

This PDF is auto-generated for reference only. As such, it may contain some conversion errors and/or missing information. For all formal use please refer to the official version on the website, as linked below.

Artificial Influence: Exploring AI's Impact on Political and Social Realities

<https://www.e-ir.info/2024/06/04/artificial-influence-exploring-ais-impact-on-political-and-social-realities/>

CHRISTOPHER FRIEND, JUN 4 2024

The spread of disinformation campaigns challenges the integrity of public discourse, democratic processes, and social harmony. These campaigns, designed to mislead, manipulate, and sow discord, are increasingly using artificial intelligence (AI) to enhance their effectiveness and reach.^[1] The involvement of AI in disinformation campaigns marks a significant evolution from traditional disinformation campaigns, offering new levels of scalability, customization, and sophistication with lower barriers to entry. Amid the increasing proliferation of AI in asymmetric warfare, election interference, character assassinations, and more, understanding the role of AI in disinformation campaigns is a serious matter of national security.^[2] By reviewing the technical underpinnings of AI-driven disinformation campaigns, we can begin to understand the mechanics of these campaigns, recognize their signs, and develop strategies to counteract their influence. This exploration is crucial for developing defenses against the manipulation of the information environment and safeguarding the foundational principles of an open society.

This paper will provide a basic examination of AI usage in disinformation campaigns, with an emphasis on the mechanisms that facilitate the generation and dissemination of misleading content. These mechanisms include technologies such as natural language processing, deepfakes, and algorithmic content targeting. Following the technical overview, this paper will present a series of plausible and real-world case studies to illustrate the practical application of AI in spreading disinformation across various platforms and contexts. The paper will conclude with policy challenges and opportunities to combat AI-enabled disinformation campaigns.

Understanding AI Enablement in Disinformation Campaigns

The Basics of AI in Disinformation Campaigns

Disinformation campaigns have played a role in conflict, propaganda, and popular culture for centuries.^[3] The meteoric adoption of social media and democratization of the Internet for billions has allowed information – true or false – to be delivered to much of the world's population more easily than ever before.^[4] The digitization of our social networks and communication channels has led to a rise in disinformation, especially as barriers to entry (technical know-how, cost, time) are lowered.^[5] However, recent advancements in AI technology are almost certainly going to contribute to an exponential growth in disinformation across all platforms. When combined with traditional disinformation outlets and methodologies, AI-enabled campaigns have the potential to be much more effective. Large language models, image and video generation, advanced audience targeting, and more provide the means for nefarious actors to increase and improve their disinformation activities many times over.

Technical Mechanisms Behind AI-driven Disinformation

Understanding the technical mechanisms that enable AI-driven disinformation campaigns is key to conceptualizing the anatomy of such campaigns. Below we will cover just a few of the technical mechanisms available to disinformation practitioners.

Large Language Models (LLM) are sophisticated algorithms trained on large quantities of text data. These tools are

Artificial Influence: Exploring AI's Impact on Political and Social Realities

Written by Christopher Friend

adept at understanding and generating text that closely mirrors human communication. Disinformation practitioners can exploit LLMs to generate false information quickly, thereby amplifying their capacity to disseminate false information. Research shows that, when equipped with sufficient computational power and well-crafted prompts, LLMs can create high-quality fake news.^[6] The ability to create believable fake news quickly is revolutionary for disinformation practitioners. The ease-of-use and ubiquity of LLMs lower the barrier to entry for bad actors with limited resources or technical skills to create and spread disinformation.

Deepfakes use sophisticated deep learning techniques to create or manipulate audio and video content with a high degree of realism. This technology has the ability to make it appear as if an individual is saying or doing things that they never actually said or did. While the creation of realistic deepfakes is more financially, computationally, and temporally intensive, it is a tool that is available to almost anyone with a decent computer, a bit of money, and time.^[7] This relative accessibility make deepfakes another powerful tool for disinformation practitioners.

Image generation using advanced AI techniques pose a similar threat as deepfakes. Like deepfakes, disinformation practitioners can leverage AI-enabled tools to create images of fabricated scenarios, people, events, etc. that are difficult to distinguish from reality.^[8] Image generators are increasingly common, with companies like Microsoft, Adobe, Canva, and many more releasing their own tools. In recent months, images created with generative AI have become nearly indistinguishable from photographs, so much so that AI-generated images can be used to train other AI tools.^[9] This step towards hyperrealism allows disinformation practitioners with limited resources to generate images that augment text-based disinformation. Research has shown that visual disinformation is processed differently than text-based disinformation and may be more effective than text-based disinformation alone.^{[10], [11]}

Audience analysis provides information to improve content targeting for a range of applications from benign marketing to nefarious disinformation campaigns. AI-assisted audience analysis allows disinformation practitioners to exploit vast amounts of readily available data to identify specific demographics, preferences, and even susceptibilities among different audience segments. Disinformation practitioners can also use existing marketing models to perform faster and more accurate analysis to determine the most effective audience segments to target.^[12]

Sentiment analysis uses natural language processing techniques and social sciences to determine the sentiment of text. For example, a sentiment analysis tool can estimate whether a collection of social media posts from the target audience about the target topic is positive, neutral, or negative. Given a large enough audience and access to data, disinformation practitioners can use these tools to understand the impact and efficacy of their campaigns.^[13] Using these insights, disinformation practitioners may be able to adjust their campaigns to invoke the desired sentiment.

