Algorithmic Warfare: Next Generation Hybrid Warfare Strategy

Abstract

In the modern era, social media platforms serve as a weapon in many conflicts. While they are useful for disseminatingimportant information, social media can also be used in espionage, psychological operations, and disinformation campaigns. A critical component to these platforms is their recommendation algorithms, which were primarily designed to facilitate engagement and enhance the user experience. Through misuse and abuse, these algorithms have been repurposed as tools for military action and strategic advantage[15]. With the potential for social media to sow the seeds of discord and confusion in rival countries, the adoption of these next generation strategies has risen to the forefront of the strategy books for nation state actors to include Russia, China, and Iran[1]. The continued manipulation of these algorithms has shown that this behavior is no longer a byproduct of propaganda, but a deliberate strategic asset in targeting adversaries in modern warfare (Bicheng, 2018). Due to the nature of social media playing a large role in societal discourse, significant security concerns have been raised to increase the pressure for social media platforms to have algorithm transparency, regulatory oversight, and technological controls in place to lessen the impact of these actions to the civilian population. This study aims to analyze how social media algorithms enable state sponsored objectives, evaluate the effectiveness of countermeasures, and propose a multi-layered mitigation strategy to reduce the collateral effects on civilian populations[19].

Keywords: Algorithm Manipulation, Disinformation, Cyber Warfare, Engagement Algorithms, Psychological Operations (PSYOPS), Recommendation Systems, Digital Influence Operations

Introduction

As the landscape of modern conflict continues to evolve, so do the domains of warfare. The inclusion of cyberspace and information warfare to the traditional land, sea, air, and space domains is nothing new, but what constitutes the domain of cyberspace is constantly expanding and often a point of contention for the jurisdiction of the domain[16]. In the last decade social media has been deployed on the international stage of warfare in many different avenues to test its ability to augment current strategies as they continually adapt to the asymmetric advantages the technology provides in modern hybrid warfare[17]. This modern frame of warfare blends conventional kinetic military strategy with non-traditional tactics such as cyber warfare, disinformation campaigns, economic coercion, and psychological operations. One of the deadliest tools in the arsenal of hybrid warfare is the use of social media platforms as it serves as a low resource conduit for many of these facets of non-kinetic campaigns.

While the recommendation algorithms of popular social media platforms such as Facebook, TikTok, X (formerly Twitter), and YouTube were originally created to enhance the user experience and drive engagement to the platform and its content, these algorithms have become increasingly vulnerable to manipulation[6]. While these systems are opaque and most visible manipulation is less of an exploitation of the code itself, but an improper use of the platform to skew recommendation engines towards certain content by optimizing platform specific attributes of said content, these strategies are used by benign entities looking to increase their audience as well as malicious actors intending to widen their sphere of influence. This behavior is critical to its applicability in warfare to enable entities to shape discourse, influence user behavior, and

manipulate public sentiment. Unlike social media personalities looking to make money, adversarial states alike have recognized the power of these systems to conduct influence operations against adversarial states through non-kinetic means.

The recent geopolitical landscape has highlighted social media's impact in utilizing these platforms for nation state objectives. During the height of the ongoing Ukraine conflict. Russian nation state operatives leveraged platforms to spread disinformation about battlefield outcomes and weaken global support for Ukraine. While on the surface, the use of modern technology such as deepfake content in the artificially generated video of Ukranian President Volodymyr Zelenskyy's surrender may seem like the largest technological threat, the real threat is significantly more visible[2].The use of platform specific engagement mechanisms such as the inclusion of emotionally provocative thumbnails and keyword driven and post titles designed to exploit YouTube's engagement metrics and maximize click through rates were an the fuel to the viral success of Anti-Ukranian social media campaigns. These nefarious tactics also saw the main stage in the recent Isreal-Hamas conflict as platform algorithms played a key role in engaging action in global audiences utilizing popular social media and communication platforms such as TikTok to fuel this ideological dichotomy. The use of TikTok was particularly noticeable as the inclusion of popular audio and visual formats to encapsulate political narratives in otherwise non-political content created an illusion of organic non-sponsored political discourse[7]. In turn allowing this nation state messaging to be promoted and disseminated by the platform algorithm. More overtly, the Unites States and European intelligence agencies have raised concerns about China's use and direct manipulation of TikTok to influence public perception and enable state sponsored campaigns for the suppression of non-favorable content and promotion of

politically divisive content as well as content favorable to their national objectives through their direct influence over the application[8].

