Graphika

Cheap Tricks

How AI Slop Is Powering Influence Campaigns

By Dina Sadek and Margot Fulde-Hardy

11.2025

Cheap Tricks

How AI Slop Is Powering Influence Campaigns

By Dina Sadek and Margot Fulde-Hardy

Key Findings

- When generative AI tools became widely accessible, the research community anticipated that AI integration would lead to higher-quality and more sophisticated AI-enabled influence operations (IOs) that could deceive and mislead audiences, especially around elections.
- Research on AI-enabled IOs, including the case studies documented in this report, indicates that the production of these campaigns has increased in speed and scale. However, while AI tools helped to amplify AI-generated text and footage across social media platforms, the quality and sophistication of this content remain low.
- The case studies here discuss well-documented IOs, including CopyCop, Doppelgänger, Spamouflage, Falsos Amigos, Operation Overload, Operation Undercut, and a pro-India celebrity impersonation campaign. Taken together, they demonstrate that manipulative state and non-state actors are actively seeking to use AI tools to deceive and mislead target audiences by amplifying polarizing or sensational content. Their output is often low-quality, yet scalable content that poses as fake and authentic individuals and institutions.
- Using the Actor, Behavior, Content (ABC) model, we analyzed IOs attributed to state and non-state actors to identify their most common characteristics. We examined the Tactics, Techniques, and Procedures (TTPs) deployed in these operations for indicators of increased use or reliance on AI tools to conduct and scale their operations.
- The sponsors behind publicly documented IOs have delegated core functions such as content creation and persona generation to AI tooling. These campaigns then saturate the information ecosystem with low-quality content but highly scalable campaigns that blur the lines between professional and amateur actors.
- The AI-enabled IOs documented by Graphika and others show that, based on engagement metrics, their ability to reach organic audiences remains limited.

Analysis

The research community was right to suggest that AI-enabled IOs risked becoming more sophisticated and convincing. However, those expectations were largely overstated compared to the observed and documented changes in TTPs and content, which often feature poor grammar, broken text, watermarked AI-generated profile pictures, stiff, synthetic video presenters, and generic templated websites.

In this report, we present case studies, as documented by Graphika and others in the research community, along with examples of the use of AI-generated content in IOs to highlight how low-quality, mass-produced content is polluting the internet and distracting audiences' attention and focus. Using the Actor, Behavior, Content (ABC) framework, the most common characteristics we observed include:

- Manipulative **actors** seeking to deceive and sow distrust among audiences by deploying AI to pose as fictitious and legitimate individuals and institutions.
- Using automated and AI-generated tools to conduct cross-platform coordinated inauthentic and deceptive **behavior**.
- Low-quality, easily scalable, and potentially harmful AI-generated **content**.

As evidenced by the case studies below, AI-enabled operations vary in quality and sophistication, and some rely primarily on mass-produced, low-quality, and provocative content, which ultimately achieves minimal engagement based on publicly visible metrics. The names researchers give to these IOs often reflect the spammy and false nature of the operations: Bad Grammar, Doppelgänger, Falsos Amigos, and Spamouflage. These campaigns include repetitive and polarizing content that prioritizes quantity and mass targeting over quality.

The increased use of AI tools to create personas and content for social media and online platforms has made IOs easier and cheaper to create, enabling campaigns to grow rapidly. While this content often appears innocuous and mostly entertaining, some targets vulnerable groups, elections, public officials, and political institutions. Long-term consequences of these campaigns extend beyond misleading audiences to saturating the information ecosystem and poisoning large language models (LLMs).

AI and IOs: Managing Expectations

In January 2023, Georgetown University's Center for Security and Emerging Technology, OpenAI, and the Stanford Internet Observatory discussed the emerging threats posed by generative language models and automated IOs. The researchers predicted that generative AI would affect the three dimensions of the Actor, Behavior, Content (ABC) framework of influence operations (see below). These changes included an increase in the number of actors and their outsourcing,

scale of campaigns, greater efficiency, and new tactics to further credibility and persuasiveness, at the expense of discoverability.

