Testimony

NATHAN BEAUCHAMP-MUSTAFAGA

Exploring the Implications of Generative AI for Chinese Military Cyber-Enabled Influence Operations

Chinese Military Strategies, Capabilities, and Intent

CT-A3191-1

Testimony presented before the U.S.-China Economic and Security Review Commission at the hearing "Current and Emerging Technologies in U.S.-China Economic and National Security Competition" on February 1, 2024

For more information on this publication, visit **www.rand.org/t/CTA3191-1**.

Testimonies

RAND testimonies record testimony presented or submitted by RAND associates to federal, state, or local legislative committees; government-appointed commissions and panels; and private review and oversight bodies.

Published by the RAND Corporation, Santa Monica, Calif.

Limited Print and Electronic Distribution Rights

Exploring the Implications of Generative AI for Chinese Military Cyber-Enabled Influence Operations: Chinese Military Strategies, Capabilities, and Intent

Testimony of Nathan Beauchamp-Mustafaga[1]
The RAND Corporation[2]

Before the U.S.-China Economic and Security Review Commission at the hearing "Current and Emerging Technologies in U.S.-China Economic and National Security Competition"

February 1, 2024

Co-chair Helberg and co-chair Wessel, thank you for the opportunity to testify today about how the Chinese military views the prospects of generative artificial intelligence (AI) for social media manipulation, and more broadly on the evolution of its cyber-enabled influence operations (IO) efforts. My testimony will provide an overview of Chinese military strategy, capabilities, and intent, including on the topic of potential Chinese interference in U.S. elections via social media.

Overview

There are many reasons to be concerned that the Chinese military will incorporate generative AI into its existing cyber capabilities to augment its ability to conduct cyber-enabled influence operations, including operations that seek to undermine the democratic process in the United States and around the world.[3] The key breakthrough with generative AI is the dramatic

---

[1] The opinions and conclusions expressed in this testimony are the author's alone and should not be interpreted as representing those of RAND or any of the sponsors of its research.

[2] RAND is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest. RAND's mission is enabled through its core values of quality and objectivity and its commitment to integrity and ethical behavior. RAND subjects its research publications to a robust and exacting quality-assurance process; avoids financial and other conflicts of interest through staff training, project screening, and a policy of mandatory disclosure; and pursues transparency through the open publication of research findings and recommendations, disclosure of the source of funding of published research, and policies to ensure intellectual independence. This testimony is not a research publication, but witnesses affiliated with RAND routinely draw on relevant research conducted in the organization.

[3] William Marcellino, 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*, RAND Corporation, PE-A2679-1, September 2023, https://www.rand.org/pubs/perspectives/PEA2679-1.html.

improvement in authenticity and scale at a lower cost—while also reducing human labor requirements and the probability of detection. Taken together, the "critical jump forward . . . is in the plausibility of the messenger rather than the message."[4] This potential applies to any malign actors, domestic or foreign, because of the open-source nature and very low bar to adoption of generative AI technology.

Beijing has long sought targeted and tailored IO, and while social media has enabled *targeted* IO, generative AI has the potential to enable *tailored* IO. Based on an initial review of Chinese military writings, there is clear awareness by at least some in the People's Liberation Army (PLA) of generative AI's revolutionary potential. We are beginning to see some adoption of generative AI technologies by Chinese Communist Party (CCP)-state actors for cyber-enabled IO, and there is early evidence that this new technology is improving CCP IO performance. However, although we know that the PLA is already engaging in cyber-enabled IO, we have no direct evidence that the PLA is specifically adopting generative AI for this purpose yet. Regardless, I argue that the PLA is currently capable of adopting generative AI if it so chooses, and I point to a case study of a PLA-affiliated researcher, Li Bicheng, to illustrate the PLA's likely readiness for adoption. For the PLA, this could support its IO objectives of shaping foreign (and domestic) public opinion, deterring U.S. involvement in a future Taiwan conflict, and degrading U.S. and Taiwanese will to fight, among other objectives.

This testimony and supporting research are based on open-source Chinese language primary source research and builds on two recent RAND reports.[5] My research focuses mainly on how the Chinese military develops its strategy and capabilities for social media manipulation, with an effort where possible to address technical-level details to provide greater clarity. Broadly, my research is interested in the *inputs* into Chinese efforts for social media manipulation versus the *outputs* of observed Chinese behavior, which is inherently an incomplete and retrospective reverse engineering of Chinese intent and capabilities.

Chinese Military Strategy for Cyber-Enabled Influence Operations

The Chinese military's strategy for cyber-enabled influence operations is evolving to incorporate new technologies and keep up with broader PLA strategic thinking. Historically, the PLA's influence operations generally fell under the concept of the "Three Warfares" (三种战法), which were developed in the mid-2000s and included "psychological warfare" (心理战), "public opinion warfare" (舆论战), and "legal warfare" (法律战). Cyber-enabled influence

[4] Marcellino et al., 2023, p. 1.
[5] Nathan Beauchamp-Mustafaga, *Chinese Next-Generation Psychological Warfare: The Military Applications of Emerging Technologies and Implications for the United States*, RAND Corporation, RR-A853-1, 2023, https://www.rand.org/pubs/research_reports/RRA853-1.html; Marcellino et al., 2023.

operations generally fell under public opinion warfare, specifically "online public opinion warfare" (网络舆论战).[6]

The Chinese military has increasingly adopted "cognitive domain operations" (认知域作战) (CDO) as the primary operational concept for cyber-enabled influence operations since the late-2010s.[7] This evolution reflects a fundamental shift in the Chinese military's conception of the battlespace from the traditional air, sea, and land domains—with space and cyber added in the 1990s—into now viewing warfare as occurring in the physical domain (物理域), information domain (信息域), and cognitive domain (认知域). There is a group of PLA researchers, often focused on IO, who argue that the cognitive domain is the new focus of warfare.[8] However, this is not yet the official PLA view, and there are alternative conceptions within the PLA; for example, the 2020 PLA National Defense University version of *Science of Military Strategy* lists space, network, deep sea, polar regions, biology, and intelligence as new domains of warfare.[9] To summarize this group's perspective, the logical conclusion of the PLA's system-of-systems warfare is to win a conflict with as little kinetic destruction as possible and force the adversary to accept defeat short of total destruction—and thus, fundamentally, a psychological or cognitive decision to surrender, as compared with the 20th century construct of total warfare and complete physical exhaustion of adversary military capabilities and resources.[10] Within PLA military theory, the identification of a new domain thus drives the exploration of the required aspects for each domain: "cognitive warfare" (认知战), "cognitive confrontation" (认知对抗), "cognitive deterrence" (认知威慑), and "command of cognition" (制认知权), among others. None of these terms are officially defined in standard PLA authoritative texts, such as the PLA dictionary (军语), because they gained popularity after the dictionary's publication in 2011, but future editions are likely to include these now key concepts for the PLA.[11]

---

[6] Beauchamp-Mustafaga, 2023. For an authoritative PLA source, see Wu Jieming [吴杰明] and Liu Zhifu [刘志富], *An Introduction to Public Opinion Warfare, Psychological Warfare, and Legal Warfare* [舆论战心理战法律战概论], National Defense University Press, 2014.

[7] For more on CDO, see Beauchamp-Mustafaga, 2023; Nathan Beauchamp-Mustafaga, "Cognitive Domain Operations: The PLA's New Holistic Concept for Influence Operations," *China Brief*, Vol. 19, No. 16, September 6, 2019.

