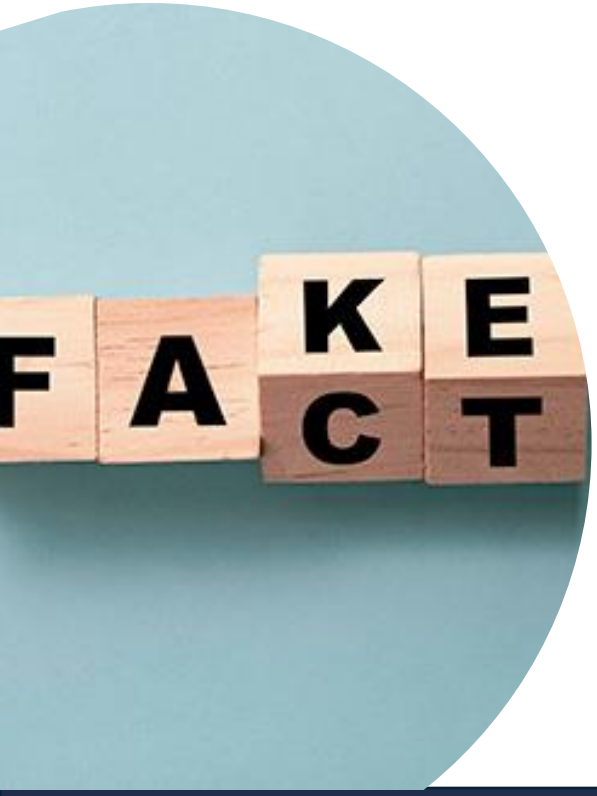
University of Virginia | National Security Data and Policy Institute

China's AI-Powered Disinformation Tactics: Threats and Implications

July 16, 2025

By Charis June Lee



Executive Summary

The People's Republic of China (PRC) government strategically uses disinformation to manipulate public perception both at home and abroad. Domestically, disinformation is a core instrument of the Chinese Communist Party (CCP), used to control public information and discussions. Case studies like the 2019 Hong Kong pro-democracy protests and the 50 Cent Army reflect this. The integration of generative artificial intelligence (GenAI) significantly enhances disinformation campaigns by reducing costs, increasing speed and scale, and creating content with more perceived authenticity. This makes attributing campaigns to state actors even more challenging. Tactics include deepfakes, astroturfing, and fake online personas to push narratives and sow division. Operations such as Spamouflage and interference in the 2024 Taiwanese presidential election reflect the growing sophistication of these efforts. This escalating threat requires a strong policy response across technological, societal, and diplomatic fronts. Policy implications include:

- **AI-generated Content Identification**: Social media platforms and users are increasingly vulnerable to AI-powered disinformation. Policymakers should consider investing in large-scale detection technology and requiring watermarks for AI-generated content.
- **Bilateral Dialogue with the PRC**: Disinformation campaigns are becoming more difficult to attribute to state actors as GenAI allows individuals to hide behind AI-generated personas. Policymakers should pursue increased dialogue with the PRC on GenAI and social media, focusing on shared rules and restrictions.
- **Media Literacy Initiatives**: The barriers to creating high quality disinformation content, like high cost and manpower, are eroding. Policymakers should consider investing in media literacy initiatives to help the public identify AI-generated content.

Introduction

This paper focuses on disinformation case studies and examines how the emergence of GenAI changes the PRC government's current disinformation capabilities. Until recently, disinformation efforts by the PRC government, were limited by scale, credibility, and cost. The emergence of GenAI significantly changes the disinformation landscape. The PRC government's ability to weaponize GenAI for the production and amplification of disinformation presents a growing threat to the United States (US) and its democratic allies. This paper concludes with policy implications for the US should the PRC government continue to integrate GenAI into its disinformation operations.

What is Disinformation?

