianism, repression, and censorship has become highly alarming and problematic. However, this study also examined the enactment of various legislations in democratic nations to counteract the misuse of AI when it contradicts the principles of human rights and democracy. Accordingly, it underscored the capability of AI to detect and remove propaganda contents, such as disinformation or hate speech, in today's online environment. The legislative efforts aimed at ensuring this, along with the obligations of social media platforms in certain democratic countries, were also discussed. This study highlighted the criticisms that have been voiced regarding the measures aimed at curbing microtargeting and removing content with the assistance of AI due to concerns revolving around freedom of expression. As one dimension of these discussions, concerns were also expressed regarding the "privatization" of state censorship authority, particularly with regard to the obligations of social media platforms. Debates concerning the relationship between AI and human supervision were also addressed. Some scholars argue that AI, which is devoid of emotions, is more reliable in identifying illegal propaganda online, while opponents contend that AI lacks certain interpretative abilities inherent to humans, thus rendering it less reliable than humans. Nevertheless, timely and effective intervention against illegal content is crucial for combating intentional misguidance of the masses and ensuring the healthy functioning of liberal democracy. Therefore, it is indisputable that AI, with both its strengths and limitations, serves as a critical instrument in the ongoing efforts to combat such content.

This study also discussed the practical performance of various regulations designed in alignment with democratic goals, including the prevention of AI misuse. The findings indicate that the DSA and similar legislation play a crucial role in addressing the misuse of AI, particularly by imposing obligations on major networks to remove illegal content (such as hate speech and disinformation). Unfortunately, major networks have not yet achieved the level of transparency anticipated by the DSA, especially in sharing the reasons behind content removal or the evaluation criteria used by their partnering organizations. Additionally, while the AI used by social networks can sometimes make highly accurate de-

tections of illegal content, it can also make serious errors. However, increased transparency in these processes could enhance accountability and public oversight, which is promising for the future. On the other hand, regulations such as the GDPR, which aim to protect personal data and prevent their misuse through AI, place significant responsibilities on relevant stakeholders. Nevertheless, the studies highlighting the practical challenges of these regulations reveal that both large corporations and small businesses are burdened with the same obligations, leading to potentially unfair situations for the latter.

Ultimately, AI serves as a tool that can both violate and uphold democratic principles in the contemporary world, a characteristic it is likely to retain in the future. Scholars interested in the subject are encouraged to explore collaborative formulations between AI and human supervision for the detection and removal of illegal propaganda on internet platforms; this would prove to be highly beneficial for the literature in the future. Empirical studies investigating the potential content and efficacy of such formulations might offer valuable contributions to the scientific community. It should be emphasized that one of the most significant findings in the literature is that we lack mechanisms to hold AI accountable for errors or ethical violations in combating illegal propaganda online. Since accountability rests with humans, for this reason alone, any discussion regarding AI must place central importance on the human factor. It can be argued that the role of the human factor in achieving desired outcomes through AI is paramount. In the forthcoming years, it is highly probable that AI will advance both in the production of more convincing disinformation content and in the more effective detection of disinformation content. Alongside the adoption of legal measures to combat the misuse of AI for misinformation purposes and the enhancement of AI capabilities, efforts to elevate societal awareness through education in propaganda literacy will prove immensely beneficial in countering the dissemination of disinformation. Enhancing the mechanisms provided to citizens by major networks for reporting illegal content and increasing transparency in areas such as removed content, the human factor in the moderation process, and the evaluation criteria of the organizations they collaborate with in this moderation will contribute to the development of a more participatory and accountable democratic culture.

Author Contributions: Conceptualization, B.A. and S.K.; methodology, B.A. and S.K.; software, B.A. and S.K.; validation, B.A. and S.K.; formal analysis, B.A. and S.K.; investigation, B.A. and S.K.; resources, B.A. and S.K.; data curation, B.A. and S.K.; writing—original draft preparation, B.A. and S.K.; writing—review and editing, B.A and S.K.; supervision, B.A. and S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

Ahmad, Taimoor, Edwin Aliaga Lazarte, and Seyedehmahsa Mirjalili. 2022. A systematic literature review on fake news in the COVID-19 pandemic: Can AI propose a solution? *Applied Sciences* 12: 12727. [CrossRef]

Armaly, Mary T., and Anna M. Enders. 2021. The role of affective orientations in promoting perceived polarization. *Political Science Research and Methods* 9: 615–26. [CrossRef]

Ashraf, Chris. 2020. Artificial Intelligence and the Rights to Assembly and Association. *Journal of Cyber Policy* 5: 163–79. [CrossRef]

Balkan Insight. 2023. Moldova Dismisses Deepfake Video Targeting President Sandu. Available online: https://balkaninsight.com/2023/12/29/moldova-dismisses-deepfake-video-targeting-president-sandu/ (accessed on 28 April 2024).

