Graphika

Deepfake It Till You Make It

Pro-Chinese Actors Promote AI-Generated Video Footage of Fictitious People in Online Influence Operation

The Graphika Team

02.2023

Information Operations

Deepfake It Till You Make It

Pro-Chinese Actors Promote AI-Generated Video Footage of Fictitious People in Online Influence Operation

Key Findings

- In late 2022, Graphika observed limited instances of Spamouflage, a pro-Chinese influence operation (IO), promoting content that included video footage of fictitious people almost certainly created using artificial intelligence techniques.
- While a range of IO actors increasingly use AI-generated images or manipulated media in their campaigns, this was the first time we observed a state-aligned operation promoting video footage of AI-generated fictitious people.
- The AI-generated footage was almost certainly produced using an "AI video creation platform" operated by a commercial company in the United Kingdom. The company offers its services for customers to create marketing or training videos and says "political […] content is not tolerated or approved."
- Despite featuring lifelike AI-generated avatars, the Spamouflage videos we reviewed were low-quality and spammy in nature. Additionally, none of the identified Spamouflage videos received more than 300 views, reflecting this actor's long-standing challenges in producing convincing political content that generates authentic online engagement.
- We believe the use of commercially-available AI products will allow IO actors to create increasingly high-quality deceptive content at greater scale and speed. In the weeks since we identified the activity described in this report, we have seen other actors move quickly to adopt the exact same tactics. Most recently, this involved unidentified actors using the same AI tools to create videos targeting online conversations in Burkina Faso.

Analysis

Wolf News

While tracking the latest activities of Spamouflage, a pro-Chinese political spam operation Graphika has exposed multiple times since 2019, we identified assets in the network promoting a new and distinctive form of video content on social media platforms including Facebook, Twitter, and YouTube.

This set of two unique videos shared many of the same characteristics as traditional Spamouflage content: they ranged between one-and-a-half and three minutes in length, used a compilation of stock images and news footage from online sources, and were accompanied by robotic English-language voiceovers promoting the interests of the Chinese Communist Party. One video accused the U.S. government of attempting to tackle gun violence through "hypocritical repetition of empty rhetoric." The other stressed the importance of China-U.S. cooperation for the recovery of the global economy.

However, the videos also included some new and previously unobserved characteristics:

- The videos featured male and female "news anchors," which we initially believed to be real people, likely paid actors, and later established were almost certainly created using the commercial services of an AI video production company.

- The videos used the branding of a likely fictitious media company called "Wolf News" - mirroring past Spamouflage efforts to pass as legitimate news outlets. The videos all used the same grey-and-white logo with a silhouette of a wolf, for instance, with the accompanying text: "Wolf News. Focus on hot spots and broadcast in real time."

The logo of 'news outlet' Wolf News

AI-generated people acting as 'news anchors' in Wolf News videos

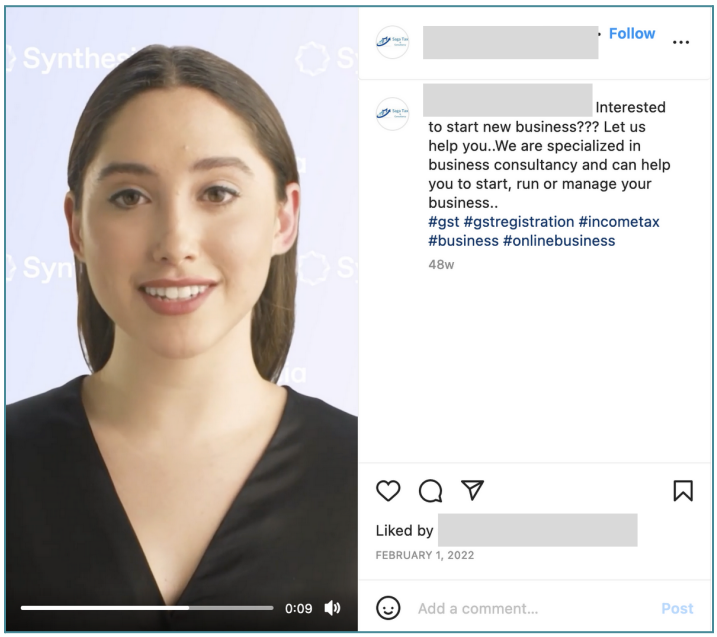
Hello, I am an avatar

At first glance, the Wolf News anchors present as real people. Our initial hypothesis was that they were paid actors that had been recruited to appear in the videos. This would align with a tactic we analyzed in June 2021, when an unconnected influence operation in Pakistan used commercial script readers to narrate fake "news report" videos.