Disinformation is commonly defined as "deliberately spreading false or inaccurate information to manipulate the opinions and actions of others".[i] A key characteristic is the perpetrator's identity and intention, often a state actor aiming to create division, spread confusion, and erode public trust in institutions. It is difficult to directly tie disinformation campaigns to suspected state actors, especially as disinformation actors hide behind fake personas. The Cybersecurity and Infrastructure Security Agency (CISA) identifies 8 common disinformation tactics, which include deepfakes, conspiracy theories, information gaps, targeted content and more.[ii] The following listed tactics have been linked to pro-China disinformation campaigns:

- **Fake online personas** are false identities created to influence others by appearing credible. This includes fake experts and institutions to lend credibility to the message.[iii]
- **Deepfakes** are images, videos, or audio created using AI that was trained on reference data. Deepfakes can be used to impersonate or falsely attribute actions to individuals.[iv]
- **Astroturfing** is mass posting by fake accounts to mimic grassroots support for a specific narrative. Astroturfing creates false credibility through artificial widespread support.[v]
- **Targeted content** is content tailored to a specific online group based on its values and interests. This can be used to gain influence and credibility within a community.[vi]

PRC Government's Disinformation Strategy

Disinformation falls under the CCP's overseas propaganda strategy.[vii] For the CCP, propaganda is both a necessary and beneficial "practice of governance".[viii] Domestically, the CCP uses censorship and disinformation to control shared information and public discussions. Internationally, CCP propaganda strategy aims to have a CCP-controlled narrative of China "prevail as the only global discourse on the CCP".[ix] Disinformation operations are also considered a component of the People's Liberation Army's (PLA) "Three Warfares" strategy – specifically its public opinion and psychological warfare sides – which aim to break down the adversary's "'will to fight'" and fuel disputes and divisions.[x] Ultimately, disinformation is a core instrument of the CCP to control public perception and is used in large-scale operations both at home and abroad.

2019 Hong Kong Pro-Democracy Protests: In 2019, the Hong Kong government introduced a controversial amendment to a criminal extradition law that would allow Hong Kong residents to be extradited to mainland China for trial.[xi] This sparked widespread protests in Hong Kong, as many protestors believed the amendment would diminish the region's judicial independence.[xii] On July 1, protestors stormed the building of Hong Kong's Legislative Assembly, marking the first move towards organized violence.[xiii] In response, Beijing launched disinformation campaigns to discredit the protestors' credibility, both domestically and abroad.[xiv] In the mainland, Chinese state media amplified violent activities of protestors, suppressed images of pro-democracy slogans, and minimized the scale of the movement and number of protestors.[xv] State media also distributed false rumors of foreign involvement in the Hong Kong Protests.[xvi] Abroad, particularly in the US, Facebook reported identifying and removing a small network of fake accounts and pages with connections to the PRC government that posted content comparing Hong Kong protestors to terrorists.[xvii] Over 15,000 accounts followed these pages.[xviii] Twitter (now X) also reported shutting down a "state-backed information operation" by suspending over 900 accounts attempting to attack the political legitimacy of the Hong Kong protest movement.[xix]

50 Cent Army: The "50 Cent Army" (or 50 Cent Party) refers to a suspected state-sponsored operation where many individuals are paid to write and post content on social media to shape public perception and opinion.[xx] It is a form of astroturfing. Social media content is posted in bursts, indicating government direction and coordination.[xxi] Posts mostly consist of comments on government sites and commercial platforms like Sina Weibo[xxii] (microblogging platform like X). The purpose of these posts is likely to paint China in a more positive light, influencing public opinion through "'positive publicity'" rather than inflammatory speech.[xxiii] The posts also serve as a barrier against online discussions that could inspire collective action by deflecting public attention away from contentious issues.[xxiv] Domestically, the 50 Cent Army posts an estimated 448 million comments on social media and government websites each year.[xxv]

AI-Powered Disinformation

Pro-China disinformation campaigns were largely limited by low quality human-generated content and unsuccessful appeals to broader audiences. However, AI-powered disinformation is cheaper, more accessible, and harder to detect. The rise of GenAI – artificial intelligence that can create new text, images, videos, and audio based on a user prompt – opens the information environment to "high-quality tailored fake text and images at scale" with "advanced, dynamic, automated distribution and coordination".[xxvi]

There are two common GenAI models: Large Language Models and Text-to-Image Models. Large Language Models (LLMs) like GPT-4 use mathematical probability to determine the most appropriate next word in a sequence or sentence, based on existing text data they were trained on.[xxvii] Text-to-Image Models like DALL-E3 take a text prompt and use statistical algorithms to iteratively transform a noisy image (essentially a random pattern of pixels) into a coherent picture.[xxviii] Building base models are costly because of the computational resources and training data requirements.[xxix] However, fine-tuning models to tailor content to a specific cause or group is far cheaper and requires less training data.[xxx] Tailored content can be used to target specific groups or individuals and increase the perceived authenticity of fake content or personas.

