

Behavioral Signals of Echo Chamber Risk Around Misleading AI-Generated Videos

Yichang Gao

yichang.gao@mit.edu.au

Shandong Normal University

Paul Harrigan

The University of Western Australia

Yuqi Wang

Johnson C. Smith University

Fengming Liu

Shandong Normal University

Article

Keywords: artificial intelligence, misleading AI-generated video, dynamic echo chamber, algorithmic exposure, platform governance, predictive modeling

Posted Date: August 5th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10367677/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

AI-generated videos are becoming an information-society concern because they can transform fabricated evidence, false authority cues, and synthetic realism into persuasive forms of communication. Rather than treating misleading synthetic video as a technical detection problem, this study examines how such content becomes embedded in algorithmically curated short-video environments and how users' participation in echo chamber (EC) communities changes over time. Drawing on data from TikTok, Instagram, Douyin, and Kuaishou, we collected 4,580 short videos during September 2025, identified 145 misleading AI-generated videos through AIGVDet screening and expert verification, and analyzed 1,039,564 comments and repost interactions. Daily EC communities were detected using Leiden clustering and an embedding-based Echo Chamber Score, allowing user EC status to be dynamically re-estimated across a 30-day observation window rather than treated as a fixed user attribute. Fourteen user-level behavioral, activity, and proxy cognitive indicators were updated daily and used to predict dynamic EC membership with ConvLSTM, ConvGRU, and CNN-LSTM architectures. The findings show that rationality-related and curiosity-related proxy indicators, together with posting frequency, were the most informative signals associated with future EC membership, while demographic indicators were less predictive. These results should be understood as predictive associations, not as evidence of causal mechanisms or stable psychological vulnerability. ConvLSTM produced the strongest predictive performance among the three models. The findings show how misleading synthetic media, platform recommendation logics, and user behavior intersect to produce dynamic patterns of enclosed information exposure. The study outlines privacy-conscious and autonomy-preserving early-warning approaches that rely on behavioral signals rather than demographic profiling.

1. Introduction

AI-generated video is changing one of our common senses: seeing is no longer believing. Generative artificial intelligence (Gen-AI) has been unsettling the trust for videos to truthfully record real events. When Gen-AI produces realistic and emotionally compelling videos with convincing authority cues, misleading information can become more credible than conventional textual misinformation [1]. Deepfakes and other AI-generated synthetic media are often difficult for people to distinguish [2]. For communication research and platform governance, the problem is not only detection and labeling, but systematically understand and monitor how misleading AI-generated videos circulate through platform environments that repeatedly organize attention, credibility, and participation.

When misleading AI-generated videos circulate on short-video platforms, recommendation systems may shape users' repeated exposure by ranking content according to prior engagement and inferred relevance. Recent studies of video and social media recommendation systems show that algorithmic curation can meaningfully affect users' information exposure, including exposure to politically aligned, low-credibility, or topically narrow content [3].

Echo chambers (ECs) offer a theoretical lens to understand this platform-mediated narrowing of exposure. Extant research commonly depicts ECs as relatively stable information environments in which

EC members are exposed to more congruent viewpoints than counterarguments [4]. However, ECs are not necessarily stable containers into which users permanently fall. In a recommendation-shaped environment, EC membership can change with user behavior, content availability, and ranking signals. This study therefore conceptualizes EC status as time-varying participation in relatively enclosed information communities around the studied misleading AI-generated videos.

This dynamic framing addresses an important gap in existing research. Prior studies have examined synthetic media detection, human ability to identify deepfakes, misinformation correction, and the structure of ECs on social media [2, 5]. Relatively less efforts has been devoted to predicting EC membership around misleading AI-generated videos from observable user signals. These signals may include behavioral activity, platform engagement, exposure history, and proxy cognitive indicators.

Against this backdrop, this study examines behavioral signals of EC risk around misleading AI-generated videos on short-video platforms. Drawing on data from TikTok, Instagram, Douyin, and Kuaishou, the study identifies misleading AI-generated videos through automated screening and expert verification, constructs daily user-level indicators, detects dynamic EC communities, and predicts future EC membership using spatiotemporal deep-learning models. The study makes three contributions. First, it shifts attention from detecting misleading AI-generated videos as isolated technical objects to examining how users interact with them over time. Second, it treats EC membership as a dynamic daily status rather than a stable user characteristic. Third, it evaluates whether behavioral, activity-based, and proxy cognitive indicators can serve as predictive signals for future EC risk, while avoiding causal claims about psychological vulnerability. The study is guided by two research questions:

RQ1: Which user-level behavioral and proxy cognitive indicators are associated with dynamic EC membership around misleading AI-generated videos?

RQ2: Can time-varying user features predict future EC membership across short-video platforms?

2. Literature Review

2.1 Misleading AI-generated video and synthetic media misinformation

Compared with other media forms of misinformation such as text or audio, AI-generated videos could be more disguising as they combine vivid realism, motion continuity, voice, narrative, and emotional cues to make fabricated contents appear like real event recordings. They can present false, misattributed, decontextualized, or fabricated claims as credible audiovisual evidence in perception, thereby complicate and distort viewers' judgments of authenticity [2]. In deepfake detection tasks, viewers often struggle to reliably distinguish authentic media from manipulated content [2, 6].

As a result, misleading AI-generated videos may pose substantial risks to audiences, platforms, and the society [6]. Some communication frameworks such as AIGVDet may be able to label whether a video is

synthetic [7], but the true consequences of misleading videos depend on how platforms recommend and distribute them, how users view and engage with them, and how repeated exposure shapes users' judgments of credibility. This study treats misleading content detection as the first step and further examines the dissemination process and patterns of user engagement.

2.2 Short-video platforms, algorithmic curation, and repeated exposure

Short-video platforms such as Instagram, TikTok, Douyin, and Kuaishou rely heavily on recommendation algorithms to categorize users' interests, push personalized contents, keep users scrolling for longer screen time, and nurture their attachment and loyalty [8–10]. Unlike traditional platforms such as Google where users actively search for information, short-video platforms cater to passive attention with continuous user analytics from behavioral traces such as viewing duration, likes, comments, reposts, follows, and post contents to infer users' characteristics and preferences. Algorithmic personalization not only shapes what users see, but also influences downstream outcomes such as user engagement, social connection, and risk awareness [11].

It is concerning that when misleading AI-generated videos attract some users' initial attention or engagement, these signals will direct more similar contents to these individuals [9]. After repeated exposure and reinforcement, users may become more receptive to the misinformation over time, even when they later encounter accurate information or opposing views [1, 12]. This makes short video platforms crucial empirical contexts for studying algorithmic exposure to misleading synthetic media.

2.3 Echo Chambers as dynamic platform-mediated information structures

ECs have commonly been studied as manifestations of like-minded exposure, ideological segregation, and selective information consumption in social media environments [13]. However, ECs should not be treated as fixed communities. In algorithmically curated environments, recommender systems continuously translate users' behavioral traces into changing exposure opportunities, producing feedback loops between user activity, content ranking, and future information diets [3]. This platform logic is especially visible on short-video platforms, where social connection is increasingly organized around algorithmically inferred interest clusters rather than stable interpersonal networks [8].

