edam
Centre for Economics
and Foreign Policy Studies

CYBER GOVERNANCE & DIGITAL DEMOCRACY PROGRAMME

AUGUST 2023

THE ROLE OF TECHNOLOGY: NEW METHODS OF INFORMATION MANIPULATION AND DISINFORMATION

Akın Ünver, Associate Professor, Ozyegin University

INTRODUCTION

The contemporary technological landscape has witnessed unprecedented advancements that have revolutionized the way information is created, consumed, and manipulated. Emerging technologies have paved the way for sophisticated disinformation campaigns, where AI-generated content can blur the lines between fact and fabrication, leaving unsuspecting audiences vulnerable to deception. A lot has happened on this front since the well-known Cambridge Analytica scandal and cases of election interference in Western democracies, as technologies designed to sow discord and confuse populations through digital communication have advanced rapidly. Recent years have witnessed an ever-growing repertoire of technologies used in information manipulation and disruption beyond the most frequently studied countries in the West, owing to the advances in Artificial Intelligence (A.I.), Generative Adversarial Networks (GAN) and deep fakes.

For example, the 2022 national election in Brazil was undermined by the malicious use of AI-generated deepfakes, where fabricated images and videos were used to spread misinformation, showing leading candidates involved in various scandalous and compromising situations.[1] This deception was masterfully orchestrated by leveraging advanced AI and sophisticated editing tools to mimic the candidates' voices and facial expressions accurately, creating a result that was highly convincing, yet belonged to an entirely fictitious ecosystem of narratives that destabilized public trust and skewed the electoral environment. In Myanmar, military personnel led a systematic campaign on Facebook to spread hate speech and fake news against the Rohingya Muslim minority.[2] The campaign involved setting up seemingly innocuous lifestyle pages that slowly began posting anti-Rohingya content, capitalizing on algorithmic amplification to reach a wider audience. This disinformation contributed to an atmosphere of hostility that culminated in acts of ethnic cleansing. As a result, Facebook faced international criticism, leading to more substantial efforts to combat hate speech and misinformation on the platform.

In India, a country rich in languages and dialects, the capabilities of Natural Language Processing (NLP) were deployed for manipulating political debate at scale. Through 2019-20, competing AI systems were specifically programmed to generate disinformation predominantly on Facebook, tailored to different regions, taking into account region-specific cultural and linguistic tones.[3] The AI used local dialects and intricately woven culturally nuanced narratives that perfectly fit the regional context. This tactic was disconcertingly effective, playing into regional biases and presenting the fabricated news in a manner that seemed more authentic and relatable to the local communities. In a more insidious example in

[1] Vasconcellos, Paulo Henrique Santos, Pedro Diógenes de Almeida Lara, and Humberto Torres Marques-Neto. "Analyzing Polarization And Toxicity On Political Debate In Brazilian TikTok Videos Transcriptions." In Proceedings of the 15th ACM Web Science Conference 2023, pp. 33-42. 2023.

[2] Mathew, Binny, Ritam Dutt, Pawan Goyal, and Animesh Mukherjee. "Spread of hate speech in online social media." In Proceedings of the 10th ACM conference on web science, pp. 173-182. 2019.

[3] Bali, Aasita, and Prathik Desai. "Fake news and social media: Indian perspective." Media Watch 10, no. 3 (2019): 737-750.

Russia, AI-based information manipulation systems were leveraged to manipulate public sentiment in Ukraine on a broad scale.[4] An extensive disinformation campaign was brought to light in 2022, revealing that bots, powered by NLP, were impersonating real users across a variety of social media platforms. These AI entities participated in public discussions, stirred up conflict, and propagated counterfeit news stories, intensifying political divisions and contributing to social unrest in the run-up to the second Russian invasion of Ukraine.

Since President Rodrigo Duterte assumed office in the Philippines on June 30, 2016, there has been a marked increase in state-sponsored disinformation campaigns, particularly on the pervasive social media platform, Facebook.[5] The government has purportedly mobilized 'troll armies', an alarming strategy wherein large groups of individuals or automated bots are deployed to systematically spread propaganda. Their primary objectives appear to encompass the manipulation of public sentiment to sway opinion in favor of the administration, to discredit and harass the opposition, and to divert attention from controversial issues. These disinformation campaigns are neither haphazard nor spontaneous. They involve targeted and calculated dissemination of content based on intricate data profiling. Their focus is generally on critics of the administration and influential opposition figures, including human rights activist, Maria Ressa, who has faced online abuse and legal harassment since her news organization, Rappler, started reporting on the disinformation tactics and alleged extrajudicial killings under Duterte's administration.[6]

Simultaneously, across the world in Australia, amidst the catastrophic bushfires that occurred from late 2019 to early 2020, a surge of misinformation was disseminated across social media platforms.[7] This misinformation inaccurately asserted that the bushfires were predominantly set by arsonists, despite official reports stating that most fires were caused by lightning and only around 1% were intentionally set. This wave of false information was mainly propagated by bots and trolls, which strategically amplified the misinformation based on big-data profiling to target susceptible demographics, including climate change skeptics and anti-government groups. Moreover, in Sweden, in the summer of 2023, the Ministry of Defense was inundated with a large-scale, automated information manipulation operation. This used deepfake technology – sophisticated artificial intelligence that can create hyper-realistic fake images or videos – to attribute a series of Quran burnings that occurred in June and July 2023, to the Swedish government. Overwhelmed by the enormity of this disinformation campaign, the Swedish Ministry of Defense was compelled to launch an organized counter-campaign to live fact-check and refute these accusations.[8] They adamantly denied government

[4] Stukal, Denis, Sergey Sanovich, Joshua A. Tucker, and Richard Bonneau. "For whom the bot tolls: A neural networks approach to measuring political orientation of Twitter bots in Russia." Sage Open 9, no. 2 (2019): 2158244019827715.

[5] Ong, Jonathan Corpus, and Ross Tapsell. "Demystifying disinformation shadow economies: fake news work models in Indonesia and the Philippines." Asian Journal of Communication 32, no. 3 (2022): 251-267.

[6] Tandoc, Edson C., Karryl Kim Sagun, and Katrina Paola Alvarez. "The digitization of harassment: Women journalists' experiences with online harassment in the Philippines." Journalism Practice 17, no. 6 (2023): 1198-1213.

[7] Weber, Derek, Lucia Falzon, Lewis Mitchell, and Mehwish Nasim. "Promoting and countering misinformation during Australia's 2019–2020 bushfires: a case study of polarisation." Social Network Analysis and Mining 12, no. 1 (2022): 64.

[8] Karlidag, Ilgin. 2023. "Sweden's Quran Burnings Put Freedom of Expression Law to Test." BBC News, July 27, 2023. https://www.bbc.com/news/world-europe-66310285.

involvement in these provocations, underscoring the sophistication and potential harm of such manipulative digital tactics.

However, while technology has fueled the rise of disinformation, it also holds the key to potential solutions. AI and NLP, once harnessed by disinformation creators, can now be wielded as powerful tools for detection. Machine learning and data analytics empower researchers and fact-checkers with the ability to sift through vast amounts of information, uncovering patterns that reveal traces of disinformation. For instance, the emergence of GPT-powered AI fact-checkers showcased how AI can be used to rapidly cross-reference claims against reliable sources and debunk falsehoods in real time.[9]

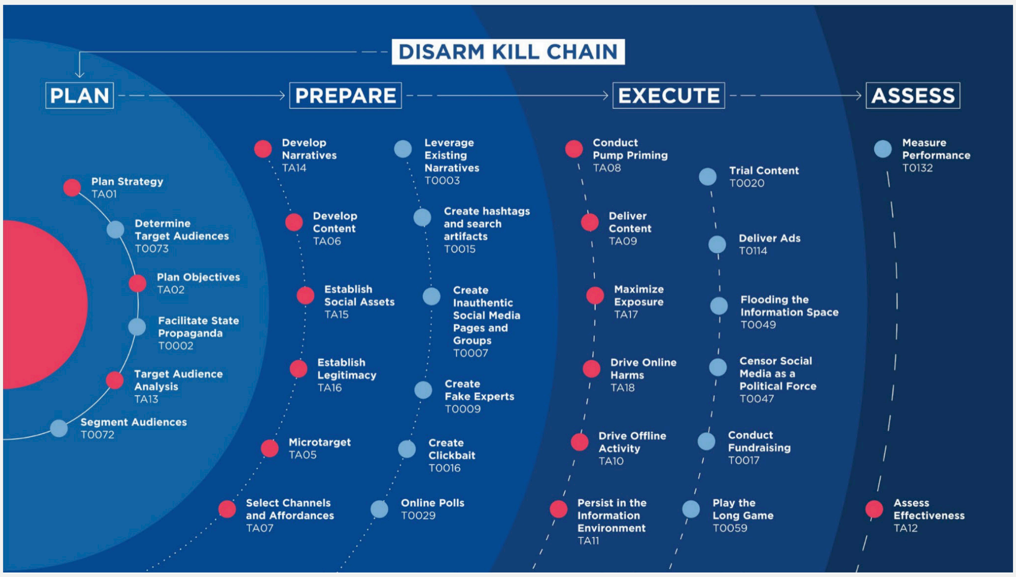
The transformative potential of blockchain technology, for example, also presents a promising path towards tamper-resistant information dissemination, safeguarding content from unauthorized alterations and ensuring its provenance. The tamper-proof nature of blockchain ensures that the data presented is authentic, helping to counter misinformation and enhance trust in online information sources. By analyzing networks and user behavior, researchers can reverse-engineer orchestrated disinformation campaigns and the role of automated bot accounts, enabling the development of targeted counter-strategies. In the aftermath of the well-known 2016 Brexit referendum, for example, researchers identified an extensive network of Twitter bots that disseminated misleading information and manipulated public sentiment. The exposure of this orchestrated campaign later proved to be instrumental in understanding the mechanics behind disinformation propagation and devising strategies to disrupt such coordinated efforts.[10]

[9] Banas, John A., Nicholas A. Palomares, Adam S. Richards, David M. Keating, Nick Joyce, and Stephen A. Rains. “When machine and bandwagon heuristics compete: Understanding users’ response to conflicting AI and crowdsourced fact-checking.” Human Communication Research 48, no. 3 (2022): 430-461.

[10] Nicoli, Nicholas, Soulla Louca, and Petros Iosifidis. “Social Media, News Media, and the Democratic Deficit. Can the Blockchain Make a Difference?.” tripleC: Communication, Capitalism and Critique 20, no. 2 (2022).

DISINFORMATION, INFORMATION SUPPRESSION and FOREIGN INTERFERENCE

The phenomena of disinformation, misinformation, malinformation, information suppression, and foreign information manipulation and interference (FIMI) present complex challenges to societies and governments worldwide. With their distinct definitions and overlapping impacts, terms such as propaganda, disinformation and foreign interference require in-depth understanding through real-world examples, as they are often confused, or used interchangeably.


