Ad Americam. Journal of American Studies 25 (2024):
ISSN: 1896-9461, https://doi.org/10.12797/AdAmericam.25.2024.25.10

Helena Jańczuk
College of Europe, Warsaw, Poland
helena.janczuk@coleurope.eu

Risks of Using Artificial Intelligence in Creating the Image of Politicians and in Electoral Campaigns

In the light of the rapid development of advanced technologies in recent years, many questions have been raised about the future application of available technological solutions in various spheres of life, including politics. An important issue that should be discussed in this field concerns the risks associated with the use of artificial intelligence algorithms in creating the public image of politicians and in electoral campaigns. This paper is based on the concept of eroded epistemics, which is a part of Existential Risk Analysis for AI research. Using the AI Safety Research perspectives of monitoring and systemic safety, it examines the potential risks of using AI in politics and ways to minimize them. The analysis is based on the examples of actions of American politicians. Firstly, the threats of using deepfake technology in creating and manipulating the image of politicians such as Nancy Pelosi, Barack Obama, and Donald Trump, are presented. The second part of the paper discusses user profiling and microtargeting strategies and how they may form opinions and influence voters' decisions. Finally, examples of present-day solutions that are being developed to combat these risks are described.

Keywords: Artificial intelligence (AI), X-Risk Analysis, AI Safety Research, deepfake, user profiling, microtargeting, electoral campaigns

1. Introduction

The 21st century is marked by the rapid development of advanced technologies. In recent decades, considerable advances have been made in fields such as the Internet of Things (IoT), robotics, artificial intelligence (AI), 5G networks, as well as augmented and virtual reality (AR and VR). The revolution in these areas is both optimistic – in terms of new possibilities for development and invention – and concerning, when taking into account the probable future applications of available technological solutions in various spheres of life. In the case of AI, the emergence over the past five years of numerous programs trained on language models (LMs), such as BERT, GPT-2, GPT-3, and Switch-C (Bender et al. 610), has provoked discussion about the possible risks and misuse of AI-based systems.

1.1. X-Risk Analysis and AI Safety Research

Apart from technical aspects of artificial intelligence enhancement, research also focuses on AI Risk Analysis, which is a part of a wider concept – Existential Risk (X-Risk) Analysis. As Bucknall and Dori-Hacohen (120) claim, X-Risk Analysis is believed to have started with Nick Bostrom's 2002 article "Existential Risk: Analyzing Human Extinction Scenarios and Related Hazard." The article enumerates machine intelligence and "a badly programmed superintelligence [that] takes over" (Bostrom, "Existential Risks…" 12) as possible causes for human extinction. The publication of several books, such as Bostroms' *Superintelligence…* (2014), Stuart Russell's *Human Compatible…* (2019), Toby Ord's *The Precipice…* (2020), as well as numerous articles (see, e.g., Bucknall and Dori-Hacohen; Hendrycks and Mazeika; Gabriel; Ngo et al.), has made the alignment problem – defined as ensuring that AI is properly aligned with human values (Russell 137) – an important topic of artificial intelligence discourse.

Building upon X-Risk Analysis assumptions, Hendrycks and Mazeika (4) distinguish four areas of AI safety research: robustness, monitoring, alignment, and systemic safety. Research on robustness equips AI systems to endure various challenges, encompassing adversarial scenarios, uncommon situations, and unforeseeable events. Monitoring research aids in detecting potential hazards, including malevolent utilization, concealed model functions, and unforeseen objectives and behaviors. Alignment research strives to mitigate the perilous aspects of AI systems by addressing issues like power-seeking inclinations, deceitfulness, and hazardous objectives. Systemic safety research focuses on minimizing risks at the system level, which encompasses the misuse of AI for malicious purposes and inadequate epistemic capabilities.

Moreover, Hendrycks and Mazeika (5) introduce the concept of eroded epistemics, one of eight so-called "speculative hazards and failure modes," which constitutes a part of AI X-Risk Analysis. This notion assumes that governments, political groups, and various entities employ technology to sway and persuade individuals toward their political convictions, belief systems, and narratives. In such a case, advanced artificial intelligence could enable the orchestration of tailored disinformation campaigns on a large scale. Furthermore, "humanity could have a reduction in rationality due to a deluge of misinformation or highly persuasive, manipulative AI systems" (Hendrycks and Mazeika 5).