Barrett, Brian, Davin K. Errickson, and Daniel Kreiss. 2021. The capricious relationship between technology and democracy: Analyzing public policy discussions in the UK and US. *Policy & Internet* 13: 522–43. [CrossRef]

Baviera, Tomás, Lorena Cano-Orón, and Dafne Calvo. 2023. Tailored messages in the feed? Political microtargeting on Facebook during the 2019 General Elections in Spain. *Journal of Political Marketing*. [CrossRef]

Black, Jane, and Carl Fullerton. 2020. Digital Deceit: Fake News, Artificial Intelligence, and Censorship in Educational Research. *Open Journal of Social Sciences* 8: 71–88. [CrossRef]
Büthe, Tim, Christian Djeffal, Christoph Lütge, Sabine Maasen, and Nicole Von Ingersleben-Seip. 2022. Governing AI—Attempting to Herd Cats? Introduction to the Special Issue on the Governance of Artificial Intelligence. *Journal of European Public Policy* 29: 1721–52. [CrossRef]
Chertoff, Michael, and Anders Fogh Rasmussen. 2019. The Unhackable Election: What It Takes to Defend Democracy. *Foreign Affairs* 98: 156.
Chika, David M., and Edeh Stanley Tochukwu. 2020. An analysis of data protection and compliance in Nigeria. *International Journal of Research and Innovation in Social Science (IJRISS)* IV: 377–82.
Chiu, Mingming, and Young Wook Oh. 2021. How fake news differs from personal lies. *American Behavioral Scientist* 65: 243–58. [CrossRef]
Chun-Chih, Chen, and Lin Thung-Hong. 2020. Autocracy Login: Internet Censorship and Civil Society in the Digital Age. *Democratization* 27: 874–95. [CrossRef]
Claussen, Volker. 2018. Fighting hate speech and fake news: The Network Enforcement Act (NetzDG) in Germany in the context of European legislation. *Media Laws* 3: 110–36.
Cobbe, Jennifer. 2021. Algorithmic censorship by social platforms: Power and resistance. *Philosophy & Technology* 34: 739–66.
Cortés-Cediela, Maria E., Andrés Pardo-Bosch, Clara Velasco-Montero, and Mario Piattini. 2023. Trends and Challenges of E-Government Chatbots: Exploratory Research in Their Application to Open Government Data and Citizen Participation Content. *Government Information Quarterly* 40: 101877. [CrossRef]
Cosentino, Giulia. 2020. *Social Media and the Post-Truth World Order*. London: Palgrave.
Daphne, Ketseridis. 2018. *Internet Platforms: Observations on Speech, Danger, and Money*. Hoover Institution's Aegis Paper Series. 1807. New York: Hoover Institution.
De Sio, Filippo Santoni. 2024. *Human Freedom in the Age of AI*. Abingdon: Routledge.
Dobber, Tom. 2023. Microtargeting, privacy, and the need for regulating algorithms. In *The Routledge Handbook of Privacy and Social Media*. Edited by Corey A. Ciocchetti and John D. Johnson. London: Routledge, pp. 237–45.
Elkin-Koren, Niva. 2020. Contesting algorithms: Restoring the public interest in content filtering by artificial intelligence. *Big Data & Society* 7: 2053951720932296. [CrossRef]
Euronews. 2023. AI Content Deepfakes Meddling in Turkey Elections: Experts Warn It's Just the Beginning. Available online: https://www.euronews.com/next/2023/05/12/ai-content-deepfakes-meddling-in-turkey-elections-experts-warn-its-just-the-beginning (accessed on 28 April 2024).
Fathaigh, Órla, Tamar Dobber, Frederik Zuiderveen Borgesius, and Joris Shires. 2021. Microtargeted propaganda by foreign actors: An interdisciplinary exploration. *Maastricht Journal of European and Comparative Law* 28: 856–77. [CrossRef]
Feldstein, Steven. 2019. *The Global Expansion of AI Surveillance*. Washington, DC: Carnegie Endowment for International Peace.
Feldstein, Steven. 2021. *The Rise of Digital Repression: How Technology is Reshaping Power, Politics, and Resistance*. Oxford: Oxford University Press.
Filgueiras, Fernando. 2022. The politics of AI: Democracy and authoritarianism in developing countries. *Journal of Information Technology & Politics* 19: 449–64. [CrossRef]
Fontaine, Richard, and Kara Frederick. 2019. The Autocrat's New Tool Kit. *The Wall Street Journal*. Available online: https://www.wsj.com/articles/the-autocrats-new-tool-kit-11552662637 (accessed on 28 April 2024).
Fontes, Cristiano, Ellen Hohma, Caitlin C. Corrigan, and Christoph Lütge. 2022. AI-powered public surveillance systems: Why we (might) need them and how we want them. *Technology in Society* 71: 102137. [CrossRef]
Funk, Allie, Adrian Shahbaz, and Kári Vesteinsson. 2023. *Freedom on the Net 2023: The Repressive Power of Artificial Intelligence*. Washington, DC: Freedom House.
Galantino, Sharon. 2023. How Will the EU Digital Services Act Affect the Regulation of Disinformation? *SCRIPTed* 20: 89–116. [CrossRef]
Garon, Jon M. 2022. When AI Goes to War: Corporate Accountability for Virtual Mass Disinformation, Algorithmic Atrocities, and Synthetic Propaganda. *Northern Kentucky Law Review* 49: 181.
Goh, Shirley, and Chong Soon. 2021. Singapore's Fake News Law: Countering populists' falsehoods and truth-making. In *The Routledge Companion to Media Disinformation and Populism*. Edited by Howard Tumber and Silvio Waisbord. London: Routledge, pp. 459–69.
Goldman, Erica, and Jack Miers. 2021. Online Account Terminations/Content Removals and the Benefits of Internet Services Enforcing Their House Rules. *Journal of Free Speech and Law* 1: 192–226.
Goldstein, Joshua A., Jason Chao, Shelby Grossman, Alex Stamos, and Michael Tomz. 2024. How Persuasive is AI-Generated Propaganda? *PNAS Nexus* 3: 34. [CrossRef]
Gonçalves, João, Ingmar Weber, Gianmarco M. Masullo, Maria Teresa Da Silva, and Jürgen Hofhuis. 2023. Common sense or censorship: How algorithmic moderators and message type influence perceptions of online content deletion. *New Media & Society* 25: 2595–617. [CrossRef]
Gong, Dawei, Y.J. Kumar, O.S. Goh, Zhen Ye, and Wei Chi. 2021. DeepfakeNet, an Efficient Deepfake Detection Method. *International Journal of Advanced Computer Science and Applications* 12: 201–7. [CrossRef]