The present study therefore uses a dynamic EC definition. EC communities are detected daily using Leiden clustering and an Echo Chamber Score (ECS), and user EC status is re-estimated across the 30-day window [14, 15]. This means the outcome is dynamic EC membership or time-varying EC status, not a permanent user state.

3. Methods

3.1 Research pipeline

This study uses an observational, predictive design to examine how users move into and out of EC communities around misleading AI-generated videos on short-video platforms. We first identified misleading AI-generated videos on TikTok, Instagram, Douyin, and Kuaishou through automated screening and expert verification. We then collected public comments, reposts, and related user activity around these videos and organized the data into daily platform-specific interaction records. On this basis, we constructed daily interaction networks and re-estimated each user's EC status at the user-day level, rather than treating EC membership as a stable attribute. User-level demographic, behavioral, activity-based, and proxy cognitive indicators were then calculated as time-ordered predictors of subsequent EC membership. To limit temporal and train-test leakage, dynamic features were constructed only from information available before the outcome being predicted, and feature selection was performed using the training data only. The analysis therefore supports predictive rather than causal claims: it identifies observable signals associated with later EC membership, but does not claim that these signals cause EC participation or reveal stable psychological vulnerability.

3.2 Data

The initial corpus consisted of 4,580 short videos collected from TikTok, Instagram, Douyin, and Kuaishou between September 1 and September 30, 2025. New research accounts were created on each platform, and publicly accessible videos recommended by the platforms were collected daily during the observation period. Audio was transcribed from these videos, so that spoken language could be assessed together with visual content [16]. The analysis was limited to Chinese- and English-language content. AIGVDet was then applied to the full corpus to identify candidate AI-generated videos, using a video-level threshold of $P \geq 0.5$ [7]. Because the study required sufficient interaction data for subsequent comment and repost analysis, only flagged videos with more than 1,000 comments were retained for expert verification.

Three postdoctoral experts in media content analysis and computational communication independently reviewed the retained candidate videos. The verification procedure was informed by recent fine-grained annotation protocols for AI-generated video assessment, with particular attention to visible generative artifacts such as spatial distortions, object inconsistencies, unnatural motions, temporal discontinuities, and violations of physical or commonsense logic [17]. Experts were also asked to distinguish likely AI-generated artifacts from low resolution, compression noise, motion blur, and platform-related degradation. Each candidate video was rated for its AI-generation status and misleadingness, and disagreements among experts were resolved through adjudication.

A video was included only if it was verified as both AI-generated and misleading. Expert coders assigned two categorical labels: A for AI-generation status and M for misleadingness, each coded as Yes, No, or Uncertain. A video was labeled $A = \text{Yes}$ when the primary visual subject was synthetically generated, a recognizable identity was reconstructed or replaced, synthetic or cloned speech carried key content, or generative editing altered factual elements of the video [2, 18]. A video was labeled $M = \text{Yes}$ only when it contained an externally verifiable factual proposition that was false, misattributed, missing necessary context, or supported by fabricated evidence. Positive misleadingness labels required corroboration from

at least two independent credible sources, following established verification practices in journalism and fact-checking [19]. Inter-coder reliability was calculated before adjudication using Fleiss' κ . Agreement was high for both the AI-generation label, $\kappa=0.87$, and the misleadingness label, $\kappa=0.81$. The final sample included 145 confirmed misleading AI-generated videos: 41 from Douyin, 37 from Kuaishou, 32 from TikTok, and 35 from Instagram.

3.3 User features extraction

3.3.1 User features

To examine which users are more susceptible to ECs driven by misleading AI-generated video content, we constructed a set of 14 user features that capture key behavioral patterns, content preferences, and cognitive tendencies. The user feature extraction followed two principles. First, all indicators had to be observable from publicly accessible profile or behaviors. Second, when the underlying behavior changed over time, the corresponding features were recalculated at the user-day level that form dynamic EC membership. Table 1 summarizes the 14 user-level features covering demographics, account visibility, platform activity, content exposure, and personality proxies related to judgment, rationality, curiosity, and self-disclosure.

Table 1
Summary of the features

Feature	Feature Name	Feature Description
Demographics		
Age	Age	Age information self-disclosed by the user in their profile, collected based on the principle of voluntary disclosure.
GU	Gender	Gender information self-disclosed by the user in their profile, collected based on the principle of voluntary disclosure.
Account and behavioral measures		
FirstVidTime	First video posting time	Time when the user posted their first short video; coded as 0 if the user has not posted any short video.
NFollower	Number of followers	Number of users who follow this account, reflecting its visibility and potential influence.
NFollowing	Number of followees	Number of other accounts followed by this user, reflecting the diversity of their information sources.
NPost	Total number of posts	Cumulative number of posts published by the user, used to measure overall activity level.
PostFreq	Video posting frequency	Average number of short videos posted by the user per day within a given observation window.
NT	Number of interest domains	Number of distinct domains the user has engaged with or continuously followed within one year.
EAI	Exposure to AI content	Frequency with which the user views or interacts with content related to AI.
AveSVPop	Average short video popularity	Average popularity of the user's short videos, for example aggregating likes, comments, and shares.
Personality proxies		
UR	Rationality	Rationality score constructed from the sentiment and controversy of the user's historical posts.
UC	Curiosity	Curiosity score computed from the semantic dissimilarity between the current post and the user's historical posts, with historical posts weighted by an exponential time-decay function.
Judg	Judgment ability	Judgment ability estimated from the user's sentiment toward short videos related to news or public issues; sentiment closer to neutral, reasoned, and non-extreme indicates higher judgment ability.
WSD	Willingness to Self-Disclose	Overall tendency of the user to reveal personal information on social media.

Following the principle of voluntary disclosure [20], demographic and profile variables were collected only when users had published the information. Age and gender were not inferred from names, profile images,

or other indirect cues.

Account and activity measures included first video posting time, number of followers, number of followees, total number of posts, video posting frequency, number of interest domains, exposure to AI-related content, and average short-video popularity. These variables captured differences in visibility, activity, and breadth of information engagement across the observation window.

Personality proxies describe how users encountered and responded to AI-, technology-, policy-, or news-related videos. The judgment-related proxy (*Judg*) captured whether users' responses to AI- or news-related content tended to remain neutral rather than emotionally extreme [20, 21]. The rationality-related proxy (*UR*) was adapted from Huang et al. [22] and Xie et al. [20]. It draws on sentiment intensity and controversy in historical replies as indicators of emotionally expressive or polarizing engagement. The curiosity-related proxy (*UC*) drew on novelty-seeking theory [23] and was adapted from Xu et al. [24], measuring the semantic distance between a target video and a user's recent interaction history with time-decay weighting. Willingness to self-disclose (*WSD*) was measured from the visibility of users' profile modules, including videos, followers, followees, and likes. Drawing on prior work on self-disclosure in interpersonal and social media contexts [13, 25, 26], users were assigned to four disclosure levels according to the privacy settings available on each platform.

These indicators should be understood as observable behavioral and communicative proxies rather than direct measures of user traits. Feature extraction was conducted after the target video-centered interaction networks and daily EC labels had been established. Full coding rules, formulas, normalization procedures, and preprocessing decisions are reported in **Supplementary Materials S1**.