ABC	Potential Change Due to Generative AI Text	Explanation of Change
Actors	Larger number and more diverse group of propagandists emerge.	As generative models drive down the cost of generating propaganda, more actors may find it attractive to wage influence operations.
	Outsourced firms become more important.	Propagandists-for-hire that automate production of text may gain new competitive advantages.
Behavior	Automating content production increases scale of campaigns.	Propaganda campaigns will become easier to scale when text generation is automated.
	Existing behaviors become more efficient.	Expensive tactics like cross-platform testing may become cheaper with language models.
	Novel tactics emerge.	Language models may enable dynamic, personalized, and real-time content generation like one-on-one chatbots.
Content	Messages grow more credible and persuasive.	Generative models may improve messaging compared to text written by propagandists who lack linguistic or cultural knowledge of their target.
	Propaganda is less discoverable.	Existing campaigns are frequently discovered due to their use of copy-and-pasted text (copypasta), but language models will allow the production of linguistically distinct messaging.

Table 6: How Language Models May Influence the ABCs of Influence Operations

Breakdown of changes likely to occur across the ABC framework. Source: Georgetown University/OpenAI/Stanford Internet Observatory.

In 2024, a record number of elections took place, and industry players warned about the risks AI posed. For example, in the context of the November 2024 U.S. election:

- OpenAI said in January 2024 that it was working to prevent abuse, provide transparency on AI-generated content, and improve access to accurate voting information. In the month leading up to U.S. Election Day, ChatGPT rejected over 250,000 requests to generate DALL·E images of U.S. politicians.
- In February 2024, Anthropic said it expected to see "surprising uses of AI systems - uses that were not anticipated by their own developers."

However, by the end of the year, the same industry actors recognized that the threat may have been overstated when compared to the volume of activity targeting elections.

- In December, Anthropic stated that "election-related activity constituted less than 0.5% of overall use, ticking up to just over 1% of total usage in the weeks leading up to the U.S. election."
- After monitoring the impact of AI on elections, the risk of widespread deepfakes, and AI-enabled disinformation campaigns throughout the year, Meta concluded that "these risks did not materialize in a significant way and that any such impact was modest and limited in

scope." Meta's Imagine AI image generator rejected 590,000 requests to generate images and election-related deepfakes of U.S. politicians, the company said.

Despite limited evidence of impactful, AI-based campaigns, fears of potential threats to the information ecosystem have not disappeared in 2025. The European non-profit AI Forensics warned that 25% of TikTok's top search results contain synthetic AI imagery, and that over 80% of this TikTok AI content comes from agentic AI accounts. Meanwhile, a recent leak revealed how GoLaxy, a Chinese company that Pamir Consulting alleges has links to the Chinese military, claims to have integrated AI into its systems to collect data, identify user profiles, and craft tailored messaging and content in large-scale IOs.

While the GoLaxy Papers suggest that threat actors strive to develop sophisticated, large-scale, AI-powered IOs, we have not yet observed compelling evidence of a campaign of this nature. Instead, the IOs documented publicly by the research community reveal that the actors behind them have primarily used AI tools to create spoofed websites, slop content, and easily scalable, shareable scripts with large audiences.

Common Traits of AI-Enabled Campaigns

The increased accessibility of generative AI tools has contributed to the rapid growth of a new category of IOs. In seeking to leverage AI to achieve their goals, these campaigns are characterized by low-quality, mass-produced AI-generated content and/or scaled AI-generated infrastructure. The result is improbable, often laughable content with messaging that lacks credibility. This material is distributed through an infrastructure that allows researchers to easily identify signs of inauthenticity, and in some cases, the likely sponsors or operators of the IO.

Using the ABC model, we analyzed IOs attributed to state and non-state actors to identify their most common characteristics. We examined the TTPs deployed in these operations, as documented by the research community, for indicators of increased use or reliance on AI tools to conduct and scale their operations. Using a representative sample of documented IOs as case studies, we observed that these operations are using low-quality, spammy, and polarizing content.

The most common characteristics of the AI-enabled IOs we examined are as follows:

- **Actor:** Using generative AI to impersonate trusted media sources, politicians, and political institutions, creating spoofed and fake websites and social media accounts that feature manipulated videos and AI voiceovers. This is one of the most common uses of AI in IOs.

  Case Studies: CopyCop, Doppelgänger, and Spamouflage.

- **Behavior:** Using generative AI to increase the scale of content production and dissemination, with little attention paid to the quality of the content. A frequent example of this behavior involves actors using automated tools to post content across multiple platforms simultaneously.

Case Studies: Falsos Amigos, Shadow Play, and Sneer Review.

- **Content:** Using generative AI to exploit trending topics and produce highly polarizing, spammy, sensationalist content. AI-generated thumbnails, cartoons, and short videos are common types of content designed to capture the audience's attention.