Disinformation practitioners can combine several of these AI tools and techniques to improve the efficacy of their campaigns.^[14] We will explore this topic in greater detail in chapters below.

The Anatomy of AI-driven Disinformation Campaigns – A Hypothetical

Effective disinformation campaigns have a few fundamental components: a targeted audience, a “voice” or “face” of the campaign to lend credibility, strategically crafted content, methods for distribution, measures of effectiveness, and strategies for improving impact. Research into effective influence operations frameworks supports these six fundamental components (Table 1). AI can augment each of these components. This chapter will present a hypothetical scenario to highlight how a moderately capable disinformation practitioner could exploit AI-enabled tools for their campaign. We will analyze real-world case studies in chapter three.

Let's set the stage for this hypothetical: disinformation practitioners from the fictional country of *Adversaria* do not agree with the political situation in the fictional, English-speaking country of *Friendville*. *Adversaria* has decided to sow discord within *Friendville* amid a tumultuous election by waging a disinformation campaign to discredit *Friendville's* ruling party's policies and governance. *Adversaria* is a mid-sized, non-English-speaking country with moderate technical and financial resources. *Adversarian* disinformation practitioners have decided to use AI to enhance and accelerate their campaign.

Artificial Influence: Exploring AI's Impact on Political and Social Realities

Written by Christopher Friend

Target Identification and Audience Analysis

With the growing amount of data available about groups and individuals on social media, advertising data exchanges, and other areas, even a well-trained practitioner may miss something amongst the volume of information. According to the US Army's Mad Scientist Laboratory, "[Influence operators] must adapt their processes to include the use of AI/ML platforms to augment human instinct and intuition as practitioners conduct target audience analysis to develop effective influence and persuasion products and actions".^[15] It is, therefore, acceptable to assume *Adversarian* disinformation practitioners would do the same.

In our hypothetical, *Adversarian* practitioners have developed advanced AI/ML models capable of scraping and screening hundreds of thousands of *Friendville* social media user profiles. *Adversarian* practitioners were able to build a clustering module using unsupervised learning techniques to put users into groups based on similar characteristics.^[16] After narrowing down the potential target groups to just a few, *Adversarian* practitioners used open-source, AI-enabled sentiment analysis tools on recent user posts to determine which groups expressed slightly positive or neutral opinions about *Friendville*'s ruling party. Since these groups did not have strong opinions for or against *Friendville*'s ruling party, *Adversarian* practitioners determined they were the most important and likely easiest to sway through a comprehensive disinformation campaign.^{[17], [18], [19]}

Crafting a Credible Face of the Campaign

Since *Adversaria* is at odds with *Friendville*, it is reasonable to assume they cannot co-opt an individual deemed trustworthy in *Friendville* to be the face of the disinformation campaign. As a result, *Adversarian* disinformation practitioners decide to take advantage of open-source deepfake technology and generative AI to create numerous fabricated images and videos of a popular *Friendville* celebrity.^{[20], [21]} This fabricated media shows the celebrity doing or saying things that disparage the current government of *Friendville*. Although the fabricated media is not perfect, *Adversarian* practitioners know that the target audience will have a hard time identifying the deepfake as fabricated.^[22] Now *Adversaria* has a seemingly credible conduit through which they can distribute and amplify their disinformation.

Content Creation and Distribution

Next, *Adversarian* disinformation practitioners need to create content with compelling message characteristics.^[23] The content needs to resonate with the target audience and appear to be coming from a legitimate source. There are cultural, linguistic, and other nuances that, as a non-English-speaking country, *Adversaria* may not understand about the English-speaking targets of *Friendville*. Research suggests that AI chatbots may be able to help improve English as a foreign language skills; however, some AI chatbots lack the strong cultural, humor, or empathy functions.^{[24], [25]} Nevertheless, lacking any linguists on their team, *Adversarian* disinformation practitioners use AI-enabled tools like Google Translate and ChatGPT (an LLM) to craft content for their disinformation campaign.

After a week of using ChatGPT and Google Translate to craft politically-focused, disinformation campaign messages, *Adversarian* practitioners are banned from ChatGPT.^[26] As a result, *Adversarian* practitioners decide to use "unrestricted" LLM chatbots like WormGPT, which, among other things, allows users to automate the creation of phishing emails and other harmful content.^[27] Using this tool, *Adversarian* practitioners create a steady flow of false and misleading text-based content which they augment with their deepfakes and fabricated images. Now they are ready to distribute their content.