3.3.2 Feature Selection

Feature selection was used to reduce redundancy and retain the most informative signals before prediction. Because the 14 candidate features varied in relevance and could overlap with one another, retaining the full set risked adding noise without improving model performance. We applied a bootstrap-based permutation importance procedure, following established feature evaluation approaches in machine learning [27–29].

LightGBM was used as the base classifier for this step because it performs efficiently with mixed tabular features and is stable under repeated bootstrap sampling. In each bootstrap iteration, out-of-bag prediction error was compared before and after each feature was permuted. Features that produced larger increases in prediction error were treated as more informative for predicting future EC membership. Importance scores were normalized, ranked, and filtered using an elbow-point threshold. The reduced feature set was then used as input for the ConvLSTM, CNN-LSTM, and ConvGRU models. Full mathematical notation and implementation details are provided in **Supplementary Materials S2**.

3.4 EC and EC members

We define an EC as a subset of users in a social network who share largely homogeneous ideological stance and interact predominantly within their own group [30]. Hallmark properties include high internal

cohesion, low external connectedness, content homogeneity, and attenuated outward influence. To identify such structures, we first recovered the meso-level organization of each daily interaction network using the Leiden community detection algorithm. Leiden was selected over alternatives such as Louvain because it guarantees well-connected communities, improves partition quality, and scales effectively to large social networks [15]. These properties make it suitable for identifying candidate ECs, which are expected to be densely connected and internally coherent.

After detecting candidate communities, we used the Echo Chamber Score (ECS) framework to assess whether each community exhibited EC-like characteristics [14]. ECS quantifies the strength of the EC effect in an embedding space by jointly measuring intra-community cohesion and inter-community separation. Specifically, it evaluates whether users within the same community are more similar to one another than they are to users in nearby external communities. Higher community-level ECS values indicate stronger within-community similarity and clearer separation from other communities. The graph-level ECS was then calculated by averaging community-level scores across all detected communities.

To avoid relying on an arbitrary fixed cutoff, we used a data-driven thresholding procedure based on Gaussian Mixture Modeling (GMM). The empirical distribution of community-level ECS values was modeled as a two-component mixture representing open, loosely connected communities and closed, self-reinforcing ECs. The component with the higher mean was treated as the EC subpopulation, and the classification threshold was defined at the point where the posterior probabilities of the two components were equal. This equal-posterior rule provides a statistically grounded and reproducible decision boundary under equal misclassification costs [31]. In this dataset, the learned cut points clustered within a narrow range. Although the ECS framework can support a three-level distinction between strong ECs, weak ECs, and non-ECs, the present study used a binary classification because the analytical objective was to determine whether users were inside or outside ECs on a given day. Communities with ECS values greater than 0.50 were coded as ECs, and all other communities were coded as non-ECs. Users located in EC-classified communities on a given day were coded as EC members for that daily observation. Full details on the Leiden, ECS, and GMM procedures are reported in **Supplementary Materials S3**.

3.5 Prediction models

To predict dynamic daily EC membership around misleading AI-generated videos, we compared three spatiotemporal deep learning models: Convolutional Long Short-Term Memory (ConvLSTM), Convolutional Neural Network Long Short-Term Memory (CNN-LSTM), and Convolutional Gated Recurrent Unit (ConvGRU), across the four major social media platforms. ConvLSTM extends LSTM by replacing fully connected state transitions with convolutional operations, allowing the model to capture long-range temporal dependencies while preserving local structural patterns in the user-platform space [32]. CNN-LSTM combines convolutional feature extraction with LSTM-based sequence modeling [33], whereas ConvGRU uses a simpler gated structure that can reduce parameter complexity while retaining temporal and spatial sensitivity [34]. To ensure comparability, all three models were trained and evaluated using the same input representation, prediction task, and evaluation protocol.

The input data were organized as a four-dimensional tensor, $t_e \times n_p \times n_b \times c$, integrating temporal variation, user-level features, sampled users, and platform information. The observation window was set to 30 days because prior research finds misinformation spreads rapidly and extensively across social networks [35–37]. Such a window captures short- to medium-term changes in EC-related exposure patterns from formation, reinforcement, to decay. Users do not necessarily remain EC members throughout the observation period. This design allowed the models to capture users’ movement into and out of EC membership. It avoided treating EC status as a fixed user attribute.

All models used a comparable high-level architecture consisting of an input layer, three stacked spatiotemporal recurrent convolutional layers, and an output layer. In this design, convolutional operations were applied over a structured user-platform space rather than a physical spatial grid, enabling the models to learn local dependency patterns and interaction regularities across users and platforms while preserving temporal continuity. To support sequential prediction under a binary user-level EC classification, we adopted a sliding-window design. Preliminary analyses tested window lengths from 1 to 7 days, $s \in \{1, 2, 3, 4, 5, 6, 7\}$, and showed the strongest predictive performance when the window length was set to 1 day; this setting was therefore used for all subsequent analyses. The four platforms were incorporated jointly through the platform dimension of the input tensor, rather than being trained as four independent platform-specific models.

The dataset was divided into training and testing sets using an 80%–20% user-level split, ensuring that all sequences from the same user remained in the same partition [38]. Results were reported across five random seeds. Models were trained with the Adam optimizer, with an initial learning rate of $\eta = 5.0 \times 10^{-5}$, because Adam is suitable for high-dimensional spatiotemporal learning due to its adaptive learning-rate behavior [39]. A learning-rate decay factor of 0.5 was applied to support stable convergence [40]. The batch size was set to 3, the number of training epochs was fixed at 100, and 16-bit mixed-precision computation was used to balance accuracy, computational efficiency, and memory constraints [41–43]. Because the target was binary daily EC membership, each model used a sigmoid output layer and class-weighted binary cross-entropy loss, with class weights estimated from the training set only. Predicted probabilities were converted into binary labels using a validation-set threshold optimized for F1 score. A callback mechanism monitored validation loss and reduced the learning rate when no improvement was observed for 10 consecutive epochs. Full details about the machine learning models, tensor notation, and testing are provided in **Supplementary Materials S4**.

3.6 Model evaluation

Because the prediction target was binary daily EC membership, model performance was evaluated as a binary classification task on the held-out test set. Each model generated a predicted probability of EC membership, $\hat{p}_{u,p,t+h}$, for each user-platform-time observation. In the main analysis, these probabilities were converted into binary labels using the threshold selected on the validation set to maximize the F1 score. As a sensitivity check, we also applied the conventional 0.50 threshold, under which observations with predicted probabilities of 0.50 or higher were classified as EC-positive and those below 0.50 were classified as non-EC. Performance was assessed using accuracy, balanced accuracy, precision, recall,

specificity, F1 score, ROC-AUC, PR-AUC, and the confusion matrix. Accuracy captured overall correctness, while balanced accuracy was included to account for potential class imbalance between EC-positive and EC-negative observations. Precision, recall, specificity, and F1 score were used to evaluate classification quality for EC and non-EC cases, and ROC-AUC and PR-AUC assessed threshold-independent discrimination, with PR-AUC treated as especially informative for the EC-positive class under class imbalance. Confusion matrices were also reported to show true positives, false positives, true negatives, and false negatives, which is important for evaluating the practical utility of EC early-warning prediction because false positives may lead to unnecessary intervention, whereas false negatives may fail to identify users moving into EC membership. Detailed metric definitions, threshold-selection procedures, and sensitivity checks are provided in **Supplementary Materials S5**.