DISARM KILL CHAIN
PLAN
PREPARE
EXECUTE
ASSESS
Plan Strategy TA01
Determine Target Audiences T0073
Plan Objectives TA02
Facilitate State Propaganda T0002
Target Audience Analysis TA13
Segment Audiences T0072
Develop Narratives TA14
Develop Content TA06
Establish Social Assets TA15
Establish Legitimacy TA16
Microtarget TA05
Select Channels and Affordances TA07
Leverage Existing Narratives T0003
Create hashtags and search artifacts T0015
Create Inauthentic Social Media Pages and Groups T0007
Create Fake Experts T0009
Create Clickbait T0016
Online Polls T0029
Conduct Pump Priming TA08
Deliver Content TA09
Maximize Exposure TA17
Drive Online Harms TA18
Drive Offline Activity TA10
Persist in the Information Environment TA11
Trial Content T0020
Deliver Ads T0114
Flooding the Information Space T0049
Censor Social Media as a Political Force T0047
Conduct Fundraising T0017
Play the Long Game T0059
Measure Performance T0132
Assess Effectiveness TA12

European External Action Service scenario depicting a blue team – red team scenario of information manipulation. Source: 1st EEAS Report on Foreign Information Manipulation and Interference Threats Towards a framework for networked defence. February 2023
https://euvsdisinfo.eu/uploads/2023/02/EEAS-ThreatReport-February2023-02.pdf

The oldest of these terms, propaganda, has been extensively defined as “the management of collective attitudes by the manipulation of significant symbols.”[11], “a consistent, enduring effort to create or shape events to influence the relations of the public to an enterprise, idea or group”[12], “An expression of opinion or action by individuals or groups deliberately designed to influence opinions or actions of other individuals or groups with reference to predetermined ends”[13], and “A process which deliberately attempts through persuasion-techniques to secure from the propagandee, before he can deliberate freely, the responses desired by the propagandist”.[14]

Let’s consider the expansive propaganda machinery utilized by the Chinese Communist Party (CCP), employed both domestically and internationally to control

[11] Lasswell, Harold D. “The theory of political propaganda.” American Political Science Review 21, no. 3 (1927): 627-631.
[12] As cited in: Hobbs, Renee, and Sandra McGee. “Teaching about propaganda: An examination of the historical roots of media literacy.” Journal of Media Literacy Education 6, no. 2 (2014): 56-66.
[13] Cantril, Hadley. “Propaganda analysis.” The English Journal 27, no. 3 (1938): 217-221.
[14] Henderson, Edgar H. “Toward a definition of propaganda.” The Journal of Social Psychology 18, no. 1 (1943): 71-87.

narratives and shape public opinion. Internally, the CCP maintains firm control over all state media outlets, facilitating the propagation of party narratives and censoring dissent. The "Chinese Dream," a narrative encompassing national rejuvenation, social progress, and improved living standards, is an exemplar designed to promote national pride and support for the government's agenda. Externally, China has been asserting its influence over global media to modify its international image. A striking example of this is the China Global Television Network (CGTN), a multi-lingual global broadcaster delivering global news from a Chinese viewpoint. Furthermore, the Chinese government invests heavily in "Confucius Institutes" worldwide, serving as both cultural organizations and tools for projecting soft power and promoting a positive image of China.[15]

Through such comprehensive use of propaganda, the CCP secures control domestically while also seeking to sway perceptions internationally, thereby highlighting the pervasive role of propaganda in shaping public and political landscapes. The contemporary North Korean regime also epitomizes the utilization of propaganda, with its government exercising strict control over all information within the country. This control shapes public perceptions of the external world, the country's leadership, and the prevailing political ideology.[16]

Disinformation, on the other hand, is often associated with short-term gain. It is a term that refers to deliberately misleading or biased information, manipulated narratives or facts, or false information that is spread with the specific intent to deceive, mislead, or confuse.[17] It usually involves the deliberate creation and sharing of false or manipulated information with the intent to cause harm, mislead the recipient, or create false perceptions about a person, organization, or a country. Disinformation can come in many forms, such as fake news, deepfakes, or misleading narratives. The post-2014 conflict between Russia and Ukraine provides a modern instance of disinformation. Here, both the Russian state and its proxies extensively used disinformation to mask the reality of the situation, delegitimize Ukraine's government, and justify Russia's actions to the global community.[18]

Disinformation is not limited to the political domain and can manifest in non-political debates, such as health-related inaccuracies, as well. For instance, during the COVID-19 pandemic, most nations faced hurdles in combating health misinformation. The rampant spread of rumors, misconceptions, and falsehoods about the virus and its treatment across social media platforms led to panic, the propagation of unscientific remedies, and an undermining of public health responses.[19]

[15] Zhu, Yanling. "China's 'new cultural diplomacy'in international broadcasting: branding the nation through CGTN Documentary." International Journal of Cultural Policy 28, no. 6 (2022): 671-683.

[16] Sukin, Lauren. "Why "cheap" threats are meaningful: Threat perception and resolve in North Korean propaganda." International Interactions 48, no. 5 (2022): 936-967.

[17] Kapantai, Eleni, Androniki Christopoulou, Christos Berberidis, and Vassilios Peristeras. "A systematic literature review on disinformation: Toward a unified taxonomical framework." New media & society 23, no. 5 (2021): 1301-1326.

[18] Erlich, Aaron, and Calvin Garner. "Is pro-Kremlin disinformation effective? Evidence from Ukraine." The International Journal of Press/Politics 28, no. 1 (2023): 5-28.

[19] Grimes, David Robert. "Medical disinformation and the unviable nature of COVID-19 conspiracy theories." PLoS One 16, no. 3 (2021): e0245900.

In grappling with the multifaceted and unique landscape of information manipulation by foreign or state-sponsored entities, the concept of Foreign Information Manipulation and Interference (FIMI) has begun to grow more popular.[20] This term encompasses a myriad of complex phenomena, the understanding of which requires a closer examination of its key characteristics and their manifestation in real-world scenarios. The defining feature of FIMI is the participation of foreign entities or governments, setting it apart from conventional forms of disinformation or misinformation that might originate within a nation's borders. A salient example of this can be seen in Australia during the 2017 power grid crisis, where reports surfaced about alleged Chinese interference in Australia's power grid.[21] Accusations were directed towards Chinese entities for spreading disinformation on social media platforms to create panic and confusion, ultimately aiming to undermine public trust in the Australian government's ability to maintain critical infrastructure. The alleged orchestrated campaign involved spreading false narratives about the extent and cause of the power grid failure, as well as exaggerating the government's inability to handle the situation. This incident presented a clear case of foreign involvement in the manipulation of information, highlighting the impact of FIMI on national security and public confidence. Equally critical to FIMI is its underpinning by geopolitical motives. It is often employed as a tool to achieve broader geopolitical aims, such as influencing elections, fomenting discord, destabilizing societies, or weakening democratic institutions in target countries. This aspect was apparent in the 2021 Ugandan elections, where foreign-based digital campaigns sought to influence public sentiment and stir up social divisions.[22]

Unlike traditional disinformation and misinformation campaigns which primarily aim to influence public opinion, FIMI specifically targets the governance and political processes of foreign countries. Moreover, FIMI campaigns are generally part of broader strategic efforts coordinated to fulfill foreign policy objectives. They often involve a mix of disinformation, cyberattacks, social media manipulation, and other psychological tactics that are reinforced through digital communication channels and advanced technologies. The suspected foreign influence campaigns in Australia illustrate this well, where 'patriotic trolling' – an approach involving the inundation of social media platforms with pro-government messages and harassment of individuals critical of the state – seemed to be part of a coordinated strategy to manipulate public discourse and perception.

[20] 1st EEAS Report on Foreign Information Manipulation and Interference Threats: Towards a framework for networked defence, European External Action Service (EEAS), viewed 24 July 2023, <https:// www.eeas.europa.eu/sites/default/files/documents/2023/EEAS-DataTeam-ThreatReport-2023. pdf>.

[21] "China Accuses Australia and Canada of 'disinformation' over Jet Encounters." The Guardian, 7 June 2022, www.theguardian.com/world/2022/jun/07/china-accuses-australia-canada-jet-encounters.

[22] Abrahamsen, Rita, and Gerald Bareebe. "Uganda's fraudulent election." Journal of Democracy 32, no. 2 (2021): 90-104.

EMERGING TECHNOLOGIES OF INFORMATION MANIPULATION: A REVIEW

In the context of FIMI, newer technologies bolster 'Techniques, Tactics and Procedures' TTPs that encompass the specific actions taken, methods employed, and standard processes followed by state actors or foreign entities to achieve their information manipulation objectives, whether they involve spreading disinformation, manipulating social media, conducting cyber attacks, or engaging in other information warfare activities. These TTPs are part of a broader strategy to influence public perception, shape narratives, and advance geopolitical interests. Newer technologies and their impact on FIMI TTPs can be summarized under three main headings: A.I.-based technologies, social media and algorithm-based technologies, deepfakes and audio-visual manipulation, and micro-targeting and ad-based techniques.

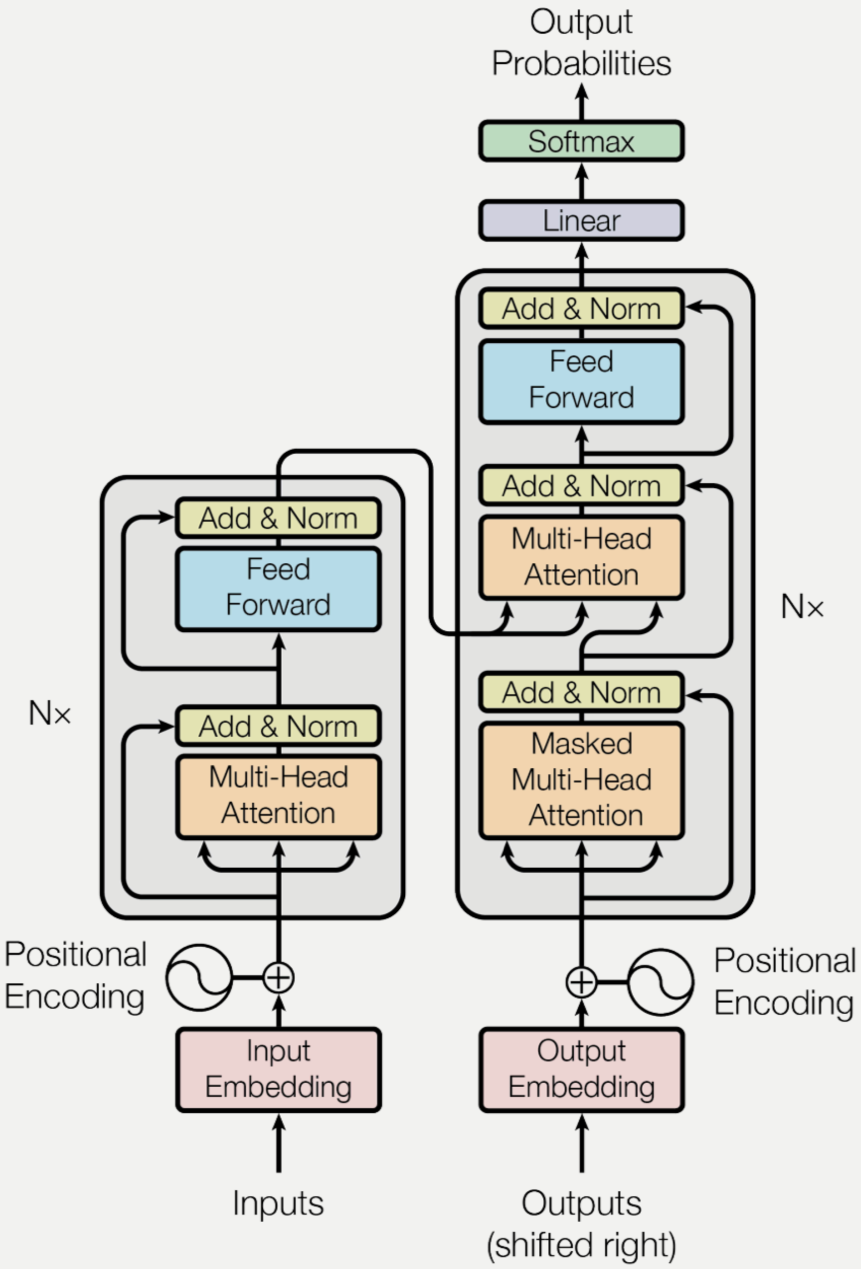
A.I.-BASED TECHNOLOGIES
1. AI-Generated Content:

The development of advanced artificial intelligence (AAI) has given rise to several powerful AI-Generated Content Generators (AI-GCGs), each capable of having a significant impact on the field of foreign interference.[23] A noteworthy model among these is OpenAI's GPT-variants, a formidable language model that boasts more than 100 trillion machine learning parameters. The robust capabilities of GPTs are demonstrated in their ability to write essays, answer questions, translate languages, and even compose poetry with a degree of fluency that borders on human-like. This technology, however, is a double-edged sword in the context of FIMI. For instance, a malicious state actor could exploit it during a politically charged standoff, using it to disseminate an influx of social media posts designed to stoke nationalist sentiments. Such a campaign could exacerbate existing tensions and potentially precipitate conflict, underscoring the potential threat that advanced AI poses in the hands of those seeking to manipulate information.

Another significant model is Microsoft's Turing NLG (T-NLG), the tech giant's most advanced language model. The AI can generate substantial paragraphs of text that could easily be mistaken for human-written content. This opens the door to potential misuse, where, say, during an election, a foreign entity could use T-NLG to churn out articles riddled with false information about a candidate, thereby influencing public opinion and undermining the democratic process. Google's T5 (Text-to-Text Transfer Transformer), although not primarily a text generator, adds to the array of tools available for potential misuse. Its flexibility and power across a range of natural language processing tasks - from translation

[23] Cao, Yihan, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S. Yu, and Lichao Sun. "A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt." arXiv preprint arXiv:2303.04226 (2023).

to summarization - make it a versatile weapon in the FIMI arsenal. Finally, BART, developed by Facebook's AI research group, adds another layer to the landscape of AI-Generated Content Generators. As a denoising autoencoder designed for pretraining sequence-to-sequence models, BART can generate coherent, high-quality text. A crisis in a developing nation with unequal, or slow information dissemination could provide an opportunity for an external state to deploy BART-powered bots, spreading rumors on social media platforms and thereby causing confusion and further destabilizing the already fraught situation.



