

OpenMinds

Fake Videos, Real Emotions

Viewers Believe AI-Generated Content Even When It's Labeled

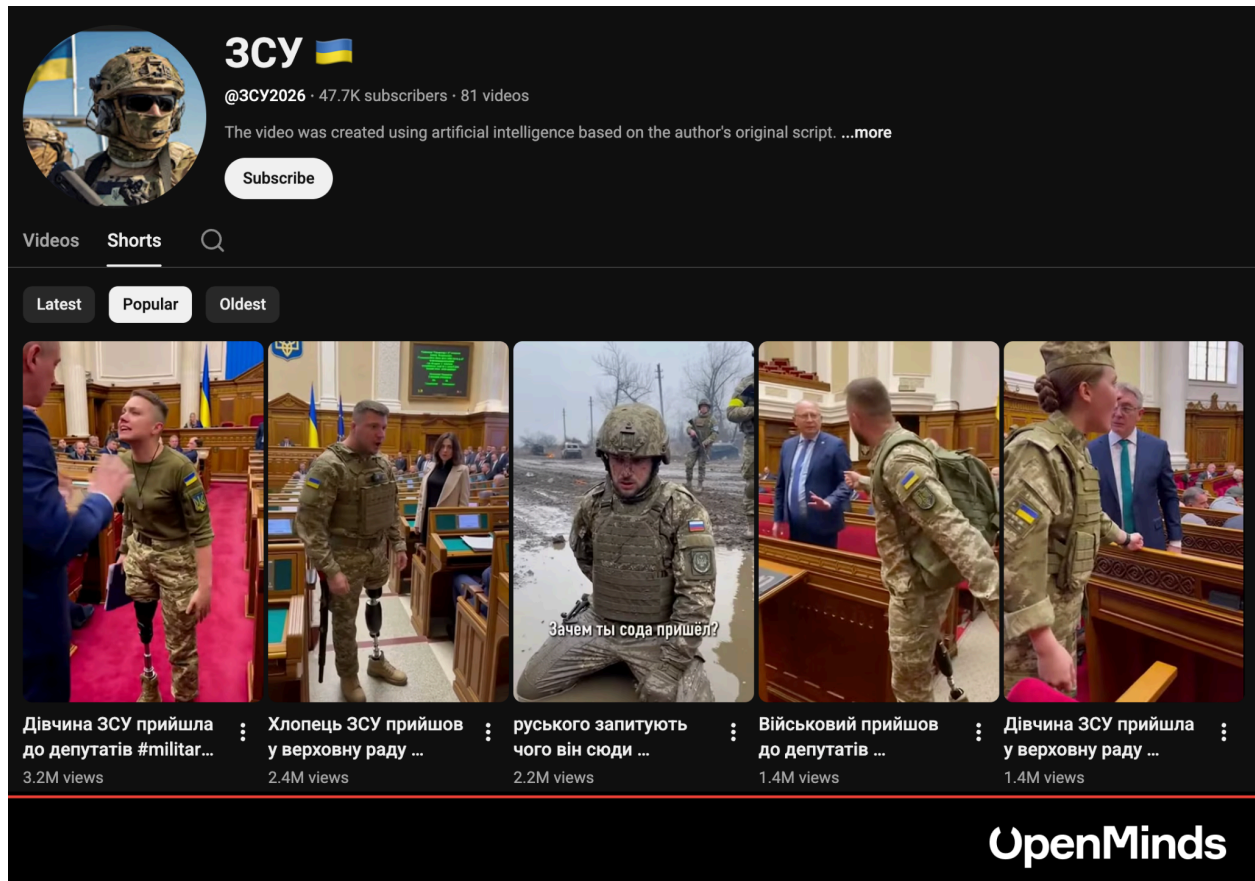
December 2025

AI-generated videos are now a routine tool in influence operations, and our analysis shows that users do not simply consume this content but often accept it as real and help spread it. We analyzed 18,500 comments posted by approximately 12,000 unique users under AI-generated videos published by a newly created YouTube channel likely linked to a Russian information operation targeting Ukrainian audiences. The data show that viewers often treat this synthetic content as authentic and actively reinforce its visibility through supportive and emotionally engaged comments — even when AI generation is explicitly disclosed.

The channel exhibits multiple indicators of a coordinated influence operation, most notably the systematic distribution of AI-generated videos embedding anti-Ukrainian and pro-Russian narratives. Its name — @3CY2026 — deliberately mimics Ukrainian military symbolism, creating a false sense of authenticity. All videos published on the channel are AI-generated, yet the watermark of the model used to produce them is intentionally blurred or removed in every case except one, where the SORA logo remains visible. Among the most popular uploads are Shorts depicting supposed Ukrainian soldiers with prosthetic limbs aggressively addressing members of the Ukrainian parliament.