[8] See, for example, Chen Dongheng [陈东恒], "Command of Cognition: An Important Support for Winning the War" ["制认知权: 战争制胜重要支撑"], *PLA Daily*, April 19, 2022; Zhao Quanhong [赵全红], "Cognitive Domain Operations: The Key to Winning Modern Warfare" ["认知域作战：现代战争的制胜关键"], *PLA Daily*, July 14, 2022; Pu Duanhua [濮端华], Li Xiwen [李习文], and Xiao Fei [肖飞], "Getting It Right on How Cognitive Penetration Influences Multi-Domain Operations" ["把准认知域渗透影响多域作战的规律"], *PLA Daily*, January 19, 2023.

[9] Xiao Tianliang [肖天亮], ed., *Science of Military Strategy* [战略学], National Defense University Press [国防大学出版社], 2020, pp. 142–180.

[10] Jeffrey Engstrom, *Systems Confrontation and System Destruction Warfare: How the Chinese People's Liberation Army Seeks to Wage Modern Warfare*, RAND Corporation, RR-1708-OSD, 2018, https://www.rand.org/pubs/research_reports/RR1708.html.

[11] All Army Military Terminology Management Committee [全军军语管理委员会], *Military Terminology of the Chinese People's Liberation Army* [中国人民解放军军语], Military Science Press, 2011.

PLA researchers often attribute to the U.S. military the conception of the cognitive domain as a domain of military struggle, but in reality, PLA interest is fundamentally about CCP regime security, which is the PLA's number one priority as the armed wing of the CCP.[12] This is part of a much longer trend of CCP concerns that information (typically foreign information) could undermine CCP regime control over the domestic Chinese population, sometimes euphemistically described as "ideological security," "cultural security," or, more recently, "cognitive security." Most directly, the Arab Spring and the wave of social media–driven overthrows of authoritarian regimes drove the PLA and broader CCP to view social media as a threat and led to renewed attention on how to influence perceptions and behavior.

CDO appears to be intended as a technologically driven update to bring the "Three Warfares" concept—developed in the information-driven era of informatization (信息化)—into the new AI-driven era of intelligentization (智能化). The U.S. Department of Defense (DoD) explains that CDO "combines psychological warfare with cyber operations to shape adversary behavior and decision making," with the likely intention to "use CDO as an asymmetric capability to deter U.S. or third-party entry into a future conflict, or as an offensive capability to shape perceptions or polarize a society."[13] DoD adds that

> [t]he PLA's goals for social media influence operations include promoting narratives to shape foreign governments' policies and public opinion in favor of the PRC's [People's Republic of China's] interests and undermining adversary resolve. The PLA views social media through the prism of information dominance, and during a crisis could use digital influence operations to undermine enemy morale and confuse or deceive adversary decision makers.[14]

As an overarching military operational concept for military activities in the cognitive domain, CDO includes four main aspects: "reading the brain" (读脑), "controlling the brain" (制脑), "resembling the brain" (类脑), and "strengthening the brain" (强脑).[15] "*Reading the brain*" focuses on understanding how others are thinking, "*resembling the brain*" is about using the human brain as inspiration for designing better computers, and "*strengthening the brain*" is about improving one's own cognition and performance. "*Controlling the brain*" focuses on influencing or even controlling adversary thinking and behavior. Although some PLA discussions of "*controlling the brain*" are futuristic, a more practical example is PLA interest in non-lethal, non-kinetic body-targeted weapons, such as directed energy capabilities like the U.S. military's Active Denial System.

It is clear that CDO is the most prominent operational concept for current PLA cyber-enabled IO. This is best represented in a 2018 article by researchers at the PLA Strategic Support Force's (PLASSF) Base 311, the PLA's only known operational unit dedicated to IO, which addressed

---

[12] For more on this threat perception, see Nathan Beauchamp-Mustafaga and Michael S. Chase, *Borrowing a Boat Out to Sea: The Chinese Military's Use of Social Media for Influence Operations*, Foreign Policy Institute at John Hopkins University School of Advanced International Studies, 2019; Beauchamp-Mustafaga, 2023.
[13] Office of the Secretary of Defense, *Annual Report to Congress: Military and Security Developments Involving the People's Republic of China*, U.S. Department of Defense, 2023, p. 156.
[14] Office of the Secretary of Defense, 2023, p. 158.
[15] See Beauchamp-Mustafaga, 2023, pp. 52–67.

the hardware requirements for CDO and specifically focused on social media manipulation.[16] It is likely that this article reflected PLA preparations for social media interference in Taiwan's 2018 elections based on Taiwan's later claims of PLA involvement, Base 311's historical focus on Taiwan, and explicit references in the article to social media platforms popular in Taiwan.[17] PLA discussions of CDO dovetail with related PLA discussions of "intelligentized public opinion warfare" (智能化舆论战) that focus on leveraging AI, big data, social media, and social bots, among other emerging technology, for improved messaging and targeting effectiveness.[18] This also dovetails with broader Chinese Party-state efforts to become the leading AI power by 2030.[19]

Chinese Military Views on the Applications of Generative AI for Cyber-Enabled Influence Operations

PLA researchers recognize the potential improvement offered by generative AI for cyber-enabled IO, although most PLA discussions apply the new technology to existing tactics. Despite OpenAI releasing ChatGPT in November 2022, it appears that the PLA did not really pay attention until March 2023. PLA researchers generally focus on the advantages of generative AI for content (mostly text) generation but also address the benefits from automation to improve content distribution, as well as reduce labor requirements and cost. This aligns well with existing PLA interest in content generation—as the PLA sometimes describes it, "synthetic information" (合成信息)—namely, producing inauthentic content based on some amount of original information, as well as with content distribution using social bots (社交机器人)—namely, algorithmic agents for social media.[20]

Much of the surveyed PLA writings are focused on the potential cyber and AI-driven information threat from the United States. For example, there is a lot of euphemistic discussion about concerns for CCP regime security stemming from AI technology, ranging from the fact that Western large language models (LLMs) are inherently biased toward Western values to the risks of the United States and other Western countries using generative AI against the CCP. However, this focus on the threat is common for PLA writings, and this should not be exculpatory of potential future PLA embrace of generative AI for offensive purposes. The PLA

---

[16] Liu Huiyan [刘惠燕], Xiong Wu [熊武], Wu Xianliang [吴显亮], and Mei Shunliang [梅顺量], "Several Thoughts on Promoting the Construction of Cognitive Domain Operations Equipment for the Omni-Media Environment" ["全媒体环境下推进认知域作战装备发展的几点思考"], *National Defense Technology* [国防科技], Vol. 39, No. 5, October 2018.
[17] Nathan Beauchamp-Mustafaga and Jessica Drun, "Exploring Chinese Military Thinking on Social Media Manipulation Against Taiwan," *China Brief*, Vol. 21, No. 7, April 12, 2021.
[18] Sun Yixiang [孙亦祥] and Yu Yuanlai [余远来], "A Brief Discussion on 'Intelligentized Public Opinion Warfare'" ["刍议'智能化舆论战'"], *Military Correspondent* [军事记者], January 2022.
[19] "New Generation AI Development Plan," People's Republic of China State Council, July 20, 2017. Translation available via Graham Webster, Rogier Creemers, Elsa Kania, and Paul Triolo, "Full Translation: China's 'New Generation Artificial Intelligence Development Plan' (2017)," DigiChina, August 1, 2017.
[20] Marcellino et al., 2023.

is indeed already engaged in cyber-enabled IO and, at a minimum, these surveyed writings reveal an understanding by PLA researchers of generative AI's potential.[21] A more forward leaning article in June 2023 specifically argued that generative AI will make it easier to conduct offensive CDO because it enables broader scope and quicker operations that will thus make it easier to push one's desired narratives.[22] The article argues that "the improvement in deepfake technology brought by ChatGPT makes offensive operations in cognitive domain operations more covert" and will make it easier to "integrate offensive actions . . . into daily life."