One field where the abovementioned areas of AI safety – especially monitoring and systemic safety – and the concept of eroded epistemics seem particularly important to examine is political reality. Electoral campaigns, understood as the self-presentation of candidates as well as activities aimed at winning the support of voters, are particularly susceptible to misuse of AI. Two main threats posed by the use of artificial intelligence algorithms in electoral campaigns are the spread of disinformation and the influence on individual voters' decisions by third parties. An increasingly common practice is the publication of false content, so-called deepfakes, by candidates and the media to manipulate the image of politicians or their opponents running in elections (Somers). Additionally, voters' decisions can be influenced by techniques such as user profiling and microtargeting (Kertysova). Tailoring content to individual audiences results in the production of so-called ideological frames, also known as filter bubbles, which can have a significant impact on voters' worldviews.

1.2. Research Questions

This paper aims to examine the potential dangers of using AI-supported solutions in politics that correspond to the concept of eroded epistemics, with particular emphasis on the image creation of candidates and influencing voters' opinions during electoral campaigns. The main purpose of the work will be to answer the following questions:

- What are the risks associated with the use of artificial intelligence algorithms in creating the public image of politicians and in electoral campaigns?
- What measures are being taken to prevent further aggravation of this problem?

The above questions will be answered with regard to research perspectives proposed by Hendrycks and Mazeika (4), that is, monitoring for risk analysis, and systemic safety for potential solutions. The analysis will be based on already existing artificial intelligence solutions that are being applied or will soon be applied in various areas of life, including electoral campaigns. Firstly, the dangers stemming from using deepfake technology in recreating images of politicians will be discussed. Examples of the problem at hand will be a manipulated video of Nancy Pelosi, an artificially generated image of Barack Obama, a fake picture of Donald Trump posted on Twitter (presently known as X), and two pieces of experimental installation art. Secondly, strategies for influencing voters' decisions – user profiling and microtargeting – as well as the role they might have played during presidential elections in the US will be analyzed. The last part of the paper will present already existing and proposed solutions to the discussed problem. As can be seen, the analysis will be centered around examples of conducting promotional activities for politicians in the United States of America. However, it is crucial to note that the issues discussed in this paper are universal and apply to all centers, state and local, that base their policy-making activities on the latest achievements in the area of artificial intelligence.

1.3. Definitions

The scope of discussion in this paper demands working on different levels and across different branches. For example, deepfake, user profiling, and microtargeting – concepts that will be discussed later in this article – incorporate elements of

various fields: technology, engineering, media studies, psychology, and marketing. Given the complexity of the presented problem, a basic vocabulary is being provided to simplify further analysis.

Artificial intelligence (AI) can be defined as "the design, implementation, and use of programs, machines, and systems that exhibit human intelligence, with its most important activities being knowledge representation, reasoning, and learning" (Whitson). It uses the achievements of computer science to interpret huge data sets and solve problems and may be used in voice recognition, image identification, natural language processing, expert systems, neural networks, planning, robotics, and intelligent agents. Artificial intelligence is also used in the area of machine learning (ML) and deep learning (Whitson).

The term *deepfake* is used to describe various kinds of forged media, in which "an image or recording is convincingly altered and manipulated to misrepresent someone as doing or saying something that was not actually done or said" (Merriam-Webster)[1]. The term was first coined by a Reddit user who published pornographic videos using open-source face-swapping technology on the social media platform (Somers).

User profiling determines the area of interest of users of websites, social networks, etc. This information can be used to improve search results to ensure the satisfaction of the person using the website and to recommend content that best matches the user's preferences and interests (Kanoje et al.). Also related to the concept of user profiling is a marketing strategy known as *microtargeting*. Its goal is to maximize the matching of advertising messages to potential customers. To achieve this, advertisers appeal to information about the personal characteristics and preferences of their audience, such as personality, political views, or sexual orientation (Lorenz-Spreen et al.).

2. Deepfake – Using AI to Manipulate Public Images of Politicians

Deepfake technology is a dangerous tool that can be used to spread disinformation and manipulate images of public figures. According to the concept of eroded epistemics, an AI-based system creating forged media might have the capacity to generate exceptionally compelling arguments that trigger basic human instincts and incite masses, and in effect subvert collective decision-making, leading to the radicalization of individuals, impeding ethical advancement, or eroding the shared understanding of reality (Hendrycks and Mazeika 13). This section exemplifies the potential threats of using AI within the concept of eroded epistemics through three forged audiovisual materials of US politicians – Nancy Pelosi, Barack Obama, and Donald Trump – as well as experimental use of deepfake technology in art.