Adversarian disinformation practitioners decide to build a seemingly real news site that will automatically upload fabricated content, including the deepfakes and LLM-generated content.^[28] Then *Adversarian* practitioners create a number of social media bots to automatically amplify disinformation content.^{[29], [30]} Through a mix of traditional marketing and inorganic engagement, certain target populations begin organically interacting with the false content. Research is varied but suggests that a user's personality may have an impact on their likelihood to organically interact with disinformation on social media.^{[31], [32]} However, now that the content has begun to spread through

Artificial Influence: Exploring AI's Impact on Political and Social Realities

Written by Christopher Friend

organic conduits, it will continue to spread through their sub-networks.^[33]

Feedback Loops and Campaign Optimization

As the disinformation spreads, it is important for *Adversarian* disinformation practitioners to evaluate its effectiveness and optimize their campaign. Similar to the process for identifying *Friendville* target audiences, *Adversarian* practitioners scrape and analyze users' social media posts using sentiment analysis models to determine if sentiment amongst the target population is shifting.^{[34], [35], [36]}

Using the analysis, *Adversarian* practitioners can adjust the campaign to improve the spread, believability, and efficacy of the disinformation. Based on similar tools for digital marketers, there is a potential that *Adversarian* practitioners can use AI to analyze features of past campaigns and offer suggestions to improve new campaigns. These AI-enabled marketing tools have shown to improve engagement metrics compared to campaigns that did not use AI-enabled tools.^[37]

The Potential Impact of this Campaign

Exploring the political and social impact of this hypothetical campaign is out of the scope of this paper. However, numerous studies have shown that disinformation is an effective tool at swaying individuals' and population's perceptions.^{[38], [39]} AI-enabled disinformation campaigns almost certainly have the chance to be just as or more effective than traditional disinformation campaigns.

Had *Adversaria* not used AI throughout its campaign against *Friendville*, it is likely that the financial and temporal cost of the campaign would be much greater. *Adversaria* would have had to hire more linguists and content writers, hire advanced video and audio editors or take time to co-opt a credible *Friendvillian*, hire individuals to manually interact with content, hire data scientists and digital marketers to analyze the efficacy of the campaign, and much more. With AI-enabled tools, small, agile teams can lead disinformation campaigns with limited resources, thereby lowering the barrier to entry for would-be disinformation practitioners.

Case Studies: AI in Action

There are real-world instances of actors using AI-enabled technologies to spread disinformation. Not all of these examples are explicitly nefarious, but each demonstrates the capability and scalability of modern AI-enabled tools to spread information.

Voter Suppression Robocalls – January 2024

In January 2024, New Hampshire residents received a call ostensibly from the treasurer of a local political committee supporting the Democratic party. When residents answered the call, President Joe Biden's voice told residents, "your vote makes a difference in November, not this Tuesday." However, according to the New Hampshire Department of Justice, this was a spoofed, artificially generated call.^[40] Disinformation practitioners likely designed the campaign to reduce primary participation for the Democratic party amongst targeted individuals.

Artificially generated audio is created using a variety of highly-complex machine learning techniques and can very often be difficult to catch.^[41] In this case, the deception was identified relatively quickly and its perpetrators caught up with it.^[42]

Fabricated Video of President Zelensky Surrendering – March 2022

Early during the 2022 Russian invasion of Ukraine, a deepfake video emerged of Ukrainian President Volodymyr Zelenskyy asking his supporters to lay down arms and surrender.^[43] While it is not clear who released the original video, Ukrainian defense officials released a warning the same month about Russian-led disinformation campaigns using deepfakes to sow discord.^[44]

Artificial Influence: Exploring AI's Impact on Political and Social Realities

Written by Christopher Friend

Since this video was released in 2022, deepfake technology has improved and can produce significantly more realistic fabrications. This technology has also been made available to anyone with an Internet connection. LipSynthesis is one such company that allows users to craft “flawless deepfake lip-syncing videos” with realistic results.^{[45], [46]} Sometimes these deepfakes are used for benign purposes, such as recreating actors’ likeness in movies, but experts say that identifying deepfakes are becoming more pervasive and are beginning to threaten elections and, in the case of Ukraine, impact wars.^{[47], [48]}

AI-generated Fake News – November 2023

Shortly after the start of the ongoing Israeli-Gaza conflict, a sensational story about Israeli Prime Minister Netanyahu’s psychiatrist went viral. Although the article is now labeled “satire”, it was shared widely across media platforms, including on Iranian TV.^[49] NewsGuard, an American-operated counter-misinformation company, reported that the source of the article was an AI-generated website. This is an example of a website determined by NewsGuard to be among many Unreliable AI-Generated News Sites (UAINS), which are sites that predominantly publish news generated by AI, with minimal human editorial oversight and without transparently disclosing this practice to their readers.^[50] These websites are not one-offs, either. According to NewsGuard, there were 125 websites that published only or mostly AI-generated content as of May 2023.^[51] These websites are often money-making ventures, using sensational articles to drive engagement and ad revenue. However, disinformation practitioners can develop these websites for a number of purposes, including influence operations.