A generic flowchart of language transformer models. Source: Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N., Kaiser, Lukasz, and Illia Polosukhin. "Attention Is All You Need." ArXiv, (2017). Accessed July 31, 2023. /abs/1706.03762.

2. Text Summarization and Content Manipulation:

The evolution of Natural Language Processing (NLP) has paved the way for the development of sophisticated techniques for text summarization and content manipulation. These techniques have significantly transformed the landscape of FIMI, creating a new arsenal of tools for malicious actors. Abstractive summarization has emerged as a potent technique in this domain. Unlike its counterpart, extractive summarization, which merely picks the most important sentences from a text, abstractive summarization crafts new sentences to convey the same information.[24]

Language models such as GPTs or T5 are adept at creating abstract summaries that preserve the context and sentiment of the original content. However, in the context of FIMI, this capability could be exploited to craft misleading summaries of complex issues or political events. An external state involved in an armed conflict, for instance, could manipulate abstractive summarization to spin narratives to its advantage, thereby influencing the morale and perceptions of battlefield dynamics by combatants. Alongside abstractive summarization, sentiment manipulation presents another potential avenue for misuse. The ability of advanced NLP models to subtly change the sentiment of a text without altering its overall content is a potent tool in the hands of those intent on sowing disinformation.[25]

Furthermore, modern NLP models have the capability to paraphrase or rewrite sentences or even entire texts while maintaining the original meaning. While this function can be beneficial for tasks such as plagiarism detection or text simplification, it also opens the door for exploitation in disinformation campaigns. For example, during a territorial dispute between two littoral states, an external entity could manipulate genuine news articles through subtle rephrasing, skewing public perception to favor its perspective and adding to audience costs and escalation patterns. Another noteworthy development in NLP is the technique of automated fact-checking, which aims to verify the factual accuracy of a text automatically. However, sophisticated actors might find ways to circumvent these systems by generating content that, while technically accurate, is misleading or taken out of context. These attempts will wreak havoc during disasters or emergencies, where information can get highly technical and confusing, opening the way for large-scale manipulation by foreign actors at scale.

The advent of these advanced techniques has considerably amplified the potential impact of FIMI TTPs, offering malicious actors sophisticated tools for information manipulation. These techniques enable the creation of not just realistic content, but also narratives that are contextually and emotionally attuned to the biases and expectations of the target audience, making the disinformation more convincing and difficult to detect. Consequently, this could result in greater social and political impact, highlighting the importance of developing advanced countermeasures,

[24] Winata, Genta Indra, Andrea Madotto, Zhaojiang Lin, Rosanne Liu, Jason Yosinski, and Pascale Fung. "Language models are few-shot multilingual learners." arXiv preprint arXiv:2109.07684 (2021).

[25] Mathew, Leeja, and V. R. Bindu. "A review of natural language processing techniques for sentiment analysis using pre-trained models." In 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), pp. 340-345. IEEE, 2020.

including AI-powered detection systems, digital literacy campaigns, and policy interventions, to combat this evolving threat.

3. Language Translation and Cross-Lingual Disinformation:

NLP has experienced substantial advancements that have culminated in the development of sophisticated language translation and cross-lingual disinformation techniques. These advancements are of utmost importance in the context of FIMI tactics, given the ubiquity of the internet and its multilingual user base. A paramount development in this domain is Neural Machine Translation (NMT).[26]

An excellent example of this is Google's Neural Machine Translation system, a pioneer in end-to-end learning approaches to machine translation. Unlike traditional translation systems that break a sentence into individual words or phrases, NMT translates entire sentences, thereby ensuring more accurate and fluent translations. However, the misuse of this technology for spreading disinformation cannot be overlooked. Picture a significant geopolitical event where a state actor uses NMT to translate and distribute misleading information or propaganda in multiple languages simultaneously. This could potentially manipulate public opinion on a global scale. Furthermore, the emergence of the Multilingual BERT (mBERT), a powerful tool for cross-lingual understanding, has deepened the potential for manipulation.[27] mBERT is adept at understanding semantic similarities between sentences in different languages, a capability that could be harnessed to spread disinformation that maintains its deceptive intent across various languages. An example could be a public health crisis where a malicious actor uses mBERT to circulate misleading health advice in several languages, thereby causing confusion and potentially harmful behavior in different countries.

Zero-shot translation, another cutting-edge development in NLP, could significantly broaden the scope of disinformation. This feature, found in advanced models such as Facebook's LASER or OpenAI's GPTs, enables a model to translate between language pairs it hasn't seen during training. In the hands of malicious actors, a zero-shot translation could be used to spread disinformation in less common languages or dialects, targeting more vulnerable or less digitally literate populations. Consider India, a country with intra-ethnic or intra-religious tensions, where an external entity utilizes zero-shot translation to incite violence by disseminating inflammatory messages in local dialects. The technique of Cross-lingual Transfer Learning has also found its way into the realm of FIMI.[28] This method involves processing text in a different language using a model initially trained

[26] Stahlberg, Felix. "Neural machine translation: A review." Journal of Artificial Intelligence Research 69 (2020): 343-418.
[27] Pires, Telmo, Eva Schlinger, and Dan Garrette. "How multilingual is multilingual BERT?." arXiv preprint arXiv:1906.01502 (2019).
[28] Schuster, Sebastian, Sonal Gupta, Rushin Shah, and Mike Lewis. "Cross-lingual transfer learning for multilingual task oriented dialog." arXiv preprint arXiv:1810.13327 (2018).

in another language. It could be used to generate fake news in one language and then adapt the same model to generate similar content in other languages. An illustrative scenario would be a contentious international climate summit where a state actor exploits this technique to circulate misinformation about the environmental policies of participating countries, potentially sowing discord and undermining international cooperation.

In essence, these advanced A.I. translation techniques considerably boost the reach and effectiveness of FIMI TTPs. They dismantle language barriers, enabling disinformation campaigns to reach global audiences and exploit linguistic and cultural differences. This expansion of techniques underlines the urgency of developing robust multilingual disinformation detection strategies and digital literacy programs that account for linguistic diversity.

4. Sentiment Analysis for Emotional Manipulation:

The integration of sentiment analysis into the realm of FIMI introduces a worrisome shift in the tactics employed. Often referred to as opinion mining, sentiment analysis is a process used to identify and categorize opinions articulated in a text. This technique is used predominantly to discern an author's attitude–positive, negative, or neutral–towards a subject. An intricate approach to sentiment analysis, termed Fine-grained Sentiment Analysis, expands beyond the fundamental positive, negative, or neutral classifications and encapsulates emotions such as joy, anger, sadness, and more.

State-of-the-art models like Google's BERT or OpenAI's GPTs can be adapted for fine-grained sentiment analysis.[29] In the landscape of FIMI, this technique might be leveraged by state actors to heighten specific emotions, like fear or anger, towards particular issues or groups, which can intensify societal polarization. Consider a contentious election, where a foreign entity might exploit this technique to provoke anger and division among the voters. On the other hand, Aspect-based Sentiment Analysis (ABSA) focuses on sentiment analysis concerning specific aspects or attributes within a text, rather than on the overall content.[30] By targeting certain aspects of a situation or individual, this technique could disproportionately influence public sentiment, a clear path to public opinion manipulation. To illustrate, a foreign actor might use ABSA during a public health crisis to magnify public discontent with a government's crisis management approach, thereby inciting discord and unrest.

The advent of Deep Learning techniques, such as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), has significantly improved the

[29] Mathew, Leeja, and V. R. Bindu. "A review of natural language processing techniques for sentiment analysis using pre-trained models." In 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), pp. 340-345. IEEE, 2020.

[30] Nazir, Ambreen, Yuan Rao, Lianwei Wu, and Ling Sun. "Issues and challenges of aspect-based sentiment analysis: A comprehensive survey." IEEE Transactions on Affective Computing 13, no. 2 (2020): 845-863.

accuracy of sentiment analysis.[31] However, these methods can also be exploited to optimize emotional manipulation by creating and disseminating content tailored to trigger specific emotional reactions from the public. A case in point could be an international military conflict where a foreign actor uses these techniques to generate messages that stir up sympathy for one side and hostility towards the other, consequently swaying public opinion and potentially influencing diplomatic outcomes.

Finally, Multimodal Sentiment Analysis stands as an advanced technique that combines inputs like text, audio, and video to enhance sentiment prediction accuracy.[32] This technique is especially potent in social media settings where content often spans multiple modes. From a FIMI perspective, a state actor could harness this technique to construct emotionally resonant, multifaceted disinformation campaigns aimed at societal disruption. For example, amidst negotiations over a contentious trade agreement, foreign entities might utilize multimodal sentiment analysis to sway public sentiment against the agreement, thereby influencing policy outcomes.

In summary, these advanced sentiment analysis techniques pose the potential to significantly amplify the impact of FIMI tactics by manipulating public emotions with increased precision and contextuality. The emotional nature and polarization induced by such disinformation pose significant challenges in counteracting its effects, thereby underscoring the necessity for the development of advanced detection techniques, public education initiatives, and comprehensive regulations to combat these evolving threats.

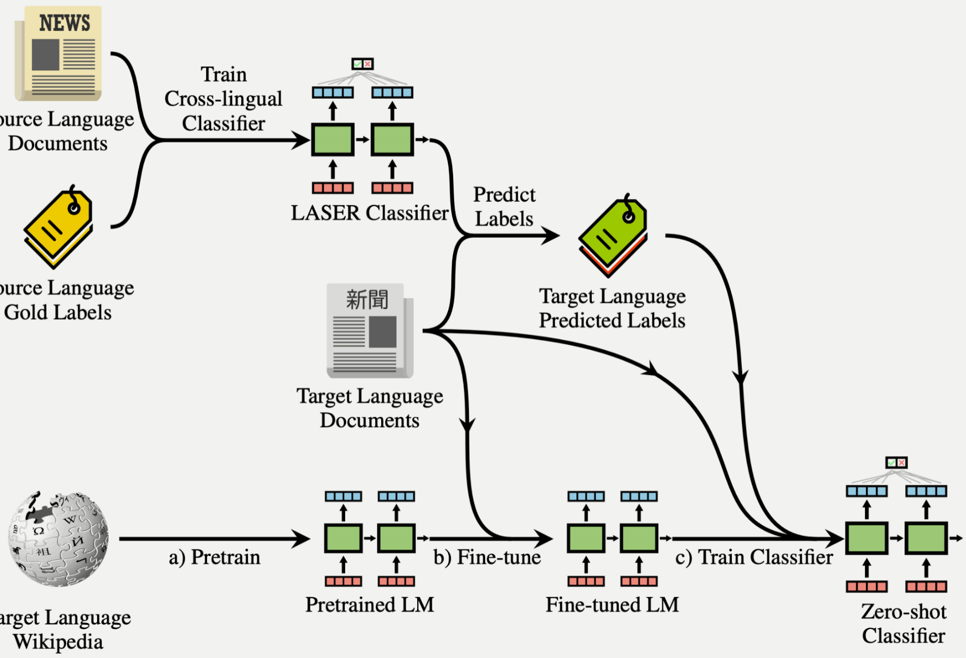
5. Language Model Fine-Tuning:

The utilization of Language Model Fine-Tuning techniques has been on the rise in a myriad of NLP applications. This process involves adapting pre-existing models on specific datasets to make them more apt for particular tasks, significantly improving the efficiency and effectiveness of language models. Nonetheless, these advancements also unlock new avenues for misuse, particularly in the sphere of FIMI.

Among the advanced techniques in this domain is Domain-Specific Fine-Tuning. This approach enables a model to better understand and generate text in a specific domain by fine-tuning it on data within that area.[33] Consider a model fine-tuned on political discourse; it would prove more efficient at generating persuasive political propaganda. A state actor engaged in FIMI might fine-tune a model on the language and cultural aspects of a target country to craft disinformation that is more persuasive and contextually fitting.

[33] Cui, Yin, Yang Song, Chen Sun, Andrew Howard, and Serge Belongie. "Large scale fine-grained categorization and domain-specific transfer learning." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4109-4118. 2018.

Another tactic is Adversarial Fine-Tuning, which involves tweaking a model to produce outputs that are misleading or biased.[34] In a FIMI scenario, this technique could be used by a state actor to create disinformation engineered to slip through content moderation algorithms. During a political crisis, adversarial fine-tuning can be utilized to generate narratives that incited ethnic tensions without setting off automated moderation systems. This is a highly concerning technology that nullifies the effect of automated content moderation in all social media platforms, and enables sophisticated FIMI efforts to bypass these barriers.