4. Results

4.1 Descriptive statistics

Table 2 summarizes the descriptive statistics of the data. We systematically collected 4,580 short videos from four major platforms: Douyin, Kuaishou, TikTok, and Instagram. Using the AIGVDet framework developed by Bai et al. [7], we initially identified 236 videos as potentially AI-generated. After expert verification of both AI-generation status and misleadingness, 145 videos were retained as confirmed misleading AI-generated short videos. Across platforms, the final sample included 41 videos from Douyin, 37 from Kuaishou, 32 from TikTok, and 35 from Instagram. Some videos showed cross-platform similarity, but similar videos appearing on different platforms were retained because they represented distinct platform-specific interaction contexts, as the users who commented, reposted, and engaged with it differed. Each video-platform instance was therefore analyzed separately to preserve platform-specific patterns of user participation and EC formation.

During preprocessing, only Chinese and English content was retained. The final dataset included 1,039,564 comments linked to 145 confirmed misleading AI-generated videos. Usernames and profile identifiers were anonymized during data collection and replaced with unique numerical IDs.

We used a 30-day observation window because misleading AI-generated videos may be removed or downranked soon after posting, which can cause data loss over longer sampling periods [44]. This window also captured the short-term dynamics of short-video engagement, as comments and discussions around misleading content often accumulate rapidly within days or weeks [45, 46]. To support daily feature construction, users were included only if they had commented on at least 15 days within the 30-day window. This criterion ensured sufficient longitudinal behavioral records while maintaining cross-platform coverage and computational feasibility. The final modelling dataset included 4,826 users: 1,326 from Douyin, 1,241 from Kuaishou, 1,176 from TikTok, and 1,083 from Instagram.

After the final video sample was established, EC communities were identified using Leiden clustering and the Echo Chamber Score (ECS), with EC membership coded as a binary daily outcome. EC-positive

observations represented user-day cases classified as EC membership, whereas EC-negative observations represented user-day cases classified as non-EC membership.

Each day within the 30-day observation period was treated as an independent temporal slice. For each retained user, the fourteen user-level features were updated daily using a rolling extraction procedure. In parallel, the EC identification algorithm was re-run for each daily interaction network to capture changes in community structure and user composition. This design allowed EC membership to be treated as a dynamic daily status rather than a fixed user attribute. Consequently, a user could be classified as EC-positive on some days and EC-negative on other days, depending on their daily interaction context and the structure of the corresponding platform network.

The final prediction dataset contained 144,780 user-day observations. Among these, 49,698 observations were EC-positive and 95,082 were EC-negative, corresponding to an EC-positive rate of 34.3%. This distribution indicates a moderately imbalanced binary classification task. Therefore, as described in Section 3.6, model performance was evaluated using not only accuracy but also balanced accuracy, precision, recall, specificity, F1 score, ROC-AUC, PR-AUC, and confusion matrices. This evaluation strategy was necessary to ensure that model performance reflected predictive utility for both EC-positive and EC-negative cases rather than being driven primarily by the majority class.

Table 2
Summary of data collection, filtering, and modeling observations

Data construction step	Number	Notes
Initially collected short videos	4,580	TikTok, Instagram, Douyin, and Kuaishou
Videos initially flagged by AIGVDet	236	Candidate AI-generated videos
Confirmed misleading AI-generated videos	145	Expert-verified AI-generation and misleadingness
Douyin confirmed videos	41	Platform-specific video instances
Kuaishou confirmed videos	37	Platform-specific video instances
TikTok confirmed videos	32	Platform-specific video instances
Instagram confirmed videos	35	Platform-specific video instances
Comments	1,039,564	Chinese and English content only
Final retained users	4,826	Users with comments on at least 15 days
Douyin retained users	1,326	After eligibility filtering
Kuaishou retained users	1,241	After eligibility filtering
TikTok retained users	1,176	After eligibility filtering
Instagram retained users	1,083	After eligibility filtering
Final user-day observations	144,780	4,826 retained users $\times$ 30 daily temporal slices
EC-positive observations	49,698	User-day cases classified as EC membership
EC-negative observations	95,082	User-day cases classified as non-EC membership
EC-positive rate	34.3%	Indicates a moderately imbalanced binary classification task

4.2 User Feature Importance

To assess and mitigate potential multicollinearity among the selected user features, we applied a two-step diagnostic procedure combining pairwise correlation analysis and variance inflation factor (VIF) evaluation [47]. First, we examined the correlation structure among all demographic, behavioral, and proxy cognitive variables to identify underlying dependency patterns that might influence downstream modeling. Second, we quantified multicollinearity using VIF values to determine whether any feature exhibited redundancy or excessive shared variance that could bias parameter estimation or reduce model interpretability. The results of these analyses are summarized in Fig. 1.

Figure 1a shows the pairwise correlation structure among the 14 features. The lower triangular heatmap displays Pearson correlation coefficients, with color intensity reflecting both magnitude and direction. Several meaningful patterns are observed: activity related features (e.g., NPost, PostFreq, NFollower)

exhibit moderate positive correlations with one another, whereas age, self-disclosure indicators, and personality constructs User rationality (*UR*), Curiosity (*UC*), and Judgment ability (*Judg*) show weaker and more heterogeneous relationships. Notably, the highest absolute correlation between any pair of features is below 0.6. As widely documented in the literature, $|r| < 0.6$ indicates low inter feature correlation that does not significantly affect downstream regression models [48]; thus, multicollinearity arising from pairwise correlations is negligible in the current feature set. Figure 1b shows the Variance Inflation Factors (VIFs) for all features, visualized using a radial bar plot. The plot clearly illustrates the degree of multicollinearity in the feature set, with the commonly used threshold of $VIF > 5$ marked by the red circle [49]. As can be seen from Fig. 1, all features have VIFs well below 5, with the highest value being 1.60, indicating that overall multicollinearity is mild. Therefore, all 14 features will proceed to the feature selection phase.

Feature selection results are shown in Fig. 2. We first quantified the contribution of the 14 features, which capture behavioral patterns, content preferences, and cognitive tendencies, using the bootstrap based permutation importance framework described in Section 3.3.2. We set the number of bootstrap replications to $B = 30$ to balance the stability of the importance estimates and the computational cost, following the empirical choice in [27]. LightGBM was used as the base classifier. For each bootstrap sample, a model was trained and the permutation importance $Score_n$ of feature n was computed as the increase in out of bag (OOB) prediction error induced by randomly permuting that feature. The resulting importance scores were then normalized to $[0,1]$. Figure 2a presents the normalized importance scores. User rationality (*UR*), curiosity (*UC*), and short video posting frequency (*PostFreq*) obtain the highest scores, indicating that these behavioral and psychological features contribute most to predicting which users are more susceptible to ECs driven by misleading AI-generated video content. In contrast, gender (*GU*) and first video posting time (*FirstVidTime*) show much lower importance.

