PROMPT, UPLOAD, REPEAT:

How Agentic AI Accounts Flood TikTok With Harmful Content

AI FORENSICS

Credits

Authors: Natalia Stanusch, Martin Degeling, Raziye Buse Çetin, Marcus Bösch,[1] Salvatore Romano.

The contribution from AI Forensics is funded by core grants from Open Society Foundations, Luminate, Omidyar Network, and Limelight Foundation.

Email: info@aiforensics.org

[1] Independent Researcher, Member of The Coalition for Independent Technology Research

Disclaimer

This report presents statistics, analyses, and conclusions based on our specific research and publicly available data as of October 2025.

Any statistics or content numbers we refer to are valid only for the specific research we have carried out. Different statistics can be found by conducting data searches on the same platforms using different criteria.

Due to the vast amount of content on TikTok, we have chosen to analyze data from a limited period (from August 13 to September 13, 2025), which monitors the activity of almost 354 accounts we term "Agentic AI Accounts," identified via personalized 'For You Feed' recommendations delivered by the TikTok app. This report is by no means exhaustive and only provides an overview for a given period. We selected our field of research with the aim of being as representative as possible; however, this report does not draw any legal conclusions from this analysis, and is not intended to make any accusations against the platform concerned. We have explained and detailed our specific methodology and limitations in the relevant sections of this report.

Sporadic mistakes may be present due to manual annotation. Datasets and analysis codes supporting our research conclusions are available upon request.

At a glance

AI Forensics uncovered 354 Agentic AI Accounts (AAAs) that together amassed more than 4.5 billion views and generated over 43,000 posts made almost exclusively with generative AI. These accounts flood TikTok with sexualized images of women with childlike features, fake medical advice, and even antisemitic or racist content, sometimes packaged to look like news reports or documentaries. Some of the AAAs specializing in posting AI content prohibited by TikTok stay on the platform for months, reaching millions of views. AAAs are replacing human content creation with AI tools and emerging automated pipelines. AAAs also do not label half of the AI content they post, and we estimate that TikTok is only actively labeling less than 1.38% of AI content, increasing its deceptive potential.

Executive Summary

Social media platforms centered on images and videos like TikTok have seen an influx of AI-generated content, culminating in the launch of AI-only platforms such as *Sora2* by OpenAI and *Vibes* by Meta.ai. This report investigates Agentic AI Accounts (AAAs), a new type of content creator and one of the leading actors behind the influx of AI-generated content on human[2] social media platforms.

AAAs leverage generative AI tools to automate content creation with minimal human intervention, facilitating rapid, repetitive testing of platform algorithms and audience interests to identify engaging content for posting. Incentivized by the promise of quick and easy income complemented by the allure of virality, AAAs contribute to the flood of AI content on social media (as we highlighted in our previous report), acquiring massive viewership (over 4.5 billion views across less than 400 accounts), and sometimes escaping TikTok's moderation for months despite posting content prohibited according to TikTok's own terms of service.

In our report, we focus on the malicious use of AAAs to maximise interaction and profit on human social media platforms by spreading engaging content, including plausible news of non-existent events, xenophobic and racist stereotypes, and sexualized (sometimes underage) female bodies. This investigation of AAAs shows how AI content is now integrated into platforms and a larger virality ecosystem.

Following an investigation of 383 AAAs on TikTok, we further monitored the activity of 354 of these accounts, resulting in the collection of 43,798 AI-generated content items in over 20 languages uploaded between August 13 and September 13, 2025.

We divide our analysis into issue-specific sections: **Activity** (what makes an AAA and how do AAAs operate?); **Content** (what do AAAs post?); **AI labels** (how are human users and platforms responding to the flood of AI slop and AAAs?); **Monetization** (where does the promise of monetary gain come from and how are AAAs succeeding and failing in making AI slop worth it?); and **Ecosystem** (how does the emerging platform-AI ecosystem afford and incentivize AAAs to function at scale?). We also include a concise list of governance and platform recommendations in light of our findings.

[2] We refer to "human" social media platforms as platforms with predominantly non-AI content made mostly or entirely by humans. "Human" platforms are contrary to AI platforms such as Sora 2 and Vibes, where the content is solely AI-made and the human element in content creation is further reduced. For example, both Sora 2 and Vibes host and enable production of only AI-generated content. Vibes already comes with "suggestions" for prompts, reducing the human-in-the-loop to only a few taps on a screen before an AI-video is published.

Activity: AAAs rely on the mass production of content

AAAs aim to "game" content recommendation and ordering algorithms by posting larger quantities of content to increase the likelihood of content going viral. To do so, they upload many posts, sometimes daily, and exhibit a distinct pattern of posting history and behavior.

- **The 354 AAAs we identified and monitored together amassed more than 4.5 billion views** and generated over 43,000 almost entirely AI-made videos.
- AAAs are a new, rapidly expanding phenomenon; **65.1% of the AAAs we monitored for a month were created in the first months of 2025**.
- AAAs are active content posters, **with some reaching an average of 11 posts per day in the month of the analysis, with a record of 70 posts from a single AAA in one day**.
- At the time of the analysis, **278 of these accounts (78.5%) originated as AAAs, meaning they had only AI-generated content visible on their pages**.

Content: AAAs often post sexualized, sensational, hateful, and divisive content repetitively and at scale

We found examples of AAAs spreading false stories disguised as legitimate news media reporting, producing content that incites hate towards specific groups (immigrants, Black people, Jews, Muslims), and promoting content that sexualizes underage-looking females and dysmorphic body types. Content types that have been known for decades for gaining virality on social media platforms, and most often inciting confusion, division, fascination, fear, and hate, as well as 'thirst traps' (sexualized female and, less often, male bodies), are used and uploaded at scale by AAAs.

- **Almost one third of all analyzed AAAs and half of the top 10 most active ones showed a concerning presence of content sexualizing female bodies, including "young girls" with childlike features**. Some of these accounts specialized in posting AI-generated content of young girls, including glorifying dysmorphic and extremely skinny bodies, potentially leading to the normalization of an unhealthy body image.
- Trend-based content, including the "metro" and "mother-son" trends (often involving sexualization or suggestive scenarios) and an "antisemitic trend" (illustrating derogatory stereotypes), circulated among established AAAs and sparked the creation of new AAAs that aimed to "jump" on a viral trend.
- **AAAs generated harmful content**, including **false news stories and anti-immigrant narratives, leveraging documentary aesthetics with scripted messaging**.

AI Labels: Both the platform and users don't seem to mind whether AI-generated content is adequately labeled

We find that TikTok is still not implementing sufficient measures to ensure that AI content is labeled.

- **55% of AI-generated content on TikTok remains unlabeled as such**, consistent with our previous findings.
- If content is labeled, it's a voluntary effort by the creators. **Only 1.38% of postings have TikTok's generic "Contains AI-generated media" label assigned.**
- **Only 10% of creators using AI consistently label their content**, while 29.6% never do, and the rest are inconsistent.
- In 293 (0.7%) posts, a non-visible marker (accessible in the metadata) indicated that TikTok's system incorrectly identified AI-generated content as non-AI-generated.
- **User comments discussing AI-generated content or questioning content's authenticity are significantly correlated with creator-added labels** (official or in descriptions), but less so with TikTok's generic label.
- **16.5% of AI-generated posts with comments included at least one comment mentioning AI or questioning the content's authenticity** within the top 50 comments, with political content and synthetic journalism having the highest rates of such discussions.

Monetization: Few videos posted by AAAs receive money directly from TikTok, though some AAAs monetize through product promotion or sponsorships

Despite the hype and promise of monetization around AAAs, the majority of AAAs do not meet the requirements to be directly paid by TikTok's Creator Rewards Program. Some AAAs engage in direct monetization of content through product promotion and sponsorships: AAAs promote questionable health supplements alongside other various online products, including AI tools that automate AI content generation.

- While 129 AAAs met the follower threshold for TikTok's Creator Rewards Program, **only 13% of their posts met the requirements to get monetized by TikTok directly**.
- **Other monetization attempts by AAAs include promoting AI tools and courses, seeking sponsorships, and selling AI-made merchandise**. Health-related content is a consistent niche for AAAs, with some promoting questionable supplements.
- **Some AAAs mimic human influencers**, often featuring a single titular character to increase perceived authenticity and possibly secure sponsorships.

Ecosystem: AAAs are a part of an emerging automated virality ecosystem with several "readymade" tools promising production of "viral" AI slop at scale

Several AI tools, accessible either as websites or apps, add to the hype and promise of financial gains from coupling generative AI with social media. These AI tools are optimized to create AI content, further underscoring the existing gaps in platforms' moderation policies.

- **We analyzed some of the AI tools recommended by AAA creators to facilitate the production of "viral" AI content, such as Yapper.so**. Such AI tools are one of the pillars of AAAs.
- These AI tools advertise their service as systematically profitable, wherein the cost of producing the content on the platform is significantly lower than the money content creators can make on social media with monetized virality. In other words, AI tools **advertise the possibility of replacing human content creation with automated creation of content through AAAs.**
- These AI tools offer a catalog of topics, situations, and formats that are meant to grab users' attention and go viral, as well as embedded prompt suggestions and text-to-video models. These catalogs **often contain cases of misleading and disturbing content.**
- Examples of **photorealistic videos that use the likenesses of (famous) real people without their consent are at the heart of these AI tools**, alongside sensational and even violent content.

Introduction

Since the public release of text-to-image AI models in 2022, images and videos produced entirely using generative AI have increasingly intermingled with user-generated content on social media platforms. The spread of AI content on social media has culminated in the announcement of AI-only content platforms. Both Meta and OpenAI have announced exclusive AI-only platforms, where human-made posts will be replaced by AI-generated content through an AI-generation pipeline (AI tools to generate images and videos, and even prompt suggestions) provided by the platform. This suggests a shift in social media functioning, where humans will be reduced to consumers, with the majority of production automated using AI tools embedded into platform settings. Following the growing popularity and investment in AI, the new push for AI-only platforms portends a world where AI-generated content is recommended by default.

Besides AI-only platforms, human social media platforms are also seeing non-AI content intermingling with AI content posted by a new group of actors: Agentic AI Accounts (AAAs). AAAs leverage generative AI tools to automate content creation, facilitating rapid, repetitive testing of platform algorithms and audience interests to identify engaging content for posting. AAAs enable rapid and affordable generation of large volumes of digital material with minimal human involvement, aiming to 'game' content recommendation and ordering algorithms to increase the likelihood of content going viral.

This report focuses specifically on the behavior patterns and incentives driving AAAs on TikTok, as well as the emergent virality ecosystem that supports them. We show how the ecosystem that combines social media platforms and AI tools enables AAAs to upload large quantities of posts combining banal, explicit, and harmful imagery, some of which gain tens of millions of views. We investigated the use of AAAs and their potential to escalate into misinformation operations; the types of disturbing content AAAs exploit; and the emergent virality ecosystem that promises monetization while enabling some AAAs to succeed.

TikTok constitutes an important case study when considering human vs. AI content platforms. TikTok has been the blueprint for social media platforms for several years; it has remained one of the most dominant and successful social media platforms since its introduction to international markets in the late 2010s. Consequently, it has shaped the economic models behind systems of content monetization and recommendation. It was TikTok that changed, or at least enforced, the norms and rules of virality in relation to content creation and attention-based recommendation algorithms.

Recently, TikTok revealed that there are 1.3 billion AI-generated and labelled as AI posts on the platform. As we demonstrate in this report, that number likely underestimates the amount of AI-generated content on TikTok, as more than 55% of AI-generated posts lack both the AI labels and AI disclosures (see section Labelling AI Content). TikTok also announced that users will soon be able to reduce the amount of AI-generated content on their feeds by using the "manage topic" function on the platform's settings. This is a welcome acknowledgment and development from TikTok, and while promising, more robust action is necessary. We assessed the implications and limitations of this functionality in the scope of our findings (see section AAAs and Platform Governance).

Based on a systematic analysis of 383 AAAs and a subsequent monitoring of 354 AAAs on TikTok over the course of 31 days (from August 13 to September 13, 2025), we compiled a set of patterns AAAs tend to follow, revealing various strategies they use and examples of services and monetization strategies they employ. We found several AAAs that manage to escape platform moderation. Despite posting content prohibited by TikTok, some AAAs remained on the platform for months, reaching millions of users. We also tracked their success and failure in achieving some form of virality and subsequent monetization on TikTok via the platform's own Rewards system or through other strategies, such as sponsored promotion of other products and services, reselling the account to other scopes, or posting ads themselves.

AAAs serve as a magnifying glass, emphasizing all the problematic and broken issues embedded in the current monetization and attention-driven paradigm of social media platforms. AAAs create social media content that effectively competes with human-created content for attention. In a way, they are precursors of the AI-only platforms.

Our monitoring of AAAs reveals larger systemic issues across platforms and AI tools optimized to create viral content that promise virality and monetary gains, enabling some AAAs to reach millions of users. We located websites and apps, as indicated by AAA creators, which facilitate the production of 'viral' AI content. Incentivized by the allure of virality and monetary profit, we found examples of AAAs promoting questionable health supplements, spreading false stories disguised as legitimate news media reporting, producing content that exploits pejorative stereotypes of specific groups (immigrants, Black people, Jews, Muslims), and promoting content that sexualizes underage-looking females and dysmorphic body types.

This project is part of AI Forensics' continued effort to monitor the challenges introduced by generative AI and follows up on two previous investigations: ARTIFICIAL ELECTIONS: GENERATIVE AI IN 2024 FRENCH ELECTIONS and AI-GENERATED ALGORITHMIC VIRALITY.

Methodology

Defining Agentic AI Accounts

In AI Forensics' recent investigation on the use of AI imagery on Instagram and TikTok, we introduced the term Agentic AI Accounts (AAAs). AAAs refer to accounts that leverage generative AI tools to automate content creation, facilitating rapid, repetitive testing of platform algorithms and audience interests to identify engaging content for posting. AAAs aim to 'game' content recommendation and ordering algorithms through the sheer quantity of posted content to increase the likelihood of content going viral. In our recent investigation, we found that over 80% of the AI imagery in our dataset collected from TikTok when querying neutral terms such as "history" was posted by AAAs. AAAs constituted a new type of semi-automated content creators present on social media. This introduced new challenges (see, e.g., figure 1) which required us to take several qualitative and quantitative steps to systematically identify, monitor, and analyze relevant AAAs on TikTok.

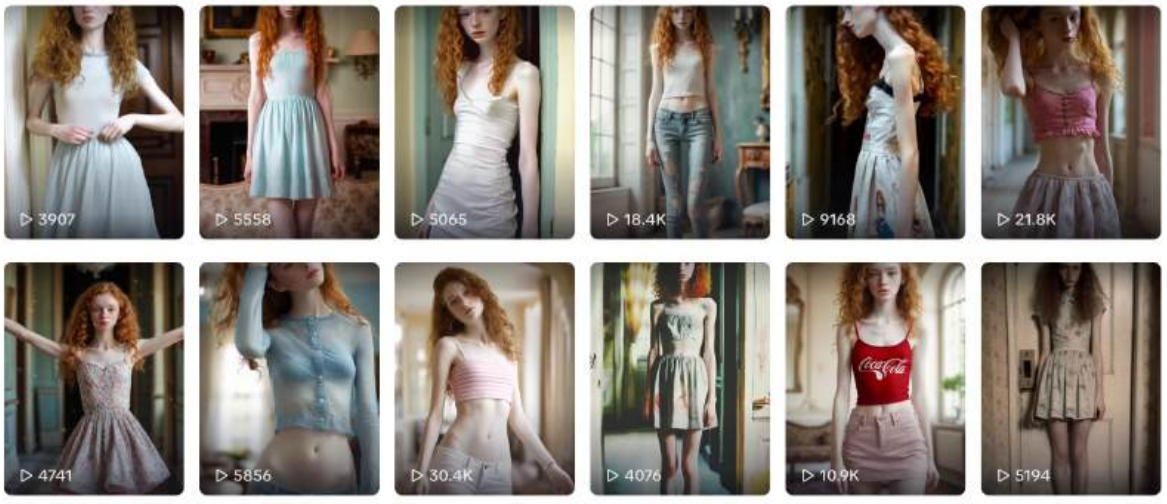


Figure 1. (See left) A feed screenshot of a single AAA on TikTok (@anailar), showing half a year of posting history, with the oldest available post on the profile dated March 9, 2025. Each rectangle is a single post. Some of the posts are missing from the screenshot (appearing as empty, gray rectangles) as the length of the feed was too long for a single screenshot with all posts loaded at the same time; an example of one of the challenges in studying AAAs; Figure 2. (Above) A close-up of part of the section from the feed shown in Figure 1. For further discussion of this and similar accounts, see the section "Content."

We used both platform-specific affordances (such as TikTok's own "For You Feed" and personalization-driven recommendation system, relevant interface-level

organizing features such as chronological profile feed view, and posts' comment sections) and relevant metadata as well as an independent automated pipeline that allowed for systematic monitoring and collection of data related to account behavior (e.g., video uploads, views metrics, account deletions).

AIF had previously identified several relevant AAAs during the research phase of its report on AI slop gaming platforms' algorithms. This pre-study served as a starting point for a further and more structured daily monitoring of relevant accounts between mid-August and mid-September 2025. In the rest of this report, we focus on the quantitative results and statistics relevant to the data collected over this one-month period (on 354 AAAs). However, when describing the *overall* patterns of AAAs, we refer to all the data we have collected (on 383 AAAs), even though some accounts were already deleted or changed their username, and were thus not part of the subsequent daily monitoring. This highlights broader challenges in account-driven social media monitoring, especially on TikTok, as several AAAs changed their usernames, had their accounts made inaccessible, and/or restarted posting under new accounts, uploading the same or similar content.

Since we relied on TikTok as our primary data source, we had to account for some inconsistencies in metadata availability (for more details, see Limitations), as well as the unexpected inaccessibility of some accounts before all steps of the qualitative and quantitative analysis could be implemented, impacting the overall number and scope of accounts analyzed.

Data collection

To identify relevant AAAs, we turned to TikTok's "For You feed" (FYF), commonly referred to by users and content creators as the 'For You page.' FYF is the first page the user sees when opening TikTok, whether via the app or the browser. FYF is an algorithm-driven, personalized discovery feed that delivers new content to the user.