[1] An example of such a video material is the deepfake created by artists Bill Posters and Daniel Howe, depicting Mark Zuckerberg, the founder of Facebook, speaking optimistically about unrestricted access to the private data of his portal's users (see Cole). Sophie Wilmès, prime minister of Belgium from 2020 to 2022, was also a victim of video manipulation, with a statement linking the COVID-19 outbreak to the climate crisis attributed to her (see Galindo).

A crucial event illustrating the dangers of using deepfake technology in political discourse was the publication of manipulated videos of Nancy Pelosi, then the speaker of the US House of Representatives, by Fox Business in May 2019 (Watson). The forged videos were later posted by US President Donald Trump on his official Twitter (now known as X) account, bearing the comment "Pelosi stammers through media conference" (Trump). Although neither face-swapping nor fake audio footage was used – the recordings were simply slowed down to accentuate the stammers – the incident highlights the potential impact of deepfakes on the reception of politicians' public images. There is no doubt that the purpose of posting the forged footage was to denigrate Pelosi by undermining her mental capacity or suggesting problems with alcohol abuse. Media comments also treated the situation as a sexist attack on the speaker (Watson). If any form of deepfake technology had been used, the consequences could have been even more severe. With current artificial intelligence capabilities, Pelosi's utterances could be altered to show her insulting other politicians or making statements inconsistent with the political party she represents, which could be treated as a lack of integrity or acting against state interest.

Politicians are relatively easy targets for deepfake creators due to the ample availability of audiovisual material. High-quality video and sound, uniform lighting, and placement of figures in the frame, as well as the sheer amount of available footage, make it increasingly easy to create a fake video. In 2017, this was proven by researchers at the University of Washington, whose task was to generate a video depicting President Barack Obama based on a recording of his voice (Suwajanakorn et al.). Crucial to the credibility of the created image was the accurate reproduction of Obama's mouth movements. The researchers used about 17 hours of recordings of his weekly presidential addresses available in the public domain, synthesized by a neural network-based AI algorithm, resulting in a series of images and a video almost impossible for humans to identify as fake, as is best shown below[2]:

Currently, the existing tools and the varied availability of audiovisual material restrict creating deepfakes to well-known individuals, such as politicians, celebrities, and social activists. However, the deepfake creation process is becoming less complex and no longer requires advanced IT skills or professional equipment. Nowadays, virtually anyone can create and publish forged images or videos of public figures, spreading misinformation. An example is a short video, kept in a light tone, in which Barack Obama, or rather his AI-generated model using actor Jordan Peele's voice, utters the words "President Trump is a total and complete dipsh*t" (BuzzFeedVideo). The ability to alter or recreate one's utterance using deepfake technology may soon become easier than ever before (Kertysova). As Suwajanakorn et al. (12) claim: "a single universal network could be trained from videos of many different people, and then conditioned on individual speakers, e.g., by giving it a small video sample of the new person, to produce accurate mouth shapes for that person." Hence, the creation of deepfakes with the help of artificial intelligence algorithms remains an increasingly serious threat.

[2] See Supasorn Suwajanakorn.