To systematically assess the impact of feature selection on model performance, we introduced a set of selection thresholds τ on the normalized importance scores. For each τ , we retained only features with $Score_n^{\text{Norm}} \geq \tau$ and retrained the EC susceptibility model. Figure 2b shows the corresponding values of the coefficient of determination (R^2). When $\tau = 0$, all features are retained and no selection is applied. As τ increases from 0 to approximately 0.4, R^2 shows an overall increasing trend, which suggests that removing redundant or weakly informative features improves predictive accuracy and stability. When τ exceeds 0.4, R^2 decreases, indicating that overly aggressive pruning removes predictive signal. Considering both predictive performance and parsimony, we set $\tau = 0.4$ and used the resulting reduced feature matrix H_{sel} (of 10 features) as input to the downstream spatiotemporal deep learning models (ConvLSTM, CNN LSTM, and ConvGRU).

4.3 Model performance comparison

Table 3 shows the performance of CNN-LSTM, ConvLSTM, and ConvGRU on the held-out test set across five random seeds. The reported metrics include accuracy, balanced accuracy, precision, recall, specificity, F1 score, ROC-AUC, and PR-AUC. Accuracy captures the overall proportion of correctly

classified user-day observations, whereas balanced accuracy accounts for the moderately imbalanced distribution between EC-positive and EC-negative cases. Precision, recall, specificity, and F1 score further indicate whether the models are practically useful for identifying EC-positive observations without producing excessive false positives or false negatives.

As shown in Table 3, ConvLSTM achieved the strongest overall classification performance across the main evaluation metrics, with the highest accuracy (0.866 ± 0.007), balanced accuracy (0.845 ± 0.009), F1 score (0.799 ± 0.011), ROC-AUC (0.925 ± 0.006), and PR-AUC (0.864 ± 0.008). ConvGRU also performed competitively, achieving an F1 score of 0.769 ± 0.012 and ROC-AUC of 0.904 ± 0.007 , but it did not exceed ConvLSTM on the main classification metrics. CNN-LSTM showed the weakest overall classification performance, with an F1 score of 0.679 ± 0.014 and ROC-AUC of 0.848 ± 0.010 . This suggests that separating convolutional feature extraction from temporal sequence modeling was less effective for capturing the dynamic user-platform patterns associated with daily EC membership.

Table 3
Model classification performance

Model	Accuracy	Balanced accuracy	Precision	Recall	Specificity	F1	ROC- AUC	PR- AUC
CNN- LSTM	0.787 ± 0.010	0.756 ± 0.012	0.702 ± 0.016	0.658 ± 0.018	0.854 ± 0.011	0.679 ± 0.014	0.848 ± 0.010	0.738 ± 0.013
ConvLSTM	0.866 ± 0.007	0.845 ± 0.009	0.819 ± 0.012	0.780 ± 0.014	0.910 ± 0.008	0.799 ± 0.011	0.925 ± 0.006	0.864 ± 0.008
ConvGRU	0.846 ± 0.008	0.822 ± 0.010	0.793 ± 0.013	0.746 ± 0.015	0.898 ± 0.009	0.769 ± 0.012	0.904 ± 0.007	0.834 ± 0.010

Note. Values are reported as mean $\pm$ standard deviation across five random seeds. These classification metrics were calculated using the F1-optimized validation threshold.

Table 4 reports the confusion matrices for the three models, calculated using the F1-optimized threshold selected on the validation set. This analysis assesses whether each model balances false positives and false negatives, which is critical for EC early-warning prediction. The results show that ConvLSTM achieved the best balance between identifying EC-positive cases and controlling false alarms. Compared with CNN-LSTM, it substantially reduced false negatives, indicating higher sensitivity to EC-positive observations. Compared with ConvGRU, it produced fewer false positives and false negatives, suggesting greater classification stability. These findings support the use of ConvLSTM as the main model for subsequent interpretation.

Table 4
Confusion matrices of the three models on the held-out test set.

Model	True negatives	False positives	False negatives	True positives
CNN-LSTM	16,243	2,777	3,396	6,534
ConvLSTM	17,308	1,712	2,185	7,745
ConvGRU	17,080	1,940	2,523	7,407

Note. The confusion matrices are based on the median-performing seed. The held-out test set contained approximately 28,950 user-day observations, corresponding to 965 users observed across 30 daily temporal slices. Among these observations, 9,930 were EC-positive and 19,020 were EC-negative.

4.4 Computational cost

To assess the computational efficiency of the three models, we compared their training and prediction times under the same input representation, selected feature set, temporal window, hardware environment, and training configuration described in Section 3.5. Prediction time refers to the total inference time on the held-out test set. Because the four platforms were jointly incorporated through the platform dimension of the input tensor, the reported training time refers to one full cross-platform training run for each model under a single random seed, rather than the sum of four separate platform-specific models. The runtime reported here excludes data collection, EC detection, feature construction, and feature-selection preprocessing.

The results are shown in Table 5. CNN-LSTM required the shortest training time, taking 1,416.84 seconds, followed by ConvLSTM with 1,537.85 seconds. ConvGRU required the longest training time, taking 2,956.54 seconds. A similar pattern was observed for prediction time. CNN-LSTM produced the fastest prediction, requiring 3.29 seconds, while ConvLSTM required 4.61 seconds. ConvGRU was substantially slower, requiring 11.83 seconds.

Table 5
Training and prediction time of the three models.

Model	Training time (seconds)	Prediction time (seconds)
CNN-LSTM	1,416.84	3.29
ConvLSTM	1,537.85	4.61
ConvGRU	2,956.54	11.83

The computational cost comparison shows that CNN-LSTM was the most efficient model, but this came with weaker classification performance. ConvGRU improved predictive performance relative to CNN-LSTM but required substantially longer training and prediction times. ConvLSTM offered the best overall trade-off in the present setting. Compared with CNN-LSTM, it increased training time by only approximately 8.5% while achieving a markedly higher F1 score. Compared with ConvGRU, it required

approximately 52.0% of ConvGRU's training time and 39.0% of its prediction time, while maintaining stronger overall classification performance.

5. Discussion

The societal risks posed by misleading AI-generated videos cannot be understood by examining content alone. On recommendation-driven short-video platforms, misleading synthetic videos are not isolated pieces of information, but components of evolving exposure systems shaped jointly by platform algorithms and user behavior. Everyday interactions such as viewing, commenting, reposting, and following generate behavioral signals that recommendation systems use to prioritize future content. Consequently, the influence of misleading AI-generated videos extends beyond their audiovisual realism. Their broader societal impact lies in their repeated circulation through dynamically evolving echo chambers, where algorithmic recommendation and user engagement continuously reinforce one another.

This study extends existing research on synthetic media by shifting attention from detecting misleading AI-generated videos to predicting how they become embedded within recommendation-driven information environments. Automated detection and expert verification remain essential for identifying synthetic and misleading content. However, they cannot explain why some misleading videos receive sustained attention while others quickly disappear. By reconstructing daily echo chamber communities and predicting future EC membership, this study demonstrates that the dissemination of misleading AI-generated videos is not solely a content problem but also a dynamic process shaped by interactions among synthetic media, recommendation systems, and user behavior. This perspective broadens current research beyond technical detection toward understanding the social dynamics of synthetic misinformation.