GenAI enables disinformation actors to mass-produce high quality content at a lower cost.[xxxi] Previously, producing better and more content required increased manpower.[xxxii] With GenAI, these labor and cost barriers are much lower for non-state actors and even non-expert individuals to conduct a disinformation campaign.[xxxiii] GenAI-produced content is also much harder to distinguish from human-generated content. Publicly available LLMs like GPT-4 can mimic text ranging from informal social media commentary to professional research papers.

GenAI provides substantial advancements to common disinformation tactics, but especially for astroturfing and deepfakes. While PRC astroturfing operations largely rely on human-generated content, GenAI's fluency in human-generated social media posts enables disinformation actors to produce content with higher quality and quantity.[xxxiv] To create deepfake audio, AI only needs hours or even minutes of audio clips to create a convincing replica.[xxxv] In July of 2025, an unidentified disinformation actor created a Signal account impersonating Secretary of State Marco Rubio.[xxxvi] They used AI to create a deepfake of Secretary Rubio's voice and communicated with government officials.[xxxvii] Similar examples are likely to become more common as more actors gain expertise in incorporating GenAI into disinformation operations.

Applications of AI in Pro-China Disinformation Campaigns

Given the emerging nature of GenAI, analysis of the PRC government's use of AI-powered disinformation primarily reflects projected capabilities and strategic intent. However, AI-powered disinformation has been on the PRC government's radar for years. In 2020 – years before the release of GPT-3 – Dr. Li Bicheng published a paper detailing a model for "AI-enabled public opinion manipulation".[xxxviii] It combines a sophisticated system of content-producing models and a model to take a prompt on public-opinion influence strategy.[xxxix] The theorized model would be able to generate and publish content based on human guidance.[xl] While this model is yet to be successfully implemented by the PRC government, with AI's rapid progress and China's growing AI sector, the question is when – not if – these capabilities will be attained.

Spamouflage: Spamouflage refers to a disinformation campaign – with suspected ties to the PRC government – that produces and distributes fake content on a large scale to influence public opinion. The campaign targets contentious political issues in the US and paints them "as examples of how the U.S. political system had failed".[xli] Within the US, Spamouflage-associated accounts pretended to be American voters and criticized prominent political figures from both major political parties.[xlii] In 2022, Spamouflage accounts posted deepfake news videos with AI-produced news anchor avatars on Facebook and Twitter.[xliii] The videos criticized US gun policy and praised China's geopolitical successes.[xliv] This marked the first known use of deepfake videos features artificial personas in a "state-aligned information campaign".[xlv] Despite obvious flaws like audio-lip sync issues, the deepfake videos were made with commercially available technology requiring only a script.[xlvi]

2024 Taiwanese Election: Several AI-powered disinformation campaigns have been attributed to PRC actors leading up the Taiwanese Presidential Election in 2024. Examples include deepfake videos of Lai Ching-Te, the candidate for Taiwan's Democratic Progress Party (DPP), expressing support for pro-Beijing political opponents.[xlvii] The DPP supports cultivating relationships with other democratic nations, while the opposing Kuomintang Party believes Taiwan is best served by amicable relationships with the PRC.[xlviii] These disinformation efforts are reflective of the CCP's strategy to promote pro-China sentiment and values. Undermining pro-independence politicians and supporting their pro-Beijing opponents aligns with the CCP's political warfare tactics.[xlix]