NEWS
Source Language Documents
Source Language Gold Labels
Train Cross-lingual Classifier
LASER Classifier
Predict Labels
Target Language Predicted Labels
新聞
Target Language Documents
Target Language Wikipedia
a) Pretrain
Pretrained LM
b) Fine-tune
Fine-tuned LM
c) Train Classifier
Zero-shot Classifier

A hypothetical model of a language fine-tuning classifier algorithm. Source: Eisenschlos, Julian M., Ruder, Sebastian, Czapla, Piotr, Kardas, Marcin, Gugger, Sylvain, and Jeremy Howard. "MultiFiT: Efficient Multi-lingual Language Model Fine-tuning." ArXiv, (2019). Accessed July 31, 2023. / abs/1909.04761.

Prompt-Based Fine-Tuning is a more recent approach that refines the model's responses to specific inputs or prompts, allowing for greater control over the model's output.[35] In the FIMI context, this could be exploited to generate disinformation tailored to cause maximum disruption. For example, during an international environmental crisis, a foreign actor might use this technique to fabricate messages that contradict scientific consensus and breed public skepticism. A similar technique is Few-Shot Learning. It involves fine-tuning a model on a small number of examples to enable it to perform a specific task.[36] In the realm of FIMI, this approach could be used to swiftly adapt a model to new topics or trends, making disinformation timelier and more pertinent. A case in point is the swift dissemination of disinformation during a global health emergency, where a

[34] Chen, Tianlong, Sijia Liu, Shiyu Chang, Yu Cheng, Lisa Amini, and Zhangyang Wang. "Adversarial robustness: From self-supervised pre-training to fine-tuning." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 699-708. 2020.
[35] Scao, Teven Le, and Alexander M. Rush. "How many data points is a prompt worth?." arXiv preprint arXiv:2103.08493 (2021).
[36] Bansal, Trapit, Rishikesh Jha, and Andrew McCallum. "Learning to few-shot learn across diverse natural language classification tasks." arXiv preprint arXiv:1911.03863 (2019).

foreign entity used few-shot learning to spread fabricated narratives about the origin and spread of a virus, fostering discord and fear among the global populace.

These advanced fine-tuning techniques significantly enhance the effectiveness of FIMI tactics, as they empower malicious actors to create disinformation that is more contextually appropriate, convincing, and adaptive.

6. Contextual Analysis and Mimicry:

The advancements in the realm of artificial intelligence, particularly in the areas of Contextual Analysis and Mimicry techniques, have had profound implications in a variety of fields. These tools are especially instrumental in the domain of NLP. While these developments open new frontiers of possibilities, they can also be misused, especially in the arena of FIMI.

Most advanced AI models come equipped with a capability known as Contextual Entity Recognition. Essentially, these models can comprehend the context in which entities - such as people, organizations, and places - are mentioned in a piece of text.[37] For example, in the sentence "Apple has just released a new product," these models would identify "Apple" as a tech company, not a fruit, based on context. Unfortunately, this nuanced understanding can be exploited in FIMI scenarios to customize disinformation that fits seamlessly into ongoing conversations or debates, thereby making the false narratives more convincing. For instance, during the 2019 Amazon Rainforest wildfires, manipulated narratives suggesting intentional burning by environmental NGOs were contextually inserted into social media discussions, leading to increased spread and acceptance of the false narrative.[38]

Simultaneously, AI models have reached a degree of sophistication that allows for Contextual Mimicry. These models can emulate the style, tone, and contextual nuances of specific sources or individuals.[39] In a FIMI scenario, this could culminate in the creation of false statements or articles seemingly originating from trusted figures or institutions. This has been seen in countries like Myanmar, Malaysia, and Indonesia, where AI-generated statements mimicking prominent politicians were disseminated, causing widespread confusion and mistrust among voters during election periods.[40] Lastly, as these AI models continue to evolve, they are beginning to demonstrate Contextual Adaptation capabilities. They can adjust their output to suit different contexts, enhancing their flexibility and mimicry capabilities. This feature can be harnessed in FIMI to adapt disinformation to different platforms, cultures, or demographics, increasing its effectiveness. For example, during the

[37] Cucchiarelli, Alessandro, and Paola Velardi. "Unsupervised named entity recognition using syntactic and semantic contextual evidence." Computational Linguistics 27, no. 1 (2001): 123-131.

[38] de Moraes, Rodrigo Fracalossi. "Demagoguery, populism, and foreign policy rhetoric: evidence from Jair Bolsonaro's tweets." Contemporary Politics 29, no. 2 (2023): 249-275.

[39] Hess, Ursula, and Agneta Fischer. "Emotional mimicry: Why and when we mimic emotions." Social and personality psychology compass 8, no. 2 (2014): 45-57.

[40] Tan, Netina. "Electoral management of digital campaigns and disinformation in East and Southeast Asia." Election Law Journal: Rules, Politics, and Policy 19, no. 2 (2020): 214-239.

ongoing territorial disputes in the South China Sea, disinformation narratives tailored to local languages and cultural nuances have been used to inflame regional tensions, showcasing the potential dangers of these advanced AI technologies when misused.[41]

SOCIAL MEDIA PLATFORMS AND ALGORITHM-BASED TECHNOLOGIES

In the last decade, social media platforms have demonstrated their power to spread disinformation rapidly. Algorithmic feeds can prioritize sensational or engaging content, regardless of its accuracy, potentially contributing to the viral spread of false information. Social media platforms and algorithms play a significant role in shaping the landscape of disinformation and FIMI. The technical advancements in these platforms and algorithms have enabled the rapid dissemination and amplification of deceptive content, making them potent tools for malicious actors seeking to manipulate public opinion. Here's an extended analysis of their impact:

1. Algorithmic Amplification of Disinformation:

Algorithmic Amplification, a potent technique that takes advantage of social media platform algorithms to propagate and promote content, plays a fundamental role in the dissemination of disinformation. This technique's significance becomes especially stark in the setting of FIMI as it equips malevolent actors with the tools to broaden their reach and amplify their influence. A myriad of sophisticated tactics has evolved within this space.

One such tactic, Microtargeting, zeroes in on specific user groups determined by their demographic profiles, interests, or behavioral patterns.[42] The strategic deployment of disinformation takes advantage of microtargeting to deliver tailored messages to the individuals most likely to be influenced. A remarkable case of this tactic surfaced during the Brexit campaign in 2016 when allegations of microtargeting emerged. Specified voter groups were exposed to misleading information concerning the implications of the United Kingdom's departure from the European Union, ranging from overstated claims about financial savings to unfounded assertions about immigration control.[43]

The digital landscape has also witnessed the rampant utilization of Social Bots and Cyborgs. Social bots are automated accounts on social media that

[41] Nguyen, Dennis, and Erik Hekman. "A 'new arms race'? Framing China and the USA in AI news reporting: A comparative analysis of the Washington Post and South China Morning Post." Global Media and China 7, no. 1 (2022): 58-77.

[42] Barbu, Oana. "Advertising, microtargeting and social media." Procedia-Social and Behavioral Sciences 163 (2014): 44-49.

[43] Šimunjak, Maja. Tweeting Brexit: social media and the aftermath of the EU referendum. Routledge, 2022.

interact with users or other bots, while cyborgs represent accounts that blend automated activity with human input.[44] These entities are utilized to augment the reach of disinformation by engaging with the content, which can trick social media algorithms into perceiving the content as popular, leading to its wider dissemination. During the presidential election in France in 2022, it was reported that social bots were employed to amplify false narratives around election fraud, fostering uncertainty and political tension.[45]

Another technique, Hashtag Poisoning (or hijacking), involves co-opting trending hashtags to disseminate disinformation.[46] This allows malevolent actors to expose their content to a larger, unsuspecting audience. During the Chinese COVID protests in 2022, the Chinese government was accused of hijacking hashtags to flood out protest content and divert attention away from users following or using protest-related hashtags.[47] Deepfake Amplification has emerged as a significant challenge. Deepfakes, which are hyper-realistic videos created with AI, can be disseminated extensively through social media and promoted by algorithms, especially when the content is controversial or sensational. During the Australian federal election in 2023, a deepfake video of a prominent candidate was widely shared, leading to confusion and mistrust among voters, demonstrating the potential harm of deepfake technology when misused.[48]

Lastly, Engagement Baiting is a technique where content is crafted to incite user engagement, such as likes, shares, and comments, potentially resulting in algorithmic amplification.[49] When disinformation is paired with sensational or divisive content, it can elicit strong reactions and consequently spread more rapidly. This tactic was evident during the general election in Nigeria in 2023, where false information was often paired with emotionally charged or sensational content, intended to drive engagement and thus magnify its reach.[50]

2. Personalized Content Delivery:

Social media algorithms are engineered to curate content that aligns with each user's preferences and behavioral patterns. This personalized content delivery often leads to the presentation of information that echoes users' existing beliefs and ideologies, thus fostering a digital environment referred to as echo chambers. This selective delivery of information can reinforce confirmation biases and leave

[44] Gorwa, Robert, and Douglas Guilbeault. "Unpacking the social media bot: A typology to guide research and policy." Policy & Internet 12, no. 2 (2020): 225-248.

[45] Shahid, Wajiha, Yiran Li, Dakota Staples, Gulshan Amin, Saqib Hakak, and Ali Ghorbani. "Are you a cyborg, bot or human?—a survey on detecting fake news spreaders." IEEE Access 10 (2022): 27069-27083.

[46] VanDam, Courtland, and Pang-Ning Tan. "Detecting hashtag hijacking from twitter." In Proceedings of the 8th ACM Conference on Web Science, pp. 370-371. 2016.

[47] China accused of flooding social media with spam to crowd out protest news. The Guardian. 4 Dec 2022. https://www.theguardian.com/world/2022/dec/04/china-accused-of-flooding-social-media-spam-covid-protests

[48] Vasist, Pramukh Nanjundaswamy, and Satish Krishnan. "Deepfakes: an integrative review of the literature and an agenda for future research." Communications of the Association for Information Systems 51, no. 1 (2022): 14.

[49] Zhang, Wanjiang Jacob, Jingjing Yi, and Hai Liang. "I cue you liking me: Causal and spillover effects of technological engagement bait." Computers in Human Behavior (2023): 107864.

[50] Nigerian elections 2023: False claims and viral videos debunked. BBC Africa. 28 February 2023. https://www.bbc.com/news/world-africa-64797274

individuals more susceptible to disinformation that corresponds to their viewpoints. Personalized Content Delivery, which leverages user data to tailor content to specific individuals or groups on social media platforms, is a common technique.