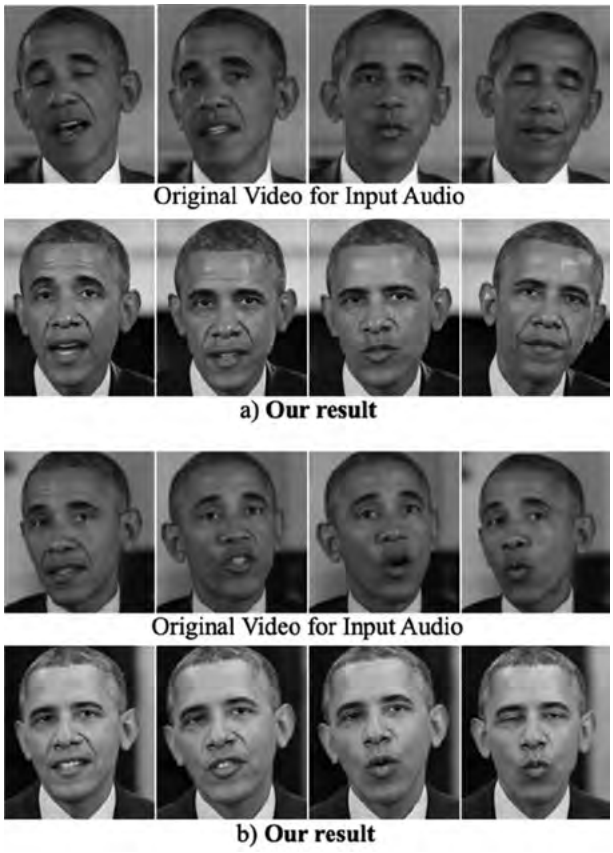

Photo 1: Comparison of the original video (Original Video for Input Audio) with the artificially generated image of Barack Obama (Our result); source: Suwajanakorn et al. 11.

A much easier task, requiring less input and effort than recreating audiovisual material, is to create false images using AI-based systems. There are plenty of programs based on artificial intelligence algorithms, such as Midjourney and DALL-E, that allow anyone to create any image based on a short text command. The impact of this type of deepfakes on public opinion and a politician's image was seen in March/April 2023. Anticipating the indictment of Donald Trump, Eliot Higgins, the founder of the investigative journalism group Bellingcat, used Midjourney to create images of former US president supposedly resisting a violent arrest (Stanley-Becker and Nix). These images were posted on Higgins' Twitter (now known as X):

It is worth noting the popularity of the post. By the end of May 2023, the falsified images were viewed by 6.5 million Twitter (currently X) users, setting the stage for a debate on how easy it is to spread such content and on its potential to confuse news outlets. The episode also highlighted the lack of standards or government regulations regarding the use of artificial intelligence to create and disseminate false information (Stanley-Becker and Nix).

The use of artificial intelligence in many projects of artistic and experimental nature raises many questions about its potential use in various fields, including

politics. *The Infinite Conversation* project created in 2022 by Giacomo Miceli, for instance, uses an AI algorithm to generate an endless conversation between Bavarian director Werner Herzog and Slovenian philosopher Slavoj Žižek based on their real-life statements. As Miceli claims on the official site of the project: "[it] aims to raise awareness about the ease of using tools for synthesizing a real voice (…) [t]his changes our relationship with the media we consume online and raises questions about the importance of authoritative sources, breach of trust and gullibility" (Miceli). Similar questions seem to be posed by the video installation *Sow the Wind, Reap the Whirlwind* created in 2022 by Andrzej Wasilewski. Using deep face fake technology, Wasilewski generated faces of nonexistent people, who communicate by quoting selected philosophical texts by Friedrich Nietzsche, Emil Cioran, and Jean-Paul Sartre. These artistic endeavors not only highlight the advanced state of deepfake technology, but also present its potential applications in widely understood political reality – in political debates, press conferences, and candidate advertising spots. For example, using the concepts and technical solutions of Miceli's *Infinite Conversation* and Wasilewski's *Sow the Wind, Reap the Whirlwind*, one could use AI to create a deepfake debate between Donald Trump and Barack Obama.

Photo 2: Eliot Higgins (@EliotHiggins) "Making pictures of Trump getting arrested while waiting for Trump's arrest," source: Twitter (now known as X), 20 March 2023.

The examples discussed above illustrate the potential dangers of using artificial intelligence algorithms to create the images of politicians, particularly in electoral

campaigns. The main problem with modern AI technologies is the spread of disinformation. It is estimated that by 2030 creating human-identifiable deepfakes will be easier than ever (Kertysova). The eroded epistemics scenario (Hendrycks and Mazeika 5) appears to be coming true, with a real risk that politicians will deliberately use AI-created or falsified audiovisual materials in their campaigns to disinform, disadvantage opponents, and manipulate their own media images. Moreover, the use of these technologies may lead to reduced trust in media and other information sources on which people rely to form judgments and make decisions (Goldstein and Sastry), such as electing candidates for federal or state office. Possibly, the ability of propagandists inside and outside the country to manipulate unsuspecting voters will increase (Goldstein and Sastry). According to the 2019 Worldwide Threat Assessment of the US Intelligence Community warning: "[a]dversaries and strategic competitors [of the US] probably will attempt to use deepfakes or similar machine-learning technologies to create convincing – but false – image, audio, and video files to augment influence campaigns directed against the United States and our allies and partners" (Coats 7). Additionally, Easterly et al. claim that "generative AI will amplify cybersecurity risks and make it easier, faster, and cheaper to flood the country with fake content."