Artificially-generated Images Spreading on Social Media – March 2024

Disinformation practitioners can use artificially generated images to augment their campaign. These images have a habit of going viral on social media. Researchers from Stanford University identified several dozen social media pages posting almost exclusively AI-generated content and images, sometimes using deceptive practices to hide the source of the content or inorganically grow their follower count.^[52] The same research found that some pages “used clickbait tactics and attempted to direct users to off-platform content farms”. These accounts are incredibly effective, too. One post the researchers analyzed was among the top-20 most popular posts on Facebook in Q3 2023, reaching over 40 million users.

The simplicity of creating AI-generated images provides disinformation practitioners an easy “win” that they can use to augment text-based campaigns. The campaign’s content will sway the target, but the images will bring them to it.

Countering AI-driven Disinformation: Challenges and Strategies

Detecting AI-generated disinformation is core to companies’ and policy makers’ ability to combat its spread. Lucky, the same AI-enabled technologies that can enhance disinformation campaigns can also be used to identify misinformation. Once disinformation can be identified, counter-disinformation practitioners can develop strategies (political, legal, technical, or otherwise) to deter campaigns. In this chapter we will briefly review the technical mechanisms used to identify disinformation, discuss the legal and political opportunities to deter misinformation, and summarize ways the everyday social media user can detect and deter disinformation.

Detecting AI-generated Disinformation with AI

There is an ongoing arm race between disinformation creators and detectors. However, many AI tools used to create fabricated content leave small breadcrumbs that help investigators find and combat their spread. Below is a short summation of potential AI-enabled technical methods to detect and/or counter disinformation:

Multimodal Fake News Analysis Based on Image–Text Similarity ^[53]: Using natural language processing, researchers discovered that there is a higher degree of similarity between an article’s text and its accompanying images in false articles compared to credible articles. This data can be used by others to improve existing disinformation detection models.

Artificial Influence: Exploring AI's Impact on Political and Social Realities

Written by Christopher Friend

Categorization of Accounts Using Unsupervised Machine Learning^[54]: Researchers used machine learning to categorize types of social media bots. According to the research, “having the ability to differentiate between types of bots automatically will allow social media experts to analyze bot activity, from a new perspective, on a more granular level. This way, researchers can identify patterns related to a given bot type’s behaviors over time and determine if certain detection methods are more viable for that type.” This builds the foundation upon which detection methods can be built, and once detected, these bots can be banned or their origins traced.

Towards Effective and Robust Detection of AI-Synthesized Fake Voices^[55]: Using deep neural networks, researchers developed a novel approach to identifying AI-generated fake voices. The approach is also adept at identifying fake voices even when deception techniques, like voice conversion or additive real-world noises, are present. In situations where high-compute costs are appropriate to identify deepfakes, this tool offers an important chance for counter-disinformation practitioners to detect high-visibility or high-impact deepfakes. Due to the compute cost, it is not likely to be appropriate for ubiquitous use across social media, for example.

Sentiment Analysis for Fake News Detection^[56]: Researchers studied the impact of sentiment analysis on disinformation detection and found that there are several different automatic and semi-automatic analysis tools, some of which were enabled by AI. However, the research also identified several challenges with using sentiment analysis to detect disinformation, including addressing biases, managing multilingualism, and explainability of the results.

The battle for superiority between creators and detectors is actively evolving. The initiatives above, and the many dozens more out there, embody the multifaceted approach necessary to combat disinformation effectively. Continued innovation will be necessary for counter-disinformation practitioners to keep pace with those creating and spreading disinformation.

Mitigating Disinformation Through Policy, Education, and Public Awareness

Although technical solutions are important to detecting disinformation, policy, education, and awareness are very likely to be the greatest tools against the widespread adoption and efficacy of disinformation. The meteoric speed at which practitioners have improved disinformation campaigns with AI-enabled tools has not been matched by the speed at which policy is being implemented to address the challenge. However, national-level policies addressing AI ethics and disinformation across the world are beginning to emerge. For example, the US Federal Communications Commission recently banned AI-generated robo calls. This ban “expands the legal avenues through which state law enforcement agencies can hold these perpetrators accountable under the law.”^[57] India also created a helpline to help WhatsApp users fact-check information, an effort to curb disinformation on India’s most popular messaging platform.^[58] Impactfully, US President Joe Biden signed the AI Accountability Act of 2023 which requires companies to assess the impacts of their AI systems, improves transparency about when and how AI systems are used, and empowers consumers to make informed choices with how they interact with AI systems.^[59] The EU has also implemented similar policies, including the EU Artificial Intelligence Act. Policymakers in the US and EU designed both of these major policies to complement or replace self-regulation.^[60] Overall, policy is slowly catching up to the technical realities; however, no policy can account for all eventualities – this is where education and public awareness play a major role in combating the spread of AI-enabled disinformation.