Biased LLMs: LLMs trained on data with pro-China biases reflect those biases within the responses. Large-scale disinformation campaigns, under the direction of the CCP, pour disinformation content onto the internet, which becomes part of data used to train LLMs.[l] Moreover, the PRC government requires that AI chatbots reflect socialist values, which companies must comply with to access Chinese markets.[li] The five most popular AI chatbots are "OpenAI's ChatGPT, Microsoft's Copilot, Google's Gemini, DeepSeek's DeepSeek-R1 and X's Grok" – all of which demonstrate the influence of CCP censorship and biases to some degree.[lii] Of the five, only DeepSeek-R1 was developed in the PRC. When all five chatbots were tested with the prompt "'Hong Kong Freedom'", only DeepSeek-R1's response indicated that Hong Kong residents had fully guaranteed freedoms.[liii] When prompted with the equivalent prompt in Chinese, all five chatbots' responses either "downplayed" assertions of limited freedoms, avoided a direct answer, or used traditional Chinese (mostly used by Taiwanese residents as opposed to simplified Chinese used in the mainland).[liv] Similar disparities between the English and Chinese prompt responses were seen in response to questions about the Tiananmen Square Massacre and Uyghurs' oppression.[lv]

Currently, the use of AI in PRC disinformation campaigns appear to have limited success and influence in the US. However, the CCP censorship's influence on AI chatbot responses highlight how such efforts can reach beyond the intended audiences and platforms. As AI becomes more accessible, AI-powered disinformation campaigns continue to amplify CCP values and views. Case studies within China reflect the PRC government's capabilities to censor and restrict information and speech, even without the full integration of AI. AI will only strengthen and obscure these tactics.

Policy Implications

AI-powered disinformation brings advantages in scale, credibility, and cost of disinformation campaigns, presenting both new and magnified challenges for policymakers. It undermines existing avenues of information sharing and presents a national security risk during pivotal moments like elections when information integrity is paramount. A strong policy response should seek to build systemic resilience to disinformation and mitigate its influence.

Social media platforms and its users are more vulnerable than ever to AI-powered disinformation. Disinformation gains momentum and credibility when it aligns with people's existing beliefs, comes from a trusted source, or is repeated enough and widespread.[lvi] PRC disinformation tactics aim for all three areas of vulnerability, exemplified by the Hong Kong protests where Chinese state media spread false narratives of violence-oriented protestors and foreign involvement in the pro-democracy movement. By utilizing a source of authority, the PRC reinforced negative domestic sentiment towards the protestors and amplified this through coordinated state media channels. The scale and quality enabled by GenAI escalates this risk. GenAI enables high quality content to be tailored for specific audiences. AI-generated content and online personas can be altered to align with a group's beliefs and values to gain credibility and influence. Potentially, this could be achieved on a much larger scale with a greater number of online communities being targeted with tailored content. On the technological front, policymakers should consider investing in "defensive technology" that can identify AI-generated content on a large scale.[lvii] Policymakers can also require social media companies to watermark AI-generated content, though regulations should carefully limit the burden placed on platforms. The goal should be to target and identify the sources of disinformation rather than punish public discourse platforms.

Diplomatic responses are limited by the challenge of concretely attributing a disinformation campaign to a state actor. AI exacerbates this issue as disinformation can be produced and distributed by bots instead of hired individuals that can be linked to a state. Furthermore, disinformation campaigns are within gray zone of operations that "fall below the threshold of armed conflict".[lviii] This ambiguity only increases mutual distrust and further complicates US-China relations. Both the US and the PRC have accused each other of conducting disinformation campaigns against the other.[lix] Policymakers should consider increasing dialogue with the PRC government about the intersection of GenAI and social media, focusing on identifying common ground.[lx] Discussions should prioritize potential restrictions, particularly by "domestic non-state actors".[lxi]

The ability to create AI-generated media is no longer exclusive to well-funded state actors. Traditional barriers to entry like high costs, technological expertise, and significant manpower are weakening. The emergence of GenAI technology significantly reduces the resource requirement for creating high quality disinformation content. This shift provides a dangerous opportunity for both state and non-state actors to conduct influential disinformation campaigns with low investment. The deepfake incident involving Secretary Rubio, where AI-generated voice recordings and text messages were used to impersonate a high-ranking US government official, highlights the increasing risk that public officials and institutions face. Likewise, Spamouflage, which used AI-generated news anchors to spread pro-China and anti-US propaganda on social media, shows how commercially available technology can be used to conduct disinformation campaigns. Policymakers should consider investing in media literacy initiatives to inform the public of rampant disinformation campaigns and educate people on adopting practices such as identifying AI-generated content and thoroughly checking sources.