Grad, Kathryn, and Andrew Turnbull. 2021. Harmful Speech and the COVID-19 Penumbra. *Canadian Journal of Law and Technology* 19: 1–35. [CrossRef]
Heawood, James. 2018. Pseudo-public political speech: Democratic implications of the Cambridge Analytica scandal. *Information Polity* 23: 429–34. [CrossRef]
Helbing, Dirk, Bruno S. Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen Van Den Hoven, Roberto V. Zicari, and Andrej Zwitter. 2019. Will Democracy Survive Big Data and Artificial Intelligence? In *Towards Digital Enlightenment: Essays on the Dark and Light Sides of the Digital Revolution*. Edited by Dirk Helbing. Cham: Springer, pp. 73–98.
Hellmeier, Sebastian, Ruth Cole, Sandra Grahn, Petar Kolvani, Jean Lachapelle, Anna Lührmann, Seraphine F. Maerz, Stefan Pilla, and Staffan I. Lindberg. 2021. State of the World 2020: Autocratization Turns Viral. *Democratization* 28: 1053–74. [CrossRef]
Hunter, Larry Y., Craig D. Albert, Josh Rutland, Kristen Topping, and Christopher Hennigan. 2024. Artificial Intelligence and Information Warfare in Major Power States: How the US, China, and Russia are Using Artificial Intelligence in Their Information Warfare and Influence Operations. *Defense & Security Analysis* 40: 235–69. [CrossRef]
Jamil, Saad. 2021. The rise of digital authoritarianism: Evolving threats to media and Internet freedoms in Pakistan. *World of Media–Russian Journal of Journalism and Media Studies* 3: 5–33. [CrossRef]
Kaplan, Anat. 2020. Artificial Intelligence, Social Media, and Fake News: Is This the End of Democracy? In *Digital Transformation in Media and Society*. Edited by A. Ayça Gül, Y. D. Ertürk and Penny Elmer. Istanbul: Istanbul University Press Books, pp. 149–61.
Kaşlı, Erhan. 2023. Kolluk Faaliyetlerinde Kişisel Verilerin Korunması: Avrupa Birliği 2016/974 Sayılı Europol Tüzüğü ve 2016/680 Sayılı Kolluk-Ceza Adaleti Direktifi Işığında Bir İnceleme. *Ankara Avrupa Çalışmaları Dergisi* 22: 97–115. [CrossRef]
Kertysova, Katarina. 2018. Artificial Intelligence and Disinformation: How AI Changes the Way Disinformation is Produced, Disseminated, and Can Be Countered. *Security and Human Rights* 29: 55–81. [CrossRef]
König, Pascal D., and Georg Wenzelburger. 2022. Between Technochauvinism and Human-Centrism: Can Algorithms Improve Decision-Making in Democratic Politics? *European Political Science* 2022: 1–18. [CrossRef]
Laterza, Vito. 2021. Could Cambridge Analytica Have Delivered Donald Trump's 2016 Presidential Victory? An Anthropologist's Look at Big Data and Political Campaigning. *Public Anthropologist* 3: 119–47. [CrossRef]
Liebowitz, Jonathan, ed. 2021. *A Research Agenda for Knowledge Management and Analytics*. Cheltenham: Elgar.
Lin, Herbert. 2020. On the Organization of the US Government for Responding to Adversarial Information Warfare and Influence Operations. In *Information Warfare in the Age of Cyber Conflict*. Edited by Christopher Whyte, A. Trevor Thrall and Brian M. Mazanec. New York: Routledge, pp. 166–85.
