

"The US & Japan: Allied Against Disinformation: Next Generation Voices Speak"

Speakers: Jonathon Marek, Taro Nishikawa, Rachel Brooks, Ryohei Suzuki, Thomas Ramage Moderator: Rob York

Key Findings

Strategies for Creating and Spreading Disinformation

The proliferation of disinformation is a cognitive warfare tactic with four goals: demoralize, destabilize, create crisis, and normalize disinformation. This is a divide-and-conquer tactic which results in radicalization and polarization of the masses. The information cocoon created by social media algorithms and the influence that authoritarian powers have over these is a significant risk as they make countering the spread of disinformation extremely difficult. The exact specifics of social media algorithms are kept secret from the government; however, they work by showing users similar content to what they currently interact with. This recommended content is often content that agrees with the user's own preferences and viewpoints. Disinformation campaigns target this aspect of the algorithm by creating targeted messages towards specific audiences.

These audiences are created from a segmentation process which considers a user's online behavior, their social status, their connections, etc. Generative AI and deepfakes are taking a greater role in producing the content sent out to these audiences as they can create vast quantities of highly targeted disinformation. The generated disinformation slowly encroaches into the individual's information cocoon and skews it over time. This disinformation often has political goals, and it can have significant impacts on the individual's cognitive processing. This process is even more evident when the digital ecosystem is dominated by an authoritarian regime.

For example, China has influence over TikTok and WeChat, which are state-linked, or stateowned, respectively, entities. Not only have they been found to host large numbers of disinformation campaigns, but their cooperation in targeting disinformation is varied. Disinformation campaigns from threat actors in democratic states like Taiwan have been identified by TikTok, as well as non-PRC authoritarian states like Russia, but have demonstrated a large and willful blind spot for those from the PRC. Notably, WeChat did not take down a disinformation campaign targeting an elected official in Canada. Ineffective and negligent responses to disinformation make mitigating the effects significantly more difficult in an online ecosystem where information cocoons and segmentation strategies elevate the effectiveness of disinformation campaigns.

Disinformation Mitigation Strategies

The free and open digital ecosystem is valuable and worth protecting, and US-Japan cooperation should be at the forefront of its defense. The US and Japan must bolster the bilateral capacity for strategic communication and share intelligence of adversaries, which will allow the states to effectively respond to disinformation campaigns. A joint platforms analysis center, modeled on the Huawei Cybersecurity Evaluation Center and undergirded by US-Japan cooperation could identify potential influencer coercion, investigate the spread of disinformation, and provide recommendations to various digital platforms, ultimately limiting the spread of authoritarian state-linked disinformation. The center would be universally applicable, transparent and evidence based. It is also imperative to mitigate adversary capabilities, while ensuring accountability of the preventive entities.

Agreements currently in place, such as the US-Japan Digital Trade Agreement—are not designed to facilitate counter-disinformation cooperation. Reforms could allow not only their governments, but also independent agencies and researchers as well, to evaluate the potential for weaponization of the platforms' algorithms in support of state-linked disinformation and monitor authoritarian-influenced platforms. Like-minded partners beyond US-Japan cooperation could also be included in cooperative efforts to prevent the spread of disinformation. The discussion of foreign information manipulation should be promoted with G7, the Quad, and the Indo-Pacific Economic Framework for Prosperity (IPEF). Lastly, incorporating a Trade and Technology Council (TTC) into US-Japan-ROK relations will create another valuable communication space to address and discuss relevant disinformation issues. Effective cooperation among the US and Japan is crucial to creating a free and open digital space rid of manipulation and disinformation, and support from like-minded partner nations would only reenforce these efforts.

The Role of Public Enterprise in Countering Disinformation

Public education and public broadcasting can play a major role in countering disinformation. The goal is to utilize public goods and trusted institutions to promote information literacy and prevent the success of disinformation campaigns across all populations. Trusted institutions like NHK are the first line of defense and they have the capabilities to engage in a variety of fact-checking work. The advantage of using public enterprises is that they are responsible to the national population and are not state controlled. NHK has the resources and public trust necessary to be a consistent and reputable source for fact-checking and disinformation mitigation activities. In public education the work is done by having a variety of exposure across a student's learning experience to give students a toolkit for understanding disinformation work at all levels from establishing specific values (fact vs. opinion etc.) to doing a more detailed analysis of strategies used by disinformation campaigns. The advantage of using public goods in disinformation mitigation is that they are accountable and accessible to the public while not having the associated concerns when government agencies are involved in fact-checking and debunking disinformation.

This document was prepared by Patrick Kovacs and Elijah Collier. For more information, please contact Rob York (<u>rob@pacforum.org</u>), Director for Regional Affairs at Pacific Forum. These preliminary findings provide a general summary of the discussion. This is not a consensus document. The views expressed are those of the individual speakers and do not necessarily reflect the views of all participants. The speakers have approved this summation of their presentations.