Pacific Forum - Session I of the US-Japan "Allied Against Disinformation" Series
Apr. 20, 2023 | 2:00pm HST | Virtual (Zoom)

"The US & Japan: Allied Against Disinformation, Part I: Definitions & Defenses"

Speaker: Christopher Paul
Moderator: Crystal Pryor

Draft Key Findings

Defining & Distinguishing Disinformation
Social media and modern internet culture has made spreading falsehoods, whether intentionally (disinformation) or unintentionally (misinformation), easier than ever before. This can be understood as a megaphone-shaped process with three distinct stages: production, redistribution, and consumption. Disinformation begins with production, where disingenuous actors such as government agents in Russia, China, or North Korea, advance an agenda by manipulating information to fit a persuasive narrative. This can be done through wholly or partially fabricated media, selective use of facts, deliberate obfuscation, exploitation of an appeal to authority, or the application of rhetorical fallacies such as false equivalency and strawmanning. Further proliferation of this disinformation by "bot" accounts and compromised actors sharing the material, is called redistribution. Finally, consumption is the end stage where a falsehood reaches its target audience, often influencing their opinions and affecting sentiment on an issue.

Frameworks for Countering Disinformation
Humans are demonstrably poor at discriminating truth from falsehood, so it is imperative to equip people in free and open internet societies like the US or Japan with frameworks to dismantle pernicious social media campaigns where possible. Governments, platforms, and civil society each have a responsibility to combat disinformation, but a balance between all three needs to be maintained to protect the rights of civil society and maintain a free media industry. While governments may consider regulating the production and distribution of disinformation on social media, they are likely too slow and clumsy to implement this effectively in the ever-evolving environment. The threat of regulation may be a more productive force than regulation itself, by incentivizing social media platforms to self-regulate for users' benefit. Actions that platforms may take in countering disinformation include revising terms of service, enforcing terms of service, fact-checking efforts, warning labels, algorithm reworks, and investment in moderation. Civil society for its part can combat disinformation with responsible social media habits, and reporting known bad actors while promoting credible voices.

Artificial Intelligence as a Double-Edged Sword
Japan's 2022 National Security Strategy included both the intent and structure for greater functions in countering disinformation, so the government recently announced a new secretariat dedicated to this end. Emerging new capabilities in AI have supercharged the ability of those on the disinformation "offensive"

to produce at volume. While the defense is stuck playing catchup, AI may also prove to be a powerful tool for platforms. Rapid automated responses such as consumer warnings or flagging mechanisms may assist moderation systems in stopping the spread of bad information early on. Generative Pre-trained Transformer (GPT) type AI tools have incredible potential but still possess shortfalls. First, these tools can occasionally produce false information themselves, and will assert their falsehoods with undue confidence. Second, is the black box problem. Though an AI tool may propose a solution, it cannot fully explain how or why it reasoned that the given solution was appropriate or best. While AI will almost certainly magnify efforts both good and bad in combating falsehoods online, it can still only make a given claim more persuasive to an extent. Disinformation can challenge a weakly-held belief, or exacerbate an already strongly-held belief, but it is unlikely to dissuade someone from a strongly-held belief.

This document was prepared by Brandt Mabuni. For more information, please contact Rob York (rob@pacforum.org), 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 speaker and do not necessarily reflect the views of all participants. The speaker has approved this summation of their presentation.