Much of this behavior happens "behind the scenes" and is not as visible to the platform users as a whole. The use of strategies such as thumbnail optimization, engagement farming, and hashtag hijacking are all methods used to manipulate the metrics that recommendation engines use as input on their decisions. On Platforms such as Instagram and Facebook, posts are generally optimized to be visually engaging and embed emotionally charged messaging to promote the manual distribution of the content into personal spheres of influence through sharing. On other platforms such as reddit, coordinated groups and artificial bot farms are used to fabricate trending discussion through the use of comment seeding and metric manipulation such as upvote or like farming[18]. The use of these techniques by nation state actors highlights the fact that integration into social media is no longer an unintended consequence of digital involvement, but an evolving strategy worthy of nation state resources and attention. The ability to remotely influence discourse and fracture societies from within without the need for kinetic measures and minimal oversight underscores the danger of this domain of warfare.

This paper argues that recommendation algorithms are no longer neutral tools but have become a key component to modern hybrid warfare strategies. It also seeks to examine the nature of the use of this technology in military strategy and proposes a mitigation framework to reduce the asymmetric potency of these operations on civilian populations.

The Ins and Out of Algorithmic Warfare

The use and weaponization of social media and its recommendation algorithms marks the turn for a new generation of hybrid warfare. The recommendation engines that drive popular platforms are often driven by opaque machine learning models that curate and prioritize consent

based on a weighted inference of user behavior. Because these algorithms prioritize engagement to retain user attention, they analyze metrics such as watch time, user interaction, and sharing to build feedback loops that prioritize content that incentivizes these behaviors. Although the underlying mechanisms of the algorithms are opaque, these metrics are apparent making the platforms themselves vulnerable to an actor's optimalization of content to satisfy the inputs for promotion.

The low resource overhead of social media platforms in rival nation engagement is a large attractor to its success and proliferation in geopolitical conflict. Many of the underlying platforms have specific optimization strategies that state actors have been seen to utilize to promote their content such as Russian state actors' use of YouTube's click-through rate and session duration optimization. This includes both optimizations to promote search query prioritization and relevant video recommendations through relevant keywords in titles, descriptions and transcripts. This is often combined with the inclusion of emotionally charged thumbnails depicting provoking facial expressions or violence combined with provoking titles to spark interaction. These methods promote the content as it is perceived as engaging and therefore disseminates it to a wider audience of many times unsuspecting users. Mozilla's YouTube Regrets report documented numerous cases where users were recommended extremist political content unrelated to their original searches, which highlights the effect that this type of manipulation can have on the user online experience as a whole[9].

Due to its relatively recent rise in popularity TikTok has been a high value target for Chinese affiliated influence campaigns. The "For You Page" (FYP) is one of the primary ways for TikTok users to discover new content and has been the primary target for campaigns to encapsulate their political messaging into popular trends, memes, or audio snippets. These

campaigns used trending mechanics to capitalize on engagement-based visibility. This can come to a head in situations such as those observed in a Newsguard study showing that users could be subject to disinformation about the Ukranian war within the first hour of signing up for the app without searching for Ukranian content[13]. This is also seen on more long-standing platforms such as Instagram and Facebook as their emphasis on visual content The use of high contrastoverlays, culturally resonant color schemes, and overall suggestive symbolism were often used in these types of operations. These themes were often packaged into memes, infographics, and as well as visual calls to action to optimizetheir ability to be easily viewed, liked, and shared. Meta's Coordinated Inauthentic behavior reports regularly document and posts these actions for accountability purposes.