3. User Profiling and Microtargeting – Influencing Voters' Decisions

The concept of eroded epistemics assumes that "AI may (…) enable personally customized disinformation campaigns at scale" (Hendrycks and Mazeika 13). This notion ties into the characteristic elements of user profiling and microtargeting strategies, which involve tailoring content to individual recipients. Using AI algorithms based on machine learning mechanisms, it is possible to create personalized media messages, such as advertisements or social media posts, in a very short time. These messages are designed to engage potential voters with increasing effectiveness, as highlighted by Kertysova. Additionally, present-day content tailoring is based not only on general demographic data, such as age, education, employment, and place of residence, but also on the behavioral data of individuals, such as personality, character traits, beliefs, needs, and weaknesses (Kertysova). Consequently, two people with the same demographic profile can be exposed to different content as they differ in terms of their psychometric profile (Kertysova), which in turn could "undermine collective decision-making [and] radicalize individuals" (Hendrycks and Mazeika 13). This section analyzes the possible application of user profiling and microtargeting strategies during the 2016 presidential election in the US, as well as privacy and data protection concerns it causes.

According to some sources (see, e.g., Kertysova; Anderson), the above-described strategies of user profiling and microtargeting were adopted during the 2016 presidential election in the United States:

> In the run-up to the 2016 US presidential election, presidential candidate Hillary Clinton used demographic segmentation techniques to identify groups of voters. In addition to demographics, Cambridge Analytica – an advertising company contracted to the Trump campaign – also segmented using psychometrics. The company amassed large amounts

> of data, built personality profiles for more than 100 million registered US voters, and then, allegedly, used these profiles for targeted advertising. (Kertysova)

Collecting voters' personal data, both demographic and psychometric (i.e., behavioral), and then using it for tailored promotional campaigns meets the definitions of user profiling and microtargeting. Arguably, it was the use of behavioral data by AI-supported Cambridge Analytica, which allowed for more accurate personalization of election campaigns, that contributed to Donald Trump's 2016 win. Targeted advertising used by the Republican candidate's team explains his victory in the so-called "swing states," which in 2016 included Iowa, Wisconsin, Michigan, Ohio, Pennsylvania, and Florida (Bomboy). Interestingly, these are the same states in which Barack Obama, the Democratic party candidate, won the previous election in 2012 (see: The New York Times, "President Map" and "2016 Presidential Election Results"). As can be seen, electoral campaigns and other political activities that are based on user profiling and microtargeting strategies supported by artificial intelligence algorithms for automated content generation can impact election results (Kertysova). In addition, automated personalization of content results in the production of a so-called filter bubble (Kertysova) – in such a situation, a web user receives only information consistent with their views and is not directed to different topics or points of view. This is a special kind of manipulation of voters' worldviews, which may translate into how they vote and deepen the radicalization of individuals.

Although user profiling and microtargeting can be treated as a form of advertising or marketing strategy, they cause several privacy and data protection concerns. According to Kertysova: "[w]hile users may believe that the encountered information is objective, spontaneous, citizen-generated, and universally encountered by other users, it is algorithms that decide what political views and information users come across online." By relying on the gathering and manipulating of user data to predict and influence voters' political opinions and election outcomes, user profiling and microtargeting can threaten democracy, public debate, and individual choices of people (Kertysova). There is no doubt that the manipulation of such a huge amount of personal data, mainly through user profiling and microtargeting strategies, violates standards and rules developed by some countries and regions. For example, according to Article 22 of the European Union 2016 General Data Protection Regulation (GDPR): "[t]he data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." EU legislation implies that under laws protecting privacy, personal data processed through automated decision-making cannot be used for political purposes (Kertysova). Nevertheless, in the United States, the issue of user profiling and microtargeting is not yet as regulated as in the EU, which makes it easier for candidates to use these strategies in their electoral campaigns. Not until 2021 did Anna Eshoo, a Democratic party member of the US House of Representatives, introduce a bill that aims to ban online platforms from distributing political ads targeting individuals (Library of Congress). However, the bill has not yet passed the legislative process and therefore has no legal force.

4. Present-day Solutions and Future Research

All the examples discussed earlier point to two main risks within the concept of eroded epistemics that stem from using artificial intelligence in political reality, namely the spread of disinformation and manipulation of voters' opinions and decisions. As AI-based programs become more prevalent, there is a need for appropriate strategies to combat and prevent what they can entail. This section will describe already existing as well as proposed solutions for AI safety, which include but are not limited to fact-checking initiatives and systemic strategies.