PLA discussions of generative AI's potential for cyber-enabled IO mainly focus on the prospects for long-term subtle influence across several different tactics: influencing public opinion via large-scale bot networks, producing intentionally biased publicly available models, and specifically degrading support for adversary leadership.[23] First, the main tactic that PLA authors discuss to artificially influence public opinion is the prospect of running large-scale networks of social bots powered by automated content generation. For example, an April 2023 article coauthored by a leading researcher for the 28th Research Institute under the China Electronics Technology Group (CETC), which provides command-related systems to the PLA, argued that generative AI will enable

> nuanced, personalized content, and not only proactively post but also respond to other users' posts and engage in long-term conversations. Therefore, after social bots based on ChatGPT are instilled with personalities, positions and tendencies, they can become invisible on the Internet and become cognitive shaping tools. They are more influential and concealed than traditional [human-run] astroturfing [网络水军].[24]

This PLA interest in large-scale bot networks builds on PLA interest in social bots and manipulated content. At least two articles by PLA researchers explicitly call for PLA adoption of social bots, as one 2022 article said,

> In the face of Western countries taking the opportunity to smear and attack [us], we must have the courage to use social bots [社交机器人] to carry out public

---

[21] Office of the Secretary of Defense, 2023, p. 158.

[22] Chen Changxiao [陈昌孝], Li Hao [李浩], Wang Zihan [王梓晗], Jiang Wenbo [姜文博], "A New Weapon in Cognitive Domain Operations: ChatGPT Cognitive Analysis and Countermeasures" ["认知域作战新利器: ChatGPT 认知剖析及对策"], *Military Digest* [军事文摘], June 2023.

[23] Chen Dongheng [陈东恒] and Xu Yan [许炎], "Generative AI: A New Weapon for Cognitive Confrontation" ["生成式 AI: 认知对抗的新武器"], *PLA Daily*, April 4, 2023. For a similar argument, see Mao Weihao [毛炜豪], "Looking at the Military Applications of Artificial Intelligence from ChatGPT" ["从 ChatGPT 看人工智能的军事应用"], *PLA Daily*, April 13, 2023.

[24] Zhou Zhongyuan [周中元], Liu Xiaoyi [刘小毅], Li Qingwei [李清伟], "ChatGPT Technology and Its Impact on Military Security" ["ChatGPT 技术及其对军事安全影响"], *Command Information System and Technology* [指挥信息系统与技术], April 2023, pp. 7–16. The lead author self-plagiarized this article shortly after in Zhou Zhongyuan [周中元], "ChatGPT's Challenges to Military Security and Countermeasures" ["ChatGPT 对军事安全的挑战与应对策略"], *Defence Science & Technology Industry* [国防科技工业], July 2023, pp. 46–48. For related articles, see Chen et al., 2023; Hua Rui [华瑞], Yang Longxiao [杨龙霄], Yang Runxin [杨润鑫], "When Generative Artificial Intelligence Heads to the Battlefield" ["当生成式人工智能走向战场"], *PLA Daily*, December 1, 2023.

> opinion [struggle], and use relevant social bots to carry out information bombing [信息轰炸] against the enemy's social network to drown it out.[25]

PLA researchers have been interested in manipulated content since at least 2005 and manipulated video since at least 2011 and have more recently begun to specifically discuss deepfakes.[26] For example, one article explained that "generative AI can quickly generate or forge the appearance, voice, emotions, expressions, and other attributes of political targets" and that it can "generate personalized content based on the preferences of political targets," or even mobilize public opinion on a desire topic.[27]

Some PLA discussions focus on the prospects of generative AI for enabling "precision cognitive attacks" (精准认知攻击), specifically highly tailored or even personalized IO against small groups or individuals.[28] For example, one article noted that "ChatGPT's powerful data processing capabilities and high autonomy enable it to conduct preference analysis and subsequent related information production and information delivery," supporting "precision cognitive attacks" based on "personalized user portraits" using big data to analyze individual preferences.[29]

Another related method is creating or reinforcing "information cocoons" (信息茧房)—specifically, information bubbles—with the intention of undermining the influence of mainstream values, further polarizing and dividing society. One article acknowledges that the traditional approach before generative AI was costly, was easy to identify because it required seizing on existing trending topics, and, thus, was easy to defeat.[30] In comparison, ChatGPT greatly improves the efficiency because it is autonomous, cheaper, and quicker and can better distribute the desired content more subtly, making it harder to detect. As the article said, "ChatGPT's precision information delivery capabilities and crowd classification capabilities will effectively and efficiently promote the formation of various small groups."

---

[25] Long Yameng [龙亚蒙] and Zhou Yang [周洋], "Research on the Application of Social Bots in Public Opinion Struggle" ["社交机器人在舆 论斗争中的应用研究"], *Military Correspondent* [军事记者], 2022. See also Wu Xiaojian [武啸剑], "How Social Bots Manipulate Public Opinion in the Age of Intelligent Communication: An Analysis of Narrative and Cognition Under Crisis" ["危机下的叙事与认知: 智能传播时代社交机器人舆论干预研究"], *Journalism and Mass Communication* [新闻界], September 2023, pp. 88–96.

[26] For more, see Beauchamp-Mustafaga, 2023; Marcellino et al., 2023.

[27] Zhang Guangsheng [张广胜], "National Security Risks of Generative Artificial Intelligence and Countermeasures" ["生成式人工智能的国家安全风险及其对策"], *Frontiers* [人民论坛·学术前沿], July 2023, pp. 76–85. For another article linking generative AI with deepfakes, see Hu Kaiguo [胡开国], "When Generative Artificial Intelligence Plays a Role in War" ["当生成式人工智能作用于战争"], *Military Digest* [军事文摘], September 2023.

[28] This discussion of precision appears to be a popular concept in PLA IO circles. See Bu Jiang [卜江] and Jiang Rilie [蒋日烈], "How to Achieve Precision Strikes in Cognitive Domain Operations" ["如何实现认知域作战精准打击"], *PLA Daily*, March 16, 2023.

[29] Chen et al., 2023. The authors self-plagiarized this article to publish the same text shortly after in Chen Changxiao [陈昌孝] and Wang Zihan [王梓晗], "Characteristics, Application, and Countermeasures of ChatGPT from a Cognitive Perspective" ["认知视角下 ChatGPT 的特征、运用及应对之策", *Political Work Journal* [政工学刊], July 2023.

[30] Chen et al., 2023.

