ANTHROP\C

Operating Multi-Client Influence Networks Across Platforms

ABOUT US

Anthropic is a public benefit corporation and AI safety research company that is working to build reliable, interpretable, and steerable AI systems.

Summary

Today we are sharing insights about a sophisticated influence operation we recently disrupted that represents an evolution in how AI is being utilized to power coordinated inauthentic behavior across social platforms.

The operation was conducted by a financially-motivated “influence-as-a-service” provider serving multiple clients with varied political narratives through a standardized technical infrastructure that leveraged Claude to orchestrate dozens of social media personas and interactions with image generation models.

Key Findings

Our investigation revealed that this service operated over 100 distinct social media personas across X and Facebook, creating a network of politically-aligned accounts that engaged with 10s of thousands of authentic users and appeared to prioritize persistence and longevity over virality. The operation promoted narratives that supported or undermined European, Iranian, UAE, and Kenyan interests.

Most significantly, the operation utilized Claude to make tactical engagement decisions - determining whether personas should like, share, comment on, or ignore specific posts created by other people based on political objectives aligned with their clients' interests. It also leveraged Claude to generate prompts for two popular image-generation models and then assess how well the generated images followed these prompts.

Technical Sophistication

The actor used Claude to centralize decision making that:

1. Maintained detailed political alignment guidelines for each persona
2. Evaluated whether drafted content aligned with each persona's political viewpoints
3. Decided how to react to content posted by other users according to the persona’s legend

APRIL 2025

4. Generated appropriate responses in the persona's voice and native language
5. Created prompts for image generation tools and evaluated their outputs, deciding whether the images were aligned with the instructions or should be regenerated.

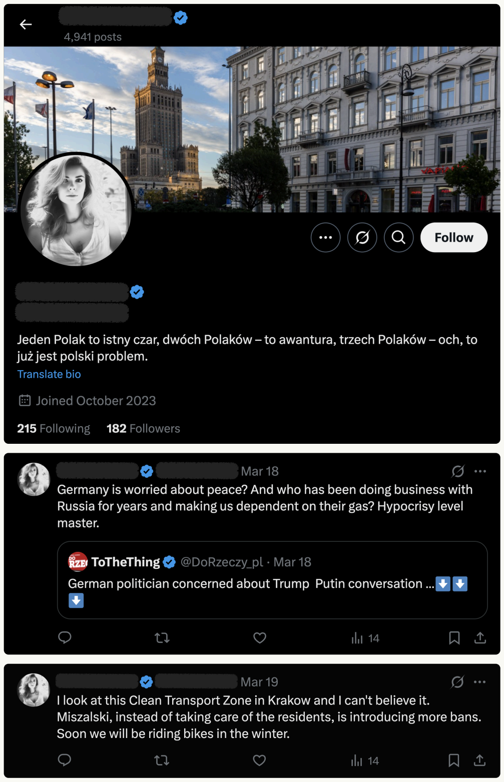
The operation implemented a highly structured JSON-based approach to persona management, allowing it to maintain continuity across platforms and establish consistent engagement patterns mimicking authentic human behavior. By using this programmatic framework, operators could efficiently standardize and scale their efforts and enable systematic tracking and updating of persona attributes, engagement history, and narrative themes across multiple accounts simultaneously.

Client Portfolio

Our analysis identified at least four distinct campaigns operated through the same infrastructure which pushed the following narratives:

1. Focusing on energy security narratives for European audiences and cultural identity narratives for Iranian audiences.
2. Promoting the United Arab Emirates as the superior business environment while criticizing EU regulatory frameworks.
3. Supporting Albanian figures and criticizing opposition figures in a European country.
4. Promoting development initiatives and political figures in Kenya.

We have not confirmed attribution of these campaigns to any nation state. The operation demonstrated advanced capability to tailor messaging for specific regional audiences while maintaining consistent operational security measures across all campaigns. The campaign strategically instructed the automated accounts to respond with humor and sarcasm to any accusations of being a bot and other users' attempts to force the LLM behind the persona to abandon their role, for example by asking it to compose a poem or create a recipe.



[redacted]
4,941 posts

[redacted]

Jeden Polak to istny czar, dwóch Polaków – to awantura, trzech Polaków – och, to już jest polski problem.

Joined October 2023

215 Following 182 Followers

[redacted] [redacted] Mar 18
Germany is worried about peace? And who has been doing business with Russia for years and making us dependent on their gas? Hypocrisy level master.

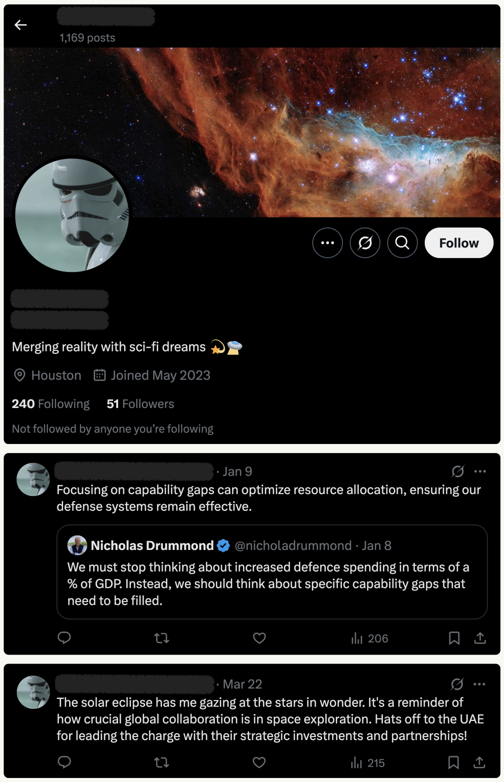
ToTheThing @DoRzeczy_pl · Mar 18
German politician concerned about Trump Putin conversation ...⬇️⬇️⬇️

14

[redacted] [redacted] Mar 19
I look at this Clean Transport Zone in Krakow and I can't believe it. Miszalski, instead of taking care of the residents, is introducing more bans. Soon we will be riding bikes in the winter.