Llansó, Emma, Joris Van Hoboken, Paula Leersen, and Jaron Harambam. 2020. *Artificial Intelligence, Content Moderation, and Freedom of Expression*. Amsterdam: Transatlantic Working Group on Content Moderation Online and Freedom of Expression.
Lührmann, Anna, and Staffan I. Lindberg. 2019. A Third Wave of Autocratization is Here: What is New about it? *Democratization* 26: 1095–113. [CrossRef]
Lynskey, Orla. 2023. Complete and Effective Data Protection. *Current Legal Problems* 76: 297–344. [CrossRef]
Maaß, Sabrina, Jil Wortelker, and Armin Rott. 2024. Evaluating the Regulation of Social Media: An Empirical Study of the German NetzDG and Facebook. *Telecommunications Policy* 48: 102719. [CrossRef]
Mariyono, Dwi, and Akmal Nur Alif Akmal. 2024. Exploring AI's Role in Supporting Diversity and Inclusion Initiatives in Multicultural Marketplaces. *International Journal of Religion* 5: 5549–82. [CrossRef]
Maroni, Miriam. 2023. 'Mediated Transparency': The Digital Services Act and the legitimisation of platform power. In *(In)visible European Government: Critical Approaches to Transparency as an Ideal and a Practice*. Edited by Päivi Leino-Sandberg, Mikael Zilliacus Hillebrandt and Inka Koivisto. London: Routledge, pp. 305–26.
Matamoros-Fernández, Ariadna. 2023. Taking humor seriously on TikTok. *Social Media+ Society* 9: 20563051231157609. [CrossRef]
Mehta, Somya, and Kristofer Erickson. 2022. Can Online Political Targeting Be Rendered Transparent? Prospects for Campaign Oversight Using the Facebook Ad Library. *Internet Policy Review* 11: 1–31. [CrossRef]
Miró-Llinares, Fernando, and Jesús C. Aguerri. 2023. Misinformation About Fake News: A Systematic Critical Review of Empirical Studies on the Phenomenon and Its Status as a 'Threat'. *European Journal of Criminology* 20: 356–74. [CrossRef]
Moats, David, and Yu-Shan Tseng. 2024. Sorting a Public? Using Quali-Quantitative Methods to Interrogate the Role of Algorithms in Digital Democracy Platforms. *Information, Communication & Society* 27: 973–1007. [CrossRef]
Monti, Maricruz. 2020. The EU Code of Practice on Disinformation and the risk of the privatisation of censorship. In *Democracy and Fake News*. Edited by Sergio Giusti and Elena Piras. London: Routledge, pp. 214–25.
Müller, Jan-Werner. 2017. *What Is Populism?* London: Penguin.
Nemitz, Paul F., and Matthias Pfeffer. 2023. *Human Imperative: Power, Freedom and Democracy in the Age of Artificial Intelligence*. Cambridge: Ethics International Press.
Netshakhuma, Nthabeleng S. 2020. Assessment of a South Africa national consultative workshop on the Protection of Personal Information Act (POPIA). *Global Knowledge, Memory and Communication* 69: 58–74. [CrossRef]
Oymak, Hande. 2020. 7253 Sayılı "İnternet Ortamında Yapılan Yayınların Düzenlenmesi ve Bu Yayınlar Yoluyla İşlenen Suçlarla Mücadele Edilmesi Hakkında Kanunda Değişiklik Yapılmasına Dair Kanun'un Getirdikleri. *Yeni Medya* 9: 129–41.
Qi, Anqi, Guoliang Shao, and William Zheng. 2018. Assessing China's cybersecurity law. *Computer Law & Security Review* 34: 1342–54.