[i] Voo, Julia. "Chapter 5: Driving Wedges: China's Disinformation Campaigns in the Asia-Pacific." International Institute for Strategic Studies, May 2024. https://www.iiss.org/publications/strategic-dossiers/asia-pacific-regional-security-assessment-2024/chapter-5/.
[ii] "Tactics of Disinformation." Cybersecurity and Infrastructure Security Agency, October 18, 2022. https://www.cisa.gov/sites/default/files/publications/tactics-of-disinformation_508.pdf.
[iii] "Tactics of Disinformation".
[iv] "Tactics of Disinformation".
[v] "Tactics of Disinformation".
[vi] "Tactics of Disinformation".
[vii] Yu, Miles. "The CCP's Strategy to Shape the Global Information Space." Hudson Institute, November 30, 2024. https://www.hudson.org/foreign-policy/discourse-power-ccp-strategy-shape-global-information-space-house-select-committee-miles-yu.
[viii] Yu, "The CCP's Strategy to Shape the Global Information Space".
[ix] Yu, "The CCP's Strategy to Shape the Global Information Space".
[x] Voo, "Chapter 5: Driving Wedges: China's Disinformation Campaigns in the Asia-Pacific".
[xi] Holz, Heidi, and Josiah Case. Episode 38: The Protests in Hong Kong. CNA Talks, 2019. https://www.cna.org/our-media/podcasts/cna-talks/2019/8/the-protests-in-hong-kong.
[xii] Holz and Case, *Episode 38: The Protests in Hong Kong*.
[xiii] Holz and Case, *Episode 38: The Protests in Hong Kong*.
[xiv] Holz and Case, Episode 38: The Protests in Hong Kong.
[xv] Holz and Case, Episode 38: The Protests in Hong Kong.
[xvi] Holz and Case, Episode 38: The Protests in Hong Kong.
[xvii] Gleicher, Nathaniel. "Removing Coordinated Inauthentic Behavior From China." Meta Newsroom (blog), August 19, 2019. https://about.fb.com/news/2019/08/removing-cib-china/.
[xviii] Gleicher, *"Removing Coordinated Inauthentic Behavior From China"*.
[xix] X Blog. "Information Operations Directed at Hong Kong," August 19, 2019. https://blog.x.com/en_us/topics/company/2019/information_operations_directed_at_Hong_Kong.
[xx] King et al., "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument," American Political Science Review 111, no. 3 (2017): 484–501, https://doi.org/10.1017/s0003055417000144.
[xxi] King et al., "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction".
[xxii] King et al., "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction".
[xxiii] King et al., "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction".
[xxiv] King et al., "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction".
[xxv] King et al., "How the Chinese Government Fabricates Social Media Posts for Strategic Distraction".
[xxvi] Marcellino, William, Nathan Beauchamp-Mustafaga, Amanda Kerrigan, Lev Navarre Chao, and Jackson Smith, The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0: Next-Generation Chinese Astroturfing and Coping with Ubiquitous AI. Santa Monica, CA: RAND Corporation, 2023. https://www.rand.org/pubs/perspectives/PEA2679-1.html.
[xxvii] Marcellino et al., *The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0*.
[xxviii] Text to Image Diffusion AI Model from Scratch - Explained One Line of Code at a Time!, 2024. https://www.youtube.com/watch?v=w8YQcEd77_o.
[xxix] Marcellino et al., *The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0*.
[xxx] Marcellino et al., *The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0*.
[xxxi] Marcellino et al., *The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0*.
[xxxii] Marcellino et al., *The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0*.
[xxxiii] Marcellino et al., *The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0*.
[xxxiv] Marcellino et al., *The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0*.
[xxxv] "Tactics of Disinformation".