The findings also contribute to echo chamber research by conceptualizing EC membership as a dynamic exposure state rather than a fixed user characteristic. A user may participate in an EC on one day but not on another, depending on evolving interaction patterns, recommendation outcomes, and changing network structures. This dynamic perspective contrasts with conventional approaches that classify users as permanently belonging to echo chambers. Consistent with this conceptualization, ConvLSTM achieved the strongest predictive performance among the evaluated models, suggesting that temporal dependency is a defining characteristic of EC formation. Together, these findings demonstrate the value of spatiotemporal modeling for understanding how recommendation systems and user behavior jointly shape changing patterns of enclosed information exposure.

The predictive models further show that behavioral indicators provide substantially greater predictive information than demographic characteristics. Posting frequency, rationality-related and curiosity-related proxy indicators were consistently among the strongest predictors of future EC membership, whereas demographic variables contributed comparatively little. These findings should not be interpreted as evidence that some users are inherently vulnerable or psychologically predisposed to misinformation. Rather, they suggest that observable behavioral patterns provide useful signals of changing exposure conditions. In recommendation-driven environments, user interactions are continuously interpreted by

algorithms as indicators of relevance, regardless of whether those interactions are supportive, skeptical, emotional, or exploratory. The findings therefore emphasize the importance of understanding recommendation feedback loops rather than attributing EC participation to stable individual characteristics.

These findings have important implications for platform governance. Existing interventions primarily focus on detecting, labeling, or removing misleading synthetic content. Although such measures remain necessary, they do not address the recommendation processes through which misleading videos repeatedly circulate within enclosed information environments. The present findings therefore support governance strategies that complement content moderation with exposure-aware interventions. Rather than relying on demographic profiling or fixed vulnerability labels, platforms could monitor emerging behavioral and network-level indicators to provide transparent recommendation adjustments, source-context reminders, verification prompts, or exposure diversification while preserving user autonomy, privacy, and informed choice.

Several limitations should be acknowledged. First, the analysis identifies predictive associations rather than causal mechanisms and therefore does not establish that the observed behavioral signals cause future EC membership. Second, the study relied on publicly observable behavioral and profile information and could not incorporate private viewing histories, recommendation logs, or proprietary ranking algorithms used by the platforms. Third, the analysis focused on a 30-day observation period, which captures short-term dynamics but may not fully reflect longer-term patterns of exposure and behavioral adaptation. Future research integrating platform-level recommendation data, longitudinal observations, and causal designs could further advance understanding of how synthetic media and recommendation systems jointly influence information exposure over time.

6. Conclusion

This study investigated the dynamic formation of EC membership around misleading AI-generated videos across four major short video platforms: TikTok, Instagram, Douyin, and Kuaishou. Misleading AI-generated videos were identified through automated screening followed by expert verification, after which EC communities were reconstructed daily over a 30-day observation period. Using user-level behavioral, textual, and profile-based indicators, the study evaluated whether future daily EC membership could be anticipated through spatiotemporal modeling. The results show that dynamic EC membership can be predicted with useful accuracy, and behavioral signals provide substantially greater predictive value than static demographic characteristics. Among the evaluated deep learning models, ConvLSTM achieved the strongest predictive performance, highlighting the importance of capturing temporal patterns in user behavior rather than treating EC membership as a fixed individual attribute.

Overall, this study provides a predictive framework for understanding how misleading AI-generated videos, recommender systems, and user behavior interact in contemporary short-video environments. The findings do not suggest that AI-generated videos inevitably trap users in ECs or that behavioral signals reveal stable personality or psychological vulnerability. Instead, they indicate that the early

development of enclosed information exposure may be detectable from observable patterns of everyday online interactions. These findings strongly support a shift of governance strategies from content detection or demographic profiling toward approaches that also consider evolving exposure patterns while preserving user autonomy, transparency, and privacy. As synthetic media become increasingly ubiquitous, understanding how misleading AI-generated content spreads through dynamic social information communities will become increasingly important for safeguarding public understanding and digital trust.

Declarations

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

The authors received no funding for this work.

Author Contribution

Y.G. conceived and designed the study, developed the methodology, conducted the data analysis and modeling, and drafted the manuscript. P.H. contributed to the study design, interpretation of the findings, and critical revision of the manuscript. Y.W. contributed to data collection, data curation, validation, and visualization. F.L. supervised the research, contributed to the conceptualization and methodology, and critically revised the manuscript. All authors reviewed and approved the final manuscript.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Clark, S. & Lewandowsky, S. The continued influence of AI-generated deepfake videos despite transparency warnings. *Commun. Psychol.* **4**, 13. <https://doi.org/10.1038/s44271-025-00381-9> (2026).
2. Groh, M. et al. Human detection of political speech deepfakes across transcripts, audio, and video. *Nat. Commun.* **15**, 7629. <https://doi.org/10.1038/s41467-024-51998-z> (2024).

3. Yu, X., Haroon, M., Menchen-Trevino, E. & Wojcieszak, M. Nudging recommendation algorithms increases news consumption and diversity on YouTube. *PNAS Nexus*. **3**, pgae518. <https://doi.org/10.1093/pnasnexus/pgae518> (2024).
4. Hartmann, D., Wang, S. M., Pohlmann, L. & Berendt, B. A systematic review of echo chamber research: comparative analysis of conceptualizations, operationalizations, and varying outcomes. *J. Comput. Soc. Sci.* **8**, 52. <https://doi.org/10.1007/s42001-025-00381-z> (2025).
5. Kozyreva, A. et al. Toolbox of individual-level interventions against online misinformation. *Nat. Hum. Behav.* **8**, 1044–1052. <https://doi.org/10.1038/s41562-024-01881-0> (2024).
6. Doss, C. et al. Deepfakes and scientific knowledge dissemination. *Sci. Rep.* **13**, 13429. <https://doi.org/10.1038/s41598-023-39944-3> (2023).
7. Bai, J., Lin, M., Cao, G. & Lou, Z. AI-generated video detection via spatial-temporal anomaly learning in Chinese Conference on Pattern Recognition and Computer Vision (PRCV) 460–470 (Springer, 2024); https://doi.org/10.1007/978-981-97-8792-0_32
8. Gerbaudo, P. TikTok and the algorithmic transformation of social media publics: from social networks to social interest clusters. *New. Media Soc.* **28**, 1019–1036. <https://doi.org/10.1177/14614448241304106> (2026).
9. He, K., Wildfeuer, J., Liu, J. & Guo, J. Diagnosing multimodal disinformation: the integrative Three-M framework of multimodality, manipulation, and malintent. *Inf. Commun. Soc.* 1–23. <https://doi.org/10.1080/1369118X.2026.2645883> (2026).
10. Starke, C., Metikoš, L., Helberger, N. & de Vreese, C. Contesting personalized recommender systems: a cross-country analysis of user preferences. *Inf. Commun. Soc.* **28**, 41–60. <https://doi.org/10.1080/1369118X.2024.2363926> (2025).
11. Eder, M. & Sehl, A. Being aware of algorithmic personalization? Insights from three European countries. *Inf. Commun. Soc.* **29**, 2158–2175. <https://doi.org/10.1080/1369118X.2025.2553015> (2026).
12. Vellani, V., Zheng, S., Ercelik, D. & Sharot, T. The illusory truth effect leads to the spread of misinformation. *Cognition* **236**, 105421. <https://doi.org/10.1016/j.cognition.2023.105421> (2023).
13. Gao, Y., Liu, F. & Gao, L. Echo chamber effects on short video platforms. *Sci. Rep.* **13**, 6282. <https://doi.org/10.1038/s41598-023-33370-1> (2023).
14. Alatawi, F., Sheth, P. & Liu, H. Quantifying the echo chamber effect: an embedding distance-based approach in Proceedings of the 2023 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2023) (eds Prakash, B. A., Wang, D. & Weninger, T.) 38–45 (Association for Computing Machinery, 2023); <https://doi.org/10.1145/3625007.3627731>
15. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233. <https://doi.org/10.1038/s41598-019-41695-z> (2019).
16. Annabell, T. Calling out disclosures, protecting followers: the new governors of influencer culture on TikTok. *Inf. Commun. Soc.* 1–19. <https://doi.org/10.1080/1369118X.2026.2674878> (2026).