Second, another tactic discussed is the potential to covertly use a publicly available model, such as ChatGPT, to influence adversary public perceptions over time. Much of this discussion focuses on the risks of Western models, trained on Western data, being thus inherently biased toward Western values and surreptitiously inculcating users with Western values. As one article explained, ChatGPT will "widely penetrate into every corner of society," presenting opportunities to "subtly influence" users over time and "unknowingly change a person's cognitive logic and behavior."[31] Although I did not find any specific PLA writings calling for Beijing to develop its own open source models and intentionally building pro-CCP bias into them, a natural result of the CCP's current regulatory model of censoring political content for publicly available Chinese models is that they are very likely to hew, by design, toward CCP-approved narratives. This would represent a relatively new opportunity for Party-state IO, comparable perhaps only to Chinese-run social media platforms, such as WeChat and TikTok, which avoid takedowns of CCP disinformation and are vulnerable to CCP efforts to artificially push favored narratives.[32]

Third, PLA authors discuss the ability to degrade popular support for adversary leadership. As one article explained, "cognitive attacks" and "precision attacks against political targets" [精准攻击政治目标] are the "general trend of current hybrid warfare," based on "generative AI, cognitive neural methods, and social media" and provided as evidence U.S. efforts to leverage AI to "demonize" Nicolás Maduro in Venezuela, reflecting recent accusations by the Chinese Ministry of Foreign Affairs (MFA).[33] This aligns with longstanding PLA discussion of "public opinion decapitation" (舆论斩首), which the PLA views as a common U.S. tactic, as seen in the wars against Iraq and Libya.[34]

PLA researchers also acknowledge some shortcomings for generative AI. For example, in March 2023, Hu Xiaofeng, a famous PLA researcher, acknowledged that generative AI "should neither be underestimated nor overestimated" and has problems, such as challenges in training data, human confidence in the models, and lack of transparency, among other issues.[35] A similar list of issues has since been repeated by several other researchers.[36] Another article argued that "there are some more subtle things, including cultural symbols, historical allusions, the processing of ambiguity (such as puns), etc., which require high context information to help AI understand" and that generative AI "fictionalizes information."[37] However, in light of

---

[31] Chen et al., 2023. For another article, see Chen and Xu, 2023.
[32] See, for example, Sapna Maheshwari, "Topics Suppressed in China Are Underrepresented on TikTok, Study Says," *New York Times*, December 21, 2023.
[33] Zhang, 2023, pp. 76–85; Chinese Ministry of Foreign Affairs, "Fact Sheet on the National Endowment for Democracy," May 7, 2022.
[34] See Beauchamp-Mustafaga, 2023, p. 140.
[35] Hu Xiaofeng [胡晓峰], "How Should We View ChatGPT?" ["ChatGPT 我们该怎么看"], *PLA Daily*, March 21, 2023.
[36] Shen Zhengzheng [申峥峥] and Shu Zhe [束哲], "Generative AI: How Far Is It from Comprehensive Application in the Military Field" ["生成式人工智能: 距离军事领域全面应用有多远"], *PLA Daily*, April 14, 2023; Hua, Yang, and Yang, 2023.
[37] "Generative AI and Science Fiction Creation" ["生成式 AI 与科幻创作"], *PLA Daily*, May 24, 2023.

longstanding PLA complaints about poor foreign language capabilities and cross-cultural understanding, generative AI will almost certainly improve current PLA capabilities.[38]

Lastly, it is difficult to overstate the depth of concern held by PLA researchers about U.S. efforts to use generative AI to undermine CCP regime security. These concerns long predate the rise of AI, but AI's dramatically increased performance appears to have exacerbated these concerns. There is a widespread belief amongst PLA researchers that the U.S. government, often specifically the U.S. military, is either developing or has already deployed such capabilities and could, or already is, targeting them at Beijing. As DoD explains,

> From the PRC's perspective, all nations—especially the United States—that use digital narratives to undermine the CCP's authoritarian system in China employ offensive influence operations. Hence, the PRC considers its influence operations that counter this perceived subversion as defensive in order to protect the party and the military.[39]

For example, a July 2023 article argued that the U.S. government is working on an "influence machine" (影响力机器) that will "combine algorithm-generated content, personalized targeting, and intensive information dissemination" in order to "achieve [U.S.] political goals of corroding, infiltrating, influencing and subverting from inside [the target country] and discrediting, containing, and blocking from the outside [of the target country]."[40] These fears find validation in public reports of U.S. IO, which are frequently recounted in PLA writings.[41] Indeed, it is these concerns that are often used to justify Chinese efforts in response.[42]

This concern may be driven in part by fears of technological inferiority because at least initial Chinese appraisals acknowledged the U.S. lead in generative AI. For example, the April 2023 article by CETC researchers stated that

> [a]t present, domestic companies such as Baidu, ByteDance, and NetEase have accumulated relevant technologies and layouts, but in terms of technical capabilities, domestic experts judge that they are about 2 to 3 years behind ChatGPT. In the military industry, although relevant companies have

---

[38] See Beauchamp-Mustafaga, 2023, pp. 112–115.
[39] Office of the Secretary of Defense, 2023, p. 156.
[40] Zhang, 2023, pp. 76–85. For other recent concerns, see Ban Wentao [班文涛] and Xie Mingxiu [谢明修], "Three Reliances: Using Artificial Intelligence Technology to Innovate Public Opinion Warfare—Research and Enlightenment on the Application of Artificial Intelligence Technology to Public Opinion Warfare by the United States and Other Western Countries" ["三个依托: 利用人工智能技术创新舆论战: 美国等西方国家将人工智能技术应用于舆论战研究与启示"], *Military Correspondent* [军事记者], January 2023; Meng Haohan [孟浩瀚] and Lan Peixuan [兰培轩], "Strategies to Generate Capabilities in Public Opinion Warfare in the Cognitive Domain: Thoughts and Warnings Brought to Us by The Public Opinion Dissemination of American and Western media" ["认知域下舆论战的能力生成之策: 美西方媒体舆论传播带给我们的思考和警示"], *Military Correspondent* [军事记者], January 2023; Wang Hejing [王鹤静] and Wu Dan [吴丹], "Strategic Tactics of the U.S. Cyber Army" ["美国网军的战略战法"], *Military Digest*, July 2023.
[41] See, for example, Ellen Nakashima, "Pentagon Opens Sweeping Review of Clandestine Psychological Operations," *Washington Post*, September 19, 2022.
[42] See, for example, Long and Zhou, 2022.

> accumulated some experience in natural language processing technology, the relevant models and functions lag far behind those in the industry.[43]

This view might be outdated almost a year later, but it's an interesting data point from someone who is likely very familiar with PLA technical capabilities.

Current State of Chinese Adoption

Chinese Military

Despite the evident interest by PLA researchers, so far there is no direct evidence of specific PLA adoption to operationalize generative AI for cyber-enabled IO. This might simply be a limitation of open-source research or perhaps a reflection of slow adoption of this new technology, or it could suggest that the PLA has decided generative AI is not worth pursuing. One indication of a lack of apparent movement toward PLA development is the relative lack of publications so far on the topic by PLASSF researchers, who would likely support such efforts.[44]

Regardless, there are many reasons to believe that the PLA will be able to leverage generative AI for cyber-enabled IO if it so chooses. First, the PLA could leverage any one of the open-source LLMs available—including Falcon or the leaked version of Meta's LLaMA, or even OpenAI's services—because it almost certainly has the technical sophistication to circumvent the basic restrictions on users' Internet Protocol addresses. Second, China has a robust tech sector that is busily working on generative AI, with significant government support as part of its plan to lead the world in AI development by 2030. As of September 2023, there were an estimated 130 LLMs under development in China.[45] The inevitable political constraints imposed by the CCP will likely negatively affect the overall performance of Chinese models, but Chinese companies should not be counted out, and recent comparisons between top U.S. and Chinese models suggest relative improvement by Chinese models in the second half of 2023, although U.S. models still lead.[46] Moreover, the PLA will not need the best models, and certainly even the original version of ChatGPT would be more than adequate for the PLA's basic needs.