[xxxvi] Hansler, Jennifer. "Someone Using AI to Impersonate Marco Rubio Contacted at Least Five People Including Foreign Ministers, Cable Says." CNN, July 8, 2025. https://www.cnn.com/2025/07/08/politics/marco-rubio-artificial-intelligence-impersonation.
[xxxvii] Hansler, *"Someone Using AI to Impersonate Marco Rubio"*.
[xxxviii] Marcellino et al., *The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0*.
[xxxix] Generative AI's Potential Role in Information Warfare, 2024. https://www.youtube.com/watch?v=IaZZ9hTVP78.
[xl] Generative AI's Potential Role in Information Warfare.
[xli] Johnson, Derek. "Disinfo Group Spamouflage More Aggressively Targeting U.S. Elections, Candidates," CyberScoop, September 3, 2024, https://cyberscoop.com/spamouflage-targeting-us-election-candidates/.
[xlii] Johnson, "Disinfo Group Spamouflage More Aggressively Targeting U.S. Elections, Candidates".
[xliii] Satariano, Adam and Mozur, Paul. "The People Onscreen Are Fake. The Disinformation Is Real.," Technology, The New York Times, February 7, 2023, https://www.nytimes.com/2023/02/07/technology/artificial-intelligence-training-deepfake.html.
[xliv] Satariano and Mozur, "The People Onscreen Are Fake. The Disinformation is Real".
[xlv] Satariano and Mozur, "The People Onscreen Are Fake. The Disinformation is Real".
[xlvi] Satariano and Mozur, "The People Onscreen Are Fake. The Disinformation is Real".
[xlvii] Mobilio, Sarah (Major, USMC). "GenAI in the 2024 Taiwan Presidential Election: Lessons for Democracies." The Cyber Defense Review, 2024. https://cyberdefensereview.army.mil/Portals/6/Documents/2024-Fall/Mobilio_CDRV9N3-Fall-2024.pdf.
[xlviii] Djou, Victoria. "Preserving Peace and Democracy in Taiwan." Association of the United States Army, November 1, 2023. https://www.ausa.org/publications/preserving-peace-and-democracy-taiwan.
[xlix] Mobilio, *"GenAI in the 2024 Taiwan Presidential Election: Lessons for Democracies"*.
[l] Courtney Manning et al., "Evidence of CCP Censorship, Propaganda in U.S. LLM Responses," The American Security Project, June 25, 2025, https://cdn.prod.website-files.com/67919c3b2972e57c613c2ea2/685b1a27a830fb5b6e7ff511_Sentinel%20Brief%20-%20Evidence%20of%20CCP%20Censorship%20in%20LLM%20Responses.pdf.
[li] Manning et al., "Evidence of CCP Censorship, Propaganda in U.S. LLM Responses".
[lii] Manning et al., "Evidence of CCP Censorship, Propaganda in U.S. LLM Responses."
[liii] Manning et al., "Evidence of CCP Censorship, Propaganda in U.S. LLM Responses."
[liv] Manning et al., "Evidence of CCP Censorship, Propaganda in U.S. LLM Responses."
[lv] Manning et al., "Evidence of CCP Censorship, Propaganda in U.S. LLM Responses."
[lvi] Paul Bolton et al., "Disinformation and Its Effects on Society," UK Parliament, July 16, 2024, https://commonslibrary.parliament.uk/disinformation-and-its-effects-on-society/.
[lvii] Marcellino et al., *The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0*.
[lviii] Forward Defense experts, "Today's Wars Are Fought in the 'Gray Zone.' Here's Everything You Need to Know about It.," Atlantic Council, February 23, 2022, https://www.atlanticcouncil.org/blogs/new-atlanticist/todays-wars-are-fought-in-the-gray-zone-heres-everything-you-need-to-know-about-it/.
[lix] Mengchen Zhang et al., "US Warns of Chinese Disinformation. China Says That's Disinformation," CNN, October 1, 2023, https://www.cnn.com/2023/10/01/world/china-disinformation-us-response-intl-hnk.
[lx] Marcellino et al., *The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0*.
[lxi] Marcellino et al., *The Rise of Generative AI and the Coming Era of Social Media Manipulation 3.0*.