But further investigation revealed the Wolf News presenters were almost certainly created using technology provided by a British AI video company called Synthesia. Below, we outline the indicators supporting this assessment and the investigative process used to identify them.

Using reverse image search techniques, we surfaced a wide range of marketing and promotional videos unrelated to China that used the same male and female presenters as the Spamouflage Wolf News content. These videos used multiple languages, ranging from English to Arabic, Romanian, and Spanish. In one video promoting freight broker services, the male Wolf News anchor says: "Hello, my name is Mr. Cruise. And I'm an avatar."

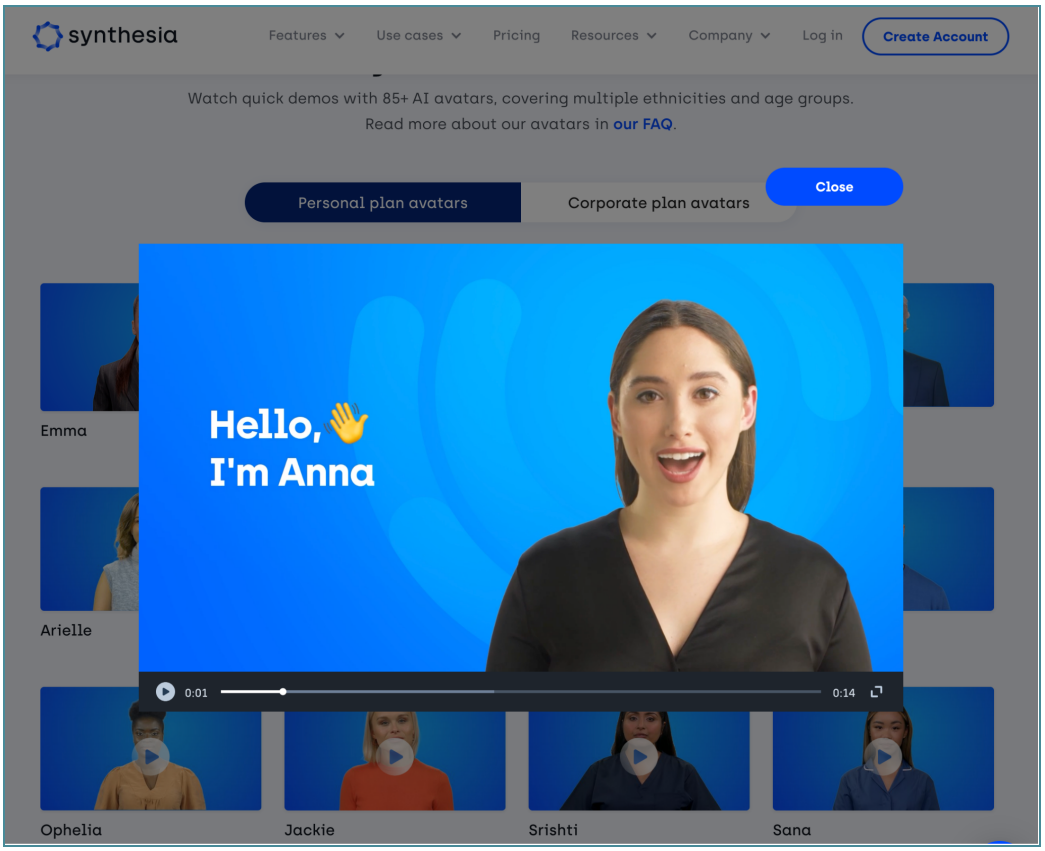
An online advertisement for freight broker services in which the male Wolf News presenter self-identifies as an avatar