17. Li, Y. et al. Skyra: AI-generated video detection via grounded artifact reasoning in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 4482–4493 (2026).
18. Sun, C., Jia, S., Hou, S. & Lyu, S. AI-synthesized voice detection using neural vocoder artifacts in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops 904–912 ; (2023). <https://doi.org/10.1109/CVPRW59228.2023.00097>
19. Picha Edwardsson, M., Al-Saqaf, W. & Nygren, G. Verification of digital sources in Swedish newsrooms: a technical issue or a question of newsroom culture? *J. Pract.* **17**, 1678–1695. <https://doi.org/10.1080/17512786.2021.2004200> (2023).
20. Xie, L., Wang, D. & Ma, F. Analysis of individual characteristics influencing user polarization in COVID-19 vaccine hesitancy. *Comput. Hum. Behav.* **143**, 107649. <https://doi.org/10.1016/j.chb.2022.107649> (2023).
21. Lu, Z. Z. & Gao, X. 网络群体极化形成机制与引导策略研究 [The study on the formation mechanism and guiding strategies of group polarization in social media platforms]. *Inf. Stud. Theory Appl.* **44**, 51–58. <https://doi.org/10.16353/j.cnki.1000-7490.2021.08.008> (2021).
22. Huang, X. J. et al. Weibo rumors real-time detection model based on fusion of multi-user features and content features. *J. Chin. Comput. Syst.* **43**, 2518–2527 (2022).
23. Berlyne, D. E., Conflict & Arousal and Curiosity (McGraw-Hill, ; (1960). <https://doi.org/10.1037/11164-000>
24. Xu, K., Mo, J., Cai, Y. & Min, H. Enhancing recommender systems with a stimulus-evoked curiosity mechanism. *IEEE Trans. Knowl. Data Eng.* **33**, 2437–2451. <https://doi.org/10.1109/TKDE.2019.2957770> (2021).
25. Jourard, S. M. & Lasakow, P. Some factors in self-disclosure. *J. Abnorm. Soc. Psychol.* **56**, 91–98. <https://doi.org/10.1037/h0043357> (1958).
26. Luo, M. & Hancock, J. T. Self-disclosure and social media: motivations, mechanisms and psychological well-being. *Curr. Opin. Psychol.* **31**, 110–115. <https://doi.org/10.1016/j.copsyc.2019.08.019> (2020).
27. Liu, B., Sun, Y. & Gao, L. Enhancing groundwater recharge prediction: a feature selection-based deep forest model with Bayesian optimisation. *Hydrol. Process.* **38**, e15309. <https://doi.org/10.1002/hyp.15309> (2024).
28. Osman, A. I. A., Ahmed, A. N., Chow, M. F., Huang, Y. F. & El-Shafie, A. Extreme gradient boosting (XGBoost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Eng. J.* **12**, 1545–1556. <https://doi.org/10.1016/j.asej.2020.11.011> (2021).
29. Vergara, J. R. & Estévez, P. A. A review of feature selection methods based on mutual information. *Neural Comput. Appl.* **24**, 175–186. <https://doi.org/10.1007/s00521-013-1368-0> (2014).
30. Morini, V., Pollacci, L. & Rossetti, G. Toward a standard approach for echo chamber detection: Reddit case study. *Appl. Sci.* **11**, 5390. <https://doi.org/10.3390/app11125390> (2021).
31. Chassagnol, B. et al. Gaussian mixture models in R. *R J.* **15**, 56–76. <https://doi.org/10.32614/RJ-2023-043> (2023).

32. Shi, X. et al. Curran Associates,. Convolutional LSTM network: a machine learning approach for precipitation nowcasting in *Advances in Neural Information Processing Systems* 28 802–810 (2015).
33. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735> (1997).
34. He, W. et al. Radar echo spatiotemporal sequence prediction using an improved ConvGRU deep learning model. *Atmosphere* **13**, 88. <https://doi.org/10.3390/atmos13010088> (2022).
35. Bak-Coleman, J. B. et al. Combining interventions to reduce the spread of viral misinformation. *Nat. Hum. Behav.* **6**, 1372–1380. <https://doi.org/10.1038/s41562-022-01388-6> (2022).
36. Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151. <https://doi.org/10.1126/science.aap9559> (2018).
37. Maertens, R. et al. Psychological booster shots targeting memory increase long-term resistance against misinformation. *Nat. Commun.* **16**, ; (2025). <https://doi.org/10.1038/s41467-025-57205-x> (2025).
38. Engelmann, J. & Bernabeu, M. O. Training a high-performance retinal foundation model with half-the-data and 400 times less compute. *Nat. Commun.* **16**, 6862. <https://doi.org/10.1038/s41467-025-62123-z> (2025).
39. Archit, A. et al. Segment Anything for microscopy. *Nat. Methods.* **22**, 579–591. <https://doi.org/10.1038/s41592-024-02580-4> (2025).
40. Nasirzadeh, M., Tahmoresnezhad, J. & Rashidi-Khazaei, P. A unified framework for detecting point and collective anomalies in operating system logs via collaborative transformers. *Sci. Rep.* **15**, 45698. <https://doi.org/10.1038/s41598-025-27693-4> (2025).
41. Amiri, M. & Khaleseh, F. Predicting drug solubility in binary solvent mixtures using graph convolutional networks: a comprehensive deep learning approach. *Sci. Rep.* **15**, 45711. <https://doi.org/10.1038/s41598-025-28272-3> (2025).
42. Malla, S. R. MallaNet residual branch merge convolutional neural network with homogeneous filter capsules for Devanagari character recognition. *Sci. Rep.* **16**, 1175. <https://doi.org/10.1038/s41598-025-30871-z> (2026).
43. Rückert, J. et al. ROCOv2: Radiology Objects in COntext Version 2, an updated multimodal image dataset. *Sci. Data.* **11**, 688. <https://doi.org/10.1038/s41597-024-03496-6> (2024).
44. Johnson, N. F. et al. The online competition between pro- and anti-vaccination views. *Nature* **582**, 230–233. <https://doi.org/10.1038/s41586-020-2281-1> (2020).
45. Lorenz-Spreen, P., Mønsted, B. M., Hövel, P. & Lehmann, S. Accelerating dynamics of collective attention. *Nat. Commun.* **10**, 1759. <https://doi.org/10.1038/s41467-019-09311-w> (2019).
46. Shao, C. et al. The spread of low-credibility content by social bots. *Nat. Commun.* **9**, 4787. <https://doi.org/10.1038/s41467-018-06930-7> (2018).
47. Alkan, B. B., Kuzucuk, S., Alkan, N. & Sinan, A. Using machine learning to predict student outcomes for early intervention and formative assessment. *Sci. Rep.* **15**, 39797. <https://doi.org/10.1038/s41598-025-23409-w> (2025).

