twosix
TECHNOLOGIES

MESSAGE MANIPULATION IN THE DIGITAL LANDSCAPE

UNDERSTANDING HOW NATION STATES USE THE INTERNET TO SPREAD MISINFORMATION, SUPPRESS DISSENT, AND SHAPE NARRATIVES

Message Manipulation in the Digital Landscape

If you've spent any time on the internet, chances are you've been a target of foreign adversary-backed content manipulation, just like you've almost certainly been influenced by social media ads and manipulated search results.

Inauthentic promotion isn't exclusive to private companies hoping to garner traffic or make a quick buck off ad revenue. Nation state bad actors frequently employ inauthentic promotion to shape domestic *and* international narratives and censorship to block sensitive discussions at home.

Modern methods of inauthentic promotion and censorship are often subtle and nuanced. Instead of approaching content like a sledgehammer, smashing down on dissent, they are far more like digital scalpels designed to shape public opinion and conversation around sensitive events and issues. Understanding what content foreign adversaries manipulate — and how — can be an indicator of foreign adversaries' overarching political, economic, and international goals.

While U.S. federal entities have strategies in place to keep a finger on the pulse of how foreign adversaries enact messaging manipulation, there is more that can be done to consider the cultural, linguistic, social, and political context behind manipulated content and netizen response, and the ever-evolving factors that motivate our adversaries.

This white paper will explore what inauthentic promotion and censorship look like on the internet today, examine the common pitfalls of studying censorship and how to avoid them, as well as explore how contextualized data on manipulated messaging can inform critical U.S. government decisions.

What is Inauthentic Promotion?

Inauthentic promotion is a digital strategy foreign adversaries use to manipulate and shape narratives around certain events, subjects, or topics. These messaging tactics seek to break down and control global discourse and discredit contrary reports and opinions. Ultimately, the goal with inauthentic promotion is to raise the profile of specific narratives by creating the artificial appearance of support for certain viewpoints and perspectives.

Inauthentic promotion can also act as a method of suppressing dissent. Amplified promotion differs from censorship in that it does not entail blocking or removing dissenting information; it simply *drowns out* dissent by flooding search engines, social media platforms, and other online forums with misinformation.

This is a particularly effective method of shaping narratives because it still allows netizens some degree of autonomy over what they can and cannot post. As such, foreign adversaries can create an illusion of free speech that — while technically there — is severely diminished.

Pro-PRC Inauthentic Campaigns

Thousands of inauthentic accounts affiliated with the government of the People's Republic of China (PRC) publish content each day on global social media platforms in a variety of languages to promote Beijing's official narratives and drown out dissent.

Starting in 2017, pro-PRC stakeholders began pumping counter-narratives about the mistreatment of Uyghurs in Xinjiang. To suppress information about the atrocities in Xinjiang, PRC authorities use a method called "astroturfing," in which they organize streams of inauthentic posts designed to create a veneer of authentic, grassroots support for certain policies and viewpoints.

More recently, inauthentic pro-PRC accounts produced content about global conflicts, such as the Israel-Hamas and Russia-Ukraine wars to promote anti-US narratives and pro-China sentiment in targeted regions.

What is Censorship?

Censorship is the explicit blocking or silencing of key terms, phrases, and information. While inauthentic promotion does not inherently prevent the sharing of dissenting information, censorship does, as automatic and manual censorship methods block certain keywords or phrases from being published or remove sensitive threads after publication.

Current models of censorship are designed to discourage netizens from finding, engaging with, or creating posts regarding sensitive or controversial events — other than those in support of an authority's preferred narrative. This can make it appear as though there is no organic or dissenting discussion around that particular topic.

Authoritarian governments intimidate people into self-censorship as a form of censorship. For example, nearly 22% of content in 2023 authored by pro-PRC inauthentic accounts on global social media aimed to intimidate overseas dissidents into silence by slandering them. This particular approach was likely successful at suppressing other potential outspoken critics, as many of these critics publicly testified how much these campaigns damaged their lives.

Who Is Responsible for Inauthentic Promotion and Censorship?

Inauthentic promotion and censorship is built on a complex web of actors and stakeholders. It is not controlled by a simple on/off switch for which a single person (or even single department) bears responsibility.

Covert manipulators are actors that are *not* official foreign government accounts. These manipulators can be bots or humans. These channels will typically push more extreme narratives that may be aligned with the foreign adversary's stance but are too controversial for official accounts to post.

Facebook information warfare: Inside Iran's shadowy operations to target you on social media

Jessica Guynn
USA TODAY

It's part of Iran's strategy to achieve security.