Education and public awareness go hand in hand and must be considered in parallel. There are a number of educational solutions that have been or are in the process of being implemented. One of which is a focus on ethical system design for current and would-be developers of AI systems. Harvard University, among others, offers an ethical intelligence systems design course for undergraduates.^[61] There are also numerous free and low-cost options available to anyone online. For the general public, fake news and disinformation detection courses and tools are being rolled out across social media platforms. For example, Facebook and X (formerly Twitter) each have resources for users to learn about fake news (Facebook’s Help Center and X’s Community Notes). Major social media companies are also announcing new policies and features to warn users about AI-generated content.^[62] More broadly, frameworks like the EU’s ABCDE framework help policy makers and individuals assess potential instances of disinformation with greater understanding of the operative factors.^[63]

Artificial Influence: Exploring AI's Impact on Political and Social Realities

Written by Christopher Friend

Conclusion

It is clear that we are at a critical point in addressing the dual-edged sword AI represents in the digital age. This paper has taken a look at the mechanisms behind AI-enabled disinformation campaigns, showcased through both hypothetical and real-world case studies, highlighting the sophisticated nature of these operations and their potential to disrupt public discourse, democratic processes, and social harmony. The role of AI in creating and spreading disinformation is a growing concern that demands a multifaceted response. The key to combating AI-enabled disinformation lies in a combination of advanced detection mechanisms, informed policies, and increased public awareness and education.

As technology develops and as the world becomes even more interconnected, the challenge of balancing the benefits and risks of AI will grow. Adapting to this challenge requires ongoing collaboration between developers, policymakers, educators, private industry, and the public to ensure AI systems enhance our digital world, not undermine it. The battle against AI-enabled disinformation is complex and evolving, but not insurmountable. With concerted effort and a proactive approach across all fronts, we can stay ahead of nefarious actors.

Proposed Components
RAND Framework
Targeted Audience Directed toward key target audiences; mindful of audience characteristics, including preexisting beliefs and attitudes.
Credible Source Make use of compelling source characteristics whose credibility, trustworthiness, or confidence make them effective spokespersons.
Effective Content Rely upon messages with compelling message characteristics (*i.e.*, those whose content, format, appeal, etc. will most resonate with the audience).
Distribution Methods Make use of the most effective combination of information channels.
Measures of Effectiveness & Improving Impact Facilitate adaptation by providing timely feedback so that components can be modified to increase their persuasiveness. **Table 1.** Proposed fundamental components of disinformation campaigns mapped to the RAND-developed framework for effective influence operations.^[64]

Notes

[1] "AI and the Future of Disinformation Campaigns" December 2021, <https://cset.georgetown.edu/wp-content/uploads/CSET-AI-and-the-Future-of-Disinformation-Campaigns-Part-2.pdf>.

[2] Zvonko Trzun, Dijana Gracin, and Tina Štengl, "ANALYZING THE COST AND NECESSITY OF THE FIGHT AGAINST DISINFORMATION FROM THE NATIONAL SECURITY PERSPECTIVE," n.d.

[3] Julie Posetti and Alice Matthews, "A Short Guide to the History of 'fake News' and Disinformation," n.d.

[4] "Digital Around the World," DataReportal – Global Digital Insights, <https://datareportal.com/global-digital-overview>.

[5] Bertin Martens et al., "The Digital Transformation of News Media and the Rise of Disinformation and Fake News," *SSRN Electronic Journal*, 2018, <https://doi.org/10.2139/ssrn.3164170>.

[6] Yue Huang and Lichao Sun, "Harnessing the Power of ChatGPT in Fake News: An In-Depth Exploration in Generation, Detection and Explanation" (arXiv, October 8, 2023), <http://arxiv.org/abs/2310.05046>.

[7] Timothy B. Lee, "I Created My Own Deepfake—It Took Two Weeks and Cost \$552," *Ars Technica*, December 16, 2019, <https://arstechnica.com/science/2019/12/how-i-created-a-deepfake-of-mark-zuckerberg-and-star-treks-data/>.

[8] Mona Kasra, Cuihua Shen, and James F. O'Brien, "Seeing Is Believing: How People Fail to Identify Fake Images on the Web," in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18: CHI Conference on Human Factors in Computing Systems, Montreal QC Canada: ACM, 2018), 1–6, <https://doi.org/10.1145/3170427.3188604>.

[9] Zuhao Yang et al., "AI-Generated Images as Data Source: The Dawn of Synthetic Era" (arXiv, October 23, 2023), <http://arxiv.org/abs/2310.01830>.