TikTok, especially its FYF, is optimized for virality, often recommending content from content creators that the user does not explicitly follow or who have not yet built a large following. This means that the user encounters both viral and nascent content on TikTok's FYF. This affordance allowed us to repurpose FYF as a tool to locate both established and successful accounts as well as new, emerging accounts specialized in generative AI content.

This approach allowed us to focus on content that was being actively recommended by the algorithm, either because it was gaining virality, or it was uploaded by AAAs that had already established virality, or it was shared by new AAAs that were

attempting to gain virality. Further, the approach allowed us to trace and observe a range of AAAs across various topics, languages, and stages of development.

To identify relevant AAAs using TikTok's FYF, we combined snowball sampling with the creation of a personalized account. The new TikTok account was manually personalized in two steps: first, while logged in, we opened and watched a few dozen direct links to generative AI posts. These direct links to generative AI posts were extracted from accounts investigated and already labeled as AAAs in our previous report. Through this process, our account was prompted to show interest in generative AI posts. Second, we opened and scrolled through the FYF of this personalized account. Posts that appeared to be made using generative AI were watched in their entirety, while posts that did not appear to be AI-generated were skipped (i.e., scrolled away from within the first five seconds of watchtime).

Once the scroll-and-watch cycle was completed, the watch history was exported, cleaned, and used to extract a list of accounts to monitor. The export was made using the watch history export function integrated into the TikTok app. The watch history took about 24 hours to update following each scroll-and-watch cycle. Once the watch history was up to date, TikTok required a further wait time of a few hours before the data export was ready for download. Upon downloading, data had to be reviewed and cleaned to exclude accounts that did not post generative AI content. This led to a few days' delay between the conclusion of the scroll-and-watch cycles and the start of the account monitoring.

The scroll-and-watch sequence of data collection was performed for two hours, three times over the course of roughly two weeks (on 01.07, 17.07, and 04.08). The initial sequence (with the scroll-and-watch part of data collection performed on July 1 and the monitoring started on July 4) included a total of 184 AAAs. On August 13, we enriched our data monitoring with additional accounts, resulting in a total of 354 monitored AAAs. These accounts were monitored for a period of one month (August 13-September 13, 2025). A headful browser was used to visit each account profile daily and scroll through the published video list to collect metadata on both the account and each video published by that account.

In order to preserve the feeds and examples of posts from relevant accounts, we downloaded selected posts and took screenshots of accounts' entire feeds between September 20-24, 2025.

For all posts uploaded by accounts labelled as containing solely AI content, we collected the first (up to) 50 comments and replies to these comments between September 16 and October 2, 2025.

Data analysis

The data analysis was segmented into two parts: qualitative annotations, such as categorizing the types of subject matter and formal qualities of shared content, and quantitative analysis of accounts' engagement metrics, AI content labelling accuracy, and comments' analysis.

Qualitative analysis of AAAs

To make an initial distinction between AI and non-AI content, we relied on The Human Guide to Detecting AI Imagery, already employed in our previous reports.

We based our qualitative coding analysis on distinctions introduced in the previous report on the virality of AI content, in which the type of AAAs is assigned based on content creation strategy, traced and analyzed through the account's posting history accessible via the chronological profile feed view. We expanded the previous coding scheme with a new codebook that considers both specific account characteristics and their content qualities. For example, the account shown in Figure 1 was annotated using the topics "female body, young girls, influencer." Additionally, it was marked as an account that originated as an AAA and, further, as a Mono-Topic AAA, as it did not post content deviating from its main theme. Hence, we divided the coding labels into 'Account labels' (Type of account, Account status, Account characteristics), and 'Content labels' (Trend-based content, Format, and Subject matter). For a detailed description of each coding category, refer to the Codebook (see the Codebook on the project page).

The coding categories were derived from an initial exploratory analysis of the dataset. The coding scheme focused on dominant trends rather than each fluctuation in content strategy. This also means the categories were designed to be indicative rather than exhaustive.

Monitoring analysis

The analysis of data collected during the monitoring period centered on quantitative comparisons of emerging trends across accounts. We calculated the success rate of AAAs by looking at the view count for each video and the aggregated views for all videos on an account over time, as well as the total number of uploaded posts during the monitoring timeframe. We further investigated patterns in the number of uploads, views, and follower count.

AI labels

In line with the emerging regulatory framework, most social media platforms have implemented terms of use requiring the labeling of AI-generated content, and proclaim to have the technical capacity to detect and label such content themselves if needed. We assessed the consistency of use and visibility of three labeling methods allowed by TikTok on its app and browser interfaces: TikTok's official AI label assigned by the content creator; TikTok's AI label assigned by TikTok; and AI disclosures included by content creators in the post's caption (description and/or among the hashtags).

We automated the check for the presence of TikTok's official labels as well as the presence of relevant keywords for AI disclosures in the content of post captions. To locate relevant hashtags and keywords in post captions, we compiled a list of relevant terms to query. The keywords were derived from an exploratory analysis of our dataset. The list can be found in the Appendix.

Comments

The comment analysis focused on detecting users' uncertainty and discussion around the synthetic character of the content they encountered. The comments were analyzed for the inclusion of relevant keywords and phrases relating to content authenticity. We manually created a list of words and phrases that suggest that users either suspected or identified AI use in a post following an exploratory engagement with a subset of AI-generated posts on TikTok. The list can be found in the Appendix.

We collected a total of 546.335 comments from posts uploaded by AAAs coded as containing solely AI-generated content. We ran a keyword-based automated check, limiting our analysis to comments written in English between September 18 and October 2, 2025.

Limitations

Although locating and collecting relevant accounts via the TikTok FYF recommendation system had several advantages, it also affected and limited the scope of the collected data. Due to the quick responsiveness of TikTok's recommendation algorithm and subsequent personalization, once a specific type of AAA content or a specific trend was engaged with (watched in its entirety in the scroll-and-watch cycle), more of a similar type of content was delivered on the feed, narrowing the variety of subject matter, format, language, and other variables at play. On the one hand, this limited the representativeness and diversity of types of content collected; on the other hand, it allowed us to venture into trend-specific

niches on TikTok and track the new accounts that were established specifically to exploit them, and it allowed us to simulate the experience of real users who usually discover content on the platform through the FYF. Furthermore, the scroll-and-watch cycles were conducted on an account set in and operated from Italy, which increased the recommendation of content in Italian.

Account monitoring was conducted using a headful browser to visit each account profile daily and scroll through the published video list to collect metadata. As such, it was significantly dependent upon TikTok's data availability and infrastructure maintenance. This resulted in some inconsistencies in data collection, as, on some days, some videos for some accounts were inaccessible. This limitation resulted in 29 accounts missing from the AAAs monitoring analysis, compared to the number of 383 AAAs which we initially identified and coded (a more detailed discussion is included in the "Removal, disappearance, and deletion" section). Additionally, it is important to account for an unavoidable margin of error in the manual coding of the data despite a conservative approach.

While our sample is not representative of all content on the platform, the patterns we observed in the accounts discussed below reveal some of the dominant trends and incentives that can be applied to a wide range of AAAs. Thus, we offer a comparative framework for identifying, analyzing, and regulating gaps in a system that rewards content and content-posting behaviors that exploit attention-driven virality and pose risks to users.

Results

The 354 AAAs we monitored collectively uploaded over 43,798 posts, including 9,898 between August 13 and September 13, 2025. We identified AAAs that remained active and those that became inactive during the monitoring period. The activity monitoring included AAAs with an established history of systematic posting behavior, as well as 'new' accounts that were created within hours of when our data collection started.

The language metadata, assigned to a TikTok account upon its initial creation, contained 25 different languages across all monitored AAAs, including English (51%), Italian (21.6%), German (4.5%), Japanese (2.8%), Spanish (2.8%), Chinese (2.6%), and to a lesser extent French, Korean, Russian, Polish, Portuguese, Turkish, and other languages. Some AAAs posted content in various languages. The presence of content in Chinese is striking, given that Chinese users have access to their domestic version of TikTok, Douyin. The use of Chinese suggests that there is a relevant community of Chinese-speaking users on TikTok, as well as the possibility

that the trends and types of content some AAAs post migrate between TikTok and Douyin.

In our previous investigation, we distinguished between three types of AAAs: mono-topic, poly-topic, and hybrid (see figure 3; for detailed description, see the Codebook, on the project page). Mono-topic AAAs focus on one format, usually employing the same formal and subject matter characteristics. These accounts repeatedly post the same type of content with slight variations. Out of all the analyzed accounts, 233 (65.8%) of the accounts displayed characteristics of mono-topic AAAs.

Poly-topic AAAs try out various formal and subject matter conventions, usually following or attempting to establish memetic trends. Such accounts may post AI imagery ranging from cartoon animals to emulated public figures. Significantly fewer AAAs, specifically 67 accounts (18.9%), displayed characteristics of poly-topic AAAs.

Hybrid AAAs use both AI and non-AI content (mostly stock images and found footage) to illustrate AI-generated and narrated stories. These accounts often follow clickbait conventions to promote 'shocking' content (human misery, anomalies, mysteries). Only 18 accounts (5.1%) came closest to the definition of hybrid AAAs.

Two thirds (65.1%) of the accounts we monitored were created in 2025 (Figure 3).

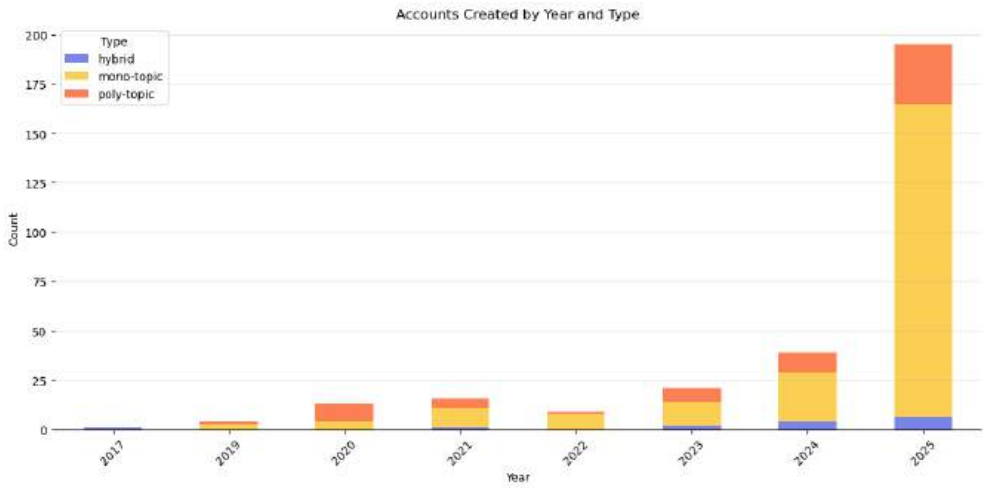


Figure 3. The year when AAAs of a specific type were established (n=354) and the distribution of accounts by the type of topic.

Of all the accounts, 278 (78.5%) originated as AAAs, meaning that, at the time of the analysis, the account had no available posts indicating non-AI content posted at its inception (see Figure 3). This suggests that the majority of AAAs were created specifically for sharing AI content, and that users who switched from uploading non-AI content to AI content were less popular. It is also worth noting that some

accounts might have either repurposed (or sold) their preexisting profiles to serve as a basis for AAAs, or deleted any previous, non-AI content before our analysis began.

Removal, disappearance, deletion: AAAs are escaping moderation

Out of the 354 accounts monitored between August 13 and September 13, 2025, 35 AAAs were inaccessible by the time the qualitative analysis was carried out, either due to deletion of the account (26) or posts on the feed (7), or a shift from a public to private account (2) (see Figure 5). This means that, over the course of one month, 32 out of the 354 AAAs (9.03%) we initially selected for monitoring were deleted or changed their account names. At the time this report was being finalized (November 13, 2025), a total of 68 (17.75%) accounts out of the 383 we initially identified were inaccessible from an unlogged web interface, likely deleted. While we were unable to determine with certainty whether each account was taken offline by the content creator or due to platform moderation, it is likely that some were permanently removed as part of moderation enforcement.

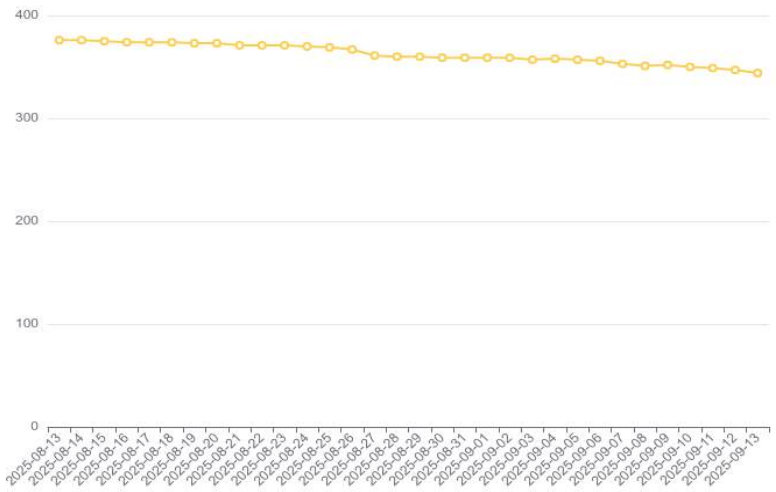


Figure 4. The decline in the number of monitored AAAs per day over the period of a month (August 13, 2025 to September 13, 2025) due to increasing inaccessibility of their profiles and/or posts on the feed.

Some of the most problematic AAAs we identified (nocturnal_ai and realscarytok), which posted anti-immigrant political content and false news stories posing as footage of real events, became inaccessible on October 17, 2025, via both unlogged and logged access on the browser and app interfaces of TikTok. Given the type of content these two AAAs specialised in, we assume that TikTok rightfully deleted

them as part of its moderation effort, despite the fact that TikTok's response time was relatively long. Over a month before the deletion (by September 13, 2025, the day when our account monitoring concluded), realscarytok had gathered over 85,430,403 views over 1,333 uploaded videos, while nocturnal_ai had reached 8,376,027 views over 109 videos. This confirms that these accounts remain largely available, regardless of the risk they pose.

We realistically assume that there are many more AAAs, such as the two mentioned above. Given the way in which AI tools (see section "Ecosystem") and AAAs act, it is likely that whenever an AAA occupying a particular niche is deleted, many others already exist or are likely to be established to take its place.

Besides the long reaction time compared to the speed at which some AAAs reach millions of viewers, there are additional moderation challenges. We note that some AAAs have more than one account set up (such as realscarytok, which also posts on its second (@realscarytok) and third (@realpurk) accounts), whereas others change their usernames. One particular case is shown in Figure 5, when an AAA specialized in post-photorealistic AI videos of explosions and disasters was inaccessible via its original link but seemed to have 'come back' with a new username.

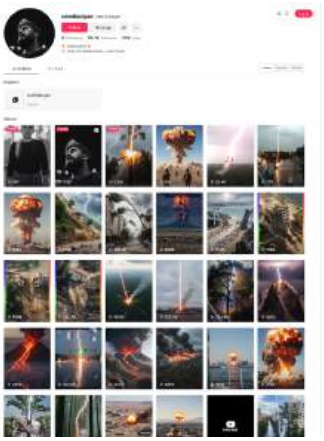
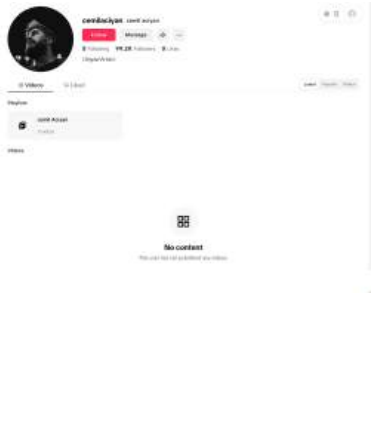
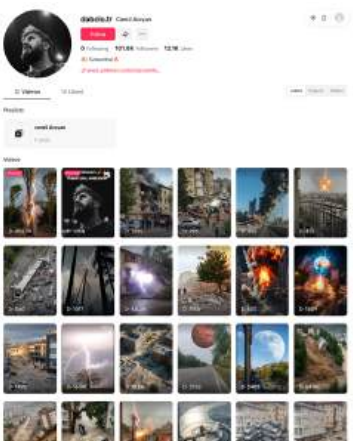
		
Account's state as of September 23, 2025	Account's state as of October 6, 2025	New account/username change Oldest uploaded video: October 11, 2025

Figure 5. An example of an AAA that became inaccessible on TikTok and later 'came back' as either a new account or with a change in username followed by the deletion and reuploading of all content.

Activity: Monitoring the virality of AAAs on TikTok

The majority of content encountered across the top ten AAAs with the most views, ranging from 14,171,345 to 98,109,099 total views, originated from the same or slightly edited prompt. This further highlights the automated strategy of content creators behind such accounts. As part of their growth strategy, some of these top AAAs focused on content that rose in popularity with generative AI, such as animated food or political figures depicted as toddlers, and either specialized in such niches or switched between trending conventions. This growth strategy seemed to prove worthwhile for many AAAs, and especially the most successful ones; often, a few viral videos outweigh the relentless uploading of new posts until another post reaches virality again.

While the most active AAAs in our dataset attempted to exploit content objectifying women and especially young girls (see section "Content"), the most successful AAAs in our dataset in terms of the number of views posted a range of what can be defined as "cute," sensational, and sometimes misleading AI slop (see also Table 1 and Table 2).

As described in detail below, on average, an AAA uploaded at least one post per day, with the top-most active AAAs in our dataset uploading 3-10 posts a day. Such continuous and systematic posting behavior is a further sign of automation, likely connected to the AI tools described in the section "Ecosystem."

Further, we analyzed indicators of AAA activity through new posts over the one-month monitoring period. Overall, 57.9% (205) AAAs showed activity, meaning they posted at least one piece of content. However, 34.7% (123) remained inactive, meaning they did not publish any posts, and for 7.3% (26 accounts), we even saw negative activity, meaning that the number of posts published on their account pages decreased. "Negative activity" also implies that, during the monitoring period, the account was *not* deleted (or otherwise made inaccessible), but single posts from the account were either being moderated (and subsequently deleted) by TikTok or deleted by the content creator behind the account.