Across the platforms, the proliferation of things like transparency reports and coordinated inauthentic behavior reports have popularized in recent times to showcase the efforts of these platforms to combat misuse. Throughout these reports it is apparent that botnets and dummy accounts are often used to fake engagement numbers and simulate fabricated consensus for these inauthentic posts. These posts aren't always taken down immediately and are often able to propagate long enough to become "trending" due to their promotion by the algorithms in question. Efforts such as the Standford Internet Observatory's Election Integrity Partnership have showcased the effect of even relatively small botnets and automated accounts increasing a snowball effect to influence algorithmic promotion and in extreme cases societal unrest and physical attacks such as the 2020 US Capitol Attack[14].

Civilian Risk and Impact

The most troubling facet of algorithmic warfare through social media is that the main target is the civilian population. Influence operations are often not targeted towards military or

government entities, but the civilian entities they protect. The impact of this targeting decision as well as the ensuing content erodes democratic institutions through sowing public mistrust and social coercion. Information pollution is described as the deliberate flooding of contradictory information and resources, therefore influencing a person's ability discern legitimate and illegitimate facts. According to the Reuters Institue Digital News Report (2023) 64% of international survey participants reported encountering misinformation on a weekly basis, with a significant portion of it being on social media[10].

Algorithmic amplification of polarizing content can lead to radicalization and social fragmentation. Controversial topics such as US involvement in foreign affairs and immigration were often found to be algorithmically promoted by engagement systems on social platforms leading to the development of extreme narratives[5]. The Pew Research Center also found that users who engage with partisan content are more likely to be exposed to misinformation and conspiracy theories due to the positive feedback loop of the engagement system (Kennedy). These situations can lead to an online community built upon like-minded individuals effectively serving as an echo chamber of extremist ideals in either direction, which in turn can lead to protests, attacks, and even political collapse.

The most apparent and quantifiable threat of algorithmic warfare is its impact of societal trust in any system political or not. A Yale Study centered on the effects of Fake news concluded that after being exposed to fake news, people often retain those beliefs even after the original information is retracted or corrected[4]. During the 2020 US election cycle, Twitter/X's transparency archives showed that false narratives about voter fraud generated significantly more user engagement than the factual corrections, this is also highlighted by Meta's Ad library which showed that divisive political ads and campaigns outperformed more neutral ones in terms of

reach and engagement. These facts highlight the fact that the truth often struggles to garner the same engagement as crafted falsehoods and in turn is less effective at shaping public opinion and sentiment.

Countermeasures

As the strategic manipulation of social media algorithms becomes integral to modern warfare, a variety of countermeasures have emerged with a varying degree of effectiveness. Many of these efforts aim to limit the impact of disinformation campaigns and protect users from manipulation but often remain inconsistent and fragmented in their approach. These countermeasures are also often reactionary and move too slowly to keep up with the constantly evolving technological landscape.

From a technological standpoint, social media companies such as Meta, Google, and X have all implemented a variety of moderation solutions. These range from the hiring of moderation staff to the integration of artificial intelligence to detect and remove content and accounts that break their terms of service and conduct malicious or otherwise inauthentic behavior. These companies have also worked to release transparency reports to expose the campaigns aimed at misusing their platform for malicious or inauthentic purposes as well as highlight their efforts to thwart this kind of behavior. Although these methods are somewhat effective at increasing the barrier of effectiveness, they are too reactionary and in turn are too slow to mitigate the effects of content before they are removed and not adaptive enough to respond to fast moving threats. This is amplified by the fact that not only are the recommendation algorithms opaque, but so are the countermeasures making it difficult to understand how moderation decisions are made or even have them to be independently audited for effectiveness and are often cited as being inconsistent.

In the regulatory space, many governments have tried initiating policies to increase the transparency and accountability of these platforms. In the European union the Digital Services Act requires these platforms to be more transparent about their algorithm and advertising systems as well as provide the facilities for opting out of algorithmically driven feeds. In the United States, the creation of the Foreign Malign Influence Center showcases an understanding of the risks to the US democratic process imposed by disinformation campaigns. Many countries nationwide have instituted policies around cyber influence operations, but many of them are either not enforceable due to jurisdiction issues or conflicts with freedoms of speech and expression or lack the appropriate punishment to incentivize compliance.