  Case Studies: Operation Overload, Operation Undercut, and Pro-India Celebrities Impersonation.

Case Studies

These case studies signal the emergence of a new class of AI-enabled, low-effort influence operations that achieve some visibility despite having glaring quality flaws. Rather than raising the bar on the level of polish to these operations, AI has lowered it: enabling cheap, fast, and outsourced IO supply chains that prioritize quantity over credibility. Sophisticated state-sponsored campaigns still exist — but many operators now rely on off-the-shelf AI tooling to delegate content creation, account generation, and messaging at scale.

IO researchers have suggested adding a "D" to the ABC framework for "Information Distribution," to understand how this content operates online. As malicious actors continue to exploit emerging tools and software to create and automate false personas and content, it is also imperative to investigate and track their distribution networks and the vulnerabilities they leverage to manipulate and deceive audiences.

Actor: Impersonation of Media Sources and Authority Figures

Case Study 1: Doppelgänger

The Doppelgänger influence operation — which Meta and the U.S. Treasury have linked to the Russian government and Russian companies, including the Social Design Agency and Structura National Technologies — uses AI-generated content to create and disseminate false information and narratives while impersonating authentic news outlets, governments, and think tanks. Ahead of the 2024 U.S. elections, the U.S. Department of Justice released an affidavit detailing Doppelgänger's alleged operations, which targeted the U.S. The court filing included examples of the IO "cybersquatting" domains of legitimate U.S. media outlets, such as the Washington Post and Fox News, featuring polarizing and sensational news headlines.

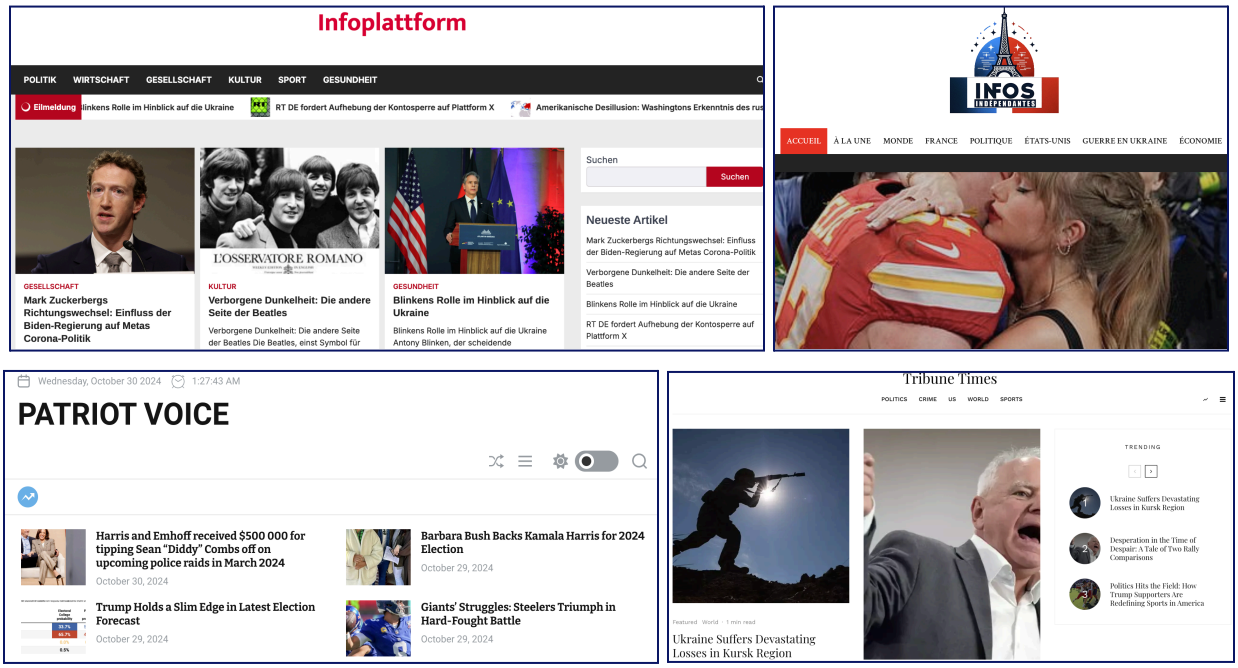
Screenshots of spoofed websites of U.S. media outlets Fox News and the Washington Post, as provided in the U.S. Department of Justice's 2024 affidavit.