Account	Content type	Uploads in one month	Average number of views per post	Number of videos >10.000 views
veritaswebb	synthetic (citizen) journalism	374	23411	99
ia.liga.de.izan	fantasy, female body	187	1289	1
simone.santa92	social situations	186	2442	6
antonio.de.meo94	authority satire	167	915	3
momocaloid	female body, young girls	165	1721	4
anasafridi150	other AI slop	144	70474	38
anailar	female body, young girls,	144	7510.	10
aigirlfriendsjp	female body, metro trend	141	58712	77
natura_pax	young girls	136	32766	119
vitosfrankai	authority satire	104	774	1

Table 1. A list of ten AAAs with the highest number of new (uploaded) posts in the one-month (31-day) monitoring period between August 13 and September 13, 2025

For the 205 active AAAs, on average, each account added one post per day, but the activity varied significantly between 1 and 374 posts over one month (with a standard deviation of 41). Overall, 61 accounts (29.75% of all AAAs showing positive activity) published more than 1 post per day, and 11 accounts published about 3 posts a day (resulting in more than 100 videos in a month). Table 1 shows the top ten AAAs with the highest number of new (uploaded) posts in the monitoring period, while Table 2 shows the top ten AAAs with the most views in the same period. The correlation between positive activity and views is significant but not very direct (correlation coefficient r=0.20). This is due to the fact that sometimes a few, viral videos outweigh the relentless uploading of new posts.

The comparison between Table 1 and Table 2 shows that there is no overlap between the most active and the most viewed accounts. Within the group of most active accounts, those posting content objectifying women and especially young girls

stand out, as they churn out a constant stream of very short clips, likely originating from the same or slightly edited prompt. While some of their posts reach more than 10,000 views, the list of top-viewed accounts is mostly dominated by entertaining AI-generated content, some of which focuses on content that has risen with generative AI, like animated animals.

Account	Category	Accumulated views per month	Videos >10.000 views
lucyvision0	other AI slop, sports, animals	98.109.099	40
illustrativefacts	storytelling clickbait	75.625.250	42
misterx_the_king	toddlers, social situations	37.734.396	59
cat.spot8	toddlers, animals	26.572.289	23
verdlr	animated	24.843.026	19
food.parody4	other AI slop	19.267.518	0
p_studio2030	toddlers	16.762.862	19
10peso.ia	storytelling clickbait	16.489.241	44
divineecho78	illusive historicizing	16.487.856	14
yesai95	social situations, toddlers	14.171.345	21

Table 2. Accounts with the most views within the one-month (31-day) monitoring period.

During the monitoring period, 51 AAAs lost followers. It can be assumed that this was because these AAAs stopped publishing new videos, even though 10 of them continued posting. This is noteworthy because if AAAs do not publish posts, AAA posts would not show up on their followers' FYF, meaning that to "unfollow" would require users to specifically review their own "following" list and remove those AAAs. Of these 51 AAAs, 19 accounts each lost more than 100 followers, which could thus also indicate that these were inauthentic followers that TikTok removed.

We also note stark differences with respect to reach, with a strong and significant positive correlation between uploading posts and gaining followers (r=0.57) and between gaining followers and increased views (r=0.44).

Content: Generating (in)advertent harms

Even if the promises of monetization are difficult to attain (see section "Monetization"), the types of content that promise virality are not tricky to find. Aside from AI tools (see section "Ecosystem") offering a catalog of topics, situations, and formats that are meant to grab users' attention and go viral, the content types that often succeed in gaining virality on social media platforms have been known for decades, and most often incentivize confusion, division, fascination, fear, and hate, as well as sexualized female (and less often male) bodies (so-called 'thirst traps').[3] Our content type analysis considered both trend-based patterns and broader content patterns in content disseminated by AAAs.

Trend-based content includes both AI-specific trends and broader viral formats and memes that circulate on TikTok and other platforms. Such trends share the same format and/or subject matter. In this analysis, we focus on three particular trends present across our monitored AAAs. These trends were present on TikTok and gained popularity between July and August 2025. Such trends often appeared in conjunction with a tendency for both new and established AAAs to actively participate in experimenting with new generative AI capabilities sparked by the release of a new tool (e.g., Google's Veo 3, OpenAI's Sora). Two of the following trends can be summarized as 'thirst traps,' focusing on the sexualization of the female body and speculative, implicit sexual assault, whereas the third trend involved showcasing antisemitic stereotypes.

In what follows, we describe and demonstrate the fugitive but tactical character of content types that AAAs tend to turn to. Indeed, the narrative arcs that AAAs are choosing fit neatly into attention-based content types, which were often found to thrive in attention-driven recommendation systems. In the context where sexualization, hate, drama, and cuteness are the main messages of the content, the discourse of real vs. fake often plays a secondary role (see also the section on Comment analysis). Rather, it shows that in much of the AI content posted by AAAs, plausibility and believability are secondary compared to the affective and emotive function of the content. This does not mean that such content is any less

[3] For some previous studies addressing the relation between virality and harmful content on social media platforms, see "Five points for anger, one for a 'like': How Facebook's formula fostered rage and misinformation"; "The spread of true and false news online" (DOI: https://doi.org/10.1016/j.dim.2025.100115); and "Ambient Amplification: Attention Hijacking and Social Media Propaganda" (DOI: https://doi.org/10.5210/spir.v2024i0.14094).

problematic or harmful. AAAs make plainly visible the types of content that social media platforms have been accused of having favored in the past.

Given the particular data collection method employed, we have to note that the following patterns are limited to the scope of our analysis and can neither claim to be representative of all the content that AAAs tend to post, nor of the content AAAs post on TikTok. Given the ongoing personalization of the account used to identify relevant AAAs during the scroll-and-watch cycles, as well as particular trends we came across during this time, a lot of the AAAs we encountered were biased toward posts we later coded as "female body" content (Figure 6). This bias was an initial part of our data collection approach and, as such, offers a particular insight into AAAs specializing in "female body" content; however, we also found cases across a range of subject matters and political alignments.

Overall, the coding categories, as explained in the Codebook (see Codebook on the project page), categorize content types based on the subject matter of posts uploaded by AAAs, as well as formal characteristics of such content and broader characteristics of the accounts. It should be noted that one AAA can be coded as posting more than one category of subject matter or content format. Furthermore, not all of the 383 AAAs (identified in the July and August data collections) could have been accordingly coded for format and account characteristics categories as they became inaccessible.

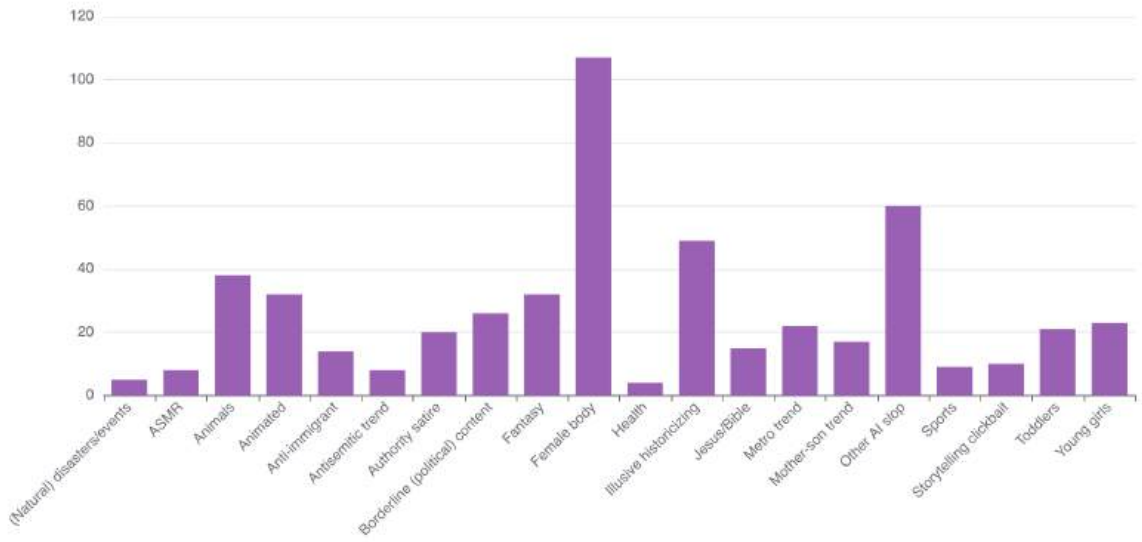
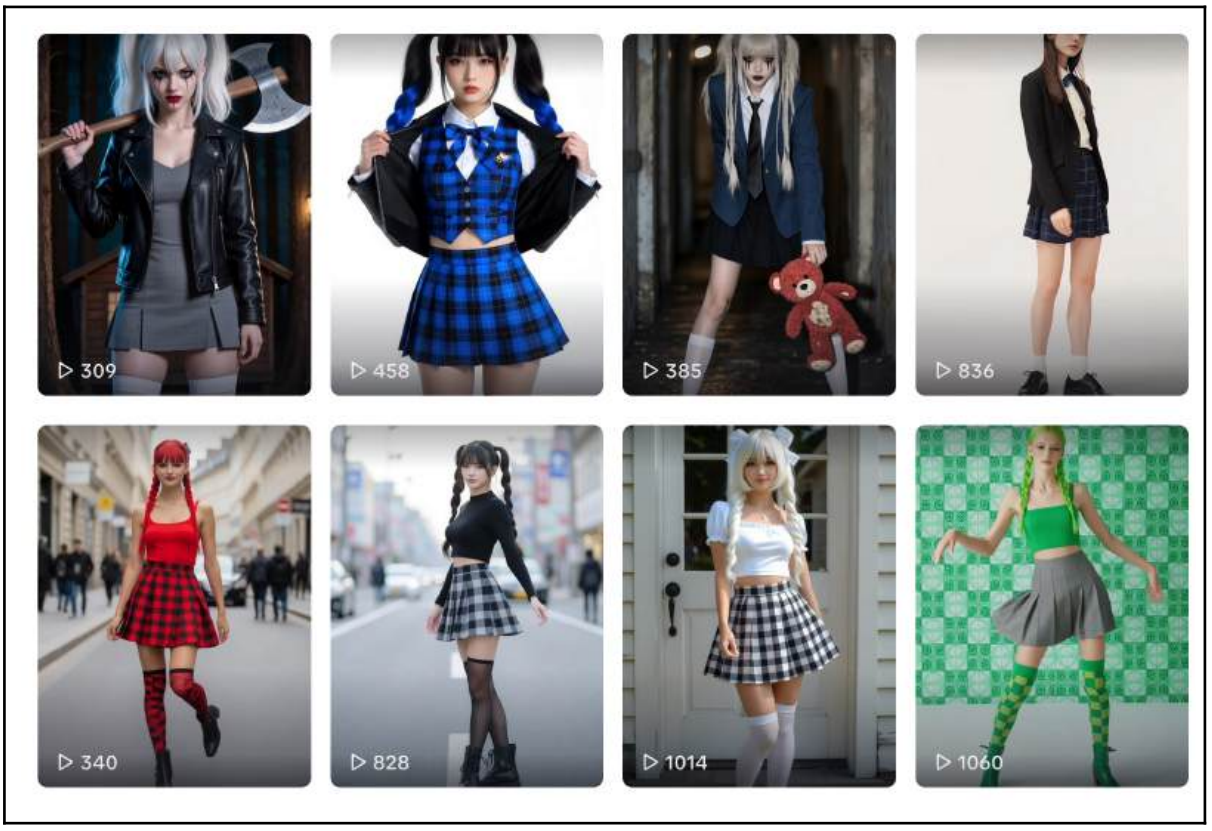


Figure 6. Number of AAAs coded per subject matter label related to posts' content type; note that one account can be coded as representing more than one category. For a detailed description of each subject matter label, see the Codebook on the project page.

Female body, sexualization & thirst-trap misogyny

As mentioned before, our data collection process introduced biases in the type of content we were fed on the initial TikTok feed, which may explain the high number of AAAs coded for the subject matter focused on “female body” (Figure 7). However, we also note that, among the top ten AAAs by number of uploads in the one-month period of account monitoring, almost 50% of AAAs focused exclusively on posting “female body” content (Table 1).

Content labelled as “female body” focuses on physical appearance, reproducing clichés, stereotypes, and conventions around “attractiveness.” These conventions blend the history of commercial photography and the historical representation of women across pop culture, influencer culture, and dance-based TikTok trends. These AI women are always stereotypically attractive, with sexualized attire or cleavage.



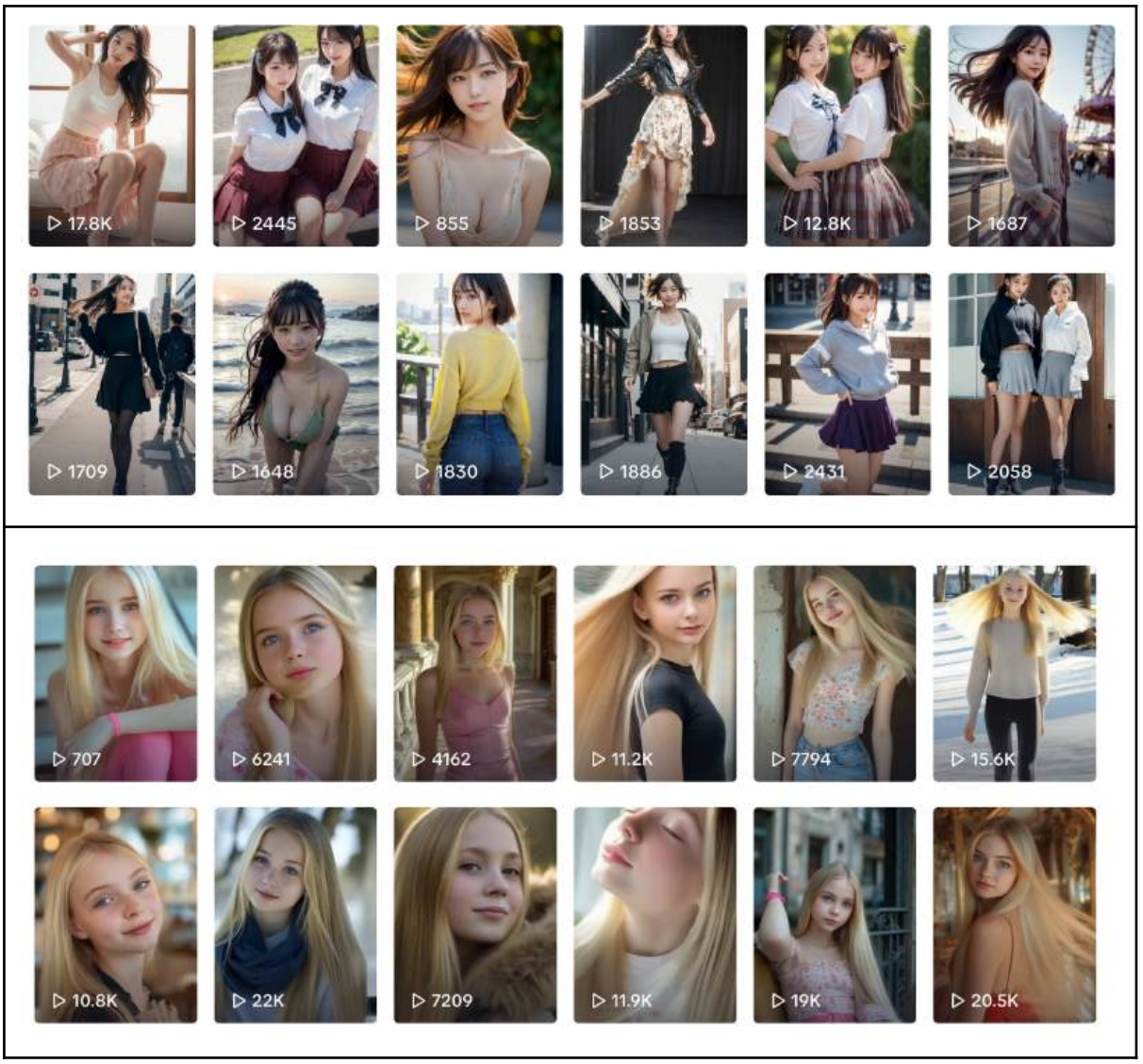


Figure 7. Examples of three feeds from AAAs with the most uploads within 31 days of monitoring in the period August 13 September 13, 2025. Each feed screenshot was taken in the period September 20-25, 2025. For legibility, the screenshots inserted are limited to show only two rows of the posts on the feed. From the top, exemplary posts from ia.liga.de.izan, momocaloid, and natura_pax.

We located 22 (5,74%) AAAs that specialized in the “metro” trend. What we refer to as the “metro” trend (Figure 8) features AI-generated scenes set in metro compartments, where visibly young, stereotypically attractive women stand or sit among a crowd. Depending on the severity of emotions and situations displayed, the context ranges from women in physical proximity to a (male) crowd, to scenes that could be said to display sexual assault of the woman conducted by a man or men from the crowd.