To better understand the potential of generative AI for the PLA, it is useful to explore the case study of Li Bicheng, a PLA-affiliated researcher who has focused on improving PLA cyber-

---

[43] Zhou, Liu, and Li, 2023, pp. 7–16.

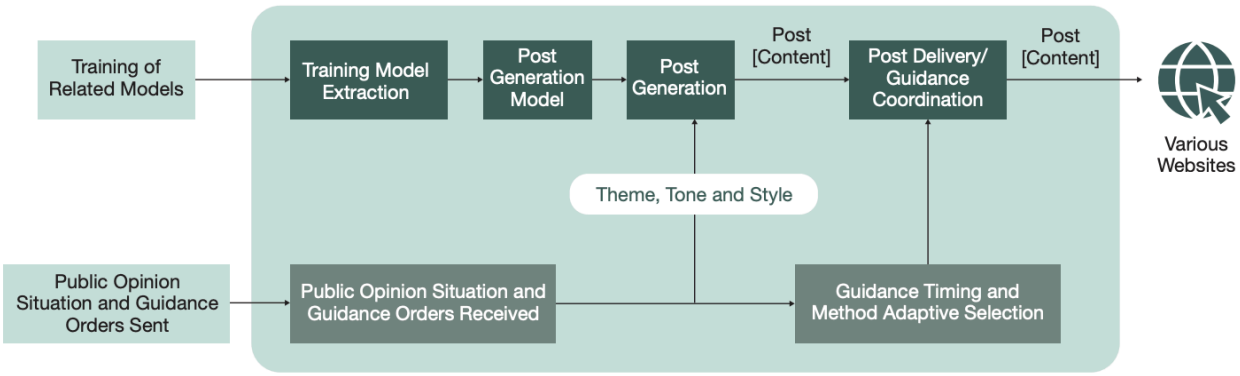
[44] For one PLASSF article that recounts a familiar litany of issues but does not suggest research and development efforts, see Fu Xiang [付翔], Wei Xiaowei [魏晓伟], Zhang Hao [张浩], Xu Ning [徐宁], "Examining and Analyzing ChatGPT from the Perspective of Digital Security" ["数字安全角度下审视和剖析 ChatGPT"], *Aero Weaponry* [航空兵器], October 2023.

[45] Josh Ye, "China's AI 'War of a Hundred Models' Heads for a Shakeout," Reuters, September 21, 2023.

[46] "Chinese Large Model Benchmark Evaluation 2023 Annual Report" ["中文大模型基准测评 2023 年度报告"], SuperCLUE, December 31, 2023. For more on the political constraints on Chinese generative AI development, see Helen Toner, Jenny Xiao, and Jeffrey Ding, "The Illusion of China's AI Prowess: Regulating AI Will Not Set America Back in the Technology Race," *Foreign Affairs*, June 2, 2023.

enabled IO and is specifically working on how to leverage AI.[47] Li began designing a system for creating and distributing inauthentic content (disinformation) for "online public opinion struggle" by at least 2016, specifically calling for the ability to conduct "online information deception" and "online public opinion guidance."[48] In 2019, Li coauthored an article with a PLA researcher from Base 311 that focused on the applications of AI to overcome the PLA's current shortcomings that its "post [content] generation is mechanized without regard for personality, occupation, and age differences[, and] there is no individuality or simulation of human characteristics, so posts are easily identified and deleted."[49] Looking forward, Li essentially foresaw the rise of generative AI and its ability to support autonomous content generation and content delivery in an end-to-end system, illustrated in Figure 1. In light of this, at least some in the PLA are very likely ready to seize on generative AI's potential and could be working toward deployment outside the public eye.

Figure 1. PLA Researcher's Vision for AI-Driven Social Media Manipulation



SOURCE: Reproduced from Marcellino et al., 2023, p. 19. Originally adapted from Li Bicheng [李弼程], Xiong Yao [熊尧], Huang Tao [黄涛], and Pan Le [潘乐], "Simulation Deduction Model and System Construction for Intelligent Online Public Opinion Guidance" ["网络舆论智能引导仿真推演模型与系统构建"], *National Defense Technology* [国防科技], October 2020.

One potential key variable for PLA adoption of generative AI is likely to be the political pressures on the output from generative AI.[50] Chinese experts recognize that generative AI can produce either factually incorrect or political unacceptable outputs; some experts have recounted asking ChatGPT about the origins of COVID-19 only to find that it (accurately) answered

[47] For more, see Marcellino et al., 2023. After spending his entire career at PLA research institutions, Li no longer claims an affiliation with the PLA and is a professor at Huaqiao University. However, he continues to coauthor with PLA researchers and continues to receive PLA funding for his research, suggesting continued ties.
[48] Li Bicheng [李弼程], "Model for a System of Online Public Opinion Struggle and Countermeasures" ["网络舆论斗争系统模型与应对策略"], *National Defense Technology* [国防科技], October 2016.
[49] Li Bicheng [李弼程], Hu Huaping [胡华平], and Xiong Yao [熊尧], "Intelligent Agent Model for Online Public Opinion Guidance" ["网络舆情引导智能代理模型"], *National Defense Technology* [国防科技], June 2019.
[50] For more consideration, see Marcellino et al., 2023.

"China."[51] Given the apparent political constraints placed on Chinese IO (even Chinese covert foreign IO) to ensure that they ultimately fall within CCP-acceptable talking points, it is possible that the PLA and other Party-state actors will shy away from embracing generative AI for fear of internal backlash. Alternatively, these actors could spend additional time and energy on developing a politically reliable series of models that they could then leverage, but this is very likely to negatively affect the overall performance of the models.

There are two important caveats for this discussion of Chinese military cyber-enabled IO against Taiwan. First, the PLA is almost certainly one of many Party-state actors involved in cyber-enabled IO, although I assume the PLA is a relatively high performer based on its overall level of technical sophistication and general lack of attribution. Other Party-state actors that are likely involved include the CCP Propaganda Department, the MFA, the Ministry of State Security (MSS), the Ministry of Public Security, the Cyberspace Administration of China, and the United Front Work Department.[52] Second, social media manipulation is only one of many ways that Beijing can interfere in Taiwanese politics and broader society.[53]

In terms of what is known about past PLA cyber-enabled influence efforts, we have a very limited understanding because attribution to specific PRC actors is very difficult.[54] In 2016, Taiwan essentially accused the PLA Air Force of disinformation for posting a photo of Chinese bombers flying close enough to Taiwan to take a picture with what was suspected to be Jade Mountain.[55] Most notably, Taipei also specifically blamed the PLASSF for interfering with its 2018 election via social media, although it did not provide any specific public evidence.[56] In this case, the effectiveness is very uncertain, although available public evidence suggests that China's overall efforts in 2020 yielded minimal results.[57]

Broader Chinese Party-State

Our best understanding of Chinese Party-state efforts at cyber-enabled IO suggests growing effectiveness after previously middling performance and that this improvement might be at least partly due to adoption of generative AI. In an April 2023 review of publicly available reporting, the Australian Strategic Policy Institute (ASPI) found that PRC "operations are now more frequent, increasingly sophisticated and increasingly effective in supporting the CCP's strategic