Doppelgänger content typically involves websites that impersonate legitimate media outlets, civil society organizations, and political institutions. These inauthentic websites promote pro-Russia narratives, which often fail to attract engagement from authentic users beyond existing pro-Kremlin communities. Meta reported in 2022 that it removed a network of accounts using its social media platforms to promote "crude ads and fake accounts" for these fake media outlets, highlighting the amplification strategy employed by this influence operation.

Doppelgänger was somewhat successful in creating convincing fake and illegitimate news sites, but ultimately failed at attracting significant engagement before U.S. authorities detected the operation. The Institute for Strategic Dialogue (ISD) linked this operation to a Moscow-based marketing company, Argon Labs, demonstrating how IOs outsource and automate content production via third parties to impersonate known and trusted actors.

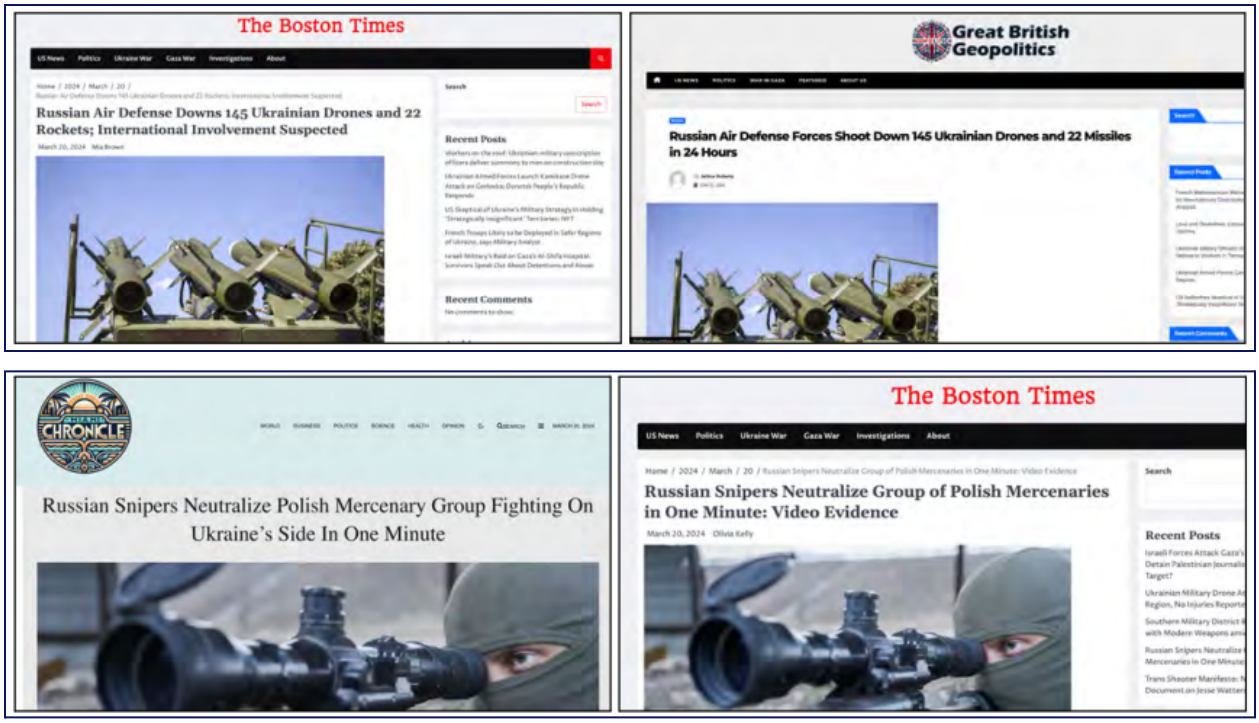
Case Study 2: CopyCop

The Russia-linked CopyCop operation has led targeted campaigns undermining Ukraine and Western interests since 2023. CopyCop operators use LLMs to publish political content at scale, plagiarizing, translating, and editing content from mainstream media outlets related to the U.S., U.K., Ukraine, Israel, and France.



Screenshots from archived pages of CopyCop's imitation news outlets in English, French, and German, posing as authentic news sites.

This operation's websites impersonate fictional local media outlets and feature fabricated and sensational news headlines in different languages, tailored to the target country. The websites often include nearly identical headlines for articles on the same topic, and sometimes include AI prompts mistakenly inserted into articles or headlines.



Nearly identical headlines were published on several spoofed CopyCop websites. Source: Recorded Future.