Although typically utilized to enhance user engagement and experience, it can also be manipulated for FIMI. Techniques such as Microtargeting and Behavioral Profiling are particularly potent tools in this arena. As discussed previously, microtargeting involves delivering specific content to select groups based on their demographics, interests, or online behaviors.[51] Its effectiveness escalates when combined with behavioral profiling, a method that forms comprehensive profiles through patterns discerned from a user's online behavior. These profiles then enable the delivery of highly targeted disinformation that resonates strongly with individual users or groups. For instance, during the Thai general election in 2019, there were reports of microtargeting tactics being employed to sway voter sentiments by spreading misleading narratives tailored to specific demographic groups.[52]

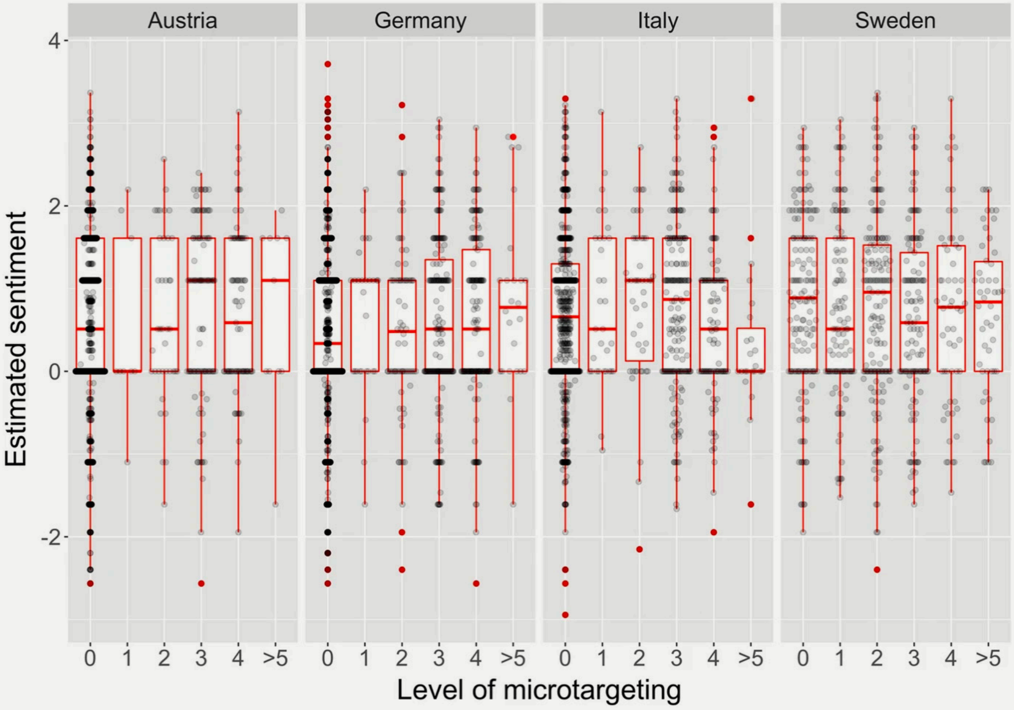
Adaptive News Feed Algorithms represent another critical aspect. Used by social media platforms, these algorithms customize a user's news feed based on their past behaviors, interests, and interactions. This can lead to the creation of "filter bubbles," where users are primarily exposed to content that reinforces their pre-existing viewpoints. Such an environment can be exploited to amplify disinformation. The specter of disinformation also looms over AI-Generated Personalized Content. With the advent of advanced AI models, it's possible to create personalized disinformation on an unprecedented scale. Each piece of content can be meticulously tailored to cater to a specific user's interests, biases, and beliefs, making it significantly more persuasive. Although no specific large-scale geopolitical event related to this has been extensively reported yet, this represents an emerging area of concern with potential future risks. Deepfake Personalization introduces an additional layer of complexity. Personalization can extend beyond text to include images and videos, particularly with the advent of deepfakes.[53] Deepfakes can be tailored to mimic familiar figures or create scenarios that align with a user's beliefs or fears. During the Argentinian general elections in 2023, for example, deepfakes of political candidates involved in contrived scenarios were widely disseminated, causing considerable confusion and misinformation.[54]

[51] Schäwel, Johanna, Regine Frener, and Sabine Trepte. "Political microtargeting and online privacy: A theoretical approach to understanding users' privacy behaviors." Media and Communication 9, no. 4 (2021): 158-169.
[52] Wanless, Alicia, and Michael Berk. "The audience is the amplifier: Participatory propaganda." The Sage handbook of propaganda (2020): 85-104.
[53] Khan, Ihtiram Raza, Saman Aisha, Deepak Kumar, and Tabish Mufti. "A Systematic Review on Deepfake Technology." Proceedings of Data Analytics and Management: ICDAM 2022 (2023): 669-685.
[54] Riera, Ariel, and Laura Zommer. "Using fact checking to improve information systems in Argentina." The political quarterly 91, no. 3 (2020): 600-604.

3. Microtargeting and Audience Segmentation:

The realm of social media platforms, with their sophisticated algorithms, has enabled microtargeting - a feature that facilitates the precise delivery of content to distinct groups of users. This capability, while beneficial in enhancing user experience and streamlining marketing strategies, also harbors potential misuse, particularly in the domain of FIMI. Malicious actors can exploit this feature to spread disinformation among targeted demographics, strategically influencing public opinion and exploiting societal divisions.

Microtargeting and Audience Segmentation serve as pivotal techniques in this regard, providing tailored content to diverse individuals or groups. A particularly potent method in this area is Psychographic Profiling, which involves the creation of detailed profiles of social media users.[55] These profiles incorporate a range of data, including online behaviors, interests, attitudes, and even personality traits. A significant application of this technique was allegedly carried out by Cambridge Analytica during the 2016 US Presidential Election. However, similar tactics have also been reported in other parts of the world. For instance, during the Brazilian Presidential Election in 2018, or Spanish elections in 2019 there were claims of psychographic profiling being used to influence voter behavior through tailored political advertisements.[56]



Psychographic profiling works best when it is used to tailor information manipulation to specific countries. However, too much microtargeting creates diminishing returns. Source: López Ortega, A. Are microtargeted campaign messages more negative and diverse? An analysis of Facebook Ads in European election campaigns. Eur Polit Sci 21, 335–358 (2022). https://doi.org/10.1057/s41304-021-00346-6

[55] Bakir, Vian. "Psychological operations in digital political campaigns: Assessing Cambridge Analytica's psychographic profiling and targeting." Frontiers in Communication 5 (2020): 67.

[56] Baviera, Tomás, Lorena Cano-Orón, and Dafne Calvo. "Tailored messages in the feed? Political microtargeting on Facebook during the 2019 General Elections in Spain." Journal of Political Marketing (2023): 1-20.

Geographical Microtargeting, another significant technique, allows for content to be tailored based on users' geographical locations. In the Indian general election of 2019, allegations emerged of geo-targeted misinformation being employed to incite regional tensions and sway voting behavior. This was especially apparent in areas already witnessing local unrest or political dissension.

Behavioral Targeting is another method which utilizes data on an individual's past behavior, like online click patterns, search history, and page visits, to predict and potentially influence future behavior. During the general elections in Nigeria in 2019, it was reported that behavioral targeting was being used to disseminate politically charged content to susceptible individuals, amplifying societal divisions and polarizing voter sentiment. Artificial Intelligence has the potential to greatly enhance the precision of microtargeting.

AI-Enhanced Microtargeting uses AI to analyze large datasets, identifying nuanced patterns that allow for highly precise audience segmentation and more efficient microtargeting. While the widespread use of this technique in FIMI has not been extensively documented, the potential risk posed by such large-scale, highly targeted disinformation campaigns cannot be overlooked. Lastly, Look-alike Audience Targeting is a technique that involves creating a "seed" audience profile–typically consisting of current supporters or a desired target group–and then leveraging social media algorithms to find and target users who resemble this profile.

While initially a marketing tool, its potential for political exploitation by identifying and targeting susceptible individuals for disinformation has become a point of concern. For example, during the Kenyan general elections in 2022, allegations emerged of Look-alike Audience Targeting being used to distribute disinformation to certain demographic groups. These advanced techniques underscore the potential for highly targeted and impactful disinformation campaigns, accentuating the necessity for increased transparency in data usage, stricter regulations on privacy, and comprehensive digital literacy education among the public.

4. Automated Bot Accounts:

Automated Bot Accounts are a fundamental instrument in social media manipulation owing to their capacity to quickly and economically produce voluminous content. These bots can be programmed to carry out various tasks, from posting and sharing content to following and befriending users. They have been utilized to warp public discourse, circulate disinformation, and amplify divisive messages. The intricacies of their application and potential harm can be dissected by investigating several advanced techniques.

Content Amplification Bots are one such technique. These bots are designed to spread specific content or messages widely across social media platforms. They operate chiefly by retweeting, liking, or sharing posts, thus boosting their visibility and perceived popularity. A striking example of this was seen during the 2016 Colombian Peace Referendum. Here, content amplification bots were used to propagate polarizing narratives and misinformation, thereby warping public discourse and potentially influencing the narrow vote outcome that rejected the government's peace deal with FARC rebels.[57]

Sybil Bots offer another illustration. Named after the well-known case of Sybil Dorsett, who was diagnosed with multiple personality disorder, these bots generate multiple false identities on social networks. They can disseminate disinformation extensively, portray a message as more popular than it truly is, and artificially inflate a user's follower count. This manipulation was clear during the South African general elections in 2019, where Sybil bots were deployed to manipulate public sentiment and discourse, skewing political conversations on social media platforms.

Social Bots constitute another category of automated accounts. They interact with human users by befriending or following them, thereby gaining access to their social networks. From this vantage point, they can introduce disinformation directly into these networks. In the Philippines, the use of social bots was reported during the 2016 presidential election. These bots infiltrated social networks and disseminated divisive content, contributing to a highly polarized political environment. Influencer Bots, on the other hand, interact with influential social media users. Their goal is to entice these influencers into disseminating their disinformation. During the Indonesian general elections in 2019, influencer bots targeted local influencers to amplify misleading messages, contributing to public opinion polarization.[58] Lastly, Botnets, which are networks of bots controlled by a single entity, have the potential to coordinate and propagate messages or disinformation more effectively. This technique was employed during the 2020 Taiwanese general elections, when Twitter dismantled a significant botnet attempting to disrupt public discourse by spreading misinformation and disinformation.[59]

These techniques highlight the escalating sophistication of bots and their potential misuse in FIMI operations. Addressing this threat necessitates a multi-faceted approach that includes technological, regulatory, and educational initiatives. As we face an increasingly digitized world, understanding and combatting these deceptive tactics becomes even more crucial to protect the integrity of public discourse and democratic processes.

[57] Gallego, Jorge, Juan D. Martínez, Kevin Munger, and Mateo Vásquez-Cortés. "Tweeting for peace: Experimental evidence from the 2016 Colombian Plebiscite." Electoral Studies 62 (2019): 102072.

[58] Uyheng, Joshua, and Kathleen M. Carley. "Characterizing bot networks on Twitter: An empirical analysis of contentious issues in the Asia-Pacific." In Social, Cultural, and Behavioral Modeling: 12th International Conference, SBP-BRiMS 2019, Washington, DC, USA, July 9–12, 2019, Proceedings 12, pp. 153-162. Springer International Publishing, 2019.

[59] Uyheng, Joshua, and Kathleen M. Carley. "Computational analysis of bot activity in the Asia-Pacific: A comparative study of four national elections." In Proceedings of the international AAAI conference on web and social media, vol. 15, pp. 727-738. 2021.

5. Social Engineering and Clickbait Tactics:

Malicious actors leverage social engineering and clickbait tactics to entice users into interacting with disinformation. Using catchy headlines, emotionally charged images, and deceptive information, they aim to boost click-through rates, thereby magnifying the visibility and impact of disinformation. Social engineering and clickbait strategies represent some of the most ingenious tactics in the world of social media manipulation, as they exploit human curiosity, biases, and trust to disseminate disinformation, stir discord, and influence public opinion.

Phishing campaigns serve as a popular form of social engineering, where misleading messages are sent, often pretending to be legitimate information requests or click prompts. While these tactics are usually used to access sensitive data, they can also assist in spreading disinformation or injecting malicious content into trusted networks. One instance of this was during Mexico's general election in 2018, where a successful phishing campaign led to a leak of sensitive documents from the leading candidate's team, triggering substantial disruption and political scandal.[60]

Another technique involves the creation of false personas or impersonation of genuine individuals or organizations. This gives malicious actors the semblance of credibility for their disinformation campaigns. This was observed during Nigeria's general elections in 2019, where Twitter detected and removed a network of accounts pretending to be Nigerian citizens but were, in fact, part of a foreign disinformation operation.[61] Malicious actors often exploit emotions and incite outrage to increase the likelihood of user engagement and sharing of disinformation. This technique was evident during the 2018 Malaysian general election, where emotionally charged and misleading content was extensively shared on social media platforms, deepening social divisions and fueling heated public debates.

Clickbait headlines and deceptive captions are another common strategy to draw attention and provoke clicks, often leading users to content that either bears no relation to or significantly distorts the true essence of the story. This tactic was heavily exploited during the Israeli legislative elections in 2019, redirecting users to websites rife with false or deceptive information, which notably distorted public discourse.[62] Pretexting is another technique, involving the creation of a plausible pretext or scenario to trick users into providing information, clicking on links, or sharing content. During the 2020 Covid-19 pandemic, this method was glaringly visible in disinformation campaigns across numerous countries, including Iran and Brazil. The crisis was exploited as a pretext to spread false information about the

[60] Mexico election: Concerns about election bots, trolls and fakes. BBC Monitoring. 30 May 2018. https://www.bbc.com/news/blogs-trending-44252995

[61] Oyebode, Oladapo, and Rita Orji. "Social media and sentiment analysis: the Nigeria presidential election 2019." In 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 0140-0146. IEEE, 2019.

[62] Mourão, Rachel R., and Craig T. Robertson. "Fake news as discursive integration: An analysis of sites that publish false, misleading, hyperpartisan and sensational information." Journalism studies 20, no. 14 (2019): 2077-2095.

disease, its origins, and treatments, leading to public confusion and obstructing effective public health responses.[63]

These tactics highlight the advanced and diverse methods deployed in social engineering and clickbait strategies within FIMI campaigns. To counter these threats, a holistic approach is required, including public education about these tactics, enhanced digital literacy, and the development and deployment of sophisticated detection and mitigation tools.