One of the most popular solutions to the problem of spreading disinformation and manipulation that results from the use of AI algorithms is so-called fact--checking. It is a set of methods that verify the veracity of information that appears online. According to data gathered by Duke Reporters' Lab, there are over 400 projects around the world whose main goal is to verify online content. To speed up the process as well as make identifying, verifying, and correcting social media content much easier, AI-assisted fact-checking is being developed. Organizations such as Full Fact, Duke Reporters' Lab, and Chequeado are working on the development of automated fact-checking (AFC) systems and tools (Kertysova). Artificial intelligence systems also prove useful in identifying illegal, questionable, and undesirable online content and detecting fake bot accounts through techniques known as bot--spotting and bot-labeling (Kertysova).

Other existing solutions that aim to increase control over content published on social media and other websites focus on authentication systems, promoting trusted sources of information, and developing digital literacy skills. To avoid the unwanted spread of AI-generated content, various human authentication tools, such as the well-known CAPTCHA test, are implemented on websites where the public can submit questions to preserve pathways for authentic human requests (Easterly et al.). These tools are still being developed and improved to make sure that Internet users are aware of when content is AI-generated. Easterly et al. claim that "[the tools] used in establishing digital authenticity, such as digital watermarking, could be extremely helpful (…) to distinguish AI-generated content from human--generated content, protect against tampering by demonstrating when content was altered after digital credentials were created, and help the public verify official content." Moreover, politicians and electoral officials are encouraged to follow the principles of transparency and credibility while communicating with local media, community leaders, and constituents to "[solidify] their role as authoritative voices" (Easterly et al.). For example, #TrustedInfo2024 initiative of the National Association of the Secretaries of State aims to "promote election officials as the trusted sources of election information during the 2024 election cycle and beyond" (National Association of Secretaries of State). What is more, Kertysova points to the importance of developing digital literacy skills, especially among election officials, elderly citizens, and marginalized and minority groups. She claims that "[i]ncreasing media and digital literacy may be one of the most efficient and powerful tools to restore a healthy relationship to information and increase the resilience of our democracies to online disinformation" (Kertysova).

Although many solutions already exist to the problem of AI-powered disinformation and manipulation, there is still a lot to be done to further minimize the risks of using artificial intelligence in politics. On some issues, the policies and procedures that would most effectively prevent the risks discussed in this paper are not yet established. For example, there is no consensus on the relationship between political institutions and Internet service providers or technology corporations. For Kertysova, such a relationship should be limited. She calls for the separation of political institutions from Big Tech companies, that is, Google, Amazon, Facebook, Apple, Microsoft, etc. On the other hand, Easterly et al. claim that the private sector, including Internet service providers, cloud service providers, and cybersecurity firms should cooperate with election officials to identify and prevent the risk of AI misuse in electoral campaigns. Moreover, Hendrycks and Mazeika (6) point out that: "research can help identify infeasible solutions or dead ends, or set new directions by identifying new hazards and vulnerabilities." Therefore, in order to safeguard the political sphere from the malicious use of AI and to establish new standards for reducing AI X-risks, further research in the area of AI safety is needed.

5. Conclusion

This paper investigated the potential risks of using AI-supported systems in politics within the concept of eroded epistemics (Hendrycks and Mazeika 5). The analysis was based on the monitoring and systemic safety perspectives of AI Safety research described by Hendrycks and Mazeika (4). Firstly, the problem of manipulating public images of politicians using AI-supported deepfake technology was discussed. The analysis was based on the example of three American politicians – Nancy Pelosi, Barack Obama, and Donald Trump as well as two artistic experiments – Miceli's *Infinite Conversation* and Wasilewski's *Sow the Wind, Reap the Whirlwind*. Then, the discussion turned to the issue of influencing voters' decisions through user profiling and microtargeting strategies and the data protection concerns it causes. To illustrate this problem, the example of the 2016 US presidential elections was used. Finally, the analysis focused on existing solutions to the risks discussed earlier and pointed out areas where regulation is still lacking. This section described already implemented ideas to prevent AI-powered disinformation and manipulation, such as fact-checking systems, as well as introducing authentication systems on websites, promoting trusted sources of information, and developing digital literacy skills. In the end, the direction for further research was suggested.