48. Hutchison, R. M., Womelsdorf, T., Gati, J. S., Everling, S. & Menon, R. S. Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques. *Hum. Brain Mapp.* **34**, 2154–2177. <https://doi.org/10.1002/hbm.22058> (2013).

49. O'Brien, R. M. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* **41**, 673–690. <https://doi.org/10.1007/s11135-006-9018-6> (2007).

Figures

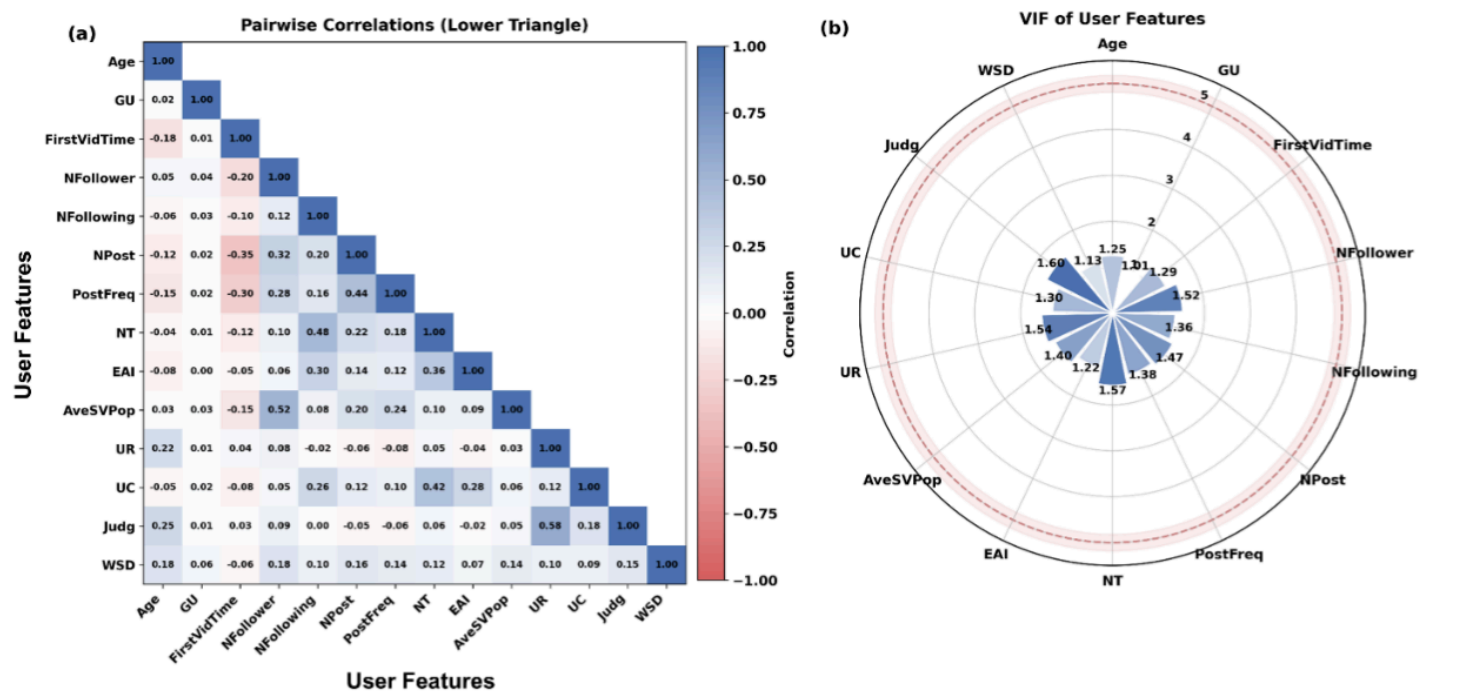


Figure 1

Pairwise correlations and VIFs for the 14 user features

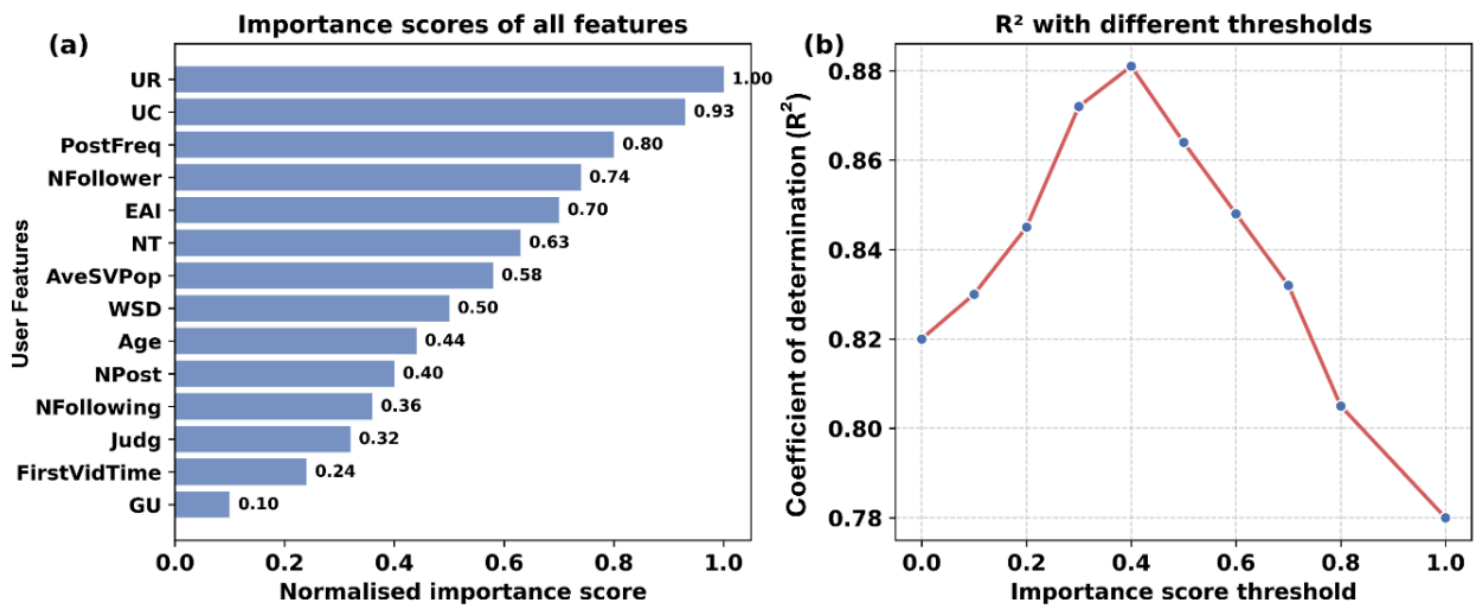


Figure 2

Importance scores of all features

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterials.docx](#)