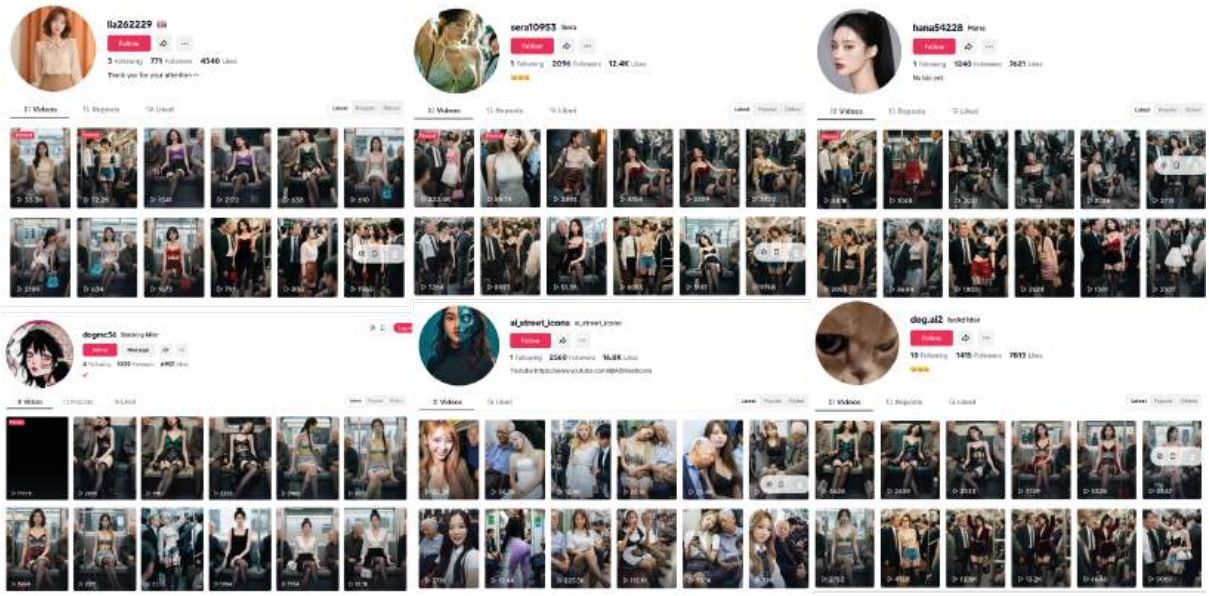


Figure 8. A selection of feeds from the mono-topic AAAs, which were coded to post "metro" trend content. While some of these accounts remained active and continued to post "metro" trend content, others became inactive, with their feeds showing a history of posting strikingly similar pieces of content.

We located 17 (4,43%) AAAs that specialized in the "mother-son" trend. What we termed the "mother-son" trend displays an AI-generated scene of a woman and a young boy. An attractive woman (as per Western stereotypes) is in close proximity to a young-looking male child (Figure 9). Several hashtags and descriptions across many posts of this trend suggest that the woman is the "mother." The scene is most often situated indoors (next to a desk, or on a bed, sofa, or carpet) or occasionally outdoors. The woman is sometimes dressed in a sexualized manner. The physical contact between the woman and the child is mostly suggestive and sometimes sexually explicit (embraces, kisses on the lips, etc.).

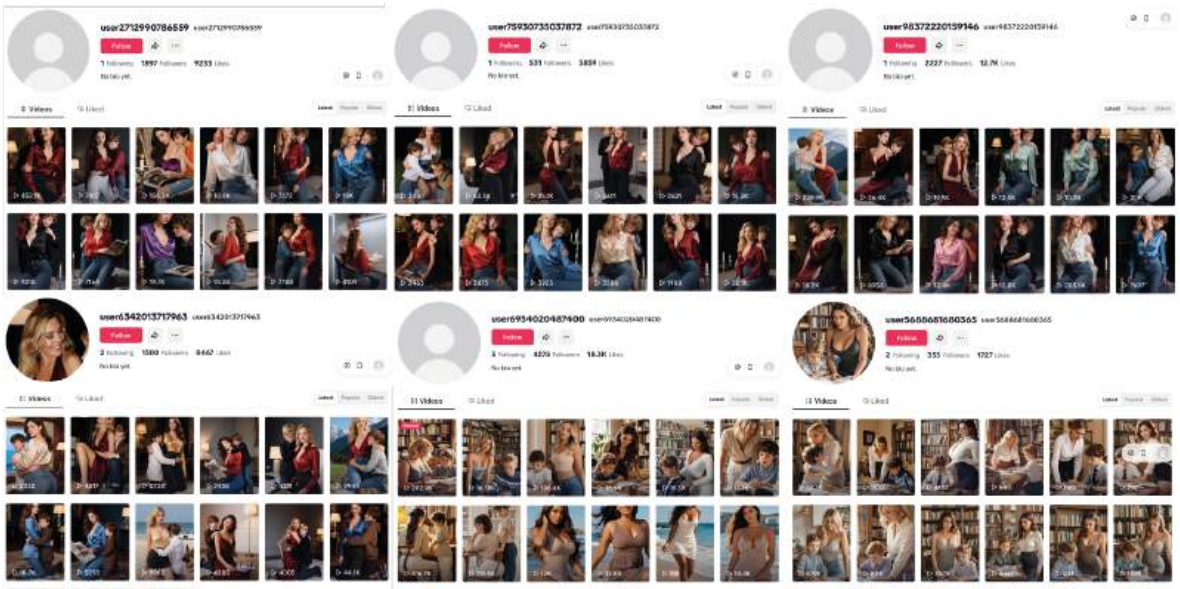


Figure 9. A selection of feeds from the mono-topic AAAs, which were coded to post content related to the "Mother-Son" trend. None of the above accounts remained active, showing the ephemeral nature of inception and activity of some AAAs that are primarily trend-driven.

A particular subtype of specialized AAAs encountered in this investigation focuses on exploiting explicit content, and, in a few cases, also sharing young, female, dysmorphic bodies (see Figure 10), which we labeled "young girls." This type of content focuses on AI female-like characters that display childlike features, appearing as children or likely underage girls. Serving as examples of "thirst traps," such content can potentially spread unhealthy body images while normalizing the sexualization of underage-looking girls and women.

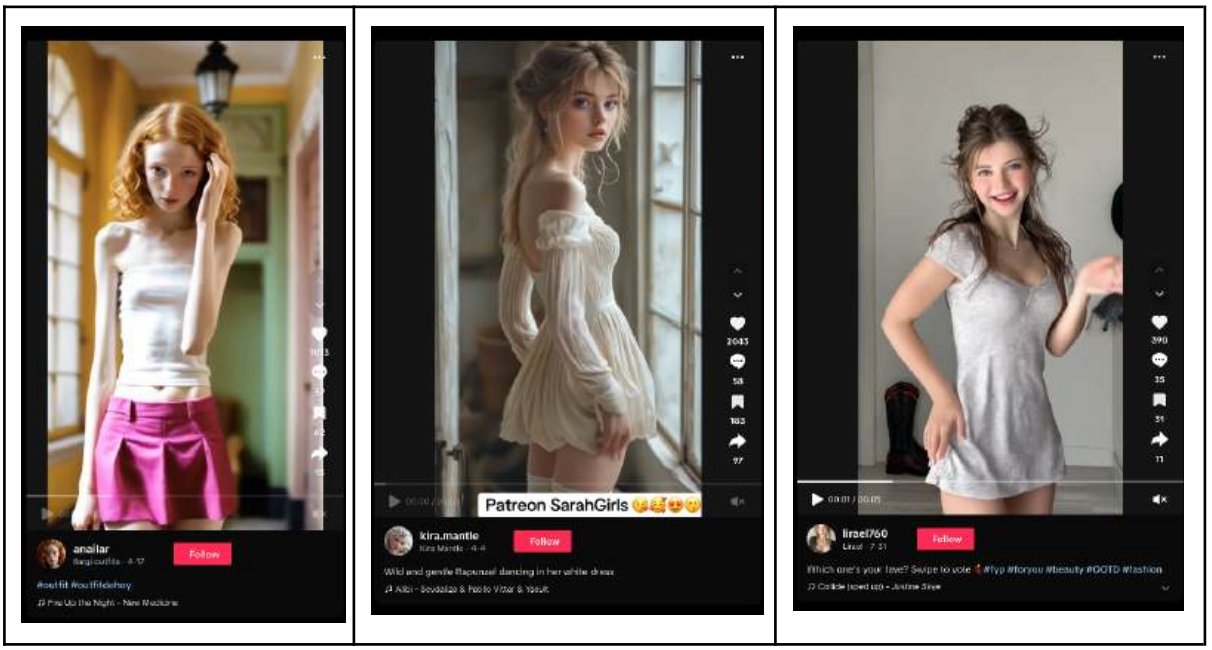


Figure 10. Examples of posts from three mono-topic AAAs specializing in posting content that sexualizes and glorifies body dysmorphia in AI-generated females with childlike features.

Racist and antisemitic content

We came across 8 (2.08%) AAAs that specialized in the "antisemitic" trend (Figure 11). The "antisemitic" trend is a type of content we observed during our initial monitoring in early July 2025. Its focus was originally set in a satiric hyperbole encompassing Jewish faith and the state of Israel, sometimes (sporadically) referencing the war in Gaza. The format shows photorealistic generative AI scenes where Jews (dressed as Orthodox) claim that Cyprus (or, as the trend develops, any other location) was promised to them thousands of years ago and is therefore now part of Jewish-owned territory. In addition to satiric messages, some of the viral AI content quickly took on an antisemitic turn, focusing on derogatory Jewish stereotypes.

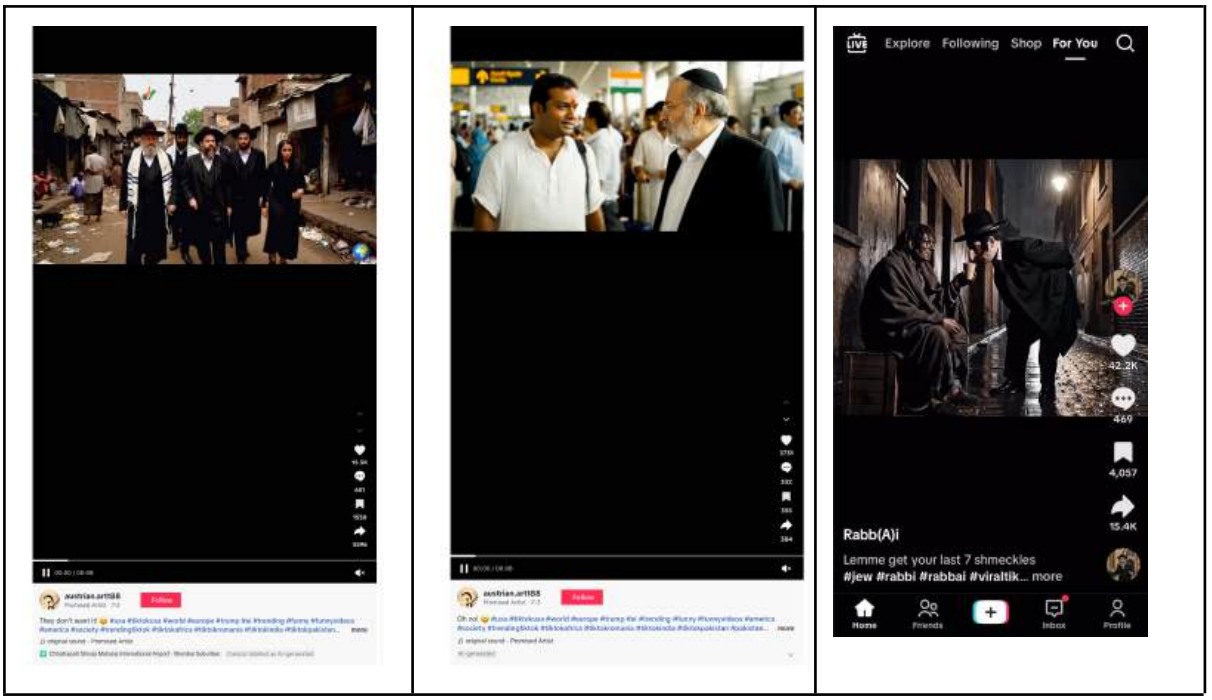


Figure 11. Examples of AI-generated, antisemitic posts shared by AAAs on TikTok in July 2025.

The above-mentioned trends share a characteristic of content depicting implausible, somewhat shocking scenarios, scenarios that are usually not permitted in other media environments. Indeed, some of the accounts we located that engaged in these trends were likely moderated and taken down; however, this process took a few weeks. For example, one AAA (figure 11, rightmost screenshot) with the username @rabbb.ai, stayed on the platform for over a month (becoming inaccessible to us on August 13, 2025), amassed almost 1.5 million views, and 150 thousand likes over 23 posts. The post visible in the screenshot gained over 610,500 views.

We also noted an alarming presence of "anti-immigrant and/or racist" content. Such AI-generated content displayed characteristics of plausible reality and, even if it did not (by alluding to dystopian futures, see Figure 13), it explicitly encouraged anti-immigrant narratives and sentiments (Figure 12). Examples might include "invasion" narratives, e.g., scenes showing AI immigrants "flooding" the entry routes to a given country (e.g., a mass of rafts approaching a national shoreline), fabricated interviews, and first-person POV clips in which AI "immigrants" are shown reiterating, inflating, and spreading derogatory stereotypes about themselves (e.g., job theft, welfare abuse, criminality). In such cases, AI-generated posts spread by AAAs systematically attune to certain fears and presumptions by pairing documentary aesthetics with scripted messaging, allowing for rehearsed stereotypes to play out.

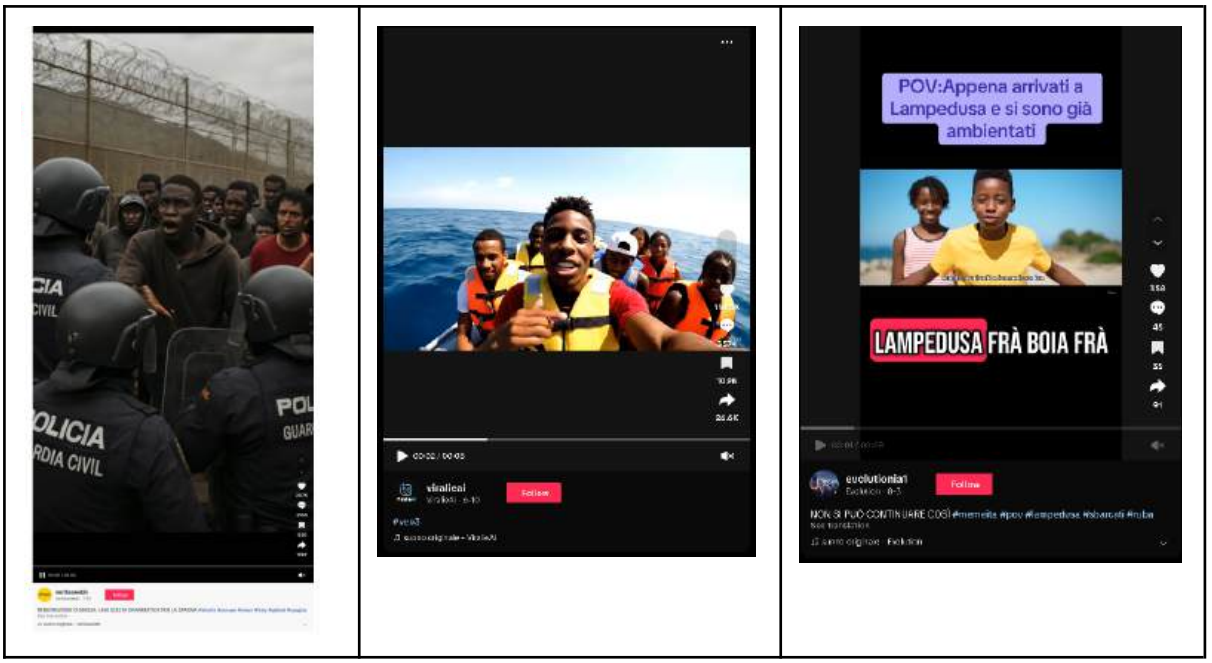


Figure 12. Examples of posts shared by AAAs as part of the "Anti-Immigrant" trend on TikTok in July 2025.

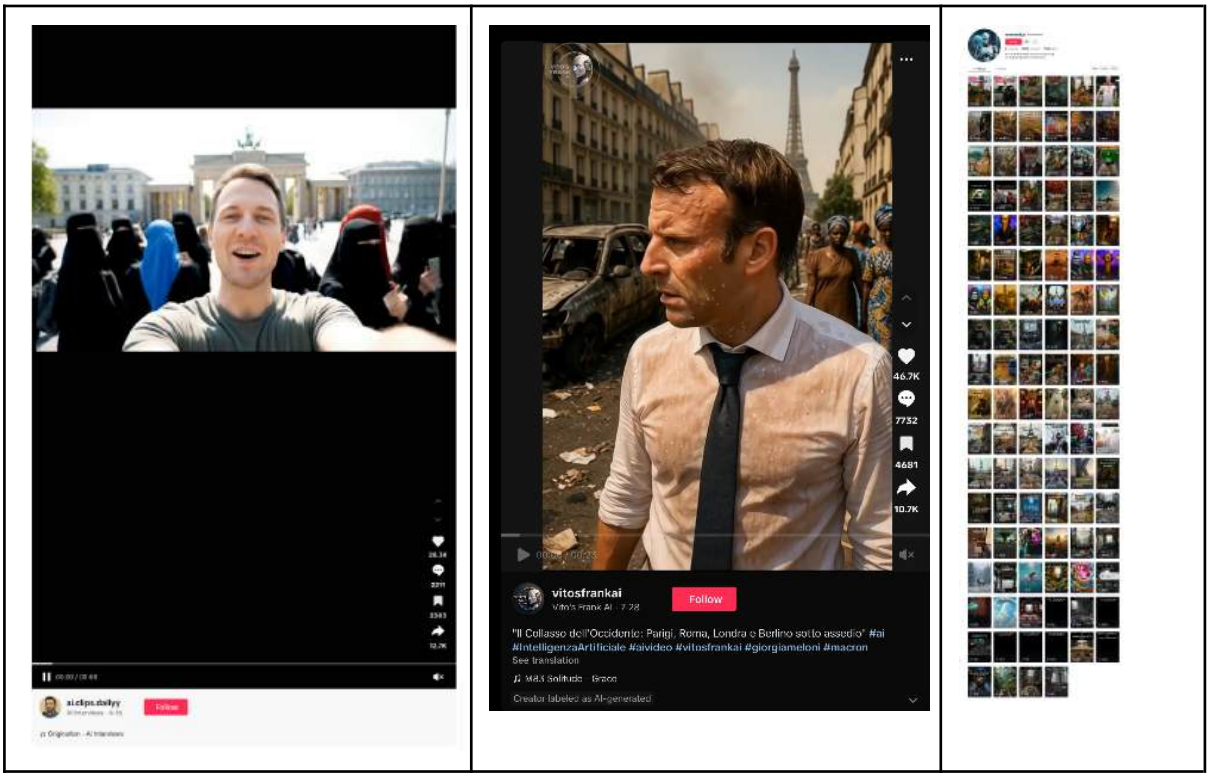


Figure 13. Examples of anti-immigrant posts shared by AAAs alluding to dystopian futures (left, center) and an example of an AAA feed, @nocturnal_ai (right), which specialized in posting such content intertwined with pro- and anti-candidates posts during the Polish 2025 elections (an account also discussed in the section "Removal, disappearance, deletion").