The examples of the (mis)use of artificial intelligence discussed in this paper – the growing popularity of deepfakes as well as user profiling and microtargeting strategies in the political sphere – are just a part of a larger problem related to the negative impact of advanced technologies on human life. They draw attention to the issue of disinformation, as illustrated by the fabricated images of Pelosi, Obama, and Trump. Moreover, they point to the problem of manipulating voters' opinions and the data protection question. As indicated earlier, the discussed risks associated with the popularization of AI should be considered in a broader context, as they are not exclusively limited to politics. Solutions to the problems discussed here

should be developed in such a universal way that they can prevent AI-assisted disinformation and manipulation in general, not just in a political or electoral context. One must remember that the fight against the risks of using artificial intelligence concerns society as a whole, not just individuals. According to Giacomo Miceli, the author of *The Infinite Conversation* project discussed earlier, "[w]e all share a duty to educate the coming generations about the new paradigm while focusing on forming compassionate individuals who would not misuse these awesome powers."

References

Anderson, Berit. "The Rise of Weaponized AI Propaganda Machine." *Medium*, 13 February 2017, https://medium.com/join-scout/the-rise-of-the-weaponized-ai-propaganda-machine-86dac61668b (10.09.2024).

Bender, Emily M., et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *FAccT'21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York: Association for Computing Machinery, 2021, pp. 610-623, https://doi.org/10.1145/3442188.3445922.

Bomboy, Scott. "What Are the Real Swing States in the 2016 Election?" *National Constitution Center*, 13 June 2016, https://constitutioncenter.org/blog/what-are-the-really-swing--states-in-the-2016-election (10.09.2024).

Bostrom, Nick. "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards." *Journal of Evolution and Technology*, vol 9, no. 1, 2002, pp. 1-36.

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press, 2014.

Bucknall, Benjamin S., and Shiri Dori-Hacohen. "Current and Near-Term AI as a Potential Existential Risk Factor." *AIES'22: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*. New York: Association for Computing Machinery, 2022, pp. 119-129, https://doi.org/10.1145/3514094.3534146.

BuzzFeedVideo. "You Won't Believe What Obama Says In This Video!" *YouTube*, 17 April 2018, https://www.youtube.com/watch?v=cQ54GDm1eL0 (10.09.2024).

Coats, Daniel R. *Statement for the Record: Worldwide Threat Assessment of the US Intelligence Community*. Washington, D.C.: Senate Select Committee on Intelligence, 2019, https://www.dni.gov/files/ODNI/documents/2019-ATA-SFR---SSCI.pdf (25.09.2023).

Cole, Samantha. "This Deepfake of Mark Zuckerberg Tests Facebook's Fake Video Policies." *Vice*, 11 June 2019, https://www.vice.com/en/article/ywyxex/deepfake-of-mark--zuckerberg-facebook-fake-video-policy (10.09.2024).

Duke Reporters' Lab. "Fact-Checking Sites." *Reporters Lab*, https://reporterslab.org/fact-checking/ (25.09.2023).

Easterly, Jen, et al. "Artificial Intelligence's Threat to Democracy." *Foregin Affairs*, 3 January 2024, https://www.foreignaffairs.com/united-states/artificial-intelligences-threat-democracy (10.09.2024).

European Union. "Art. 22 GDPR: Automated Individual Decision-making, Including Profiling." *General Data Protection Regulation (GDPR)*, https://gdpr-info.eu/art-22-gdpr/ (25.09.2023).

Gabriel, Iason. "Artificial Intelligence, Values, and Alignment." *Minds and Machines*, vol. 30, no. 3, 2020, pp. 411-437, https://doi.org/10.1007/s11023-020-09539-2.

Galindo, Gabriela. "XR Belgium Posts Deepfake of Belgian Premier Linking Covid-19 with Climate Crisis." *The Brussels Times*, 14 April 2020, https://www.brusselstimes.com/all-news/belgium-all-news/politics/106320/xr-belgium-posts-deepfake-of-belgian-premier-linking-covid-19-with-climate-crisis (10.09.2024).

Goldstein, Josh A., and Girish Sastry. "The Coming Age of AI-Powered Propaganda." *Foreign Affairs*, 7 April 2023, https://www.foreignaffairs.com/united-states/coming-age-ai--powered-propaganda (10.09.2024).

Hendrycks, Dan, and Mantas Mazeika. "X-Risk Analysis for AI Research." *arXiv*, https://doi.org/10.48550/arXiv.2206.05862.