They also may push narratives that *do not* align with a government's official stance but are intended to discredit the opinions of dissidents. For example, sources have shown that Iran-backed actors will flood Instagram with radicalized memes that don't reflect the official position of the Iranian government but are intended to influence the beliefs of U.S. social media users.

Censorship Approaches

Foreign adversaries' censorship techniques are not stagnant — they continually evolve and adapt along with the news and users' attempts to evade control. Censors must react to political changes and current events as public discourse and key words, phrases, and topics shift. For example, in China in 2020, online discussions about COVID accounted for at least 50% of all censored discussions — a massive shift in the focus of the country's censorship apparatus. Censors similarly honed in on COVID conversations in 2022 with the Zero Tolerance Policy lockdowns.

However, censorship works differently in different information spaces. And governments are often able to control online discussion with selective or targeted censorship, rather than removing all related posts or content.

Understanding China's Censorship System

The Chinese government uses various methods to censor online content, including filtering and blocking websites, deleting social media posts, and shutting down entire websites or social media accounts. Censors of social media content in China primarily aim to prevent collective action, such as street protests or widespread demands for reforms, from starting or gaining traction.

Social media platforms in China are required by law to carry out censorship to protect the government's hold on power but are incentivized by the market to refrain from censoring more than competitors' platforms. Social media platforms, therefore, walk a fine line between censoring just enough content so they don't run afoul of regulators and allowing just enough free expression so that their platform maintains a competitive edge in the market.

5 – 10%

China only needs to censor 5 – 10% of a conversation to slow down or stop it entirely. Blanket censorship on a topic is relatively rare because social media platforms do not wish to scare away users and lose profit — and as such, social media companies must employ humans to review content and make judgment calls on what is worth censoring or not.

Censorship in Russia vs. Censorship in China — A Game of Catch-Up

When the internet arrived in the USSR in the 1990s, the Soviet Union was in its final days and the region was flirting with democracy. As a result, the internet infrastructure in the region was built like that of the West with an eye toward openness and sharing.

By contrast, in China the Chinese Communist Party (CCP) was solidly in power when the internet rose to prominence, such that the Party was able to shape and design it from the ground up. The government built in tools to monitor, block, and remove content and control the flow of information from the outside world, a system often referred to as “The Great Firewall.”

This difference in internet architecture means that the Chinese and Russian governments use different tactics and tools to control their domestic information environments. While Beijing continues to use the censorship techniques it built into its information environment, Moscow is newer at censorship efforts, having leaned into it more heavily since the invasion of Ukraine. This content control is part of a broader effort to crack down on independent media and civil society that accelerated during the pandemic.

New laws criminalizing criticism of the Kremlin and associated prosecutions of those who speak out on social media have instilled a greater culture of fear, while blockages of social media platforms like Facebook and the websites of some Western news outlets have made it less likely that Russians would stumble across anti-Kremlin messaging.

Common Pitfalls in Understanding Foreign Governments' Online Content Manipulation — and How To Avoid Them

Collecting data about inauthentic promotion and censorship is critical to develop and enact the right response to foreign adversaries' aims. However, that data is only valuable if it is accurate and viewed through the lens of the foreign government's cultural, social, and operational goals.

For example, many US government-affiliated researchers studying Chinese government online manipulation only consider content the PRC publishes on the X platform — not Weibo, where Beijing conducts much more of its messaging work. While X posts are easier for U.S.-based researchers to collect and observe, and the U.S. government has fewer than 10 official accounts on Weibo, this opportunity-based collection means observers cannot see the range of Chinese government messaging, which often differs for audiences at home and abroad in important ways. Ignoring information published behind China's Great Firewall leaves the researcher without the full picture of the PRC's intentions.

A review of the publicly available research on pro-PRC content manipulation on global and domestic social media shows that there are three main pitfalls to avoid in working to understand the significance of online messaging and Beijing's true intent.

Pitfall 1: Attributing inauthentic activity to the Chinese government without enough evidence.

A good story is not an indicator. Many different researchers and observers in this field noted and debated whether or not a massive inauthentic campaign on the X platform on 1 December 2022 was sponsored by the PRC. At the time, most voices concluded that it was, because it made sense that the PRC would sponsor the Chinese-language campaign one week after it faced its largest protests in forty years. However, with more time and data collection, it now appears more likely that a private sector entity is the sponsor of the campaign. That campaign — and other similar campaigns on X — likely target PRC-based netizens to sell them services, scam them for money, or steal their data.