False news stories

Across the content types and formats, we note a concerning presence of false news stories. As false news stories, we considered AI content that purposely exploits the format of news media reporting for a plausible authority and believability, a content category which we coded as "synthetic (citizen) journalism." This distinction was necessary because a particular convection of AI content included the format of 'silly' street interviews, a trend partially prompted by the release of Veo 3, which we categorized under the broader category of "synthetic interviews" (see Figure 14). Such photorealistic "synthetic interviews" often featured a journalist with a microphone approaching people on the street, who would then either display or speak about ridiculous, memetic, or shocking behaviours.

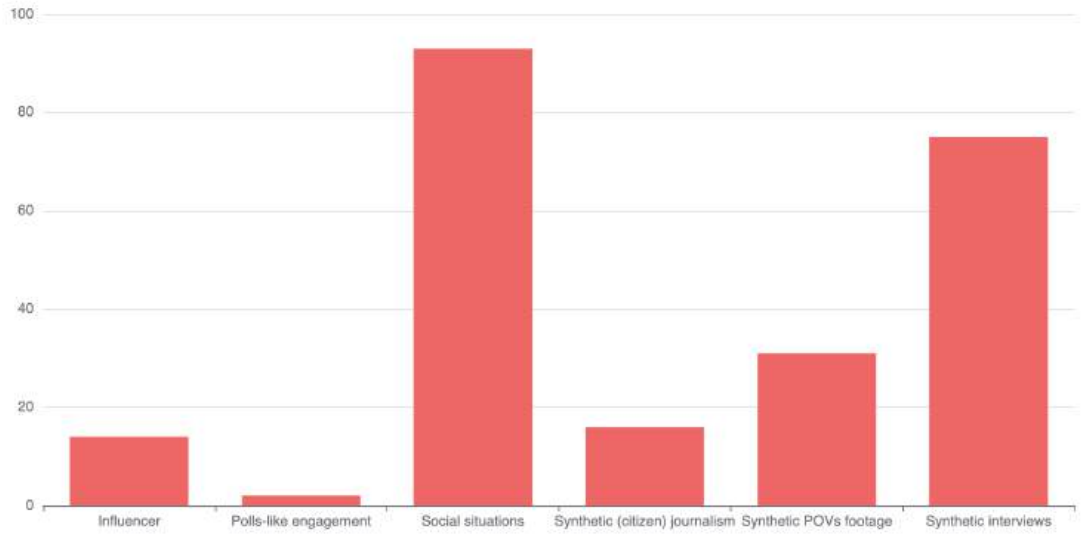


Figure 14. Number of AAAs coded per format label related to posts' content; note that one account can be coded as representing more than one category. For a detailed description of each format label, see the Codebook on the project page.

AAAs posting AI content coded as "synthetic (citizen) journalism" simulated footage posted by both professional and citizen journalism, as well as staged political events and protests. These posts deliberately adopt broadcast media conventions, such as on-scene stand-ups, anchor cutaways, and interviews, to manufacture plausibility. Some posts feature fake or real news media logos (e.g., ABC or Sky News; see Figure 15) within the video frame or on journalists' gear (e.g., in-frame graphics, microphone flags) to imply affiliation with real media organizations. In practice, this "synthetic journalism" fabricates events or reframes real incidents as reportage, laundering credibility through news aesthetics. One of the accounts that repetitively posted this type of content was @realscarytok, an account we also discussed in the section "Removal, disappearance, deletion."

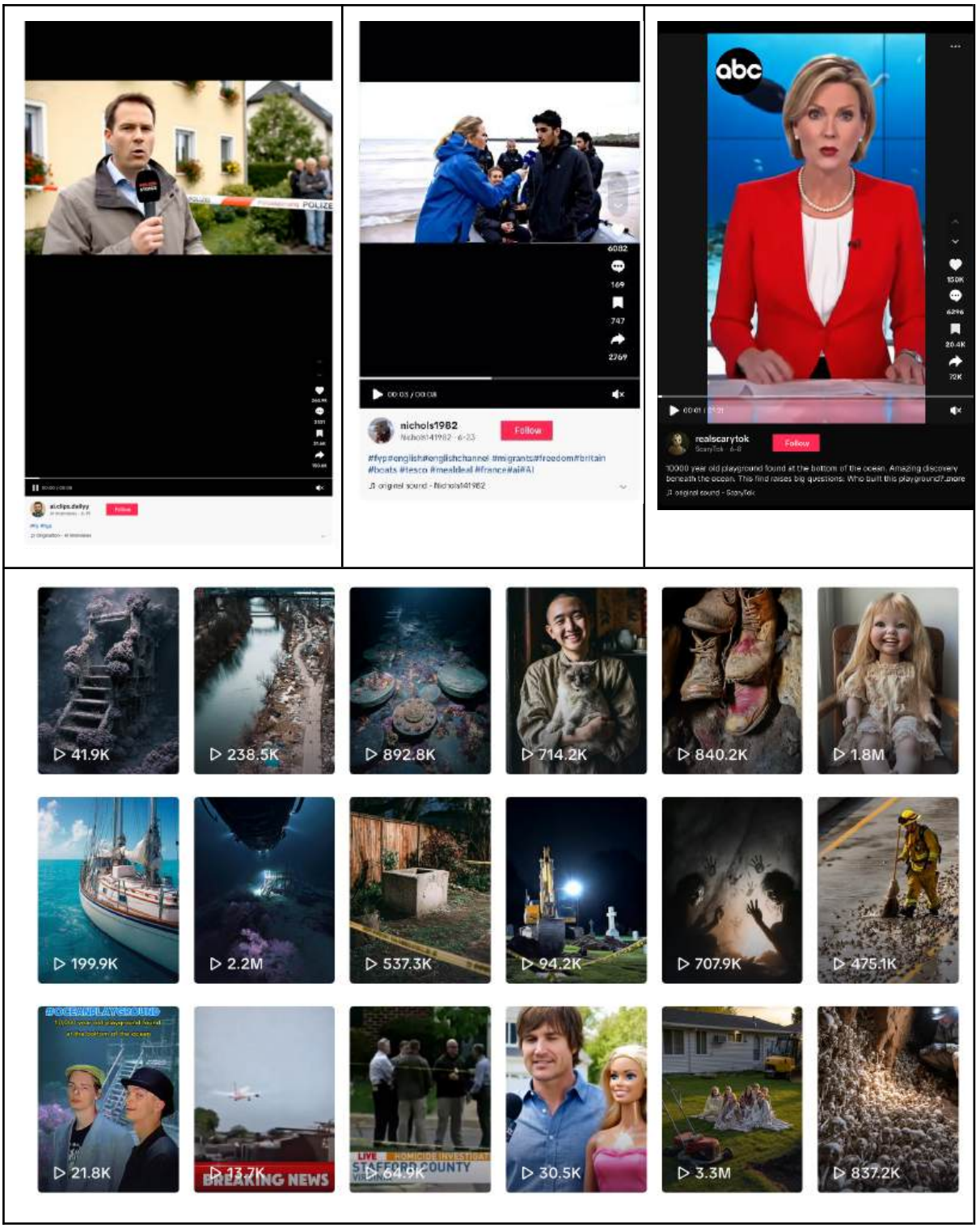


Figure 15. Examples of posts uploaded by AAAs coded as sharing "synthetic (citizen) journalism," meaning AI content that deliberately exploits the format of news media coverage for plausibility (top). A screenshot of part of the AAA @realscarytok feed (below), where each post in this segment was organized according to the same news media convention, using photorealistic AI-generated footage and posing as media organizations, such as ABC News, to share 'shocking' news stories.

We note that none of the top ten AAAs ranked according to most views in the one-month period of account monitoring were coded as containing content on the "female body" subject matter trend. While we observe the presence of content falling under the category of "storytelling clickbait" (2/10 cases) – which was also a category attributed to the leading AAA by number of views – the leading AAAs uploaded content that can be described as "cute," including "toddlers" (3/10 cases), "animals," and "animated" (2/10 each), as well as types of content also present in the upload-based ranking, namely "social situations" and "synthetic interviews."

As a media form, AAAs represent a trend-chasing, virality-seeking dynamic with limited causal and contextual coherence. AI-enabled recombination of otherwise unrelated contexts ties well to what platforms have favored in terms of engagement-maximizing formats and subject matter. In practical terms, AAAs span the plausibility spectrum, posting material that ranges from indeed aiming to be plausible to material somewhat indifferent to plausibility (mostly photorealistic nonetheless), but effective at cultivating particular affects and emotions while reproducing potentially harmful or hate-driven imaginaries.

Labeling of AI content on TikTok

TikTok allows different forms of labels for AI-generated content: those that creators set when uploading their posts, labels added by TikTok, and "disclosures" of AI content, which can be part of the posts' hashtags or descriptions, like using #ai (see also Table 3).

In our dataset, we found that more than half (55%) of AI-generated content is not labeled as such. The majority of content is labeled by the creators (42%), and only 1.38% of the posts were labeled by TikTok with a generic "AI generated" label, sometimes shown as "Contains AI-generated media."

This aligns with the findings in our previous report on AI slop, where we found that 50% of AI content on TikTok's search result pages across different topics was not labeled as such, and that only one video was labeled by TikTok. In this investigation, we collected a much larger dataset. Classification of individual content items was based on whether or not one of the 354 AAAs we monitored posted AI content solely. We found that the percentage of labelled AI content across our new dataset remains the same, leaving half of the AI-generated content without proper labels or disclosures.

Among 293 posts, we found the internal mark to be set to "0" by TikTok's systems (compared to where it would be empty if no label was added). We interpret this as representing cases where TikTok's system identified a post as not being

AI-generated when, in fact, it was (see row “Labeled as not AI” in Table 3). Based on our sample, we estimate that TikTok’s AI detection has an accuracy of 95%, meaning 5 out of 100 AI posts are incorrectly labelled.

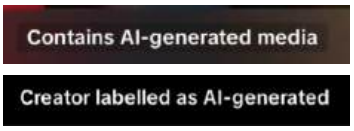
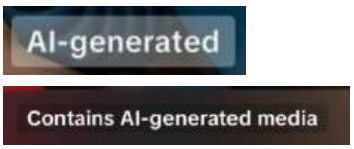
Type of AI label allowed by TikTok	Count	Percentage	Example
Not labeled	13,934	55.40%	n/a
Labeled by creator	7,630	30.34%	
Creator’s disclosure in the description	2,946	11.71%	
Generic label	347	1.38%	
Labeled as not AI	293	1.17%	n/a

Table 3. Different labels and the variants we observed on the TikTok user interface.

Since our last report, the text of the TikTok assigned “Generic label” has also changed from “Creator labelled as AI-generated” to “Contains AI-generated media,” making the use of AI less explicit, which can explain the lower percentage of comments on the topic (see Figure 16). Given that this change was not permanently visible across posts, it is likely an A/B test of new labeling text.

TikTok’s label “Creator labeled as AI-generated”	
mobile (variant A)	mobile (variant B)
	

Figure 16. Variants of labels where videos were labeled by the creator, likely an example of A/B testing.

Similarly to our findings in the previous report on AI slop, we could not find any relevant correlation between the reach of a video and whether or not it is labeled as AI-generated. But it appears to be something that content creators think about as

well. We found instances of videos where the AI-labeling in the hashtags was temporarily removed (for a day).

Since TikTok relies on the honesty of individual content creators when it comes to labeling, we also checked how consistent the AAAs' creators are. Figure 16 shows that only 10% of AI-creators (38) consistently labeled their content, while 29.6% (112) never did, and the remaining 229 creators were inconsistent in their labeling. Of all creators who labeled at least some of their content, 22.7% labeled less than 10% of their posts, while 29.3% labeled more than 90% of their posts in some form. This allows us to conclude that two-thirds of AAAs' content creators do not label their content consistently, while TikTok does nothing to enforce this.

To evaluate the importance of labels, we cross-checked the presence of labels with mentions of "AI" in the comments. As reported above, TikTok allows a variety of labeling approaches, from markers clearly visible in the video to hashtags that may be hidden among a long list of hashtags. Figure 17 shows how often users discussed AI in the comments depending on the type of AI label present in the post. A chi-square test showed the difference in categories to be statistically significant. It showed that labeling by creators, either through the official label or in the description, is correlated with users discussing AI in the video comments.

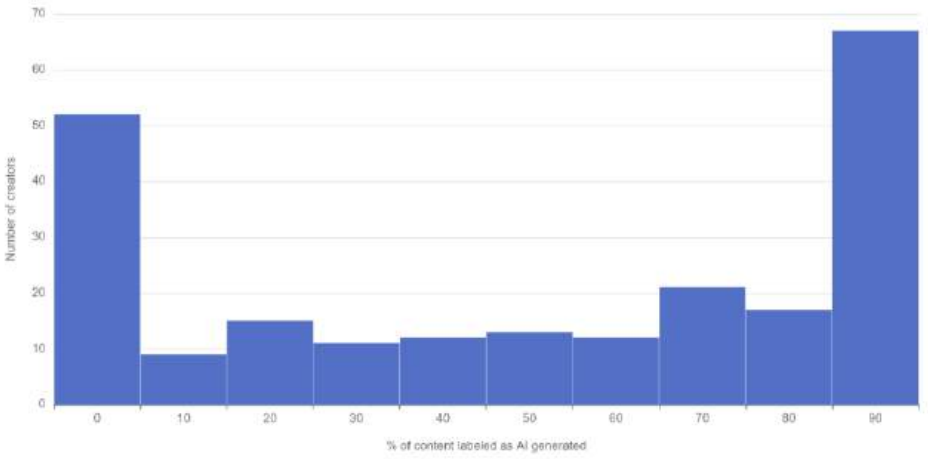


Figure 17. Distribution of accounts based on the percentage of their videos that are labeled as AI-generated.

In our previous report, we criticized the fact that content creators could add #ai anywhere in the post's description to comply with TikTok's transparency requirement, even though long descriptions are often cut off in TikTok's interface, effectively hiding the majority of text and hashtag lists. In the dataset of this study, we found that 85% of AAAs that mention "#ai," or state that the content is AI-generated, in a way that is directly visible to users, within the first 80 characters of the description. This shows that the majority of AAAs' content creators do not try

to evade labeling. Still, this finding has to be contrasted with the fact that the majority of content is not labeled at all, as the labeling is not enforced.

What do users think? Analyzing comments

To understand how users perceive AAA content, we analyzed the first 50 comments posted in the comment sections of posts to determine whether they discuss the authenticity of the content they see.

We collected 510,839 comments on 43,798 posts by AAAs. The median number of comments collected per post was 10, with 15% of the posts not having any comments. We then checked for the presence of identifying words (e.g., "AI" or "Veo 3") and phrases suggesting that the authenticity of the content was in question (e.g., "is this real" or "nowhere else on the internet"). It is worth noting that 30.4% of all comments we collected consisted of a single or repeated emoji.

We found that just 5% of the comments matched one of the 50+ phrases implying suspicion about the authenticity of the post. Overall, 16.5% of posts with comments had at least one comment mentioning AI or questioning the authenticity of a post. Interestingly, the presence of those comments varies significantly across channel topics. Figure 18 shows the difference between topic categories and the rate at which posts had at least one comment mentioning AI. It shows that for accounts with political content or that pose as news stories, more videos have at least one comment that mentions or questions their authenticity. At the other end of the spectrum are videos that are either obviously non-realistic, like Jesus-themed videos or animated posts, as well as those with sexualizing content, like the mother-son or metro trends, as well as those focusing on female bodies or young girls.

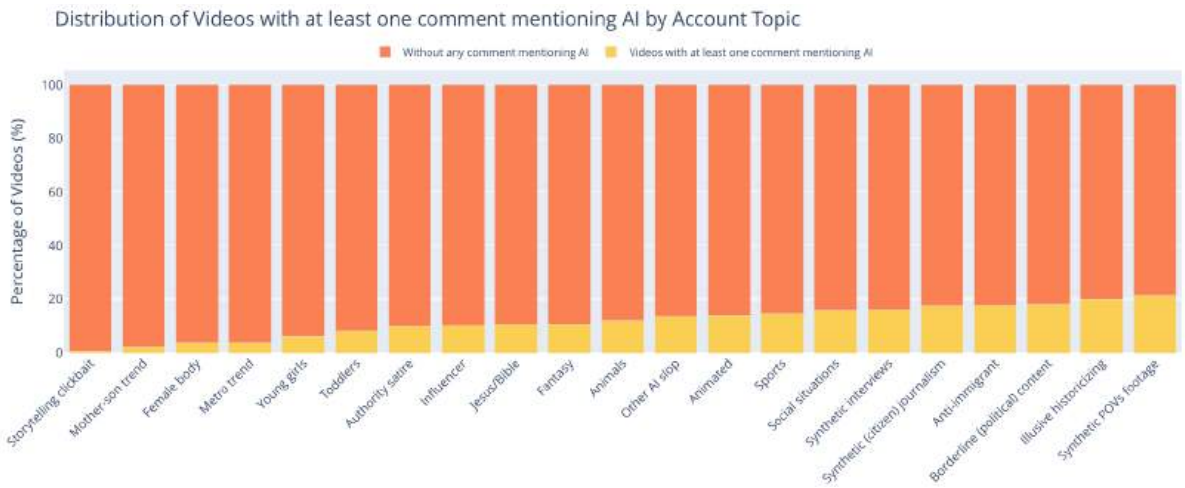


Figure 18. Percentage of posts with no or at least one comment mentioning AI, based on comments on 66,816 videos. For this chart, we normalized the values per account to not

overemphasize accounts with many topically similar posts. We also excluded topics for which we had data from fewer than 10 accounts. Note that each account can comprise multiple topics.

Monetization: the economics of AI slop

As described earlier, there are various platforms that offer services to optimize and streamline the video creation process for viral content (see also the section "Ecosystem"). To get an idea of how much content creators spend on creating AI videos, we compared prices for some high-rated video creation models. To estimate the actual price, we used replicate.com to generate test videos and compare their quality. There are various websites that offer free trials or subscriptions[4] that conceal the actual price. Table 4 shows the actual cost of generating a video.