Artificial Influence: Exploring AI's Impact on Political and Social Realities

Written by Christopher Friend

- [10] Teresa Weikmann and Sophie Lecheler, "Visual Disinformation in a Digital Age: A Literature Synthesis and Research Agenda," *New Media and Society*, 25(12), 2022. <https://doi.org/10.1177/14614448221141648>.
- [11] Jiyoung Lee, Michael Hameleers, and Soo Yun Shin, "The Emotional Effects of Multimodal Disinformation: How Multimodality, Issue Relevance, and Anxiety Affect Misperceptions about the Flu Vaccine," *New Media & Society*, April 11, 2023, <https://doi.org/10.1177/14614448231153959>.
- [12] Iman Ahmadi et al., "Overwhelming Targeting Options: Selecting Audience Segments for Online Advertising," *International Journal of Research in Marketing* 41, no. 1 (March 1, 2024): 24–40, <https://doi.org/10.1016/j.ijresmar.2023.08.004>.
- [13] Jochen Hartmann et al., "More than a Feeling: Accuracy and Application of Sentiment Analysis," *International Journal of Research in Marketing* 40, no. 1 (March 2023): 75–87, <https://doi.org/10.1016/j.ijresmar.2022.05.005>.
- [14] Katarina Kertysova, "Artificial Intelligence and Disinformation: How AI Changes the Way Disinformation Is Produced, Disseminated, and Can Be Countered," *Security and Human Rights* 29, no. 1–4 (December 12, 2018): 55–81, <https://doi.org/10.1163/18750230-02901005>.
- [15] "448. Applying Artificial Intelligence/Machine Learning to the Target Audience Analysis Model | Mad Scientist Laboratory," June 8, 2023, <https://madsciblog.tradoc.army.mil/448-applying-artificial-intelligence-machine-learning-to-the-target-audience-analysis-model/>.
- [16] Petar Korda and Pavle Vidanovic, "Machine Learning Techniques for Social Media Analysis", Master thesis submitted to Politecnico di Milano Department of Electronics, Informatics and Bioengineering, 2018 <https://www.politesi.polimi.it/retrieve/a81cb05c-9f7d-616b-e053-1605fe0a889a/Machine-Learning-Techniques-for-Social-Media-Analysis.pdf>.
- [17] Andrea Ceron et al., "Every Tweet Counts? How Sentiment Analysis of Social Media Can Improve Our Knowledge of Citizens' Political Preferences with an Application to Italy and France," *New Media & Society* 16, no. 2 (March 1, 2014): 340–58, <https://doi.org/10.1177/1461444813480466>.
- [18] Saurabh Dorle and Nitin Pise, "Political Sentiment Analysis through Social Media," in *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)*, 2018, 869–73, <https://doi.org/10.1109/ICCMC.2018.8487879>.
- [19] Anthony Fowler et al., "Moderates," *American Political Science Review* 117, no. 2 (May 2023): 643–60, <https://doi.org/10.1017/S0003055422000818>.
- [20] Lee, "I Created My Own Deepfake—It Took Two Weeks and Cost \$552."
- [21] Rachel Winter and Anastasia Salter, "DeepFakes: Uncovering Hardcore Open Source on GitHub," *Porn Studies* 7, no. 4 (October 1, 2020): 382–97, <https://doi.org/10.1080/23268743.2019.1642794>.
- [22] Nils C. Köbis, Barbora Doležalová, and Ivan Soraperra, "Fooled Twice: People Cannot Detect Deepfakes but Think They Can," *iScience* 24, no. 11 (November 2021): 103364, <https://doi.org/10.1016/j.isci.2021.103364>.
- [23] Larson and United States, *Foundations of Effective Influence Operations*.
- [24] Na-Young Kim, "A Study on the Use of Artificial Intelligence Chatbots for Improving English Grammar Skills", *Journal of Digital Convergence*, 2019, Vol 17, Issue 8 <https://openurl.ebsco.com/EPDB%3Agcd%3A13%3A25593197/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Agcd%3A138623173&crl=c>.
- [25] Chunpeng Zhai and Santoso Wibowo, "A Systematic Review on Artificial Intelligence Dialogue Systems for

Artificial Influence: Exploring AI's Impact on Political and Social Realities

Written by Christopher Friend

Enhancing English as Foreign Language Students' Interactional Competence in the University," *Computers and Education: Artificial Intelligence* 4 (January 1, 2023): 100134, <https://doi.org/10.1016/j.caeai.2023.100134>.

[26] "How ChatGPT Maker OpenAI Plans to Deter Election Misinformation in 2024 | AP News," <https://apnews.com/article/ai-election-misinformation-voting-chatgpt-altman-openai-0e6b22568e90733ae1f89a0d54d64139>.