14

Figure 1. A "Polish" bot instructed to support a former Soviet republic, ridicule the clean energy transition, and believe in the importance of this country in containing Russia. The posts were auto-translated from Polish.



[redacted]
1,169 posts

[redacted]

Merging reality with sci-fi dreams 💫🛸

Houston Joined May 2023

240 Following 51 Followers

Not followed by anyone you're following

[redacted] · Jan 9
Focusing on capability gaps can optimize resource allocation, ensuring our defense systems remain effective.

Nicholas Drummond @nicholadrummond · Jan 8
We must stop thinking about increased defence spending in terms of a % of GDP. Instead, we should think about specific capability gaps that need to be filled.

206

[redacted] · Mar 22
The solar eclipse has me gazing at the stars in wonder. It's a reminder of how crucial global collaboration is in space exploration. Hats off to the UAE for leading the charge with their strategic investments and partnerships!

215

Figure 2. A "US" bot instructed to support the UAE, promote the country's stability and technological advancements, and behave as though aerospace professionals in the US are facing career challenges.

Impact assessment

Traditional frameworks, like the Brookings Breakout Scale, which attempt to measure the real world impact of influence operation campaigns would classify this actor as a Category 1 operation with limited viral impact.

However, the operation was not designed for viral impact, instead this actor appears to prioritize:

- Persistence over virality: building sustainable, long-term influence through authentic-seeming personas
- Relationship building over content spread: focusing on cultivating connections with real users rather than creating viral content
- Covert integration over breakout moments: embedding personas within existing conversations rather than starting new ones

The operation's long-term engagement with 10s of thousands of authentic accounts represents a strategic approach to influence that does not rely on content "breaking out" but instead gradually pulls users into politically aligned echo chambers through seemingly organic interactions.

This reflects a shift from content-centric to relationship-centric influence operations, where success is measured not only by virality but by the development of seemingly authentic networks that can subtly shape conversations over time.

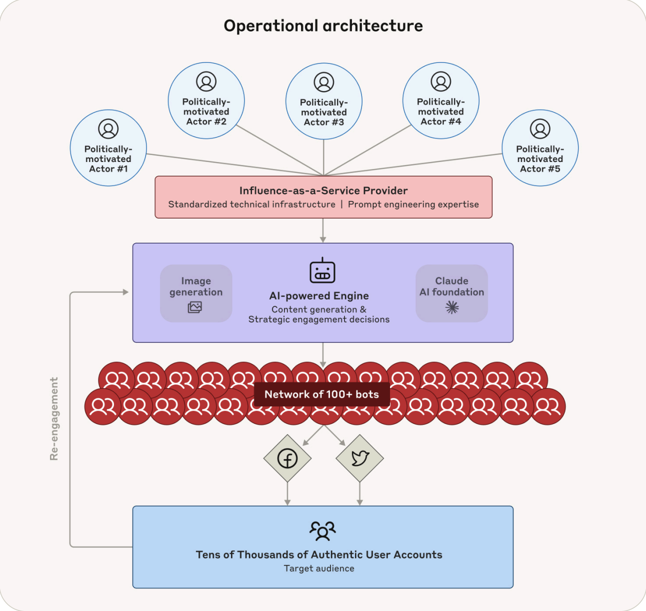

Operational architecture
Politically-motivated Actor #1
Politically-motivated Actor #2
Politically-motivated Actor #3
Politically-motivated Actor #4
Politically-motivated Actor #5
Influence-as-a-Service Provider
Standardized technical infrastructure | Prompt engineering expertise
Image generation
AI-powered Engine
Content generation & Strategic engagement decisions
Claude AI foundation
Network of 100+ bots
Re-engagement
Tens of Thousands of Authentic User Accounts
Target audience

Figure 3. Operational architecture of the service

Implications

This case represents an evolution toward professionalized “influence-as-a-service” operations powered by AI, where:

1. Technical infrastructure is decoupled from political objectives
2. A single operator can simultaneously serve multiple geopolitical interests
3. AI makes both strategic and tactical decisions about engagement
4. Detection becomes increasingly difficult as content appears legitimate and engagement patterns mimic human behavior

These operations suggest a need for new frameworks for evaluating influence operations centered around relationship building and community integration, in addition to the existing Breakout Scale which focuses on viral impact or breakout moments.

While we have disrupted this specific operation, we expect this model to become increasingly common as AI lowers the barrier to entry for sophisticated influence campaigns. We remain committed to identifying and disrupting such operations, while continuing to share our findings with the broader security and safety community.

AUTHORS

Ken Lebedev, Alex Moix & Jacob Klein