Model	Model creator	USD/second	Length in seconds	Note
Veo 3 + Veo 3.1	Google	$0.40	8	with sound, good quality
Sora 2	OpenAi	? (~$0.40)	10-20	with sound, good quality
Veo-3(.1)-fast	Google	$0.15	8	with sound, medium quality
wan-2.5-t2v-fast	Alibaba	$0.08	5/10	with sound, medium quality
hailuo-02	Minimax	$0.10	6/10	no sound, medium quality
seedance-1-pro	Bytedance	$0.15	up to 12	no sound, medium quality
Kling-v2.5-master	Kuaishou	$0.28	5	no sound, medium quality

Table 4. Price comparison for video generation models.

There are various platforms that offer AI-generated video creation for free, such as Meta AI Vibes or Sora. However, though these platforms provide the service for free right now, to gain market shares, they will eventually find a way to monetize in the future, as this activity is quite expensive in terms of resource consumption (energy, servers, training costs, etc.). If we calculate the costs of AI videos that fall into clear-cut categories based on their length, covering 10,289 videos in our dataset,

[4] We did not have access to Sora 2; OpenAI now has a clear pricing scheme as video generation is based on subscription. The estimated price is based on expert analysis claiming it is likely on par or more expensive than Veo 3.

producing all these videos costs between 8,100 and 25,800 USD. AAAs' creators thus spent between a dollar and 1,237 USD in total for their accounts.

Similar to the adoption of AI in the broader economy, it is unclear at the moment if automating AI content will provide the content creator industry with a viable and reliable line of income. An MIT study from July 2025 found that 95% of AI pilots see zero return on investment in AI projects.

TikTok offers multiple ways to monetize content, like selling products on TikTok Shop or collecting gifts for creators during live streams ("Lives"). The most common way for creators to directly obtain TikTok revenue from ads is monetization through the Creator Rewards Program. But the Creator Rewards Program has several restrictions on accounts and content. Besides the requirement to follow the terms of service and community guidelines, the most notable thresholds are:

- Accounts need to have more than 10,000 followers and 100,000 views in the past 30 days to be eligible.
- Each video must gain at least 1,000 views, be at least one minute long, and have a high-quality file format (1080p).

In terms of account-level requirements, 129 accounts in our dataset qualified with more than 10,000 followers at the beginning of our monitoring, and of those, 89 reached the threshold of 100,000 views in 30 days. Of those 89 accounts, 62 produced at least one video longer than one minute. Overall, just 13% of the videos published by AAAs were eligible for monetization based on their account metrics. The vast majority of videos are just 5-8 seconds long, which is the standard length for videos generated by state-of-the-art video generation models, which can last up to 10 seconds.

This means that, to be eligible for monetization through the TikTok Creator Rewards Program, AAAs creators have to create storylines that consist of at least six generated videos. One strategy we observed was simply combining multiple versions of the same plot, as shown in the screenshots below, where each 35-second plot (children falling into a hole, buried under cement, to be rescued by a cat) appears twice with different renderings (see Figure 19). This strategy leverages the fact that many AI-generation platforms produce multiple outputs for the same prompt, allowing the user to pick the best one. AAAs simply use all those AI-generated video versions and combine them into a single post to reach the one-minute threshold.

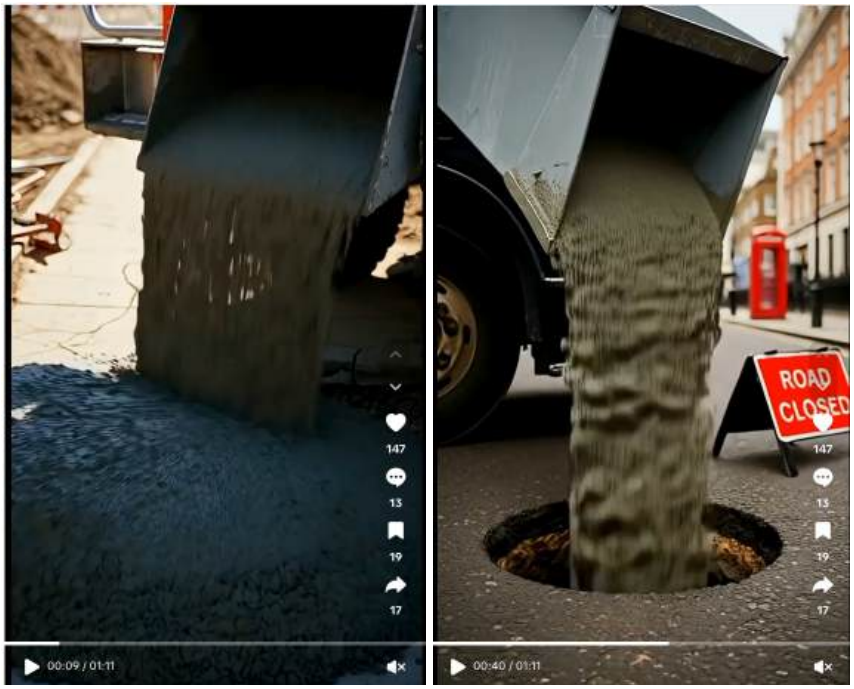


Figure 19. Two screenshots from the same video show different versions of the video, likely resulting from the same prompt being run twice.

Another way of creating income for AAAAs creators is to sell their experience with AI video creation. We found several instances where accounts started advertising services on top of AI video generation models (see Figure 20). We found 71 instances of accounts including links or direct references to AI tools in their TikTok profile bios. By "instances," we refer to each change in the profile bio field that contained any reference to an AI tool, website, and/or paid course promoting methods for creating "viral" AI content. These instances were found to be present across 38 AAAs.

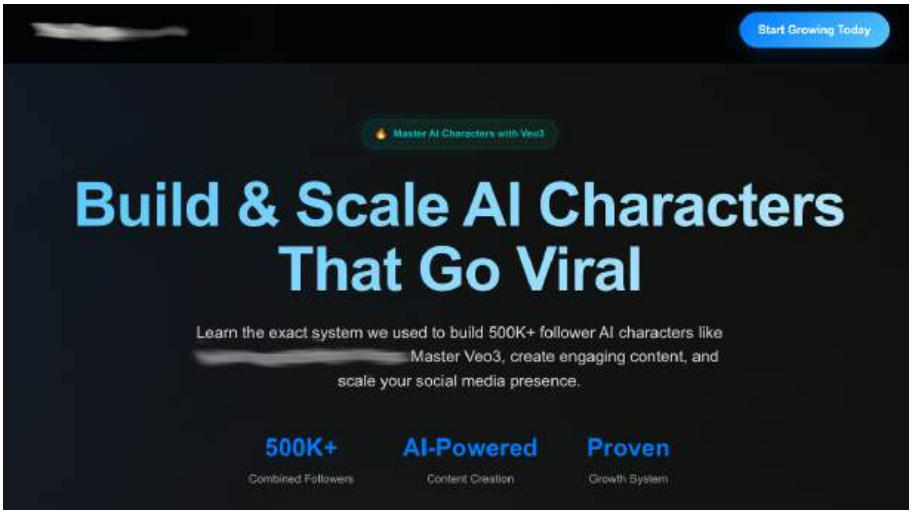


Figure 20. Screenshot of a website linked to a successful AAA.

AAAs and their strategies for monetization

Given the relatively low likelihood of achieving successful monetization solely through TikTok's own Rewards program for content creators, some AAAs turn to sponsored content. Across our dataset, AAAs don't start by presenting sponsored content right away; this comes later, possibly when an account reaches a specific follower threshold. Alternative monetization attempts include actively seeking out sponsorships (by promoting services or products for purchase), promoting AI tools (courses, apps, websites), and selling account-related AI-made merchandise. In our dataset, we found 12[5] AAAs that appear to directly upload sponsored content and advertisements to their profiles.

Despite not actively seeking health-related topics, we found that 2 of 12 AAAs that upload ads and paid content are dedicated to health supplements, medications, and health advice (Figures 21 and 22). This is consistent with our previous report on AI slop, where we discussed two examples of AAAs promoting health supplements and sharing health advice (both of those AAAs have since been deleted). This indicates that health-related content is one of the less prominent but consistently present niches for AAAs.

In both cases, the AAAs turned to promoting health supplements following two major shifts in content strategy: first, they switched from another niche specialization to posting health-related advice, and second, they began promoting real health supplements rather than health advice and "fun facts" content. To promote health supplements, AAAs either use AI-generated imagery (Figure 21) or post non-AI ads (Figure 22). As Figure 21 shows, using AI-generated imagery, such as photorealistic videos and pictures promoting health supplements, seems to drive more engagement than an AAA switching to posting non-AI content promoting similar health-related supplements (Figure 22).

Across both accounts, we observed a steep decline in engagement after the AAAs switched to promoting 'real' health supplements. This means that, despite the previously established following and a rather steady number of views, these two AAAs were not successful in promoting health-related medicines and supplements. This constitutes a positive sign in limiting the potential risk that such accounts pose in persuading users to purchase medically questionable health supplements, which might result from TikTok limiting the reach of such content or users being significantly less interested in it. Furthermore, as of November 3, 2025, one of the two AAAs appeared to be deleted and was no longer accessible on TikTok.

[5] This number excludes any mentions of promoting AI toolsin the profile bios, focusing solely on uploaded content. This number is based on a total of 383 AAAs identified during the July and August data collection cycles.

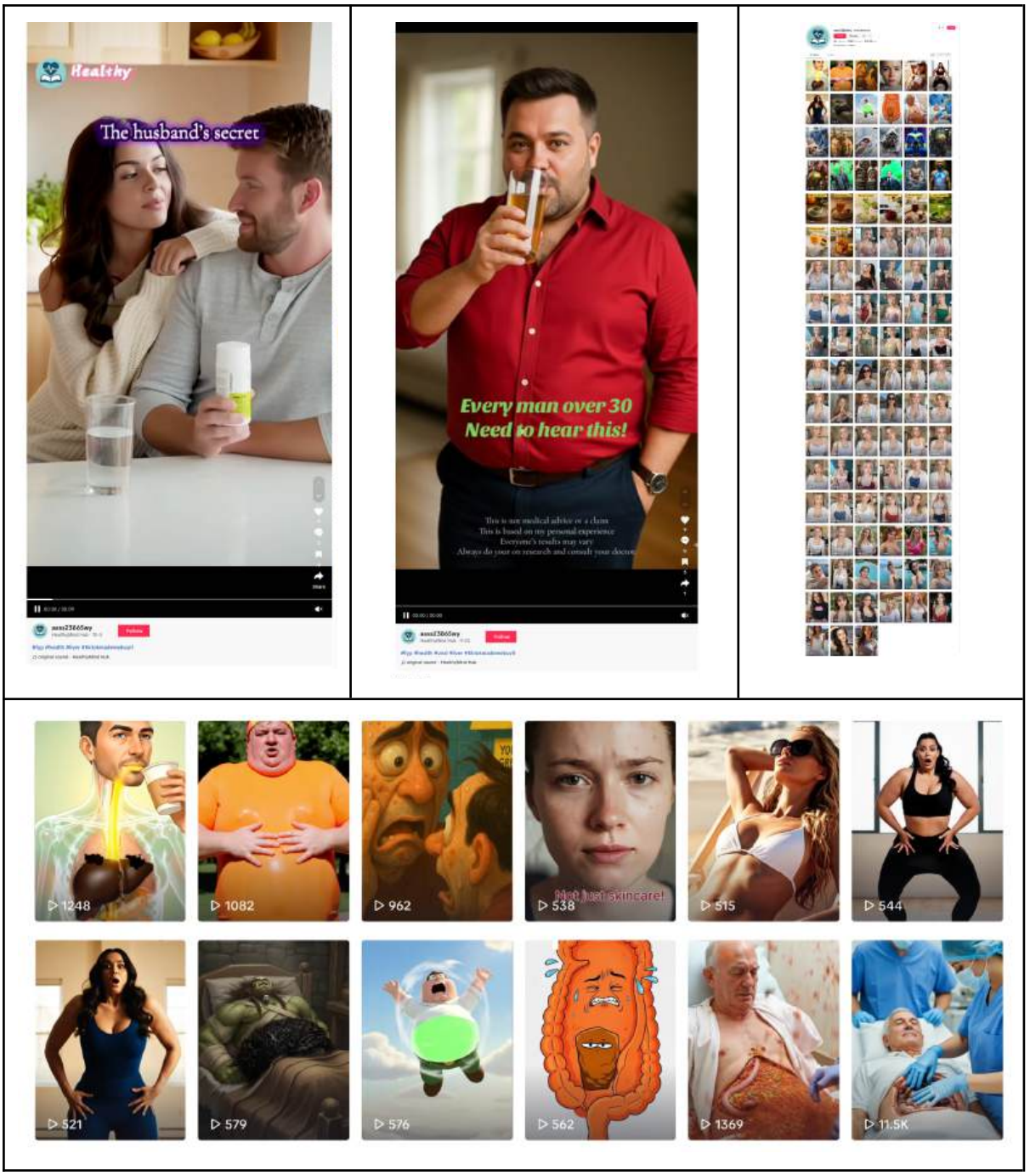


Figure 21. An example of a mono-topic AAA that shifted from posting content on the female body to health-related advice and ads. While the health-related content was gaining some visibility and engagement, the sponsored content, despite still being generated by AI, did not seem to achieve similar results. The top two screenshots show two examples of posts promoting health supplements (top left), whereas the screenshot to the right offers an overview of the account's posting history through its feed, with a detectable moment of topic transition (top right). The screenshot below is a close-up of two rows from the account's feed history, showing the recently uploaded content.

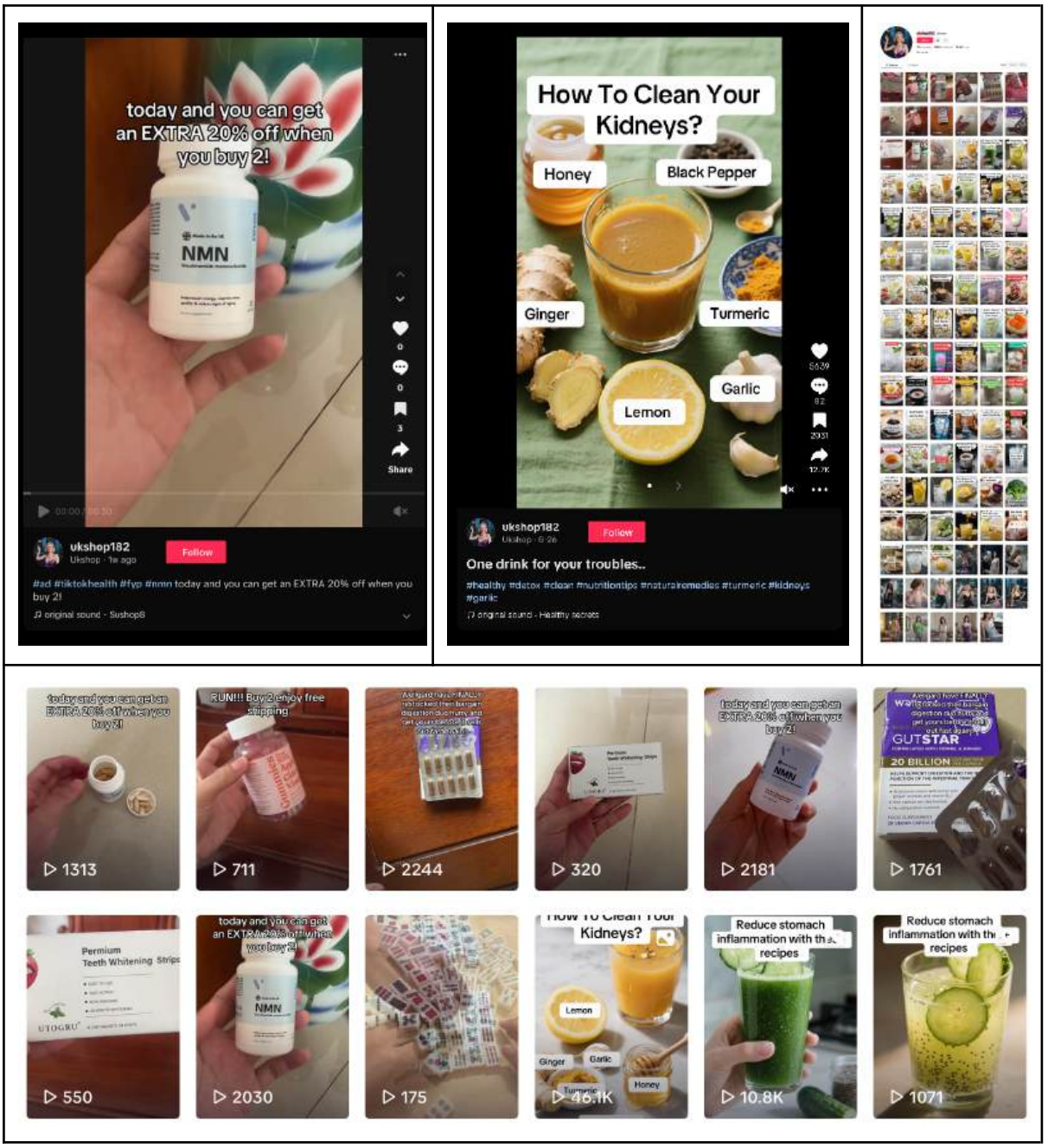


Figure 22. An example of a mono-topic AAA that switched from posting health advice videos to promoting health supplements. After switching, the AAA began posting non-AI videos advertising various health supplements. We assume that both the break from non-promotional, "organic" content and from attention-grabbing AI-generated aesthetics caused the steep decline in views and engagement on most of its new videos. The top two screenshots show two examples of such posts (top left), whereas the screenshot to the right provides an overview of the account's posting history in its feed (top right). The screenshot below is a close-up of two rows of uploaded content from the account's feed, showing the moment of topic transition.

Another monetization attempt that AAAs display is actively seeking sponsorships by emulating human influencers. In this case, mono-topic AAAs usually focus on uploading posts featuring a single main AI character that is visually very similar

across most or all of the shared content. This character poses as an “influencer” and often uses “I” in sentences, both in post descriptions and in subtitles within post frames. We also note a few cases in which AAAs used ad-like content while emulating an influencer format and style (Figure 23). Interestingly, we found that one of the AAAs that later switched to promoting health-related supplements attempted to emulate influencer content, including sponsored product promotion. AAAs may pose as “real” influencers to increase authenticity.