---

[51] Zhou, Liu, and Li, 2023, pp. 7–16.
[52] Albert Zhang, Tilla Hoja, and Jasmine Latimore, *Gaming Public Opinion: The CCP's Increasingly Sophisticated Cyber-Enabled Influence Operations*, Australian Strategic Policy Institute, April 2023.
[53] Ben Blanchard, "Taiwan Says China Has 'Very Diverse' Ways of Interfering in Election," Reuters, October 4, 2023. For a recent broader examination of CCP IO, see U.S.-China Economic and Security Review Commission, "Hearing on 'China's Global Influence and Interference Activities,'" March 23, 2023.
[54] For the most recent U.S. government report on the topic, see Global Engagement Center, *How the People's Republic of China Seeks to Reshape the Global Information Environment*, U.S. Department of State, September 2023. For a good nongovernmental roundup, see Zhang, Hoja, and Latimore, 2023.
[55] Matthew Strong, "Military Denies Yushan in China Bomber Picture," *Taiwan News*, December 17, 2016.
[56] Chung Li-hua and William Hetherington, "China Targets Polls with Fake Accounts," *Taipei Times*, November 5, 2018.
[57] Aaron Huang, *Combatting and Defeating Chinese Propaganda and Disinformation: A Case Study of Taiwan's 2020 Elections*, Harvard Kennedy School Belfer Center for Science and International Affairs, July 2020.

goals. They focus on disrupting the domestic, foreign, security and defence policies of foreign countries, and most of all they target democracies."[58] The authors find that "[t]he CCP has developed a sophisticated, persistent capability to sustain coordinated networks of personas on social-media platforms to spread disinformation, wage public-opinion warfare and support its own diplomatic messaging, economic coercion and other levers of state power."[59] They continue that "[t]hose efforts have evolved to nudge public opinion towards positions more favourable to the CCP and to interfere in the political decision-making processes of other countries. A greater focus on covert social-media accounts allows the CCP to pursue its interests while providing a plausibly deniable cover."[60]

To date, there have been two nongovernment public reports and one foreign government report that suggest that the Chinese Party-state is beginning to adopt generative AI for cyber-enabled IO, reinforcing concerns outlined in recent RAND research about the ease of adoption for malign actors.[61] In September 2023, Microsoft reported that "[s]ince approximately March 2023, some suspected Chinese IO assets on Western social media have begun to leverage generative artificial intelligence (AI) to create visual content. This relatively high-quality visual content has already drawn higher levels of engagement from authentic social media users."[62] However, this appears to be more likely early, small-scale experimentation rather than reflect rapid broader adoption by known Party-state actors. More recently in December 2023, an ASPI report identified a "new campaign (which ASPI has named 'Shadow Play') [that] has attracted an unusually large audience and is using entities and voice overs generated by artificial intelligence (AI) as a tactic that enables broad reach and scale."[63] The report explained that the "coordinated inauthentic influence campaign originat[ed] on YouTube [and promotes] pro-China and anti-US narratives in an apparent effort to shift English-speaking audiences' views of those countries' roles in international politics, the global economy and strategic technology competition."[64] The campaign reportedly employed text-to-image and likely text-to-speech generative models to generate thumbnails and voiceovers for their videos, respectively. The ASPI report states that "the YouTube campaign is one of the first times that video essays, together with generative AI voiceovers, have been used as a tactic in an influence operation."[65] Most recently, Taiwanese officials claimed that the MSS was producing videos with "artificial intelligence (AI)-generated voiceovers and fake hosts," targeting President Tsai Ing-wen in the lead-up to Taiwan's January

[58] Zhang, Hoja, and Latimore, 2023, p. 1.
[59] Zhang, Hoja, and Latimore, 2023, p. 3.
[60] Zhang, Hoja, and Latimore. 2023, p. 1.
[61] Marcellino et al., 2023.
[62] Microsoft Threat Intelligence, *Sophistication, Scope, and Scale: Digital Threats from East Asia Increase in Breadth and Effectiveness*, Microsoft, September 2023, p. 6.
[63] Jacinta Keast, *Shadow Play: A Pro-China Technology and Anti-US Influence Operation Thrives on YouTube*, Australian Strategic Policy Institute, December 2023, p. 3.
[64] Keast, 2023, p. 3.
[65] Keast, 2023, p. 4.

2024 elections, validating earlier concerns, but that the campaign reportedly failed to garner much traction online.[66]

Regardless of foreign assessments of previous PRC IO efforts, the fact that Beijing continues to invest money, time, and resources into this behavior suggests that Beijing believes it is a worthwhile endeavor. This might be because Chinese propagandists believe that their overt public diplomacy and propaganda efforts are so poorly received that any covert efforts are helpful. It could also be that bureaucratic politics and Xi Jinping's centralization of power is driving Party-state IO efforts: Everyone is trying to please Xi and wasting money at ineffective efforts because that is what they think Xi wants, or because it allows them to argue to Xi that they are working toward his goals. Recent evidence that Chinese IO efforts are gaining more traction online suggest that the United States and other like-minded countries should not become complacent and should instead work to counter these activities.

What Generative AI Could Do for PLA Cyber-Enabled Influence Operations

Generative AI, as evident in the PLA writings surveyed above, is likely to improve existing overarching IO objectives and tactics, including in a conflict with the United States over Taiwan. I authored a recent RAND report that provides more detail on Chinese psychological warfare objectives, which very likely overlap with CDO: "degrading adversary decisionmaking, weakening adversary will to fight, undermining adversary support for war, undermining adversary government from within, along with supporting deterrence."[67] The report similarly details some relevant combat methods: "propaganda for persuasion, emotional manipulation, sowing discord, driving defections, as well as achieving deterrence and deception through psychological means."[68] Generative AI has the potential to improve PLA performance for all these objectives.

Specifically considering Chinese social media manipulation against Taiwan, the other recent RAND report forecast some potential changes under generative AI adoption, which is provided in Table 1 in an appendix at the end of this testimony.[69] For example, Beijing may be less interested in buying the accounts or otherwise bribing Taiwanese influencers to push pro-China narratives because generative AI would enable Beijing to create better viral content on its own. Beijing could also ramp up swarming—namely, creating content for spamming online discussions, such as comments or hashtags, like what Beijing has done on Xinjiang content.[70]

---

[66] Tsai Yung-yao and Jonathan Chin, "China Is Posting Fake Videos of President: Sources," *Taipei Times*, January 11, 2024; Shelley Shan, "China Might Use AI to Sow Chaos: NSB," *Taipei Times*, April 27, 2023.
[67] Beauchamp-Mustafaga, 2023, p. 12.
[68] Beauchamp-Mustafaga, 2023, p. 13.
[69] Marcellino et al., 2023.
[70] Global Engagement Center, "PRC Efforts to Manipulate Global Public Opinion on Xinjiang," U.S. Department of State, August 24, 2022; Global Disinformation Index, "Suspicious Twitter Hashtag Networks Promote Pro-China Line on Treatment of Uyghurs in Xinjiang," October 27, 2021.