CopyCop established more than 200 websites in 2025. Many of these are spoofed websites that impersonate media and political entities in the U.S., France, and Canada. Others target non-Western countries such as Armenia. This influence operation also targeted audiences in other regions with this content by posing as fictional fact-checking organizations publishing content in Turkish, Ukrainian, and Swahili.

Screenshots from archived pages of fake and spoofed media websites that targeted Canada, France, and the U.S., posing as authentic news sites. Source: Archive Today and Wayback Machine.

The volume of fake websites created by CopyCop, along with the operation's links to the Moscow-based Center for Geopolitical Expertise and Russia's military intelligence agency (GRU), underscores the practice of actors impersonating trusted individuals or institutions to produce polarizing content at a low cost. Despite the minimal engagement received by CopyCop's posts, the French Service for Vigilance and Protection against Foreign Digital Interference (VIGINUM) reported that the IO had already achieved very high visibility and is anticipated to adapt its TTPs further to gain greater credibility and circumvent platforms' safety guardrails.

Case Study 3: Spamouflage

Since 2019, researchers at Graphika and other institutions have tracked Spamouflage (also known as Dragonbridge), one of the largest known pro-China influence operations. In 2023, Meta attributed Spamouflage to "individuals associated with Chinese law enforcement." This operation uses likely inauthentic or hijacked accounts to flood mainstream, fringe, and small social media platforms with identical videos, images, and posts that spread pro-China narratives.

All of the cases we provide illustrate how Spamouflage utilized AI to create credible American personas discussing U.S. politics. However, these efforts were often marked by easily detectable

AI-generated images, presenters with asynchronous body and lip movements, and scripts strewn with grammatical errors.

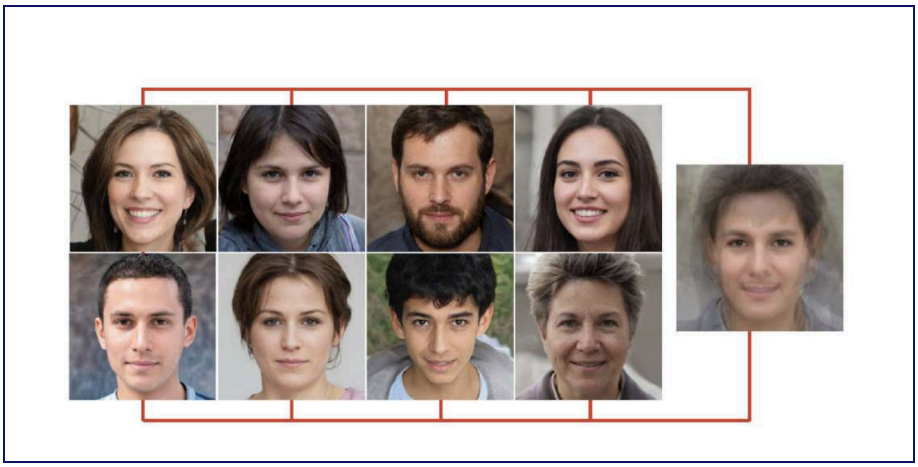
Since our initial report, we have observed Spamouflage using AI to generate profile pictures for social media accounts. Ahead of the 2024 U.S. elections, we observed Spamouflage leveraging an AI-generated avatar to build the identity of an X account and pose as a narrator in videos aimed at polarizing U.S. social and political topics.

Example of the @Harlan_RNC X account, which previously used the same likely AI-generated avatar (left) featured in a Spamouflage-linked video (right).

In 2023, we observed Spamouflage promoting videos featuring AI-generated anchors, which we assessed to be almost certainly produced using an "AI video creation platform" operated by Synthesia, a U.K.-based company.

AI-generated people acting as "news anchors" in videos promoted by Spamouflage-linked accounts.

Spamouflage has also continuously used generative adversarial networks (GANs), a class of machine-learning frameworks that enable computers to generate synthetic photographs of people who never existed.