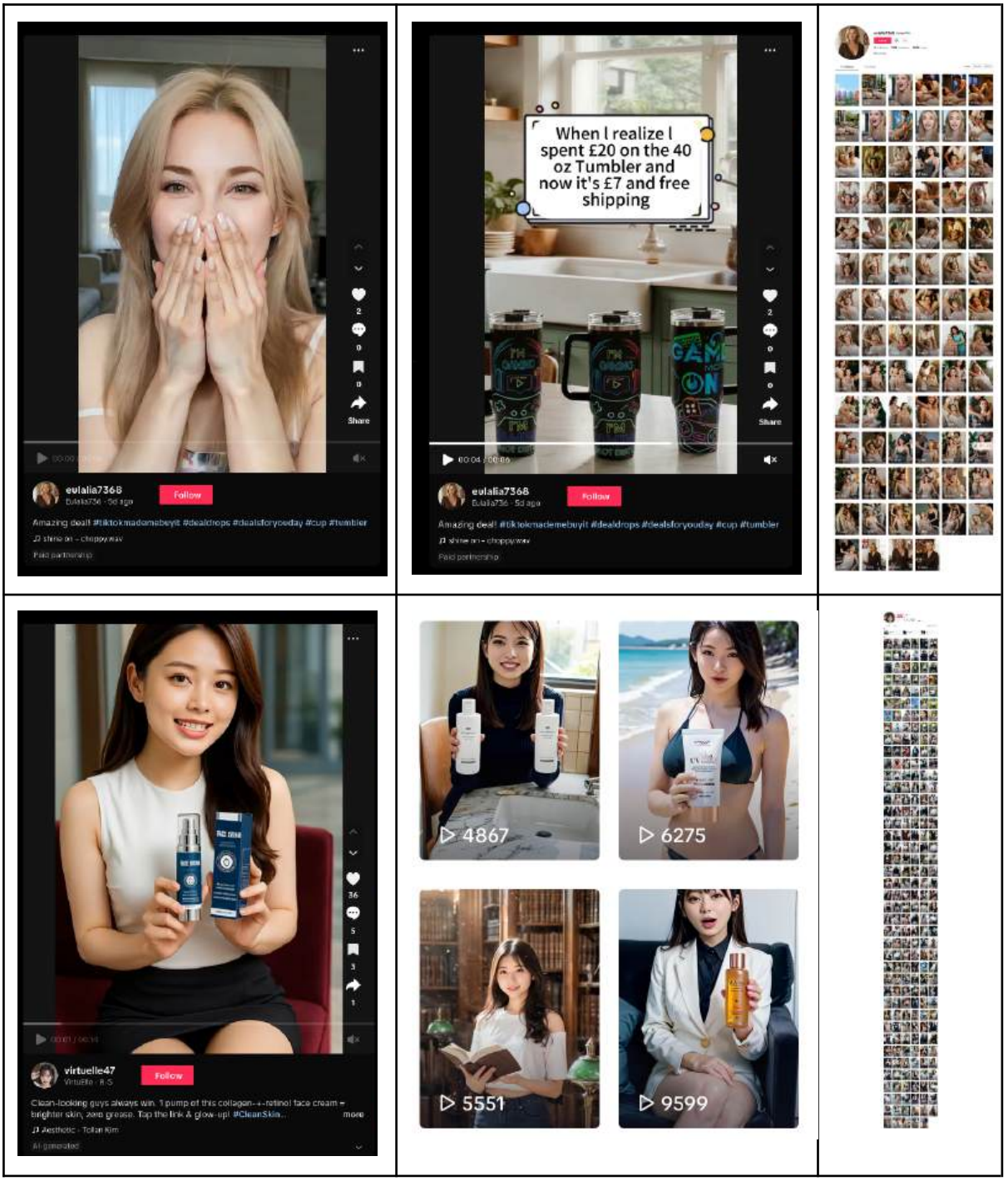


Figure 23. Examples of two mono-topic AAAs emulating the format and style of an influencer and engaged in disclosed paid partnerships (above) and undisclosed paid partnerships (below).

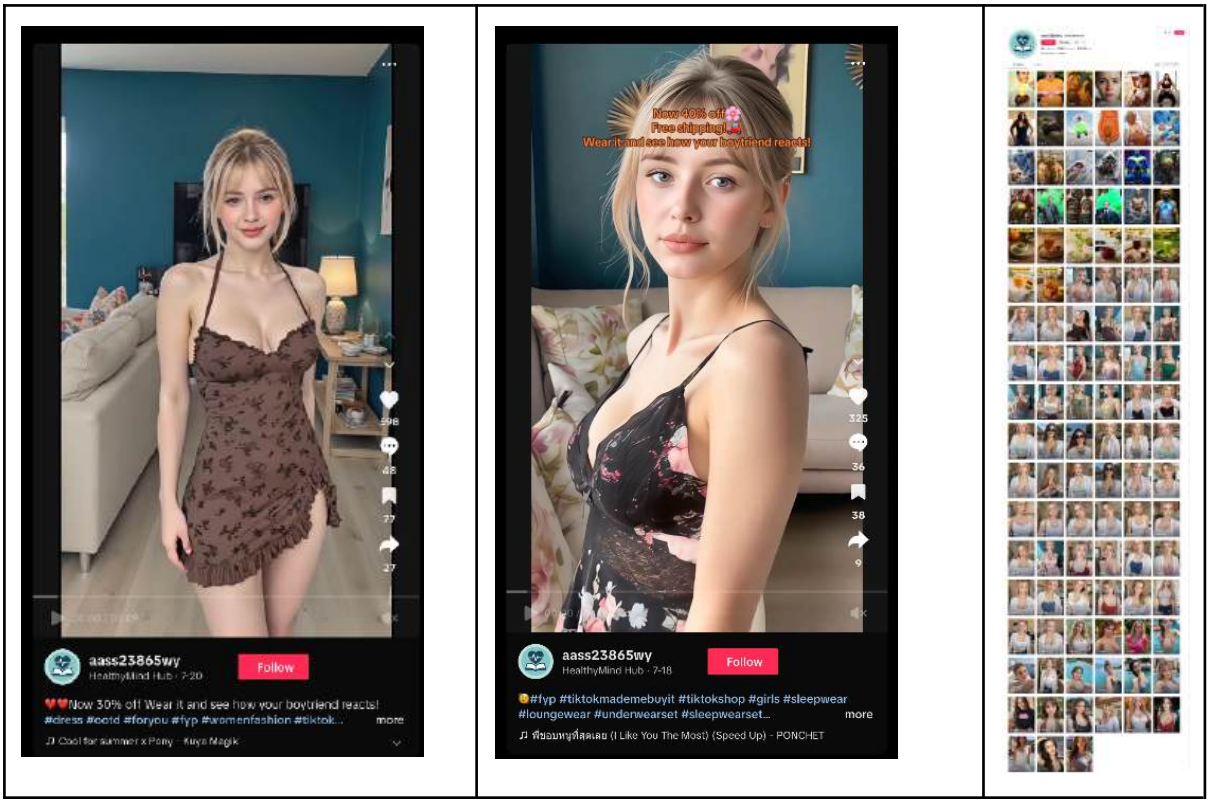


Figure 24. An example of an AAA that later switched to promoting health-related supplements (Figure 21, above), which previously attempted to emulate influencer-like content, including sponsored product promotion.

In relation to selling account-related AI-made merchandise, we found two examples of mono-topic AAAs that exclusively promote their own products (Figure 24). Such content includes using AI-made celebrity imitations to promote a product, raising questions about the use of likenesses for commercial purposes on social media platforms (Figure 25, left). Alternatively, adding a widget with a link to a store containing account-related merchandise is also observed (Figure 25, right).

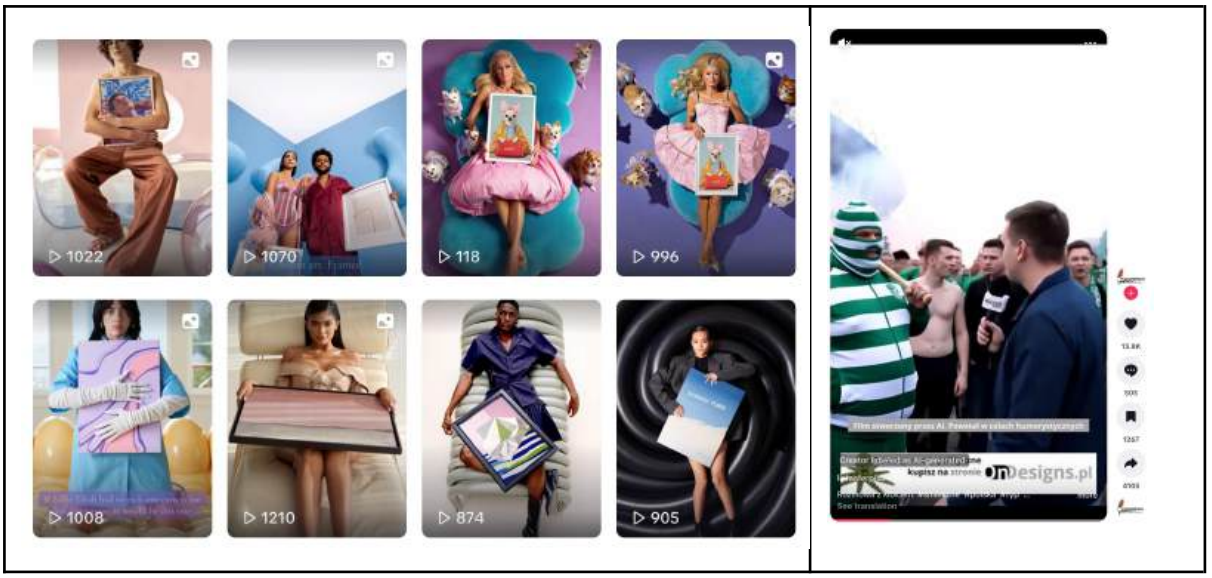


Figure 25. Examples of two mono-topic AAAs promoting account-related products (left) or merchandise (right).

Ecosystem: AI virality tools

Even before Meta and OpenAI announced their AI-only platforms Meta.ai Vibes and Sora2, a new, automated set of tools emerged that specialized in generating solely AI content. Ranging from Discord channels to proper websites, such tools are one of the pillars, if not the backbone, of AAAs. These tools aim to change how users create content by automating content creation and production with AI agents. They also play a leading role in maintaining the hype and promise of monetary gains from coupling generative AI tools with social media.

AI tools that are the pillar of AAAs offer user-friendly interfaces that facilitate the use of tailored prompts to generate different types of content. They constitute a collection of agents and pipelines that offer different scopes, including the production of 'viral' social-media-ready content. Such tools are often "readymade" in two ways: first, they provide pre-made prompts, and second, they offer direct pipelines to AI models to generate content from these prompts. By providing pre-made prompts, they take away content creators' burden to "design prompts" that "get" the model to do what "the creator wants," saving both content creators' time and the AI models' use credits. Indeed, some AI tools and models require paid credits to generate prompts; hence, the more attempts and adjustments a prompt requires before generating the "right" piece of content, the greater the monetary loss. These 'readymade' AI tools promise to take care of this obstacle, offering "verified" prompts that promise to *work* and *go viral*. Furthermore, many of the AI tools integrate leading generative AI models so that content creators can easily deploy viral prompts into production without leaving the website or app.

To further illustrate what such tools are, we highlight Yapper.so as a case study – a website whose landing page reads “Make Viral AI content” (see Figure 26). We located Yapper.so through one of the prominent AAAs in our collected dataset (@realscarytok, with over half a million followers and 10 million likes), which provided a link to a website yapper.so in its profile bio. As expected, Yapper.so is a website that promises and optimizes the generation of viral AI content and its subsequent monetization. Yapper.so is also an example of how AI tools constitute an emerging commercial business model, thieving off the promise and allure of virality and monetization. AI tools for content creators on social media platforms operate on a similar remuneration system, promising monetary and social rewards.

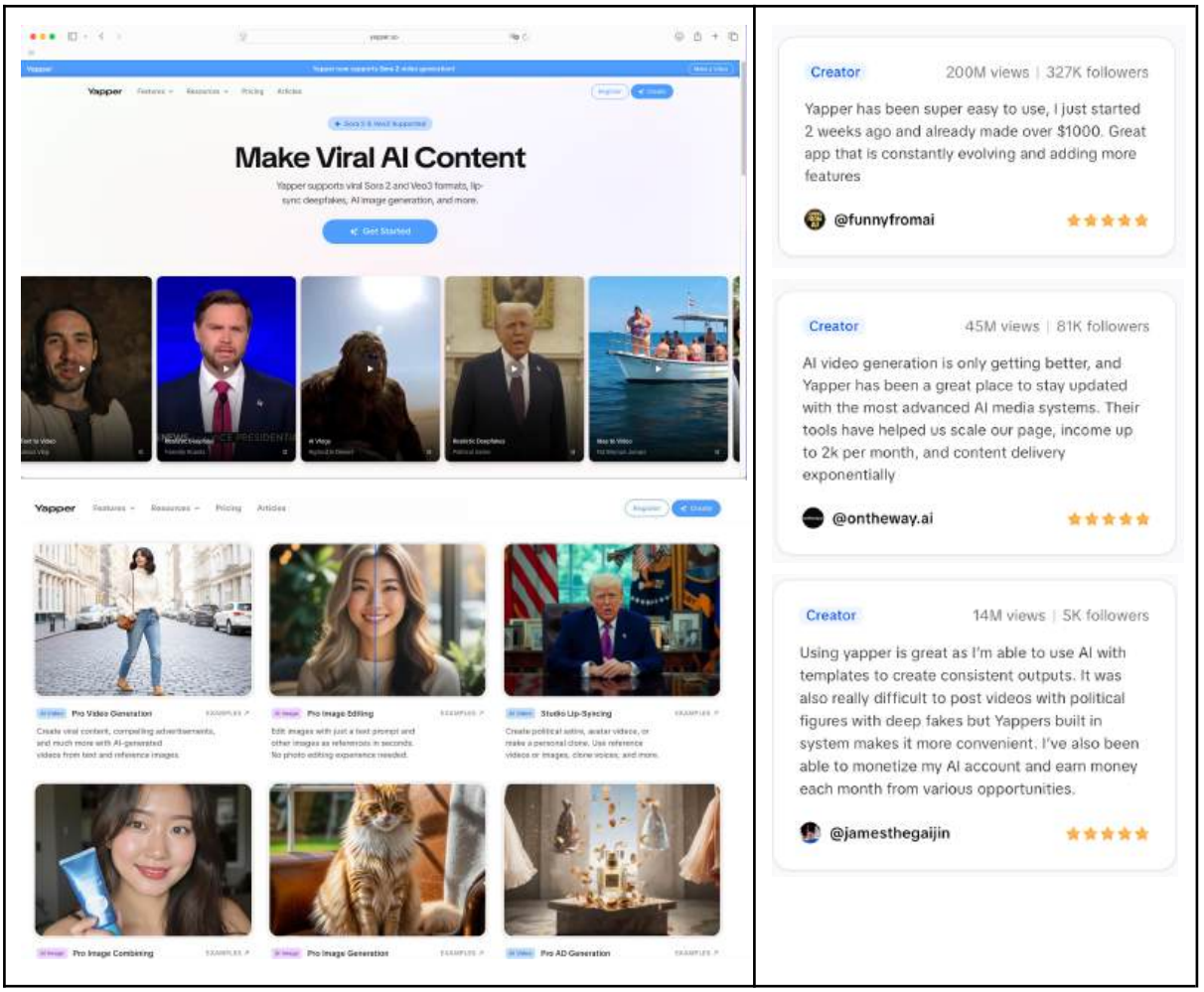


Figure 26. The two screenshots on the left show examples of AI-generated content Yapper.so shows on its landing page. The three screenshots to the right are close-ups of three of the top six exemplary customer reviews (displayed under the headline “Join thousands of happy customers”) on the same page.

The website of Yapper.so (Figure 26) indicates the range of content promoted as potentially “viral.” While scrolling through the site’s main page, one sees that deepfake-like generative AI clips of Donald Trump are used as examples three times, alongside other politicians (J. D. Vance) and celebrities (Elon Musk). It shows that

plausible, photorealistic videos that use the likeness of (famous) real people without their consent are at the heart of 'readymade' AI tools. The three examples of "reviews" (Figure 26, on the right) are screenshots of three of the top six exemplary customer reviews that Yapper.so includes on its main page under the headline "Join thousands of happy customers." The users seem to praise how their income has increased thanks to the monetization of viral AI content made using Yapper.so.

The website also offers an extensive collection of "example videos" created using its pipeline (Figures 27-29). Some examples show selected engagement metrics, including the number of views and likes for a post where the video was deployed. Upon clicking on a video example with engagement metrics, Yapper.so allows us to view the piece of content and access it on the platform where it was posted (figure 28).

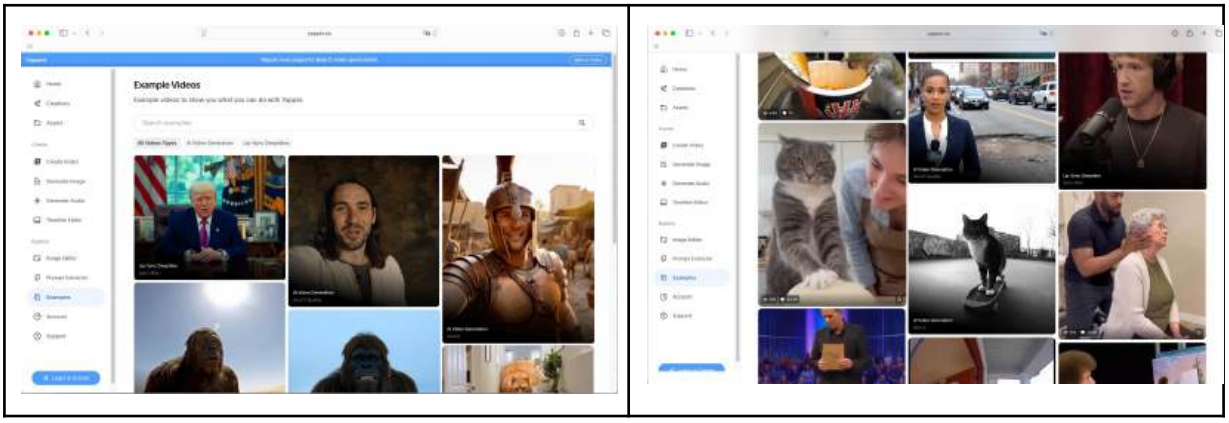


Figure 27. Examples of videos shown on the Yapper.so website, including a deepfake video of Donald Trump (top-left) as well as various other examples of AI slop (right). For a detailed example, see the subsequent Figure X which displays screenshots from a video (located in the top right screenshot) of a Black man massaging an elderly woman, and its embedding as an actual post on Instagram accessible via a direct link from Yapper.so.