Understanding Chinese Intent for U.S. Election Interference

There is growing concern by the U.S. government and broader national security community about the risks of Chinese election interference in the upcoming U.S. elections in November 2024. As the Office of the Director of National Intelligence (ODNI) said in its 2023 *Worldwide Threat Assessment*, "Beijing largely concentrates its U.S.-focused influence efforts on shaping U.S. policy and the U.S. public's perception of China in a positive direction, but has shown a willingness to meddle in select election races that involved perceived anti-China politicians."[71] This interference is not just at the federal level but also the subnational level: "Beijing has adjusted by redoubling its efforts to build influence at the state and local level to shift U.S. policy in China's favor because of Beijing's belief that local officials are more pliable than their federal counterparts."[72] This concern is furthered by a recently declassified U.S. National Intelligence Council (NIC) report on the 2022 U.S. elections that states that "China tacitly approved efforts to try to influence a handful of midterm races involving members of both U.S. political parties" and by nongovernment reports that such efforts continued into 2023.[73]

PLA research provides insights into how China may be targeting its election interference efforts. A 2021 article by PLA researchers from the National University of Defense Technology and PLASSF represented what was likely a proof of concept effort to use the social media activity of U.S. politicians to predict their favorability toward China.[74] The researchers used the activity of 21 high-profile U.S. politicians—including then-sitting senior officials in the Trump administration—on the X platform (formerly known as Twitter) as the training data to teach multiple deep learning models and then applied the models to predict the views of 20 then-sitting U.S. senators and governors, categorizing them into four U.S. political factions.[75] The researchers concluded that a fine-tuned pretrained Bidirectional Encoder Representations from Transformers (BERT) model performed best and then had "intelligence analysts" validate their assessments of individual U.S. politicians. The authors explained the value of their research as

---

[71] ODNI, *Annual Threat Assessment of the U.S. Intelligence Community*, February 6, 2023, p. 10.
[72] ODNI, 2023, p. 10.
[73] NIC, *Foreign Threats to the 2022 U.S. Elections*, December 23, 2022, declassified on December 11, 2023, p. i. This aligns with earlier nongovernment research. See, for example, Alden Wahlstrom, Jess Xia, Alice Revelli, and Ryan Serabian, "Information Operations Targeting 2022 U.S. Midterm Elections Include Trolling, Narratives Surrounding Specific Races, Politicians," Mandiant, December 19, 2022. For nongovernment reports since 2022, see Microsoft Threat Intelligence, 2023; Ben Nimmo, Nathaniel Gleicher, Margarita Franklin, Lindsay Hundley, and Mike Torrey, *Q3 2023 Adversarial Threat Report*, Meta, November 2023.
[74] Chang Chengyang [常城扬], Wang Xiaodong [王晓东], and Zhang Shenglei [张胜磊], "Polarity Analysis of Dynamic Political Sentiments from Tweets with Deep Learning Method" ["基于深度学习方法对特定群体推特的动态政治情感极性分析"], *Data Analysis and Knowledge Discovery* [数据分析与知识发现], Vol. 51, No. 3, May 2021, pp. 121–132. The authors' own translation is available in Chang Chengyang and Wang Xiaodong, "Research on Dynamic Political Sentiment Polarity Analysis of Specific Group Twitter Based on Deep Learning Method," *Journal of Physics: Conference Series*, 2020. Zhang is from the PLASSF Space Systems Department. This article was previously discussed with the commission by John Chen (John Chen, "Testimony Before the U.S.-China Economic and Security Review Commission," Hearing on "China's Cyber Capabilities: Warfare, Espionage, and Implications for the United States," February 17, 2022).
[75] The authors claim to more broadly also analyze 75 then-sitting U.S. senators and representatives, as well as 39 governors, suggesting a broader ambition.

"to assist intelligence analysts in their assessment of U.S. political inclinations and future paths for U.S.-China relations," by "using multiple types of deep learning technologies . . . to obtain relatively accurate dynamic political sentimental polarity from huge volumes of Twitter text data."[76]

To put it plainly, this research appears specifically targeted toward identifying politicians that might be favorable to China so that the PRC can support them, or likewise oppose politicians that they deem as anti-China. Indeed, this matches the NIC's assessment that "PRC intelligence officers, diplomats, and other influence actors probably viewed some election influence activities as consistent with Beijing's standing guidance to counter U.S. politicians viewed as anti-China and to support others viewed as pro-China."[77] This 2021 PLA article provides the first open-source evidence of how at least some Chinese Party-state actors could be leveraging social media not just to engage in IO but also to identify preferred candidates to support or oppose specifically for election interference. It also suggests that at least some in the PLA were considering such activities by at least 2020.

Implications and Recommendations for U.S. Policymakers

The PLA is one of China's premier actors for cyber-enabled IO already, and its adoption of generative AI would very likely improve its ability further while complicating detection and thus attribution. Furthermore, given CCP emphasis on the development of AI capabilities in general, the question is not whether the PLA will adopt generative AI in its cyber operations but how quickly and successfully it will integrate this capability. It is important to understand that Party-state actors, including the PLA, do not need to have access to *the best, cutting-edge* generative AI models to improve their cyber-enabled IO, making adoption and employment even easier. Lastly, the scope and ambition of PLA IO should be understood to likely go beyond mere military activities, such as deterrence and degrading adversary will to fight, to include foreign election interference.[78] In light of this fact, U.S. policymakers could consider the following recommendations.

Risk Reduction

Require social media platforms to label generative AI content and redouble their efforts to combat fake accounts. At a minimum, social media platforms could require users to label their own content as generated using AI. Ideally, social media platforms would lead the development of detection tools to automatically label uploaded content as appropriate. This could apply to all content on their platforms, not just political advertising.

---

[76] Chang, Wang, and Zhang, 2021.
[77] NIC, 2022, p. i.
[78] The PLA was also reportedly involved in previous efforts targeting U.S. elections. See David Jackson and Lena H. Sun, "Liu's Deals with Chung: An Intercontinental Puzzle," *Washington Post*, May 24, 1998; David Johnston, "Committee Told of Beijing Cash for Democrats," *New York Times*, May 12, 1999.

Invest in capabilities to detect generative AI-produced content, understanding that this technology will likely be a long-term investment. Although there are a growing number of tools available that claim to detect AI-generated content, tests suggest that their actual performance is inconsistent at best. However, the U.S. government should not adopt a fatalistic acceptance of the proliferation of such content and should instead invest widely in potential counters, with the goal that detection capabilities eventually catch up and surpass generative AI production. The U.S. government can also engage the private sector to support public-private collaboration. Congress could take a specific role by allocating dedicated resources to appropriate agencies' budgets and funding studies on the topic.

Promote Media Literacy and Government Trustworthiness

Mandate better media literacy training for U.S. government employees to identify inauthentic content and, especially, generative AI content. It is important for U.S. government employees to ensure that they are accessing accurate information, and generative AI makes this more challenging. This is especially true for national security agencies during a crisis or conflict, so preparations now by improving media literacy is very important. Congress could specifically take a role by incorporating this emphasis on training into relevant legislation.

Similarly, support media literacy for citizens' ability to recognize inauthentic content. The U.S. government can encourage broader educational initiatives to make the United States more resilient in the long-term against malign influence operations. Congress could specifically take a role by engaging with local organizations and constituents as part of constituent outreach, as well as by funding government agencies to similarly support such outreach.

Support blockchain, watermarking, or other similar technologies for media to improve the trustworthiness of authentic media, especially U.S. government public statements. If it is currently difficult to identify AI-generated content, then another approach would be to clarify the trustworthiness of important media. This could be pursued by the private sector, but at a minimum, the U.S. government should consider adoption if this proves promising. If this approach proves promising, then Congress could be an early adopter to ensure credible communication with constituents and the broader public.

Public Reporting

Commit now to releasing a nonpartisan declassified assessment by the U.S. intelligence community following the U.S. 2024 election. Given the heightened public attention to foreign election interference this year and the risks of politicalization, it will be very important to provide as much transparency as possible about the integrity of U.S. elections. The recent declassification of the NIC's 2022 report is useful but arguably too slow, given that it took over a year to release publicly.