Robinson, Pamela. 2022. Automation in Municipal Public Consultation Processes. In *Artificial Intelligence in the City: Building Civic Engagement and Public Trust*. Edited by Renée Sieber and Eric Champagne. Montreal: Centre for Interdisciplinary Research on Montreal, McGill University, pp. 19–20.

Ruffo, Giancarlo, Alfonso Semeraro, Anastasia Giachanou, and Paolo Rosso. 2023. Studying Fake News Spreading, Polarisation Dynamics, and Manipulation by Bots: A Tale of Networks and Language. *Computer Science Review* 47: 100531. [CrossRef]

Senate Select Committee on Intelligence. 2018. Committee Findings. Available online: https://www.intelligence.senate.gov/publications/committee-findings-2017-intelligence-community-assessment (accessed on 10 September 2024).

Shahbaz, Adrian. 2018. *The Rise of Digital Authoritarianism*. Washington, DC: Freedom House.

Shankar, Ravi, and Taimoor Ahmad. 2021. Information Technology Laws: Mapping the Evolution and Impact of Social Media Regulation in India. *DESIDOC Journal of Library & Information Technology* 41: 295–301. [CrossRef]

Sombra, Thiago Luís. 2020. The General Data Protection Law in Brazil: What Comes Next? *Global Privacy Law Review* 1: 113–30. [CrossRef]

Stanford University. 2017. Full Translation: China's New Generation Artificial Intelligence Development Plan 2017. Available online: https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/ (accessed on 28 April 2024).

Stoycheff, Elizabeth, G. Scott Burgess, and Maria Clara Martucci. 2020. Online Censorship and Digital Surveillance: The Relationship between Suppression Technologies and Democratization across Countries.Information. *Communication & Society* 23: 474–90. [CrossRef]

Tappin, Ben M., David Rand, Joshua Becker, and Jay Van Bavel. 2023. Quantifying the Potential Persuasive Returns to Political Microtargeting. *Proceedings of the National Academy of Sciences* 120: e2216261120. [CrossRef]

Taraktaş, Başak, Kadir Cihan Duran, and Susan Üsküdarlı. 2024. Do Activists Prioritize Solutions over Grievances? A Twitter Study of Black Lives Matter. *Marmara Üniversitesi Siyasal Bilimler Dergisi* 12: 1–22. [CrossRef]

Thanvi, Imran A. 2023. Challenges in Implementation of Personal Data Protection Law No. 45 of 2021: A Case Study of The United Arab Emirates. *Cyber Law Reporter* 2: 1–15.

Trengove, Macey, Ekrem Kazim, David Almeida, Ava Hilliard, Stefano Zannone, and Erik Lomas. 2022. A critical review of the Online Safety Bill. *Patterns* 3: 100544. [CrossRef] [PubMed]

Ulmer, Anna, and Angela Tong. 2023. Deepfaking It: America's 2024 Election Collides with AI Boom. Available online: https://www.reuters.com/world/us/deepfaking-it-americas-2024-election-collides-with-ai-boom-2023-05-30/ (accessed on 28 April 2024).

Verma, Prashant, and Cat Zakrzewski. 2024. AI Deepfakes Threaten to Upend Global Elections: No One Can Stop Them. *The Washington Post*, April 23. Available online: https://www.washingtonpost.com/technology/2024/04/23/ai-deepfake-election-2024-us-india/ (accessed on 28 April 2024).

Weber, Michael. 2021. Reform for Online Political Advertising: Add on to the Honest Ads Act. *Federal Communications Law Journal* 81: 81–110.

Wirkuttis, Nadine, and Hadas Klein. 2017. Artificial Intelligence in Cybersecurity. *Cyber, Intelligence, and Security* 1: 103–19.

Wu, Tim. 2019. Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems. *Columbia Law Review* 119: 2001–28. [CrossRef]

Zurth, Philipp. 2020. The German NetzDG as Role Model or Cautionary Tale? Implications for the Debate on Social Media Liability. *Fordham Intellectual Property, Media and Entertainment Law Journal* 31: 1084–152. [CrossRef]