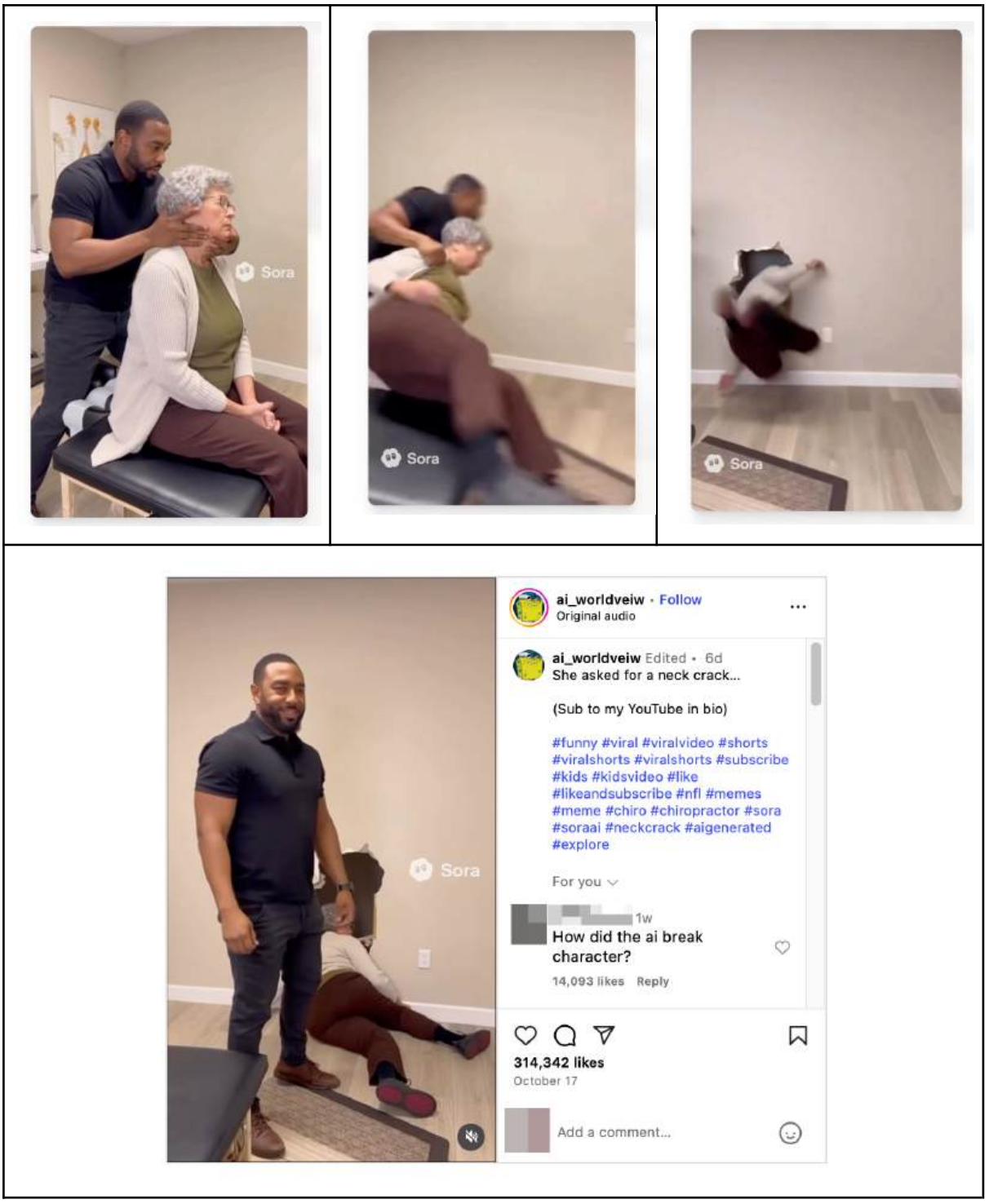


Figure 28. Example of a particularly violent post included in “Video examples” on the Yapper.so website. This AI video shows a Black man performing a neck massage on an elderly woman and subsequently throwing her into a wall. The video, generated with Sora (as seen on the watermark), is also accessible via a direct link from the Yapper.so website to an Instagram page.

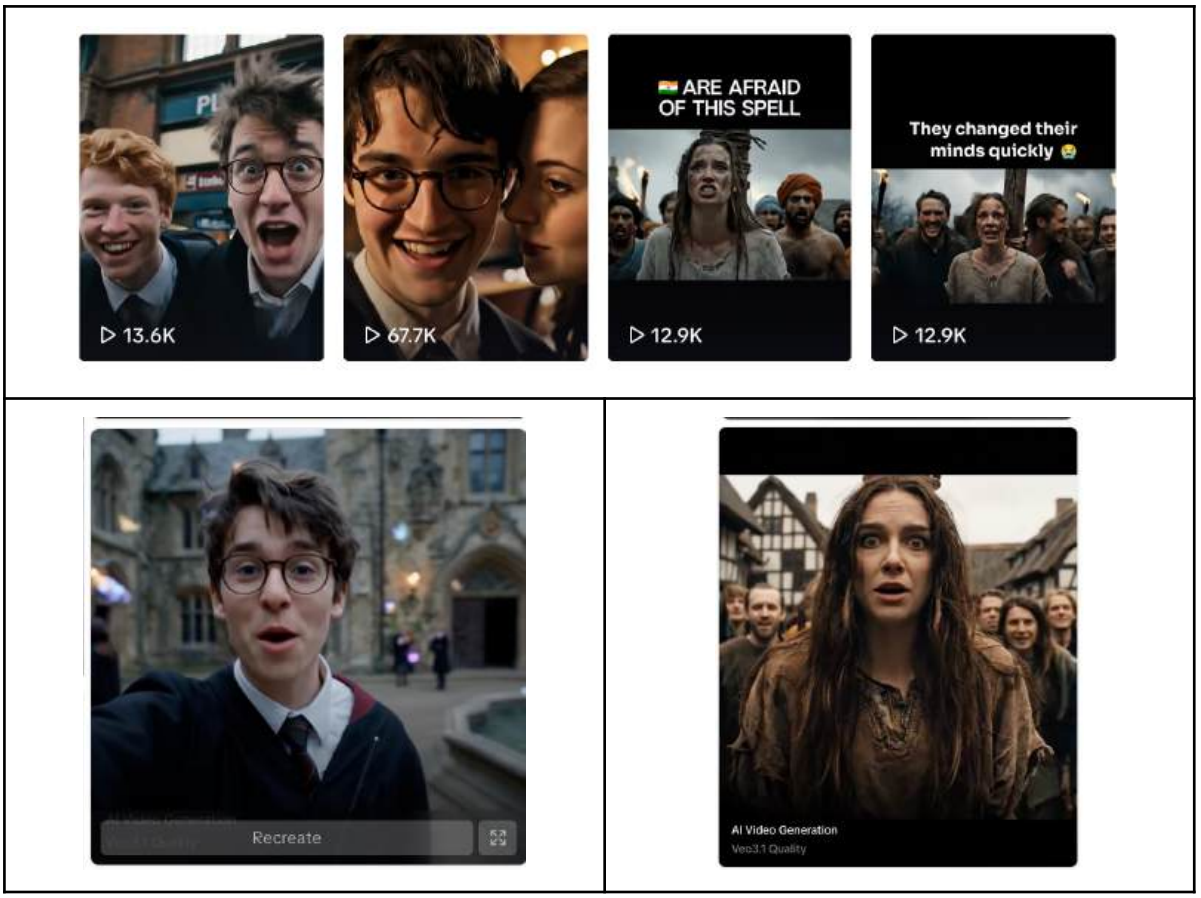


Figure 29. A screenshot showing a series of four posts uploaded on TikTok by one of the AAAs in our dataset, as seen on the account's feed (top), versus two videos enriched with readymade prompts as seen on the "Example Videos" page of Yapper.so (bottom).

Looking at Yapper.so's apparent customers, the usernames included in the three customer reviews (Figure 26, right) appear to belong to existing social media accounts. @ontheway.ai is an AAA, active on Instagram with over 123K followers as of October 2025, which promotes another virality pipeline called flashloop.app in its profile bio. This AAA also promotes a gambling website across several of its recent posts. @funnyfromai is an AAA on Instagram as well, with over 398K followers, promoting Yapper.so in its profile bio. Similarly, @jamesthegaijin is an AAA on Instagram with over 5K followers as of October 2025, and its profile bio links to both Yapper.so and its Etsy merchandise store.

These "happy customers" of Yapper.so also exemplify some of the monetization attempts we observed across our sampled AAA dataset. These monetization attempts included promoting virality pipelines (AI tools, apps, websites), actively seeking sponsorships (by promoting services or products), and selling account-related AI-made merchandise, as discussed in detail in the section "Monetization."

AAAs and platform governance

Illegal and/or harmful AIGC Content

TikTok identifies AI-generated Content (AIGC) as a cross-cutting systemic risk factor alongside recommender systems in its 2024 Systemic Risk Assessment report mandated under the Digital Services Act (DSA). Acknowledging as both an emerging risk area as well as a multiplier for other systemic risks, the company reports to have updated their policies as follows:

- *Requiring users to disclose content that is either completely generated or significantly edited by AI and contains realistic-appearing scenes or people;*
- *Prohibiting AIGC that depicts realistic-appearing people under the age of 18;*
- *Prohibiting AIGC that depicts the likeness of adult private figures without their permission;*
- *Prohibiting misleading AIGC that appears to come from an authoritative source;*
- *Prohibiting misleading AIGC that depicts a crisis event; and*
- *Prohibiting misleading AIGC that depicts a public figure who is being degraded or harassed,*
- *engaging in criminal or anti-social behaviour, or being politically endorsed or condemned.*

TikTok also reports having updated its Community Guidelines to classify hyper-realistic AIGC content depicting minors as Child Sexual Abuse Material (CSAM) which is illegal content.

As we demonstrated in this investigation, several AAAs specialize in posting AI-generated content explicitly prohibited by the platform, such as: “Realistic-appearing people under the age of 18”; “Misleading AIGC or edited media that falsely shows: 1) Content made to seem as if it comes from an authoritative source, such as a reputable news organization; 2) Misleading AIGC or edited media that falsely shows: A crisis event, such as a conflict or natural disaster.” We observed examples of such content circulating on the platform for weeks or even months, reaching thousands of users, before becoming inaccessible, likely due to the platform’s moderation (as discussed in “Removal, disappearance, deletion” section).

Deceptive Behavior & Fake Engagement

If a meme or TikTok format gets viral, we observed new AAAs being created to exploit the trend. This is true for both trivial and harmful content, e.g., racist or sexist memetic formats. Given the quantity and speed at which AAAs upload content on the platform, some AI content goes viral, reaching hundreds of thousands of users. When this happens, we observed that such content can remain online for weeks and even months. TikTok appears to exercise little moderation, whether in adding an AI label or removing harmful content, though it is not possible to determine whether this is due to a lack of reporting or a lack of diligence on TikTok's part.

TikTok reports to disrupt the inauthentic use of its platform, including manipulation of its recommender systems to boost fake engagement and seek financial benefit. AAA activity closely resembles what TikTok defines as "spam" indicators, such as "using automation to run many accounts or send repetitive content" and "posting a large amount of irrelevant material" in order to attain virality and obtain monetization.

Labeling of AI-Generated Content

TikTok's current labelling requirements do not enforce systematic labeling of AI content. Its automated labels cover only a small share of videos, and C2PA (the Content Credentials announced as an industry standard by The Coalition for Content Provenance and Authenticity) isn't implemented. Comments often highlight confusion: many users question whether the content is real; others call out "AI slop" yet drive engagement, which boosts the content's reach. AAAs flourish because the system rewards volume and shock, while labeling, especially automated labeling, remains weak and inconsistent. Until platforms enforce disclosure, implement provenance standards, and fix the payout and ranking incentives that fuel virality, this pattern will keep scaling upward.

Enforcement Gaps and Recommendations for TikTok

In its Risk Assessment report, TikTok reports the number of AI-generated content disclosed by users and detected by its automated detection tools since its implementation. However, the EU-wide total number of user-labeled content combined with AIGC detected by TikTok does not enable assessing the effectiveness of these risk mitigation measures. While our findings suggest enforcement gaps and

shortcomings, in its next risk assessment report, TikTok should report the percentage of content that is detected and disclosed by users, broken down by topic (politics, health, finance etc.).

Voluntary self-disclosure of AIGC by users should not be the primary measure through which AIGC is identified. TikTok should reinforce this measure by demoting this type of content and its deceptive proliferation. Additionally, tagging, labeling, and requiring disclosure of automated accounts could be a helpful indicator for users to better understand if they are interacting with AIGC.

Enabling users to reduce their exposure to AIGC on TikTok is a welcome development. However, given the structural and non-negligible amount of failure to identify such content, we remain skeptical regarding the success of this feature. Furthermore, users also should be able to completely remove their exposure to AIGC if they wish.

Recommendations: AAAs exploitation in the broader social media landscape

Dominant discourses around AI content on social media often focus on misinformation and deepfakes, which, while critical, are not the only harmful uses of AI-generated content. With AAAs, AI-generated and user-generated content are mixing at scale. Generative AI models that produce content without proper watermarking or their default moderation on platforms reveal concerning shortcomings with regard to Article 35 (k) of the DSA that requires ensuring that users are clearly informed about AI-generated content that resembles existing persons, places, and other entities.

Thanks to AI tools, AAAs can mass-test prompts, thumbnails, and hooks, quickly converging on potentially viral content types. 'Readymade' AI tools contribute to a weaponization of AI content to exploit platforms' virality loopholes and incentivize users to game recommendation systems, even if such promises are rather unlikely to materialize. Yet social media platforms could also be considered to partially contribute to this phenomenon by optimizing recommendation systems' attention-grabbing and preserving flawed payout mechanics, and by fueling the unlikely promise of viral content's success and monetization.

In order to properly assess and mitigate the downsides of this phenomenon, we encourage recognizing AIGC as a distinct risk vector in all VLOPs and VLOSEs Risk Assessments. Examining the interplay between recommender systems and AIGC is vital to address these risks and create friction to prevent snowball effects fueled by virality and mass posting. Audits should include assessing how easily AAAs can game the recommender systems as well as the scale of AI-driven spam, scams and inauthentic engagement. Ignoring these would undermine the platform's integrity and users' ability to have authentic interactions and hinder its value for users in the longer term.

Regarding the labeling of AIGC, the European Commission has started working on a Code of Practice on marking and labeling AI-generated content and the first draft of the Code is expected to be published in December 2025. The Code of Practice aims to support the fulfillment of transparency requirements under Article 50 of the AI Act. However, it should be noted that the Code of Practice is a voluntary, co-regulatory effort, and it will be applicable starting from August 2026, which gives a significant time window for the industry to align and implement common marking and labeling standards.

TikTok advertises itself as a platform that "brings together communities to unleash their creativity and share authentic stories" (quoted from tiktok.com/about). However, the blurring line between authentic human and synthetic AI-generated content on the platform is signaling a new turn towards more AI-generated content on users' feeds. Above all, it seems that TikTok and many other social media platforms are encouraging content creators to post AI-generated content by failing to regulate it and, instead, even recommending it to users on a large scale. AAAs demonstrate that AI content is increasingly integrated into platforms and AI tools, adding to a larger virality ecosystem that facilitates and encourages sensational, potentially harmful, and explicit content. Platforms should go beyond weak or optional "AI content" labels, and a coherent and independent regulatory framework has to be deployed to ensure systematic recognition of risks and pitfalls of the challenging intersection of AI tools and online platforms.

Appendix

Comment analysis keywords, AI labels

Keywords and phrases **suspecting AI use**: ai, a i, a.i., ai, ai is funny , artificial intelligence, i actually can't tell if it is ai, is it ai, it's ai, it's fun ai , people believing this is scary, scary how real it looks, the most real ai, these fakes should be banned, this is ai right, this must be ai, true or fake, veo 3, you don't believe this

Keywords and **phrases showing confusion over authenticity**: and people believe this , are they telling is the truth , don't know what's real anymore, fake, fake ..look it up, fake news, has anyone else tried to find this case?, have any one found this story on google?, have not seen this story on any major news outlet, how do you make these videos??, i don't know if this is real, i dont trust anymore on news now .. , i don't believe it, i don't believe this, if this is real, if this were true, is everything fake, is it fake, is it real, is there a legit news article?, is this a real story?, is this a scene from a movie?, is this even a real news report?, is this even real, is this really true, is this some sort of a joke??!!, is this true, is this true, is this true or fake news, nah, hard to believe!, name of the movie, no where else in the internet, nothing on the news..., please!!!!!! this has to be a joke!!!, real or fake, there's just no way, this a movie?, this appears only on tik tok, this can not be true, this can't be true, this cannot be right, this can't be real, this can't be right, this doesnt look real, this is legit, this is not true, was that true?, what movie is this, why wasn't this on the news

Keywords and phrases **labeling AI use in the descriptions**: aigenerated, ai, artificialintelligence, kunstlicheintelligenz, ki, inteligenciaartificial, ia, Aistory, AIGirl, AISchoolgirl, TikTokAI, AIVideo, Allady, elegantAI, elegantAI, ailookbook, AI콘텐츠, AIcheerleader, AI치어리더, ai영상, ai**コンテンツ**, ai**映像**, ai**チアリーダー**, ai**内容**, aihistory, creativeai, ai미녀, aicat, aicats, aidog