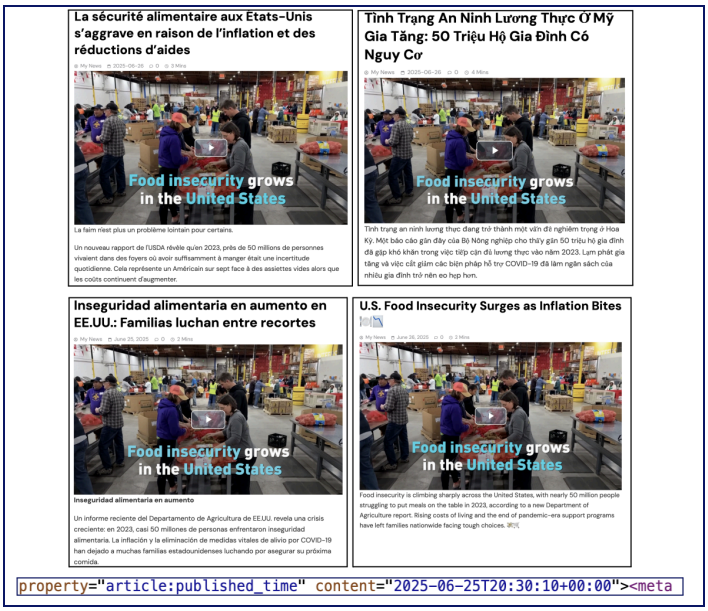
Profile picture of Spamouflage-linked YouTube accounts. Source: Graphika Report Spamouflage Goes to America.

Behavior: IO Production at Speed and Scale

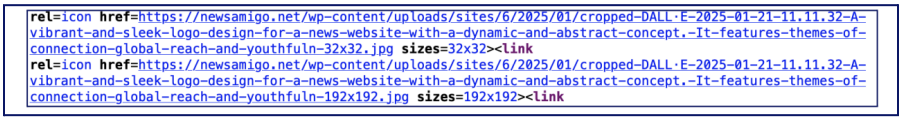
Case Study 1: Falsos Amigos

In August 2025, Graphika identified a network very likely located in China, consisting of 11 domains and 16 companion social media accounts that laundered articles from the Chinese state television network CGTN to global audiences in multiple languages, including English, French, Spanish, and Vietnamese.

This network — which we dubbed "Falsos Amigos" — almost certainly used AI to increase the reach of CGTN's content while obfuscating its origin. It simultaneously pushed AI-generated summaries or LLM-assisted English translations of the same articles, originally published by CGTN. It also used AI tools, including DALL-E, to generate infrastructure, such as favicons and logos, and replicate them across multiple social media platforms.



Example of content pushed simultaneously by the information laundering network. Source: [Graphika Report](#) Falsos Amigos.



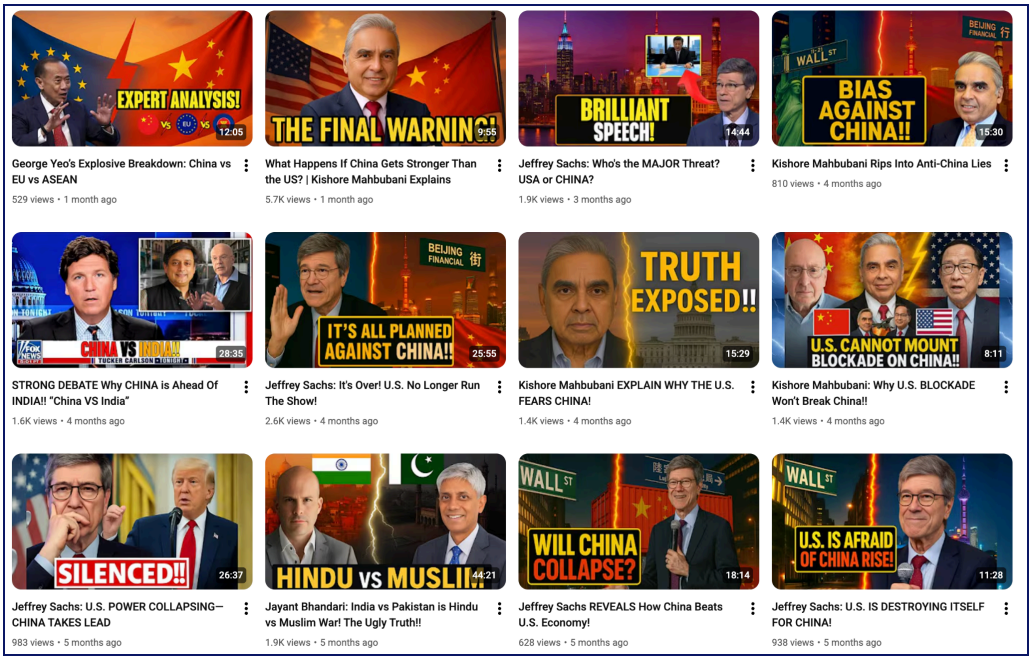
The source code for one of the network's domains included an image URL that suggested DALL-E was used to generate a "vibrant and sleek logo design for a news website with a dynamic and abstract concept." Source: [Graphika Report](#) Falsos Amigos.