DEEPFAKES AND AUDIOVISUAL MANIPULATION

Deepfakes use AI to convincingly alter videos, images, or audio to mislead audiences. Audio manipulation technologies can generate fake voices that sound remarkably real, making it difficult to distinguish genuine audio recordings from fake ones. Deepfakes and audio manipulation have emerged as powerful tools in the arsenal of disinformation campaigns. These technologies allow malicious actors to create highly realistic and deceptive content, making it increasingly challenging to discern between genuine and manipulated information.

1. Deepfake as Video Manipulation:

The advent of deepfakes and video manipulation techniques presents an escalating challenge in the domain of social media manipulation. These sophisticated technologies enable the generation of incredibly realistic falsified media that can drastically influence public sentiment. The term "deepfake" is associated with AI-based technology capable of creating or altering video and audio content to make it seem authentic. By learning from genuine footage, this technology can generate a convincing fake that can portray an individual saying or doing something they never actually did. A stark example of this occurred during the 2021 Peruvian general election, when deepfake videos of the candidates were disseminated widely, causing confusion and provoking questions about the veracity of video content.[64] In another incident related to the 2023 South Korean presidential elections, several deepfakes were released, some of which depicted candidates making inflammatory and false statements. These incidents served to confuse voters and disrupted the normal electoral discourse.[65]

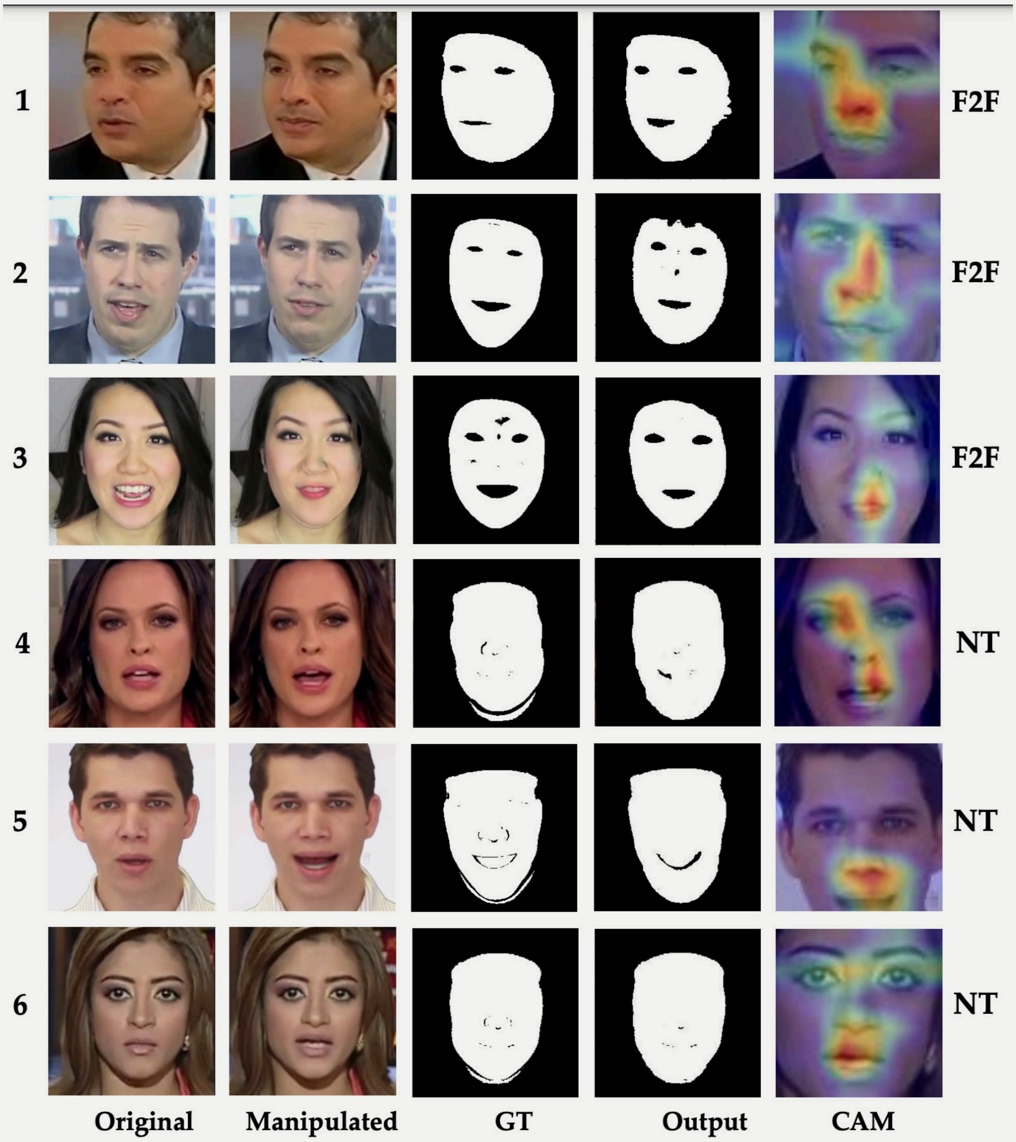
Video deepfakes are a sophisticated application of artificial intelligence and machine learning, specifically leveraging deep learning techniques. At their core,

[63] Ceron, Wilson, Mathias-Felipe de-Lima-Santos, and Marcos G. Quiles. "Fake news agenda in the era of COVID-19: Identifying trends through fact-checking content." Online Social Networks and Media 21 (2021): 100116.
[64] Pashentsev, Evgeny, and Darya Bazarkina. "Malicious Use of Artificial Intelligence and Threats to Psychological Security in Latin America: Common Problems, Current Practice and Prospects." In The Palgrave Handbook of Malicious Use of AI and Psychological Security, pp. 531-560. Cham: Springer International Publishing, 2023.
[65] Pashentsev, Evgeny. "The Malicious Use of Deepfakes Against Psychological Security and Political Stability." In The Palgrave Handbook of Malicious Use of AI and Psychological Security, pp. 47-80. Cham: Springer International Publishing, 2023.

deepfakes are the result of using deep neural networks to manipulate and alter video content to make it appear as though a person or object in the video is saying or doing something that they didn't actually say or do. The term "deepfake" comes from the combination of "deep learning" and "fake." The process of creating a video deepfake involves several essential steps. Initially, a deepfake creator needs to gather training data, which typically includes a vast amount of video footage of the target individual or object. This data serves as the foundation for training a deep neural network, which is responsible for learning and mimicking the visual and auditory characteristics of the target. The core technology used in creating video deepfakes is generative adversarial networks (GANs).

GANs consist of two primary components: a generator and a discriminator. The generator's role is to produce fake content, while the discriminator's role is to distinguish between real and fake content. Both components work in tandem, engaging in a competitive learning process. The training process starts by feeding the GAN with the real video data of the target individual, and the generator starts generating fake videos. Initially, these generated videos are of poor quality and are easily distinguishable from real ones. The discriminator is then trained on a mixture of real and generated videos to learn to differentiate between them.

Over time, the generator improves its ability to create more convincing fakes, and the discriminator becomes more adept at detecting subtle differences. This iterative training process continues until the generator produces fake videos that are difficult for the discriminator to distinguish from real ones. At this point, the deepfake creator has a well-trained model capable of generating convincing video content that appears to be authentic. To create a specific deepfake video, the trained model is given input, such as new audio or facial movements, and it generates corresponding fake video frames. For example, if the goal is to make a person in the video say something they didn't say, the deepfake model can be fed with the desired audio while retaining the original facial expressions and movements of the target individual. The generator then produces video frames that align the lips and facial features to match the new audio input, resulting in a convincing deepfake video.



First and second columns show the original images and manipulated ones respectively. The black and white images in the third column are corresponding bi- nary GT masks. Predicted masks (column 4) and generated CAMs (column 5) for manipulated images from Face2Face (row 1,2,3) and Neural-Textures (row 4,5,6) dataset. Source: Mazaheri, Ghazal, and Amit K. Roy-Chowdhury. "Detection and localization of facial expression manipulations." In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1035-1045. 2022.

2. Audio Manipulation and Voice Synthesis:

Audio manipulation techniques such as text-to-speech (TTS) synthesis and voice cloning have revolutionized the creation of synthetic voices that closely mimic human speech. TTS technology converts written text into speech that closely mirrors natural human intonation and rhythm. Voice cloning technology, on the other hand, replicates a specific individual's voice based on a relatively small amount of sample audio data. The advancement of artificial intelligence has sparked a revolution in audio manipulation and voice synthesis technologies. As a result, it's now possible to generate hyper-realistic and often indistinguishable fake audio content.

Voice cloning, in particular, stands out as one of the most prominent advancements. This technology uses AI to accurately replicate a person's unique voice. After training the AI system with a sample of the person's speech, it can generate new dialogue in the cloned voice, uttering words that the original person never said.[66] A worrying demonstration of this technology's potential for misuse was observed in 2019 when a German chief executive Rüdiger Kirsch of Euler Hermes Group fell prey to a scam. He transferred a significant amount of money, fooled by a voice cloning AI that convincingly mimicked his superior's voice. Such events highlight the urgent need to counter these rapidly evolving technological threats.[67]

Text-to-speech synthesis is another significant AI advancement in this domain. This technology transforms written text into spoken words, resulting in audio content that sounds incredibly natural and human-like. Its potential uses include creating convincing fake radio broadcasts or phone calls, providing a new platform for disinformation. For example, in 2023 during the Nigerian general elections, allegations arose of fake radio broadcasts disseminating false information, although the use of TTS technology was not conclusively proven, hinting at the problematic nature of verifying these attempts even after their perpetration.[68]

Similar to video deepfakes, audio 'deepfakes' employ machine-learning models to mimic a specific individual's voice. The distinction between voice cloning and audio deepfakes usually a lies in the level of sophistication and realism, with deepfakes typically offering a more convincing imitation. The advent of real-time voice spoofing poses a particularly significant threat.

Technological advancements have made it possible to impersonate someone else during live conversations, such as in a 2022 incident in the Philippines, where a politician's voice was faked during a live radio show, causing widespread confusion and controversy. These techniques could potentially create havoc in real-time political scenarios, enable fraud, or drive social engineering attacks. The continued advancement of audio manipulation techniques highlights their potential misuse in FIMI operations. Given their widespread availability, there is an urgent need to develop robust detection mechanisms, improve security protocols, and educate the public about the potential and impacts of these deceptive practices.

[66] Almutairi, Zaynab, and Hebah Elgibreen. "A review of modern audio deepfake detection methods: challenges and future directions." Algorithms 15, no. 5 (2022): 155.

[67] Pashentsev, Evgeny. "The Malicious Use of Deepfakes Against Psychological Security and Political Stability." In The Palgrave Handbook of Malicious Use of AI and Psychological Security, pp. 47-80. Cham: Springer International Publishing, 2023.

[68] Bola Tinubu's Nigeria election win: The rigging claims of Peter Obi and Atiku Abubakar. BBC Africa. 1 March 2023. https://www.bbc.com/news/world-africa-64802490

TARGETED ADVERTISING AND MICRO-TARGETING

Targeted advertising and micro-targeting represent two influential marketing strategies reshaping the landscape of digital advertising, but are simultaneously being leveraged for the purposes of FIMI. These techniques leverage the vast amounts of consumer data collected from various digital platforms to deliver personalized advertising content to specific demographics, individuals, or niche groups. In this technical text, we will delve into the intricate mechanics behind targeted advertising and micro-targeting, examining the technological underpinnings, data analytics methods, and the algorithms employed.

1. Data Collection and User Profiling:

Targeted advertising relies on extensive data collection to build detailed user profiles. Social media platforms and other online services gather information on users' demographics, interests, behavior, and preferences, creating a rich dataset that informs advertisers about individual users.

Social media platforms have become the central nexus of data collection and user profiling, especially with the increased sophistication of data science techniques. From everyday interactions and engagements, an enormous amount of data is collected, which is then processed and analyzed using advanced data science methodologies to create comprehensive user profiles. These profiles, when misused, can serve as potent tools for Foreign Information Manipulation and Interference (FIMI).

Here's an overview of some advanced techniques and their impact on FIMI Tactics, Techniques, and Procedures (TTPs)

- Metadata Analysis: Metadata, or data about data, can reveal a lot about user behaviors and preferences. For example, the time and frequency of posts, geolocation data, and the type of device used for posting all offer valuable insights about user habits and preferences. For instance, metadata played a crucial role during the 2019 Canadian Federal Election. Detailed analysis of users' geolocation data was used by foreign actors to identify and target specific demographic groups with personalized political content designed to influence voting behaviors.