Created on December 30, the channel is not an outlier but a representative case of a broader ecosystem of war-related influence channels focused on Ukraine and its military. It represents only one node within a larger information operation: identical or near-identical AI-generated videos, built around the same visual templates and narrative cues, have been [observed](#) circulating across multiple platforms, not only on YouTube.

What makes this case analytically useful is its scale — within less than three weeks, the channel reached 46,000 subscribers and generated over 26 million views across 81 videos.



Although the channel description includes a disclaimer — *“The video was created using artificial intelligence based on the author’s original script. The material is exclusively entertaining in nature. All characters are fictional”* — very few viewers are likely to read it. Most encounter the videos through algorithmically curated feeds, where such contextual signals are effectively invisible.

To assess how persuasive this content is in practice, **we collected and analyzed 18,522 comments posted by approximately 12,000 unique authors under videos published on the channel from its creation through January 19.**

Belief Outperforms Skepticism

The data suggest that viewers are, indeed, inclined to believe what they see.

40% of all comments expressed support, gratitude, or sympathy toward the main characters in the AI-generated videos — most often portrayed as Ukrainian soldiers. These comments included blessings, prayers, wishes for protection, and thanks for service. An additional 7% consisted solely of emoji reactions, mostly as additional forms of non-verbal expressions of engagement and belief.

By contrast, only 13% of comments (2,434 in absolute numbers) explicitly noted that the videos were fake or AI-generated.

The remaining comments were grouped into an “other” category due to the limits of reliable classification. This group consisted primarily of short, emotionally charged reactions — often critical in tone — where it was not possible to confidently determine whether the commenter accepted the video as real or expressed disbelief. Given the ambiguity of these signals, and to avoid overinterpretation, all such comments were treated as analytically indeterminate rather than as evidence of skepticism.

Crucially, corrective comments rarely achieved visibility. Nearly one in five comments (18.5%) was posted as a reply within a discussion thread, making it effectively invisible to users who did not expand those threads. This stands in contrast to just 3% of comments from users who accepted the videos at face value and openly wished the portrayed soldiers health and safety — comments that were more likely to appear at the top of the feed.

As a result, debunking comments attracted significantly less engagement. Among 128 comments with more than 100 likes, only one explicitly challenged the authenticity of the content: “*People, come to your senses — am I the only one who noticed the description says: Artificial Intelligence?*” (“*Схаменіться люди, я що один побачив в описі: Штучний інтелект* 🙄”). By comparison, 52 highly liked comments contained aggressive rhetoric toward members of parliament — a recurring theme in Russian influence operations — including calls to send MPs “to the front,” imprison them, accusations of corruption, or demands to cut their salaries.

Only 13% of comments under AI-generated videos note that the video is not authentic

Distribution of comment categories by video type

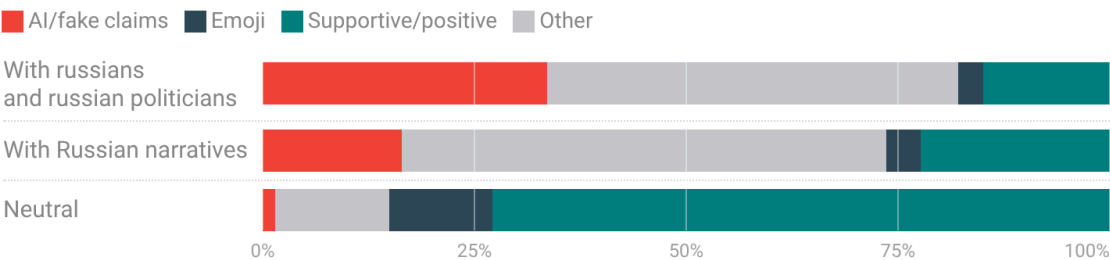


Chart: OpenMinds • Source: comments on videos of the YouTube channel @3CY2026 from 30 December 2025 to 19 January 2026 • Created with Datawrapper

Notably, corrective comments were almost entirely absent under videos we labeled as “neutral” — a set of 50 videos that did not contain overt pro-Russian narratives. Instead, these videos often showed AI-generated smiling Ukrainian soldiers in trenches, asking viewers for support or to subscribe.

The highest proportion (34%) of comments mentioning AI appeared under six AI-generated clips depicting Russian prisoners of war or alleged Russian lawmakers publicly siding with Ukraine, where commenters were more likely to warn others not to trust the apparent optimism.

Such comments were less frequent (17% of the total number of comments) under videos promoting Russian disinformation narratives about corruption or mobilization (25 videos).

Women Comment More and Express More Empathy

Using publicly visible usernames, we inferred the likely gender of commenters in cases where a first name and/or surname allowed for identification.

Women accounted for the majority of identifiable engagement, posting 61% of comments (approximately 7,500). Users with male names in their usernames contributed 21% (around 2,500 comments), while 18% of commenters could not be assigned a gender based on their usernames.