However, the increase in the operation's scale was not necessarily accompanied by an improvement in the quality of the content. For example, the summaries were characterized by particularly flowery writing and liberal use of emojis in summary texts — both of which suggest the use of an LLM or AI. Some of the thumbnail images displayed the CGTN logo, revealing the source of the articles.

Case Study 2: Shadow Play

Shadow Play, a network first [exposed](#) by the Australian Strategic Policy Institute (ASPI) in 2023, comprised at least 30 spammy YouTube channels that amplify pro-China, anti-West narratives. These videos primarily focused on China's alleged superiority over the U.S. in tech, trade, and other areas of global competition.

In 2023, ASPI reported that the network had produced more than 4,500 video essays. Each video featured an AI-generated voiceover narrating a story and advancing an argument, combined with text-to-video software, stock photos, and video. The thumbnails of the videos also displayed AI-generated imagery.



Example of Shadow Play's most recent videos, featuring AI-generated thumbnails with sensationalistic video titles to capture the viewer's attention.

The channels managed to gain a cumulative 120 million views and 730k subscribers. However, despite the apparent number of viewers, the video failed to gain organic engagement on the platform. Some videos received no comments. Others appeared to be populated with comments, but these were likely posted by inauthentic YouTube accounts.

Case Study 3: Sneer Review

In June 2025, OpenAI exposed Operation Sneer Review, an IO likely originating in China that used ChatGPT to generate short posts and associated comments in English, Chinese, and Urdu for distribution on Facebook, Reddit, TikTok, X, and various websites.

Following OpenAI's report, we observed X accounts employing this tactic in multiple campaigns targeting various audiences with both pro-China and anti-China stances, as well as pro-Russia, pro-Iranian opposition, and pro-BJP stances.

The campaigns repeated themselves weekly, sometimes with the same comments pushing hashtags. The high volume of content and contradictory stances suggest that a commercial actor is likely behind this campaign, using generative AI to operate for multiple clients simultaneously. The approach likely involves creating large numbers of accounts pushing the content, some of which also appeared to be AI-generated.

Examples of cartoons used by X accounts likely linked to the pro-China influence operation Sneer Review.

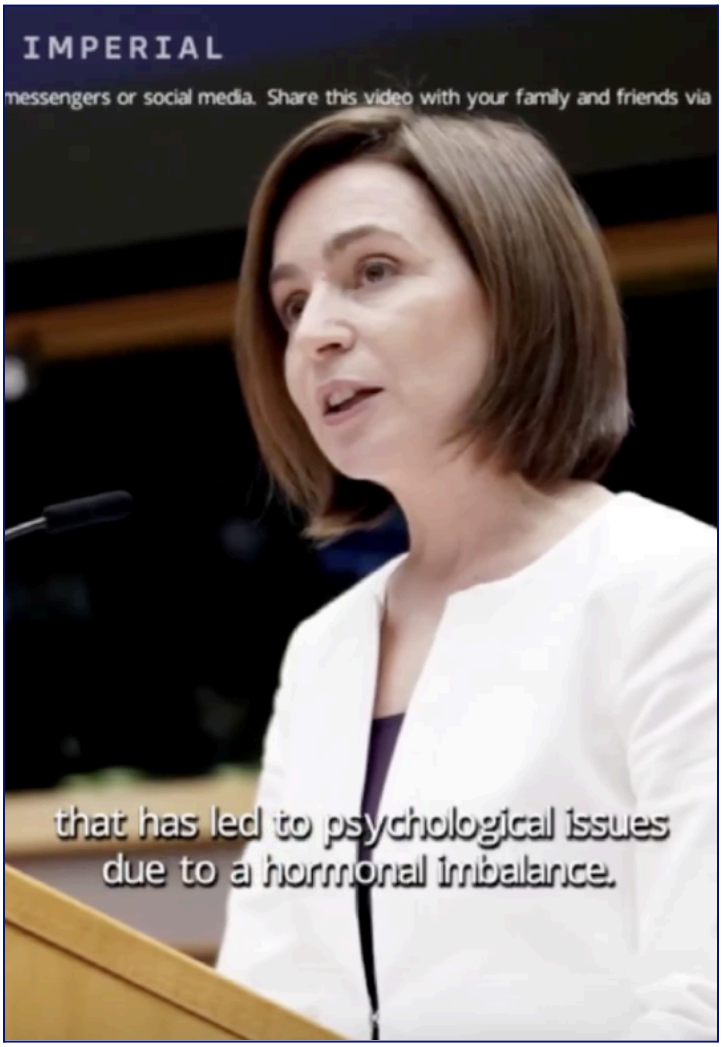
Example of likely Sneer Review-linked X accounts commenting with likely AI-generated comments under a post targeting the U.S. Redactions added by Graphika.