[27] "WormGPT – The Generative AI Tool Cybercriminals Are Using to Launch BEC Attacks | SlashNext," July 13, 2023, <https://slashnext.com/blog/wormgpt-the-generative-ai-tool-cybercriminals-are-using-to-launch-business-email-compromise-attacks/>.

[28] Robin Guess, "Analysts Warn of Spread of AI-Generated News Sites," Voice of America, February 21, 2024, <https://www.voanews.com/a/analysts-warn-of-spread-of-ai-generated-news-sites-7497011.html>.

[29] "How to Create Mass Social Media Bots," AskHandle, August 14, 2023, <https://www.askhandle.com/blog/how-to-create-mass-social-media-bots>.

[30] Chengcheng Shao et al., "The Spread of Low-Credibility Content by Social Bots," *Nature Communications* 9, no. 1 (November 20, 2018): 4787, <https://doi.org/10.1038/s41467-018-06930-7>.

[31] Tom Buchanan, "Trust, Personality, and Belief as Determinants of the Organic Reach of Political Disinformation on Social Media," *The Social Science Journal* 0, no. 0 (2021): 1–12, <https://doi.org/10.1080/03623319.2021.1975085>.

[32] Tom Buchanan and Vladlena Benson, "Spreading Disinformation on Facebook: Do Trust in Message Source, Risk Propensity, or Personality Affect the Organic Reach of 'Fake News'?", *Social Media + Society* 5, no. 4 (October 1, 2019): 2056305119888654, <https://doi.org/10.1177/2056305119888654>.

[33] Eni Mustafaraj and Panagiotis Takis Metaxas, "The Fake News Spreading Plague: Was It Preventable?," in *Proceedings of the 2017 ACM on Web Science Conference*, WebSci '17 (New York, NY, USA: Association for Computing Machinery, 2017), 235–39, <https://doi.org/10.1145/3091478.3091523>.

[34] Demitrios E. Pournarakis, Dionisios N. Sotiropoulos, and George M. Giaglis, "A Computational Model for Mining Consumer Perceptions in Social Media," *Decision Support Systems* 93 (January 1, 2017): 98–110, <https://doi.org/10.1016/j.dss.2016.09.018>.

[35] Aobo Yue et al., "Detecting Changes in Perceptions towards Smart City on Chinese Social Media: A Text Mining and Sentiment Analysis," *Buildings* 12, no. 8 (August 2022): 1182, <https://doi.org/10.3390/buildings12081182>.

[36] Yuguo Tao et al., "Social Media Data-Based Sentiment Analysis of Tourists' Air Quality Perceptions," *Sustainability* 11, no. 18 (January 2019): 5070, <https://doi.org/10.3390/su11185070>.

[37] Moumita Sinha, Jennifer Healey, and Tathagata Sengupta, "Designing with AI for Digital Marketing," in *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '20 Adjunct (New York, NY, USA: Association for Computing Machinery, 2020), 65–70, <https://doi.org/10.1145/3386392.3397600>.

[38] Salvatore Vilella et al., "The Impact of Disinformation on a Controversial Debate on Social Media," *EPJ Data Science* 11, no. 1 (December 2022): 29, <https://doi.org/10.1140/epjds/s13688-022-00342-w>.

[39] Dorje C. Brody, "Modelling Election Dynamics and the Impact of Disinformation," *Information Geometry* 2, no. 2 (December 1, 2019): 209–30, <https://doi.org/10.1007/s41884-019-00021-2>.

[40] "Voter Suppression Robocall Complaint to Election Law Unit | News Releases | NH Department of Justice,"

Artificial Influence: Exploring AI's Impact on Political and Social Realities

Written by Christopher Friend

<https://www.doj.nh.gov/news/2024/20240122-voter-robocall.html>.

[41] Zahra Khanjani, Gabrielle Watson, and Vandana P. Janeja, "How Deep Are the Fakes? Focusing on Audio Deepfake: A Survey" (arXiv, November 28, 2021), <http://arxiv.org/abs/2111.14203>.

[42] "How Investigators Solved the Biden Deepfake Robocall Mystery," *Bloomberg.Com*, February 7, 2024, <https://www.bloomberg.com/news/newsletters/2024-02-07/how-investigators-solved-the-biden-deepfake-robocall-mystery>.

[43] *Deepfake Video of Volodymyr Zelensky Surrendering Surfaces on Social Media*, 2022, <https://www.youtube.com/watch?v=X17yrEV5sl4>.

[44] Defence intelligence of Ukraine [@DI_Ukraine], "Всі ви, напевно, чули про технологію Діпфейк (англ. deepfake; поєднання слів deep learning («глибинне навчання») та fake («підробка») — методика синтезу зображення людини, яка базується на штучному інтелекті. Готується провокація РФ. <https://t.co/XYyS9WsPkK>," Tweet, *Twitter*, March 2, 2022, https://twitter.com/DI_Ukraine/status/1499157365937119235.