Beyond platform and government countermeasures, independent organizations have emerged to address the critical impacts of algorithmic warfare strategies. Groups such as the EU Disinformation Labs have emerged as investigative research point of presence on disinformation and state sponsored social media campaigns[11]. These organizations along with academic and non-profit organizations such as Tracking Exposed and Mozilla provide independent analyses of social platforms and their actions around misinformation and recommendation manipulation and bias. While these groups provide a breadth of informative and useful metrics, this is done with no additional platform visibility which limits the effectiveness of their actions and ultimately curtails their effectiveness to academic exercise and user awareness campaigns although despite the growing awareness these countermeasures are often insufficient against the scope and sophistication granted by nation state backing. Until platforms, governments, and civil societies can converge on enforceable and informed standards for accountability, these results will remain inconsistent.

Proposed Mitigation Framework

Given the fast moving and adaptive nature of the technological landscape, reactionary methods are no longer plausible for this approach. A proactive approach requires the cooperation of multiple fields to address the obvious issue of content moderation as well as the systematic vulnerabilities both within the technology and the political landscape. To be effective this framework must emphasize accountability and transparency, legal enforcement and compliance, and most of all user protection. The goal of such a framework is not simply to mitigate state sponsored manipulation, but to safeguard users and provide them with digital platforms that emphasize authenticity.

Primarily, the algorithmic decision framework requires fundamental reform. Optimizing recommendation systems for engagement fundamentally prioritizes emotionally charged content to drive intensity and time spent on platform. This favors content that is conspiratory, divisive, or otherwise provocative which should not be the fundamental goal of a global social network. Countering this does not require a complete overhaul, but an alteration to the current system that weighs the integrity of an account based on trustworthiness, factual accuracy, and commitment to the terms of service. While this approach undoubtedly raises the concern of bias, the integration of community moderation and third-party fact checking could prioritize the truth while limiting subjectivity.

Central to this mitigation strategy is the tenant of transparency. Social media platforms may not be forced to fully disclose the innerworkings of their recommendation algorithms, but they should be required to standardize the inputs required for these models as well as provide a method for reporting the weighting of their influence to better understand their metric prioritization. This is a compromise that affords social platforms the trade advantages that their algorithms grant them as well as provides the public the visibility to better understand the digital

landscape in which they reside in. Such transparency should also enable independent audits of algorithmic performance and bias detection to ensure compliance with agreed upon standards of input metrics. The European Digital Services Act provides a good starting point but is highly localized and must be amended to standardize transparency requirements across countries and platforms.

The integration of proper source verification methods should be another technological measure integrated into this system. To reduce the speed of deceptive media, instituting methods that enforce non-repudiation and identity verification should be paramount to prevent the spread of false accounts and attribute harmful content to the individuals responsible. The inclusion of cryptographic signatures for verified content can help achieve both goals as content can not only be attributed to an account, but an individual can assert the validity of their own posts. Other standards such as the Coalition for Content Provenance and Authenticity (C2PA) have already begun development and integration of similar technologies in social platforms. This opens the possibility of auto-removal of reported content that is not verified or the automated flagging and action of materials that break terms of service without user interaction.

While the technology provides a foundation for the strategy, it must be paired with educational efforts for users to understand the impact of both the dangers of misuse as well as the countermeasures in place for protection. Governments, NGOs, and academic institutions should collaborate on digital literacy initiatives that teach users things such as how recommendation algorithms work and how to evaluate online content. The collaborative nature of this strategy not only eliminates bias from a single entity in the disseminated training but also eliminates ownership which allows the information to be permissively licensed and freely spread and improved upon.

With the educational initiatives so must come the norm building in society and governance. Nation state algorithmic manipulation must be addressed as a matter of international security in the same manner that other forms of cyber and kinetic warfare are governed. There is a growing need for treaties or cooperative laws guidelines that govern acceptable use of social media and that prohibit misuse by state actors. This would include agreements on non-interference with democratic processes and information requests and sharing mechanisms for investigations. International and regional institutions such as the UN, NATO, and others should act as the arbiters of these efforts to ensure proper oversight and representation as well as establishing protocols for attribution, enforcement, and punishment.