Despite the volume of activity, these accounts failed to gain traction, except for artificial amplification. The lack of credible messaging, as well as the inability of the infrastructure to expand to other platforms in the same manner as its ability on X, may explain this lack of organic engagement.

Content: Low Quality Text and Video

Case Study 1: Operation Overload

This pro-Russia influence campaign, Operation Overload (also known as Matryoshka or Storm-1679), impersonates European authorities, legitimate entities, journalists, researchers, and fact-checkers to disseminate Russian narratives tailored to global events. The operation functions similarly to Doppelgänger but utilizes a separate infrastructure and set of accounts to amplify AI-generated content on Bluesky, Telegram, TikTok, and X.



X post by an account linked to Operation Overload targeting President Maia Sandu, claiming she is unfit to lead Moldova based on a medical condition.

Ahead of the 2024 Paris Olympics, Operation Overload amplified an AI-generated feature film called "Olympics Has Fallen," which impersonated Tom Cruise and targeted the International Olympic Committee with pro-Russia narratives.

Screenshots from X posts of the AI-generated feature film "Olympics Has Fallen," impersonating Tom Cruise and targeting the International Olympic Committee with pro-Russia narratives.

Operation Overload's content typically consists of low-quality AI-generated voiceovers and AI-enhanced videos impersonating European security and political officials. However, it largely fails to attract a broader audience outside of this network's echo chamber.

Case Study 2: Operation Undercut

Operation Undercut is a Russia-aligned IO that uses AI-generated images and AI-enhanced videos, along with hijacked hashtags, to target European countries with pro-Russia, anti-Ukraine narratives. Recorded Future exposed this campaign in 2024. During the September 2025 Moldovan parliamentary election, Graphika observed Undercut-linked TikTok and X accounts disseminating AI-generated content to claim that Moldovan President Maia Sandu and her Party of Action and Solidarity (PAS) were corrupt and attempting to interfere in the electoral process.

Examples of AI-generated cartoons (top) and AI-enhanced videos (bottom) that were spread on X by Operation Undercut and targeting the Party of Action and Solidarity and Moldovan President Maia Sandu.

This content did not appear to gain any traction, again demonstrating the minimal reach received by the spammy content produced by Undercut. Based on our observation, we assess that the engagement is limited to the account with the network and a set of accounts that amplify Operation Undercut content.

Case Study 3: Celebrity Impersonations Delivering Pro-India Content

We documented spammy YouTube channels featuring business leaders, actors, and politicians that amplify pro-India narratives in a fashion similar to Shadow Play. The majority of videos posted by these channels feature AI-generated audio of celebrities commenting on politics and world events, overlaid on stock pictures and thumbnails of the celebrities.

Individuals depicted in the AI-generated audio from this campaign include Arnold Schwarzenegger, Oprah Winfrey, Steve Harvey, Jack Ma, and former U.S. President Barack Obama, all of whom appear to comment on India's rise as a global leader and its relationship with the U.S.

The audio tends to be low quality and inconsistent with the depicted individual's voice and tone, making it easily recognizable as synthetic or altered content. However, we observed an increase in view count, despite the lack of likes, comments, or reposts.

YouTube content with AI-generated audio and video impersonation of several U.S. actors commenting on global political events.

Graphika

About Us

Graphika *is the most trusted provider of actionable open-source intelligence to help organizations stay ahead of emerging online events and make decisions on how to navigate them. Led by prominent innovators and technologists in the field of online discourse analysis, Graphika supports global enterprises and public sector customers across trust & safety, cyber threat intelligence, and strategic communications spanning industries including intelligence, technology, media and entertainment, and global banking. Graphika continually integrates new and emerging technologies into our proprietary intelligence platform and analytic services, empowering our customers with high-precision intelligence and confidence to operate in a complex and continuously evolving information environment.*

For more information or to request a demo, visit our website.