- Social Network Analysis (SNA): SNA is a method of visualizing and analyzing relationships between individuals. By examining how individuals connect and interact with each other, one can infer group dynamics, identify influential individuals, and spot patterns of information flow. In 2020, this technique was used by various actors to amplify divisive narratives during the Black Lives Matter protests in the United States, exploiting existing societal divisions for political gain.

- Machine Learning and AI Analytics: Advanced algorithms are used to process and analyze vast amounts of data, identifying patterns, trends, and correlations. This allows for highly precise user segmentation, behavior prediction, and content personalization. A case in point is the 2021 general elections in Japan, where machine learning models were used to predict voting preferences and deliver customized political advertisements.

- Sentiment Analysis: This involves using Natural Language Processing (NLP) to analyze the sentiment behind social media posts. By understanding public sentiment towards specific issues or entities, actors can design messages that exploit these feelings. During the 2020 Hong Kong protests, sentiment analysis was used by foreign actors to fuel discontent and exacerbate political tension.

- Psychographic Profiling: This involves classifying people according to their attitudes, aspirations, and other psychological criteria. During the 2020 Thai political protests, psychographic profiling was utilized by foreign actors to fuel dissatisfaction and escalate political tension.

The impact of these advanced techniques on FIMI TTPs is profound. The ability to understand, predict, and manipulate individuals' behaviors on a large scale offers powerful tools for those wishing to interfere with foreign affairs or influence public opinion. This necessitates ongoing efforts to improve data privacy regulations, develop countermeasures, and educate the public about the ways their data can be used.

2. Algorithmic Targeting and Ad Placement:

Sophisticated algorithms analyze user profiles and behaviors to determine the most appropriate content to display to each individual. These algorithms consider factors such as location, browsing history, and engagement patterns to serve personalized ads. Algorithmic targeting and ad placement on social media is a rapidly evolving field with significant impacts on information manipulation and suppression:

- Lookalike Audience Targeting is a powerful tool used to identify potential new users who might be receptive to a certain message due to their similarity to a known group of users. This technique was employed with notable efficacy during the 2016 Brexit referendum when various political entities utilized lookalike audience targeting to extend their reach to individuals who mirrored their existing supporters. This amplified their campaign messages and reinforced their influence. In a similar vein, during the 2019 general elections in India, some political parties leveraged lookalike audience targeting to expand their voter base, capitalizing on the expansive user data available on platforms like Facebook to reach potential supporters.

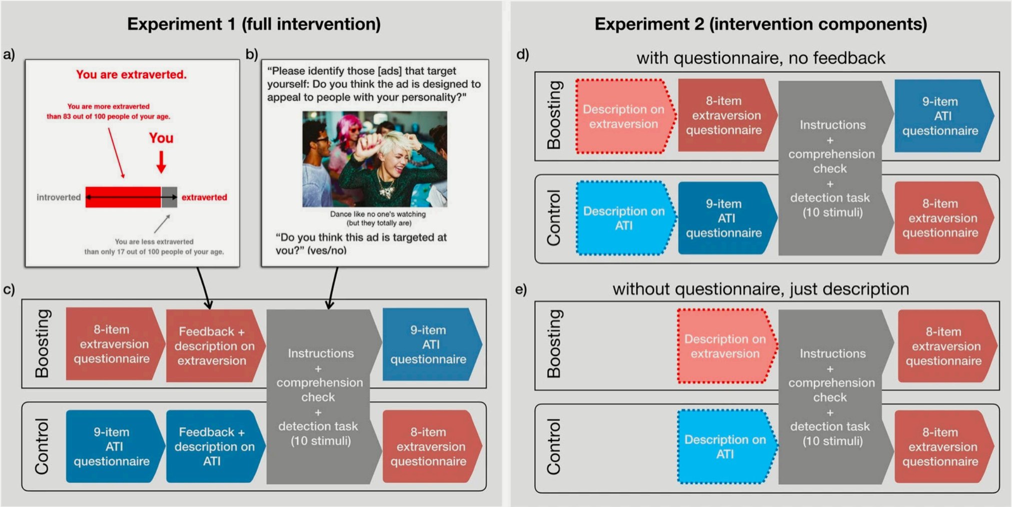
- Predictive Analytics for Ad Placement leverages machine learning and statistical algorithms to forecast future outcomes based on historical data. Advertisers harness predictive analytics to optimize their ad placements by determining which ads should be displayed to which users at what times for the maximum effect. An illuminating example is the 2018 Brazilian presidential election, where predictive analytics were purportedly deployed to deliver hyper-targeted ads designed to exploit voters' fears and biases. Similarly, during the 2020 South Korean legislative election, predictive analytics were used to tailor campaign messages to different demographic groups, maximizing the impact of the advertisements.

- Real-time Bidding (RTB) is an integral component of programmatic advertising, where advertising inventory is traded on a per-impression basis through an instant auction. This allows for precise micro-targeting and efficient ad placement. RTB was utilized extensively during the 2019 Australian federal election, where campaigns exploited the feature to bid competitively for ad slots targeting key voter groups.

- Behavioral Targeting is a strategy that segments audiences based on their online behavior patterns, including pages viewed, search queries, and links clicked. This technique was harnessed in the lead-up to the 2020 Taiwan presidential election, where personalized political ads were served to specific voting blocs based on their online behaviors. Similarly, during the 2021 Israeli general election, behavioral targeting was utilized to deliver highly targeted political messages aimed at specific demographic groups.

- Micro-targeting often includes the use of Dark Ads, which are ads aimed at specific users and are invisible to the general public. Advanced AI and machine learning algorithms, which scrutinize users' online behaviors, interests, and demographics, are utilized to create these ads. These techniques have significant implications for Foreign Information Manipulation and Interference (FIMI) Tactics, Techniques, and Procedures (TTPs), as they enable the creation of highly customized disinformation campaigns. The discreet circulation of dark ads, shielded from broader public scrutiny, allows for the subtle manipulation of individuals and groups. Notably, during the 2020 Myanmar general election, dark ads were used by various actors to disseminate divisive political messages, taking advantage of the limited visibility to bypass scrutiny.

- A/B Testing and Content Optimization involve comparing different versions of content to ascertain which one yields the best response. Micro-targeting allows for granular level A/B testing, and fine-tuning content to maximize engagement. These methods, frequently employed in the digital marketing world, have notable applications in the context of FIMI Tactics, Techniques, and Procedures (TTPs). By utilizing A/B testing and content optimization, actors can enhance the effectiveness of their disinformation campaigns. For instance, during the 2018 Mexican general election, A/B testing was used by various actors to refine their political messages and maximize voter engagement.


Experiment 1 (full intervention)
a)
You are extraverted.
You are more extraverted than 83 out of 100 people of your age.
You
introverted
extraverted
You are less extraverted than only 17 out of 100 people of your age.
b)
"Please identify those [ads] that target yourself: Do you think the ad is designed to appeal to people with your personality?"
Dance like no one's watching (but they totally are)
"Do you think this ad is targeted at vou?" (ves/no)
c)
Boosting
8-item extraversion questionnaire
Feedback + description on extraversion
Instructions + comprehension check + detection task (10 stimuli)
9-item ATI questionnaire
Control
9-item ATI questionnaire
Feedback + description on ATI
8-item extraversion questionnaire
Experiment 2 (intervention components)
d)
with questionnaire, no feedback
Boosting
Description on extraversion
8-item extraversion questionnaire
Instructions + comprehension check + detection task (10 stimuli)
9-item ATI questionnaire
Control
Description on ATI
9-item ATI questionnaire
8-item extraversion questionnaire
e)
without questionnaire, just description
Boosting
Description on extraversion
Instructions + comprehension check + detection task (10 stimuli)
8-item extraversion questionnaire
Control
Description on ATI
8-item extraversion questionnaire

There have been numerous microtargeting optimization studies in recent years that are leveraging profiling techniques with survey experiments. Source: Lorenz-Spreen, P., Geers, M., Pachur, T. et al. Boosting people's ability to detect microtargeted advertising. Sci Rep 11, 15541 (2021). https://doi.org/10.1038/s41598-021-94796-z

DISINFORMATION ECOSYSTEM IN TURKEY

A comprehensive analysis of the Turkish disinformation ecosystem was previously published by EDAM.[69] In recent years, Turkey has been grappling with the extensive spread of disinformation within its information landscape. The rise of digital technologies and social media platforms has amplified the reach and impact of false narratives, misleading content, and manipulative propaganda, threatening the very fabric of its democracy. Within the labyrinth of Turkey's disinformation ecosystem, several troubling forms of deceptive content have flourished, exploiting the vulnerabilities of modern communication channels. Manipulated media, which includes doctored images, manipulated videos, and altered audio recordings, are cleverly crafted to distort reality and fabricate false narratives. Advanced editing tools and deepfake technology have enabled perpetrators to create remarkably convincing disinformation, making it challenging for the average user to differentiate between authentic and manipulated content. Disinformation is by most measures a form of political communication, rather than an anomaly of communication in Turkey, systematically deployed both by the opposition and government channels

False claims and conspiracy theories play a strategic role in disinformation campaigns, preying on people's emotions and exploiting societal fault lines. The use of psychological targeting and microsegmentation techniques has become prevalent in shaping disinformation narratives to resonate with specific target

[69] Unver, Hamid Akin, Russian Disinformation Ecosystem in Turkey (March 8, 2019). EDAM Reports, 2019, Available at SSRN: https://ssrn.com/abstract=3534770. Kirdemir, Baris. "Exploring Turkey's Disinformation Ecosystem: An Overview." Centre for Economics and Foreign Policy Studies, 2020. http://www.jstor.org/stable/resrep26087.

audiences. For instance, in the aftermath of the 2016 coup attempt, false narratives of both pro- and anti-government sentiments were disseminated through social media platforms to incite fear and anger among specific demographic groups, leading to heightened polarization and social unrest.[70] The rise of online trolling and the proliferation of coordinated disinformation campaigns are perhaps the most insidious aspects of the problem. Well-organized groups or individuals leverage social media to disseminate false narratives with unprecedented speed and scale. The use of automation and social media bots amplifies disinformation, enabling it to reach millions of users within a short period. These campaigns often exploit algorithmic biases on social media platforms, which prioritize sensational and divisive content that garners higher engagement, inadvertently facilitating the spread of deceptive narratives.[71]

In the run-up to Turkey's closely contested 2023 elections, disinformation has emerged as a major concern. Both President Recep Tayyip Erdogan and opposition leader Kemal Kilicdaroglu have accused each other of employing deceptive tactics. In one instance, a video montage was shown at a rally, giving the impression that Kilicdaroglu was aligned with banned PKK members.[72] Kilicdaroglu has accused 'foreign hackers' recruited by Erdogan's team of preparing deepfake content to discredit rivals ahead of the election.[73] The use of disinformation has intensified on social media platforms, and both sides have made accusations against each other. Turkey's parliament has passed a law criminalizing the spread of 'fake news,' but critics argue that this law has led to a 'chilling effect' on journalists and critical voices. A flurry of manipulated images and cropped or taken-out-of-context content, disseminated on social networks and during meetings from both the government and opposition ranks. Fake campaign literature has also been employed, with one leaflet claiming to be from Kilicdaroglu's team falsely promising the withdrawal of troops from Syria and halting military operations against the PKK.[74]

Another notable case study is the dissemination of disinformation during Turkey's involvement in the Syrian conflict.[75] Throughout the conflict, various actors exploited social media platforms to advance their agendas, manipulating public sentiment and influencing policy decisions. The Syrian refugee crisis became a contentious issue within Turkey, with disinformation campaigns seeking to sway public opinion and fuel anti-refugee sentiments. False narratives were spread, falsely linking refugees to criminal activities and portraying them as an economic burden on the host country. Additionally, sensationalized and fabricated stories

[70] Akgül, Harun Güney. "Fake news as a tool of populism in Turkey: The Pastor Andrew Brunson case." Polish Political Science Review 7, no. 2 (2019): 32-51.

[71] Furman, Ivo, and Asli Tunc. "The end of the Habermassian ideal? Political communication on Twitter during the 2017 Turkish constitutional referendum." Policy & Internet 12, no. 3 (2020): 311-331.

[72] "Disinformation Adds Dark Note to Pivotal Turkish Election." France 24, 12 May 2023, www.france24.com/en/live-news/20230512-disinformation-adds-dark-note-to-pivotal-turkish-election. Accessed 3 Aug. 2023.