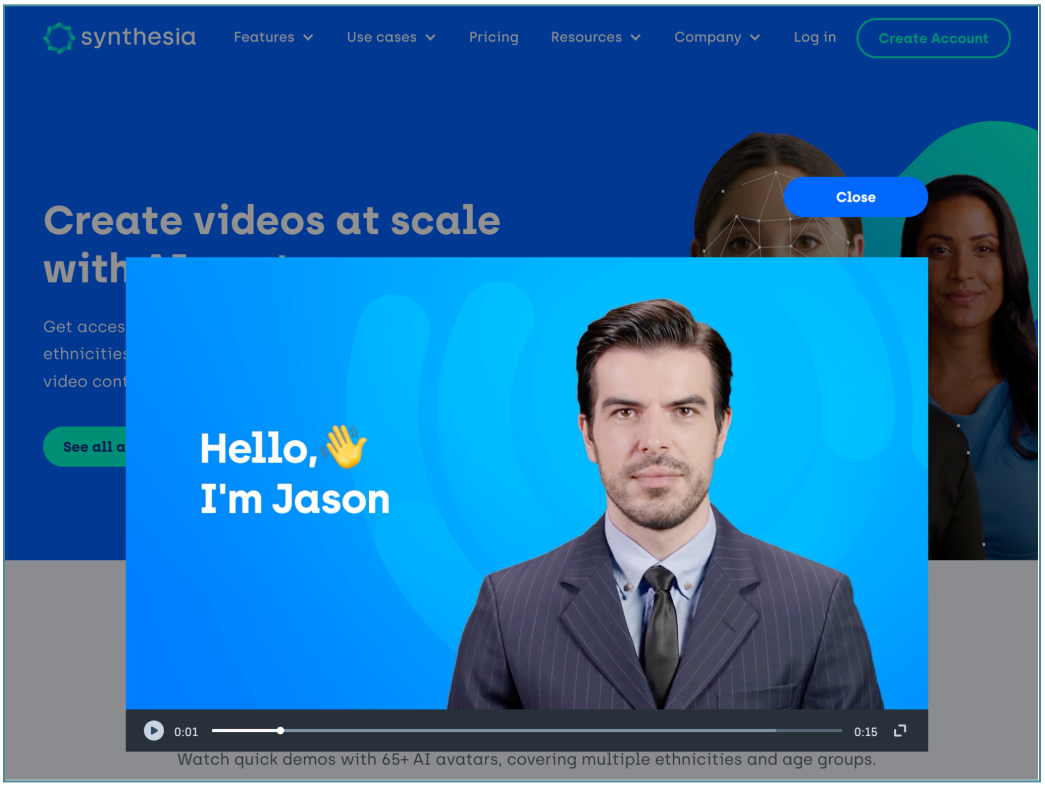


A video promoting an India-based consulting company featuring the female Wolf News presenter

We subsequently established that all the videos had been created using Synthesia products. Both the individuals that appear in the Spamouflage Wolf News content feature on the Synthesia website, for instance, and are identified by image as avatars that customers can use in their videos. The female avatar is listed as "Anna" and the male avatar is named "Jason."

On Synthesia's website, the company states that for ethical reasons, it "will not offer our software for public use" and that "all content will go through an explicit internal screening process before being released to our trusted clients." It also says "political, sexual, personal, criminal and discriminatory content is not tolerated or approved."





The Wolf News presenters are identified on Synthesia's website as avatars that customers can use in AI-created videos

Deepfakes, Real Problems

Researchers, journalists, and government officials have all raised concerns about the use of AI-generated content in political influence operations. The FBI warned in March 2021 that Russian and Chinese actors were already leveraging these technologies in their operations, and a U.S. Department of Homeland Security report that year discussed possible national security scenarios, including incitement to violence, fabricated evidence, and staged kidnappings.

Until now, however, our encounters with AI-generated media in online influence operations have been limited to GAN-generated fake faces and misleadingly edited videos showing an individual conducting fabricated actions or speech. The activity detailed in this report is the first time Graphika has observed state-aligned IO actors using video footage of AI-generated fictitious people in their operations. Based on our analysis, we offer the following assessments:

- The main benefit of this technology to the creators of the Spamouflage videos appears to be increased efficiency, specifically high-speed, low-cost content production. Synthesia's products can create AI-generated videos in a matter of minutes and subscriptions start at $30 per month. This aligns with the findings of a recent OpenAI report on the potential uses of AI language models in influence operations, which found that likely impacts include lower running costs and easier scaling of deceptive techniques.

- Despite the use of AI-generated avatars, the identified Spamouflage videos present as low-quality political spam. The English-language scripts are strewn with grammatical errors and the operation relies on fake persona accounts to amplify them online. This illustrates the limitations of employing these technologies in IO campaigns - they are one tool in an increasingly advanced toolbox, not a magic cure-all.

- The deepfake element of the videos was almost certainly created using a commercial service rather than an in-house capability, suggesting IO actors will continue to leverage the tools most readily available to them. This also raises questions about how to effectively moderate the use of these products and services.

- IO actors will continue to experiment with AI technologies, producing increasingly convincing media artifacts that are harder to detect and verify. The most effective uses to date appear to be when actors combine multiple different techniques, such as GAN-generated fake faces with edited photos, or using a language model to write a script for an AI-generated actor.

Graphika

About Us:

Graphika *is an intelligence company that maps the world's online communities and conversations. Graphika helps partners worldwide, including Fortune 500 companies, Silicon Valley, human rights organizations, and universities, discover how communities form online and understand the flow of information and influence within large-scale social networks. Customers rely on Graphika for a unique, network-first approach to the global online landscape.*

For more information, please contact: info@graphika.com

Graphika