Higgins, Eliot [EliotHiggins]. "Making pictures of Trump getting arrested while waiting for Trump's arrest." *X (Twitter)*, 20 March 2023, https://twitter.com/EliotHiggins/status/1637927681734987777 (10.09.2024).

Kanoje, Sumitkumar, et al. "User Profiling Trends, Techniques and Applications." *International Journal of Advance Foundation and Research in Computer*, vol. 1, no. 1, 2014, https://doi.org/10.48550/arXiv.1503.07474.

Kertysova, Katarina. "Artificial Intelligence and Disinformation: How AI Changes the Way Disinformation Is Produced, Disseminated, and Can Be Countered." *Security and Human Rights*, vol. 29, 2018, pp. 55-81, https://doi.org/10.1163/18750230-02901005.

Library of Congress. "H.R.4955 – Banning Microtargeted Political Ads Act of 2021." *Congress*, https://www.congress.gov/bill/117th-congress/house-bill/4955 (25.09.2023).

Lorenz-Spreen, Philipp, et al. "Boosting People's Ability to Detect Microtargeted Advertising." *Scientific Reports*, vol. 11, 2021, https://doi.org/10.1038/s41598-021-94796-z.

Merriam-Webster. "Deepfake," https://www.merriam-webster.com/dictionary/deepfake (25.09.2023).

Miceli, Giacomo. *The Infinite Conversation*, https://infiniteconversation.com/ (25.09.2023).

National Association of Secretaries of State. "#TrustedInfo2024." *NASS*, http://www.nass.org/initiatives/trustedinfo (15.01.2024).

Ngo, Richard, et al. "The Alignment Problem from a Deep Learning Perspective." *arXiw*, 2022, https://doi.org/10.48550/arXiv.2209.00626.

Ord, Toby. *The Precipice: Existential Risk and the Future of Humanity*. New York: Hachette Books, 2020.

Russell, Stuart. *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking Books, 2019.

Somers, Meredith. "Deepfakes, Explained." *MIT Sloan School of Management*, 21 July 2020, https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained (10.09.2024).

Stanley-Becker, Isaac, and Naomi Nix. "Fake Images of Trump Arrest Show 'Giant Step' for AI's Disruptive Power." *The Washington Post*, 22 March 2023, https://www.washingtonpost.com/politics/2023/03/22/trump-arrest-deepfakes/ (10.09.2024).

Supasorn Suwajanakorn. "Synthesizing Obama: Learning Lip Sync from Audio." *YouTube*, 12 July 2017, https://www.youtube.com/watch?v=9Yq67CjDqvw (10.09.2024).

Suwajanakorn, Supasorn, et al. "Synthesizing Obama: Learning Lip Sync from Audio." *ACM Transactions on Graphics*, vol. 36, no. 4, 2017, pp. 1-13, https://doi.org/10.1145/3072959.3073640.

The New York Times. "2016 Presidential Election Results," 9 August 2017, https://www.nytimes.com/elections/2016/results/president (10.09.2024).

The New York Times. "President Map," 2012, https://www.nytimes.com/elections/2012/results/president.html?mtrref=www.google.com&gwh=7B66F1AD24AF6781F080C09AD73E0D3A&gwt=pay&assetType=PAYWALL (10.09.2024).

Trump, Donald J. [realDonaldTrump]. "PELOSI STAMMERS THROUGH NEWS CONFERENCE." *X (Twitter)*, 24 May 2019, https://twitter.com/realDonaldTrump/status/1131728912835383300 (10.09.2024).

United States House of Representatives. "Speakers of the House by Congress." *History, Art & Archives*, https://history.house.gov/People/Office/Speakers-List/ (25.09.2023).

Wasilewski, Andrzej. *Kto sieje wiatr, zbiera burzę*. Wrocław: Muzeum Narodowe we Wrocławiu, 26 February 2023–4 June 2023.

Watson, Kathryn. "Trump Tweets Heavily Edited Video of Pelosi Played by Fox Business." *CBS News*, 24 May 2019, https://www.cbsnews.com/news/trump-tweets-heavily-edited-video-of-pelosi-played-by-fox-news/ (10.09.2024).

Whitson, George M. "Artificial Intelligence." *Salem Press Encyclopedia of Science*, 2023, https://searchworks-lb.stanford.edu/articles/ers__89250362 (10.09.2024).