Bibliography

1. Beauchamp-Mustafaga, N., Green, K., Marcellino, W., Lilly, S., & Smith, J. (2024). Dr. Li Bicheng, or How China Learned to Stop Worrying and Love Social Media Manipulation.
2. Boháček, M., & Farid, H. (2022). Protecting President Zelenskyy against deep fakes. arXiv preprint arXiv:2206.12043.
3. Bossetta, M. (2018). The weaponization of social media: Spear phishing and cyberattacks on democracy. Journal of international affairs, 71(1.5), 97-106.
4. Buonomano, Lydia. "Cognitive substrates of belief in fake news." Yale University. April 23 (2020).
5. Diaz Ruiz, Carlos, and Tomas Nilsson. "Disinformation and echo chambers: how disinformation circulates on social media through identity-driven controversies." Journal of public policy & marketing 42, no. 1 (2023): 18-35.
6. Dmitrievna, Ruzanova Elizaveta, and Tkacheva Ekaterina Olegovna. "IMPACT OF SOCIAL PLATFORM ALGORITHMS ON AD CONTENT DISTRIBUTION AND USER ENGAGEMENT." ЭКОНОМИКА, БИЗНЕС, ИННОВАЦИИ: АКТУАЛЬНЫЕ ВОПРОСЫ ТЕОРИИ И (2025): 44.
7. González-Esteban, José-Luis, Carmen Maria Lopez-Rico, Loraine Morales-Pino, and Federico Sabater-Quinto. "Intensification of Hate Speech, Based on the Conversation Generated on TikTok during the Escalation of the War in the Middle East in 2023." Social Sciences 13, no. 1 (2024): 49.
8. Finkelstein, Danit, Sonia Yanovsky, Jacob Zucker, Anisha Jagdeep, Collin Vasko, Ankita Jagdeep, Lee Jussim, and Joel Finkelstein. "Information manipulation on TikTok and its

relation to American users' beliefs about China." Frontiers in Social Psychology 2 (2025): 1497434.

9. McCrosky, Jesse, and Brandi Geurkink. "YouTube Regrets: A crowdsourced investigation into YouTube's recommendation algorithm." Mozilla Foundation. Retrieved November 15 (2021): 2021.
10. Newman, Nic, Richard Fletcher, Kirsten Eddy, Craig T. Robertson, and Rasmus Kleis Nielsen. "Digital news report 2023." (2023).
11. Saravyn, Hanna. "Warfare & Social Media: Disinformation and Propaganda in the Digital Age." (2023): 1-48.
12. SAYGILI, Neriman. "INFORMATION POLLUTION, MANIPULATION AND ETHICS IN THE AGE OF TECHNOLOGY AND SOCIAL MEDIA." International Online Journal of Education & Teaching 10, no. 3 (2023).
13. Serafin, Tatiana. "Ukraine's President Zelensky takes the Russia/Ukraine war viral." Orbis 66, no. 4 (2022): 460-476.
14. Vishnuprasad, Padinjaredath Suresh, Gianluca Nogara, Felipe Cardoso, Stefano Cresci, Silvia Giordano, and Luca Luceri. "Tracking fringe and coordinated activity on Twitter leading up to the US Capitol attack." In Proceedings of the international AAAI conference on web and social media, vol. 18, pp. 1557-1570. 2024.
15. Wahab, M. I. Weaponization of Social Media: Challenges and Responses.
16. Libicki, Martin C. "Cyberspace is not a warfighting domain." Isjlp 8 (2012): 321.
17. Warren, Jason W. "On "Social Media Warriors: Leveraging a New Battlespace"." The US Army War College Quarterly: Parameters 50, no. 3 (2020): 13.

18. Weerasinghe, Janith, Bailey Flanigan, Aviel Stein, Damon McCoy, and Rachel Greenstadt. "The pod people: Understanding manipulation of social media popularity via reciprocity abuse." In Proceedings of The Web Conference 2020, pp. 1874-1884. 2020.
19. Zaighum, Z., & Ali, R. F. (2025). From Strategy to Tactics: Conceptualizing Weaponization of Digital Media Platforms across Levels of Warfare. Lahore Institute for Research and Analysis Journal, 3, 39-53.