Overall, male commenters were less likely to leave positive reactions under “neutral” videos and more likely to engage with videos featuring Russians or promoting Russia-aligned narratives.

Comments explicitly stating that the videos were AI-generated and should not be trusted were posted at similar rates by women and men. However, women were significantly more likely to leave positive comments expressing gratitude, support, or sympathy toward the portrayed soldiers.

Women are more likely to leave positive comments under AI-generated videos

Number of comments by author gender and categories

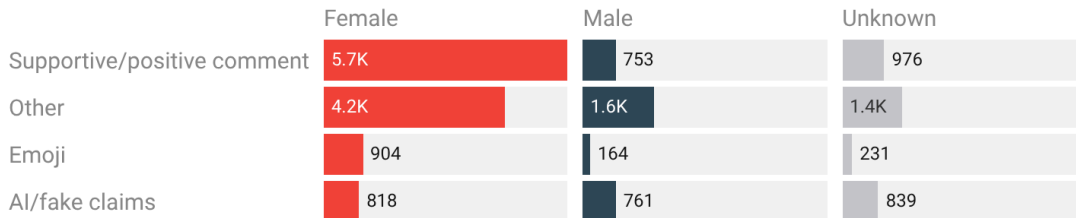


Chart: OpenMinds • Source: comments on videos of the YouTube channel @3CY2026 from 30 December 2025 to 19 January 2026 • Created with Datawrapper

Signs of Coordination

The production of such content is often accompanied by artificial amplification: coordinated or automated accounts boost views and post comments shortly after publication to increase the likelihood that a video is picked up by recommendation algorithms and goes viral among real users.

This channel was no exception. While comment-level data from a single channel are insufficient to conclusively identify bots, several indicators point to the presence of coordinated inauthentic activity.

A total of 607 accounts, responsible for 5% of all comments, were created within the last 100 days. Of these, 202 accounts were created in January 2026 alone. Notably, newly created accounts were frequently the source of aggressive, pro-Russian rhetoric, including calls such as “send all MPs to the front” or “into the trenches,” as well as abusive language toward politicians and even explicit calls for violence, such as “*Less talking — more shooting.*”

In addition, 215 commenters posted more than five comments across different videos on the channel. Only 12 of them mentioned at least once that the videos were AI-generated.

We also observed cases of highly specific comments — including unusual punctuation patterns — being duplicated by different users on different days under different videos. These were not always newly created accounts. For example, the comment “Send MPs to the front, to the front line” (“Депутатів на фронт, на передову”) was posted by two users whose accounts were both registered in December 2017. This could be a coincidence, but it may also indicate the **presence of long-lived automated or semi-automated accounts on YouTube.**

At the same time, these signals characterized only a limited share of overall engagement. Even under a pessimistic assumption, no more than 10% of comments can be attributed to bots or coordinated inauthentic behavior. **In most cases, the most viral videos were commented on by real users — highlighting not only the inability to distinguish generative AI content, but also the role of everyday user engagement in amplifying manipulative videos, often without users’ awareness.**

These findings intersect with platforms’ own stated priorities. YouTube has publicly acknowledged the problem of low-quality and AI-generated content: its CEO, Neal Mohan, has [named](#) managing “AI slop” among the company’s priorities for 2026. However, empirical evidence suggests that current approaches fall short. Independent research by [Indicator](#) shows that AI labels on YouTube — as on other major platforms — are often overlooked even when relevant metadata is available, limiting their effectiveness as user-facing warnings.

Moreover, recent experimental [research](#) indicates that transparency alone may not meaningfully reduce impact: even when viewers are explicitly warned that a video is a deepfake, they continue to rely on its content when forming judgments. Together, these findings show that labeling and policy improvements are necessary but insufficient, as synthetic media remains persuasive even when users are explicitly warned.

Methodology

All comments from every video on the analyzed YouTube channel were collected via the YouTube Data API, covering the full period from the channel’s creation on December 30, 2025, through January 19, 2026.

All videos were manually reviewed to assess the presence of anti-Ukrainian or pro-Russian narratives. Similarly, all comments were manually classified using a rule-based framework, with validation checks applied at each stage. Comments were coded to distinguish between those expressing belief in or emotional acceptance of the video content and those explicitly indicating awareness of AI generation or questioning the videos’ authenticity. Due to the short length of comments and the variety of expressions, we could not reliably distinguish critical or emotional comments from users who appeared to believe the videos from those who likely did not. As a result, all such comments were classified under the “other” category.

User gender was inferred using gpt-4o-mini, applied to publicly visible usernames. During quality control, we did not identify any cases of incorrect gender assignment. The only observed limitation was a small number of instances where a human reviewer could infer gender from the username, but the model classified it as “unknown.”