[45] "LipSynthesis 2023 | Create Ultra Realistic Lip Syncing Videos," <https://lipsynthesis.com/>.

[46] "Anderson Cooper, 4K Original/(Deep)Fake Example – YouTube," <https://www.youtube.com/watch?v=3wVpVH0Wa6E>.

[47] "How 'Rogue One' Incorporated a Dead Actor into the Cast – Cleveland.Com," https://www.cleveland.com/entertainment/2016/12/how_rogue_one_incorporated_a_d.html.

[48] *Can You Spot the Deepfake? How AI Is Threatening Elections*, 2024, <https://www.youtube.com/watch?v=B4jNttRvbpU>.

[49] News Desk, "Israeli Prime Minister's Psychiatrist Commits Suicide: Satire," *Global Village Space*, November 6, 2023, <https://www.globalvillagespace.com/israeli-prime-ministers-psychiatrist-commits-suicide/>.

[50] "AI-Generated Site Sparks Viral Hoax Claiming the Suicide of Netanyahu's Purported Psychiatrist," *NewsGuard* (blog), <https://www.newsguardtech.com/special-reports/ai-generated-site-sparks-viral-hoax-claiming-the-suicide-of-netanyahus-purported-psychiatrist>.

[51] "NewsGuard Now Identifies 125 News and Information Websites Generated by AI, Develops Framework for Defining 'Unreliable AI-Generated News' and Information Sources," *NewsGuard* (blog), <https://www.newsguardtech.com/press/newsguard-now-identifies-125-news-and-information-websites-generated-by-ai-develops-framework-for-defining-unreliable-ai-generated-news-and-information-sources/>.

[52] Stanford University, Stanford, and California 94305, "How Spammers, Scammers and Creators Leverage AI-Generated Images On," March 18, 2024, <https://cyber.fsi.stanford.edu/io/news/ai-spam-accounts-build-followers>.

[53] Xichen Zhang et al., "Multimodal Fake News Analysis Based on Image-Text Similarity," *IEEE Transactions on Computational Social Systems* 11, no. 1 (February 2024): 959–72, <https://doi.org/10.1109/TCSS.2023.3244068>.

[54] "Types of Bots: Categorization of Accounts Using Unsupervised Machine Learning – ProQuest," accessed April 7, 2024, <https://www.proquest.com/openview/918d2cc36fdbb0609eb7d7f0cf87dbfc/1?pq-origsite=gscholar&cbl=18750&diss=y>.

[55] Run Wang et al., "DeepSonar: Towards Effective and Robust Detection of AI-Synthesized Fake Voices," in *Proceedings of the 28th ACM International Conference on Multimedia*, MM '20 (New York, NY, USA: Association for Computing Machinery, 2020), 1207–16, <https://doi.org/10.1145/3394171.3413716>.

Artificial Influence: Exploring AI's Impact on Political and Social Realities

Written by Christopher Friend

[56] Miguel A. Alonso et al., "Sentiment Analysis for Fake News Detection," *Electronics* 10, no. 11 (January 2021): 1348, <https://doi.org/10.3390/electronics10111348>.

[57] "US FCC Makes AI-Generated Robocalls Illegal," February 8, 2024, <https://www.bbc.com/news/world-us-canada-68240887>.

[58] "Curbing Misinformation in India: How Does a Fact-Checking WhatsApp Helpline Work?," World Economic Forum, March 5, 2024, <https://www.weforum.org/agenda/2024/03/ai-deepfake-helpline-india/>.

[59] Wyden, Rob "Algorithmic Accountability Act of 2023 Summary" https://www.wyden.senate.gov/imo/media/doc/algorithmic_accountability_act_of_2023_summary.pdf.

[60] Jakob Mökander et al., "The US Algorithmic Accountability Act of 2022 vs. The EU Artificial Intelligence Act: What Can They Learn from Each Other?," *Minds and Machines* 32, no. 4 (December 1, 2022): 751–58, <https://doi.org/10.1007/s11023-022-09612-y>. <https://rdcu.be/dJynW>

[61] "CS 108: Intelligent Systems: Design and Ethical Challenges," accessed April 7, 2024, <https://projects.iq.harvard.edu/cs108/home>.

[62] Reuters, "Facebook and Instagram to Label Digitally Altered Content 'Made with AI,'" *The Guardian*, April 5, 2024, sec. Technology, <https://www.theguardian.com/technology/2024/apr/05/facebook-instagram-ai-label-digitally-altered-media>.

[63] James Pamment, "The ABCDE Framework," *The EU's Role in Fighting Disinformation: (Carnegie Endowment for International Peace, 2020)*, <https://www.jstor.org/stable/resrep26180.6>.

[64] Eric V. Larson and United States, eds., *Foundations of Effective Influence Operations: A Framework for Enhancing Army Capabilities*, Rand Corporation Monograph Series (Santa Monica, CA: Rand Arroyo Center, 2009).