Publish a yearly, dedicated report on foreign malign actors' efforts to influence U.S. public opinion, including via social media. The U.S. government's current reactive and selective approach to addressing foreign efforts risks inconsistency and incompleteness. A yearly report by the U.S. intelligence community or State Department's Global Engagement Center, modeled on the annual *Worldwide Threat Assessment*, would provide more details to the public

about foreign efforts and help baseline actual foreign activity over time for public discussions. This could also include specific sections focused on adoption of generative AI.

Take a more active approach to publicly calling out Chinese cyber-enabled influence operations, when attribution is available. There is certainly a balance between forcefully identifying malign activity by hostile actors versus calling more attention to their activities in the process, and the U.S. government can consider how to best weigh these trade-offs. Congress could specifically take a more proactive role by asking relevant executive branch officials about recent trends in malign activity during public hearings on a regular basis.

Diplomatically

Encourage Taiwan to increase its information-sharing, both publicly and privately, about Chinese cyber-enabled influence operations. This includes Taipei releasing a declassified report about Chinese efforts against their recent January 2024 election. Congress could take a role via its interactions with Taiwan interlocutors and also by encouraging the U.S. Department of State and other relevant government agencies to emphasize this interest in their own interactions.

Support Taiwan's engagement with other democracies to share its lessons learned and best practices combating Chinese IO. As Chinese efforts turn more global, Taipei is well-positioned to support other democracies as they work to counter Chinese malign influence, and Washington is crucial to providing appropriate platforms for such engagement. One option is the Global Cooperation and Training Framework, which provides opportunities for Taiwan to share its expertise on global issues of concern with other interested countries. Congress could take a role by supporting outreach to other global legislatures and elected officials.

Engage with allies and partners about the risks of Chinese cyber-enabled IO and cooperate on response options. This can include sharing information, publicly or privately, about Chinese malign activities, as well as sharing best practices on how to mitigate and respond to this growing threat.

Engage in dialogue with China on AI-driven social media manipulation. Despite the challenges of engaging Beijing in Track 1 or even Track 2 dialogue these days, it is worth considering whether there is any room for cooperation on limiting the adoption of generative AI for malign purposes. In October 2023, the Chinese MFA released its proposal for a "Global AI Governance Initiative" that specifically said that Beijing "[opposes] using AI technologies for the purposes of manipulating public opinion, spreading disinformation, intervening in other countries' internal affairs, social systems, and social order, as well as jeopardizing the sovereignty of other states."[79] Although this is a blatant lie, it provides an opportunity to begin a dialogue with Beijing, with the hope of reaching an agreement against such uses of generative AI.[80] The agreement on beginning a U.S.-China dialogue on AI, reached between President

---

[79] Chinese Ministry of Foreign Affairs, "Global AI Governance Initiative," October 20, 2023.
[80] Nathan Beauchamp-Mustafaga, "Biden Should Call China's Bluff on Responsible AI to Safeguard the 2024 Elections," *RAND Blog*, November 14, 2023, https://www.rand.org/pubs/commentary/2023/11/biden-should-call-chinas-bluff-on-responsible-ai-to.html.

Biden and General Secretary Xi Jinping in November 2023, lays this foundation, although it appears the current focus overlooks IO as a potential topic for inclusion.[81] Understanding that Beijing may well violate this agreement like it has in the past, a bilateral agreement would at least empower the United States to hold Beijing accountable publicly and better engage with allies and partners on the topic.

Additional Research

Fund additional research on Chinese strategy for cyber-enabled influence operations, including the PLA's CDO operational concept. There is very limited high-quality understanding of Chinese strategies, capabilities, and intentions, and public discussion would benefit from additional information on the topic.

Conduct an independent assessment of the net benefit of U.S. government information efforts. This could include DoD and other relevant agencies. At DoD, for example, this review could ensure that all DoD messaging—whether by combatant commands or others—aligns with stated DoD strategic priorities and relevant strategic messaging. The review could also consider whether U.S. activities against one adversary produced sufficient benefits weighed against potential downsides in behavior from other adversaries. Given the risk that Chinese observations of U.S. activities are driving Chinese malign activities, compared with an uncertain benefit from such U.S. activities, it is worth considering the net value.

Fund additional research evaluating the risk of malign influence operations presented by generative AI models from Chinese companies. Because this is a novel risk and currently uncertain, it is worth further consideration.

[81] Graham Webster and Ryan Haas, "A Roadmap for a US-China AI Dialogue," Brookings Institution, January 10, 2024.

Appendix. Potential Implications of Generative AI for PRC Social Media Manipulation Against Taiwan

Table 1. Potential Implications of Generative AI for PRC Social Media Manipulation Against Taiwan

Common PRC Tactic	Definition	Previous Shortcoming	Potential Implication with Generative AI
Advertising	Paid promotional content to support a cause or actor	PRC paid Taiwan influencers to promote pro-CCP content, but they sometimes were easy to identify with blatant one-off messages	May diminish; PRC no longer needs to pay others to create viral content if it is able to generate convincing, authentic text
Bots for astroturfing	Using large numbers of inauthentic (fake) accounts (bots) to create the appearance of a broad consensus on a topic	PRC has largely relied on human-generated comments, limiting quality and scale	Likely to increase dramatically; generative AI will give bots written voices that are near-indistinguishable from human-created content
Cheapfakes and recontextualized media	Supporting a campaign either with simple edits or by repurposing media (usually, images)	PRC attempts are relatively easy to identify and slow enough for Taiwan government to expose and debunk	May diminish; realistic, highly believable fakes will be far cheaper to make en masse and may not be able to be identified or may overwhelm Taiwan government response capabilities
Impersonation	Pretending to be another person in order to misrepresent their position or views	PRC relies on pressuring public individuals into creating misleading information (especially, confessions)	Now possible to (1) mass-generate text in the style of a given individual's writing (2) falsify images of an individual and produce those images en masse. There is no longer a need to actually coerce a targeted individual
Keyword squatting	Creating mass content to manipulate search engine results related to a given term, phrase, or hashtag	Past PRC campaigns on Xinjiang issues have lacked variety, making them easier to detect	Generative AI does not revolutionize keyword squatting's mechanism but permits squatters to automate mass content generation containing a given keyword
Swarming	Loosely organized groups coordinating to fill an information space (e.g., spamming a comment section)	50 Cent Army members are inconsistent in their ability to avoid detection or achieve specific narrative goals	Generative AI automates the process of creating mass unique content for spamming a comment section or otherwise drowning out a narrative

Common PRC Tactic	Definition	Previous Shortcoming	Potential Implication with Generative AI
Testimonials	Personal stories used to elicit emotional reactions or sway opinions	PRC manufactured testimonials have historically been presented on state media and appear to be relatively scripted	Generative AI is capable of writing short-form and long-form testimonials of wide-ranging content on a mass scale, representing various demographics for both broad and niche effects

SOURCES: Reproduced from Marcellino et al., 2023, p. 22. Originally adapted from Harvard Kennedy School Shorenstein Center for Media, Politics, and Public Policy, *The Media Manipulation Casebook Code Book*, version 1.4, updated January 7, 2022; Aaron Huang, *Combatting and Defeating Chinese Propaganda and Disinformation: A Case Study of Taiwan's 2020 Elections*, Harvard Kennedy School Belfer Center for Science and International Affairs, July 2020.