In these cases, observers of inauthentic activity can be fooled by the language in which the content is written. Chinese-language content from inauthentic accounts is not an automatic indicator of PRC sponsorship, and mis-attribution of this signal can lead to gravely misinformed analysis.

Pitfall 2: Assuming inauthentic messaging represents Beijing's true or only message.

Inauthentic accounts can sometimes push messages that do not align with Beijing's official message. Research shows that the PRC government and inauthentic accounts pushed different messages in response to the same events on several occasions from 2022 through 2023. Many inauthentic accounts have a level of freedom to post their own interpretation of a common message. Inauthentic messaging should not be equated with official government messaging, and analysts should not assume that such messages represent the central government's true intent.

Pitfall 3: Viewing the Chinese government as a monolith in its actions online.

Bureaucracies are not known for close, consistent, and rapid coordination across departments, in particular on covert strategies, and China's government is no exception. It's useful to know what department within the government is producing the content or sponsoring the inauthentic messaging because that will help uncover likely motives and whether something represents whole-of-government messaging.

To avoid these pitfalls in researching and understanding Chinese government intent online, organizations should work to:

- **Understand messaging from the actor's perspective.** U.S. federal entities will be more successful in decoding the intent behind a messaging campaign if they work to understand how the messaging helps support the government's political goals. The content should be read, when possible, in its original language so the intricacies and subtleties of phrasing are not lost in translation.

- **Know the actor — and know them well.** While it is difficult to determine whether some content is inauthentic or official, the distinction is important. Researchers and U.S. government organizations should strive to make decisions about foreign government intent in messaging only when they have high confidence about the source: authoritative vs. non-authoritative, inauthentic vs. authentic, diplomat vs. state media, and so on. Understanding who is talking helps you know why the content is important to that speaker, what their goals are, and the possible effect of the campaign.

- **Use baselines for reporting inauthentic activity.** When dealing with measuring the scale and significance of inauthentic activity, it's critical for teams to determine whether the activity is statistically significant.

The Benefits of Monitoring for Inauthentic Promotion and Censorship

When U.S. agencies strongly consider cultural and social context when gathering critical intelligence, they can uncover the true motivations behind foreign adversaries' message manipulation programs. Unveiling that truth helps federal entities plan accordingly and create the most effective strategic response to nation-state narratives.

With the right processes in place, government intelligence officials can leverage data in inauthentic promotion and censorship to:

Help U.S. Federal Entities Understand a Foreign Adversary's Limits

Censorship shows the limits of acceptable discourse in rival nation states. These limits are typically looser and broader than many would expect because social media platforms are businesses that want to retain users, and so must balance government policies against their goal of creating a space for people to engage in conversation.

As such, foreign governments can't implement too much censorship, otherwise it risks compromising the very platforms it leverages to shape narratives.

Censors in the PRC rarely remove 100% of targeted content for that very reason, which exposes vulnerabilities in an adversary's messaging apparatus and allows U.S. federal entities to identify and collect critical information on the status quo within that foreign nation.

> "We must adhere to the Party's management of the internet and adhere to (the principle of) making the internet work for the people." –
>
> Xi Jinping, General Secretary of the CCP (2012 - present)

Provide A Window Into the a Nation State's Perspective on Warfare

In an increasingly tech-oriented global ecosystem, digital activity can absolutely be considered a means of attack. Whether it's continued attacks on Ukrainian operational technology by Russian nation state actors or data breaches like the 2023 vulnerability exploit by Russia-backed actors, technological warfare is another method U.S. government entities must study and defend themselves against.

The Truth About Misinformation

If there's one thing you take away from reading this paper, it should be this: We have the power to understand adversaries' intentions with careful study of how they manipulate content online. Gathering meaningful intelligence requires context. Digital activity around controlling and shaping narratives offers an essential window to the aims of adversarial governments which arms U.S. government agencies with the intelligence they need to respond quickly and accordingly.

The truth about misinformation and censorship is it requires an intensive level of technical and cultural expertise to identify. More than ever, government intelligence officials need to tap into resources that can demystify many of these nuances, which can include leveraging language experts. The information age has introduced more subtle, nuanced methods of inauthentic promotion and censorship.

At Two Six, our Media Manipulation Monitor (M3) solution decodes foreign markets to uncover political, military, and economic insights on critical adversaries. We combine human-in-the-loop analysis with subject matter experts operating at the top of their fields, so U.S. government organizations can breathe easier knowing the insights we deliver will have the cultural, linguistic, and political context they need to make the right national security decisions.

M3

READY TO TURN YOUR INTELLIGENCE GOALS INTO A REALITY? LEARN MORE ABOUT M3.