[73] "Turkey's Opposition Accuses Russia of Interfering in Elections." Al Jazeera, 12 May 2023, www.aljazeera.com/news/2023/5/12/turkeys-kilicdaroglu-accuses-russia-of-interfering-in-elections. Accessed 3 Aug. 2023.

[74] Ioannou, Demetrios. "Deepfakes, Cheapfakes, and Twitter Censorship Mar Turkey's Elections." Wired, 26 May 2023, www.wired.com/story/deepfakes-cheapfakes-and-twitter-censorship-mar-turkeys-elections/. Accessed 3 Aug. 2023.Ioannou, Demetrios. "Deepfakes, Cheapfakes, and Twitter Censorship Mar Turkey's Elections." Wired, 26 May 2023, www.wired.com/story/deepfakes-cheapfakes-and-twitter-censorship-mar-turkeys-elections/. Accessed 3 Aug. 2023.

[75] Salem, Fatima K. Abu, Roaa Al Feel, Shady Elbassuoni, Mohamad Jaber, and May Farah. "Fa-kes: A fake news dataset around the syrian war." In Proceedings of the international AAAI conference on web and social media, vol. 13, pp. 573-582. 2019.

about refugee-related incidents were shared widely, deepening divisions and exacerbating tensions between different segments of the Turkish society.

Disinformation campaigns during the Syrian conflict also sought to portray terrorist groups in a particular light to serve political interests. The narratives varied depending on the perpetrators, but some groups aimed to downplay the activities of certain terrorist organizations or exaggerate the threat posed by others.. Furthermore, the spread of fake news and clickbait articles contributed significantly to the disinformation problem during sensitive political events or elections. Opportunistic actors capitalized on sensationalized stories and click-worthy headlines, designed solely to generate high levels of engagement without regard for their accuracy. For instance, during election periods, false stories about political candidates, parties, and potential outcomes were deliberately circulated to create confusion, sow mistrust, and manipulate voters' choices. During critical policy decisions related to the Syrian conflict or refugee management, disinformation campaigns were employed to influence policy makers and public opinion.[76]

Foreign state-origin disinformation campaigns in Turkey have become a concerning trend, with several notable cases involving various countries attempting to influence Turkish politics, society, and international relations. Russia, known for its expertise in information warfare, has been particularly active in conducting disinformation campaigns in Turkey. One prominent case involved Russia's attempts to shape narratives around Turkey's military involvement in Syria and its relations with other actors in the region. Russian state-controlled media outlets and social media bots have disseminated misleading content to sway public opinion, fuel anti-Turkey sentiments, and undermine Turkey's regional initiatives. These campaigns aimed to manipulate perceptions of Turkey's actions in Syria and portray the country in a negative light on the global stage.

During times of tension between Turkey and the European Union (EU), the Turkish government has accused European countries of inciting disinformation campaigns to influence public opinion and decision-making in Turkey.[77] Alleged false narratives and misleading information have been disseminated to create discord between Turkey and its European partners, and to portray the EU negatively in the eyes of the Turkish public. For example. Greece and Turkey have a long history of regional rivalry, and disinformation has been used to exacerbate tensions. In some cases, disinformation campaigns attributed to Greece aimed to distort Turkey's actions in the Aegean Sea and Cyprus and manipulate public perception to portray Turkey as an aggressor were highlighted by the Turkish government.[78] These campaigns aimed to fuel anti-Turkey sentiments within Greece and among international observers.

[76] Yerlikaya, Turgay, and Seca Toker Aslan. "Social Media and Fake News in the Post-Truth Era." Insight Turkey 22, no. 2 (2020): 177-196. Filibeli, Tirşe Erbaysal, and Can Ertuna. "Sarcasm beyond hate speech: Facebook comments on Syrian refugees in Turkey." International journal of communication 15 (2021): 24.

[77] Please see Heinrich Böll Stiftung Project on homegrown EU disinformation: https://eu.boell.org/en/homegrown-disinformation

[78] Markham, Lauren, and Lydia Emmanouilidou. "How Free Is the Press in the Birthplace of Democracy?" The New York Times, 26 Nov. 2022, www.nytimes.com/2022/11/26/business/greece-journalists-surveillance-predator.html.

Turkish-American relations and Turkey's position within NATO have also been targeted by foreign disinformation campaigns, with the aim of shaping public perception and influencing Turkish policies. In some cases, state-origin disinformation has sought to sow mistrust between Turkey and its NATO allies and to undermine support for the alliance within Turkey. False narratives have been propagated to undermine Turkey's role in regional security and portray the country as a contentious ally.[79] During periods of regional conflict, particularly in relation to the Nagorno-Karabakh conflict involving Azerbaijan and Armenia, disinformation campaigns originating from both the EU and Armenia, or sympathizers have targeted Turkey. False narratives have been propagated to demonize Turkey's role in the region and to fuel anti-Turkey sentiments.[80] These campaigns aimed to exploit historical grievances and geopolitical tensions to advance Armenia's interests and tarnish Turkey's image. Iran and Turkey have complex relations, and disinformation campaigns originating from Iran have aimed to influence Turkish public opinion and shape perceptions of Iran's actions in the region. False narratives have been used to undermine Turkey's regional influence and to portray Turkey's policies negatively. These campaigns have sought to exploit religious and sectarian differences to sow discord between the two nations.

Automated accounts and trolls played an instrumental role in these campaigns, flooding social media platforms with deceptive content that amplified certain viewpoints and undermined alternative perspectives. In one specific case, during a critical debate in the Turkish Parliament on a new refugee policy, a coordinated disinformation campaign flooded social media platforms with misleading statistics, false stories of refugee crimes, and fabricated accounts of public dissent. This campaign aimed to sway public sentiment against the new policy and to discredit the government's handling of the refugee crisis. The disinformation not only disrupted the democratic process but also deepened societal divisions and eroded public trust in institutions.[81]

Understanding the actors behind Turkey's disinformation ecosystem is vital to comprehend the motives and strategies at play. Among the primary perpetrators are state actors, who may utilize disinformation campaigns to control the narrative, suppress dissent, or shape public opinion in their favor. Political groups also play a significant role, employing disinformation to undermine rivals, sway voters, and discredit opponents. Social media bots and trolls, often funded by shadowy entities, are instrumental in amplifying specific messages and creating the illusion of widespread support for both pro- and anti-government viewpoints. Moreover, foreign actors engage in disinformation campaigns to influence Turkish politics, sow discord, and further their interests in the region.[82]

[79] Bernstein, Jonas. "US Dismisses Russian Allegations of Turkey's Involvement in Trading IS Oil." VOA, 2 Dec. 2015, www.voanews.com/a/russia-claims-to-have-proof-turkey-involved-in-is-oil-trade/3084253.html. Accessed 3 Aug. 2023.

[80] Atanesyan, Arthur. "Media framing on armed conflicts: limits of peace journalism on the Nagorno-Karabakh conflict." Journal of intervention and statebuilding 14, no. 4 (2020): 534-550.

[81] Saka, Erkan. Social media and politics in Turkey: A journey through citizen journalism, political trolling, and fake news. Lexington Books, 2019.

[82] Unver, Hamid Akin, Russian Disinformation Ecosystem in Turkey (March 8, 2019). EDAM Reports, 2019, Available at SSRN: https://ssrn.com/abstract=3534770

To effectively combat the pervasive disinformation ecosystem, Turkey can adopt a comprehensive set of proactive measures that address various aspects of the problem. Firstly, it is crucial to recognize the correlation between disinformation and government censorship. Research has shown that disinformation tends to thrive in environments where governments suppress free information flows. To tackle disinformation at scale, it is essential to foster an open and free information ecosystem. This requires safeguarding media freedom, protecting journalists' rights, and promoting an environment where diverse viewpoints can be expressed without fear of censorship or reprisal. Empowering independent media and fact-checking organizations can further contribute to countering disinformation while upholding democratic values.

Secondly, promoting and supporting independent fact-checking organizations is pivotal in verifying information and debunking false claims. Fact-checkers play a crucial role in holding media outlets and public figures accountable for spreading disinformation. By providing accurate and credible information to the public, fact-checkers help build media literacy and enable citizens to discern between reliable sources and deceptive content. Thirdly, collaboration with social media platforms is essential to enforce stricter policies against disinformation. These platforms have become major battlegrounds for the spread of deceptive content, and tech companies play a crucial role in limiting the reach of false narratives. Implementing robust content moderation mechanisms and transparent algorithms can help reduce the virality of disinformation, making it less likely to go viral and reach a broader audience.

Additionally, ensuring transparency in media ownership is vital to mitigate the influence of biased reporting and propaganda. When media ownership structures are disclosed, the public can better assess potential biases in the information they consume. This transparency fosters greater accountability and helps citizens make informed decisions about the credibility of the news sources they rely on. Lastly, fostering international cooperation in addressing cross-border disinformation campaigns and foreign interference is crucial. Disinformation threats often transcend national borders, and collaborative efforts among nations can strengthen collective resilience against such challenges. Initiatives like joint NATO disinformation tracking and attribution mechanisms can help identify the sources of disinformation and respond effectively to coordinated campaigns.

CONCLUSION

Concluding this technical examination of the impact of advanced technologies on disinformation and information manipulation, a profound realization emerges of their multifaceted influence on global geopolitics. The advent of Machine Learning (ML), Deep Learning (DL), and Artificial Intelligence (AI) has ushered in a new era of potent dual-use technology, capable of both exacerbating and mitigating disinformation challenges. This report has explored the escalating threat posed by deep fakes, synthetic media generated through complex machine learning models, and their potential to ignite societal unrest, erode trust in media, and escalate conflicts, permeating the geopolitical landscape. The sophistication of deep fakes makes them increasingly challenging to detect and can lead to the dissemination of fabricated content with severe real-world consequences. By merging the likeness of an individual with manipulated audio, these malicious tools can create videos that appear genuine, causing confusion and polarization in society.

State and non-state actors can exploit deep fakes to manipulate public perception, discredit political opponents, and incite tension between nations, amplifying the impact of disinformation campaigns. Moreover, the report has shed light on the role of AI-driven algorithms in shaping information consumption on social media platforms and other digital channels. The intricate design of these algorithms inadvertently fosters echo chambers, where users are exposed primarily to content aligned with their pre-existing beliefs. This filtering of information can lead to information silos, where individuals are isolated from diverse perspectives, enabling the unchecked spread of disinformation. By tailoring content recommendations to maximize engagement, these algorithms can inadvertently reinforce existing biases and exacerbate societal divisions.

While the challenges posed by advanced technologies in the realm of disinformation are formidable, the report also highlights the potential of AI and ML in combating these issues. AI-powered tools can be harnessed to detect and analyze disinformation patterns, thereby enabling swift identification of deep fakes and other deceptive content. By leveraging advanced data analytics, researchers and fact-checkers can detect anomalies and inconsistencies in media to mitigate the impact of disinformation campaigns. Additionally, blockchain technology holds promise in ensuring the authenticity of information by providing an immutable and transparent record of content provenance, enabling users to verify the veracity of shared information. However, the responsible deployment of advanced technologies requires careful consideration of potential risks and unintended consequences. The report emphasizes the necessity of stringent regulatory frameworks to govern the use of such technologies, preventing malicious actors from exploiting these tools to spread disinformation.

To address the transnational nature of disinformation, international collaboration is paramount, with countries working together to establish harmonized policies that foster transparency, accountability, and information integrity. Recognizing that technology is a double-edged sword, the report underscores the importance of digital literacy initiatives to empower individuals in distinguishing factual information from disinformation. By equipping users with critical thinking skills and media literacy, societies can build resilience against the manipulation of digital realities. Through public awareness campaigns and educational programs, individuals can be better prepared to discern deceptive content and make informed decisions.

In conclusion, the impact of advanced technologies on disinformation, information manipulation, and geopolitics is far-reaching and multifaceted. Addressing these challenges demands a concerted effort from governments, tech companies, researchers, and civil society. By harnessing the potential of AI, ML, and blockchain responsibly, coupled with robust regulatory frameworks and digital literacy initiatives, we can pave the way towards a more secure and informed information landscape, safeguarding the foundations of democracy and public discourse. This comprehensive investigation serves as a stepping stone towards understanding the intricacies of the evolving disinformation landscape and building effective strategies to counter this complex and ever-changing threat.

Ekonomi ve Dış Politika
Araştırmalar Merkezi

edam

Address : Hare Sokak NO:16 AKATLAR 34335 İstanbul/Türkiye
Phone : +90 212 352 18 54
Fax : +90 212 351 54 65
Email : info@edam.org.tr