UNEC Journal of Computer Science and Digital Technologies
Volume 1, № 2, 2025, pp. 43-59
https://doi.org/10.30546/UNECCSDT.2025.02.1014

Section:
Cybersecurity & Privacy

Adaptive Cross-Platform Threat Correlation Layer: A Cybersecurity Approach to Detecting Distributed Digital Violence on Social Media Platforms

Roza Misirli[1*], Khanim İmanova[2], Hikmat Mammadli[3]

[1,2,3]Department of Digital Technologies and Applied Informatics, UNEC, Baku, Azerbaijan
[1]0000-0002-7375-4300, roza-misirli@unec.edu.az, roza.misirli@uniroma1.it
[2]0009-0007-8672-2985, imanova.khanim@unec.edu.az
[3]0009-0005-8287-0222, mammadli.hikmat.faig.2025@unec.edu.az

Abstract

Background: Large-scale digital assaults and online harassment increasingly span multiple social media platforms. However, effective cross-platform threat correlation remains constrained by privacy regulations, incompatible data formats, and the absence of a shared behavioural analysis layer. Existing privacy-conscious machine-learning approaches primarily focus on platform-specific anomaly detection and fail to capture coordinated behaviour across decoupled environments. This study introduces the Adaptive Cross-Platform Threat Correlation Layer (ACTCL), a lightweight cybersecurity framework that conceptualizes digital violence as a privacy-aware behavioural correlation problem. ACTCL employs Anonymous Feature Hashing (AFH) to transform three transient behavioural features—temporal rhythm, activity concentration, and session duration—into irreversible eight-dimensional fingerprints. These fingerprints enable cross-system comparison without exposing raw or personally identifiable data. The framework was evaluated using a controlled synthetic dataset. Using a synthetic dataset of 150 users, including 10 coordinated attackers, cosine similarity analysis over the hashed behavioural fingerprints revealed a degenerate similarity cluster among coordinated attackers (mean = 1.0000, SD = 0.0000). In contrast, normal users exhibited a more dispersed similarity distribution (mean = 0.9877, SD = 0.0224), indicating distinct behavioural variability. The findings demonstrate that coordinated malicious behaviour retains stable behavioural signatures even under strong anonymisation constraints. This confirms the feasibility of privacy-compliant cross-platform threat correlation. ACTCL provides a practical foundation for future multi-platform security cooperation, with potential applications in fraud detection, coordinated misinformation tracking, botnet analysis, and other distributed digital threats.

Keywords: cross-platform security; digital violence detection; behavioural fingerprinting; privacy-preserving cybersecurity; federated threat correlation; social media security.

Received:	*Revised:*	*Accepted:*	*Published:*
1/12/2025	*16/12/2025*	*19/12/2025*	*23/12/2025*

1. Introduction

The social media are massive digital ecosystems that support billions of interactions between users on a daily basis. As they allow quick information sharing and social interaction, these systems have also become a place where the threats to the user level of cybersecurity become organised at the level of behavioural coordination and not the one that would be organised around the infrastructure. Organized types of digital violence, such as organized harassment campaigns, identity spoofing, account cycling, malicious mass reporting, and deepfake-assisted extortion, are almost always staged on more than one platform, which takes advantage of architectural fragmentation and the inability to see across platforms [4, 11, 15].

Most social media platforms process behavioural logs only on their own systems, and pipelines are separate and independent. Such architecture disconnection prevents the observation of orchestrated hostile tactics like platform hopping, synchronized posting schedules, and swift identity alternations. Simultaneously, privacy laws and ethical limitations severely limit the exchange of raw user data across platforms, and the traditional cross-platform threat intelligence methods will be legally and technically impossible [12, 18]. There are therefore no systems in place on platforms to correlate weak or ambiguous behavioural signals that are only meaningful when analysed in ecosystems.

The literature mainly focuses on platform-specific anomaly detection, metadata-related correlation, or privacy-sensitive machine learning in isolated systems. Although these strategies enhance the accuracy of detection at the individual platform level, they fail to offer a specific mechanism of privacy-aware behavioural correlation between autonomous platforms [14, 17, 19]. The behavioural signatures, though, are particularly cross-platform friendly, as it almost does not depend on language, content format, and platform-specific metadata. According to empirical evidence, the patterns of participation, including the presence of temporal rhythms and the concentration of activities, are structurally stable across platforms, despite changes in the toxicity of the content level (Figure 1). Also, behavioural patterns have generalizable properties that can be exploited in privacy-aware correlation, unlike the textual or multimedia content that is highly platform-dependent.

To address these limitations, this paper introduces ACTCL, a minimalist cybersecurity architecture that reframes digital violence detection as a privacy-preserving behavioural similarity problem. Instead of exchanging raw logs, identifiers, or platform-specific metadata, ACTCL enables platforms to share irreversible behavioural fingerprints derived from minimal activity features, thereby supporting cross-platform correlation without violating privacy constraints.

This study makes the following contributions:

1. It proposes ACTCL, a cross-platform architectural layer designed to detect distributed digital violence through behavioural correlation rather than identity resolution.
2. It introduces an Anonymous Feature Hashing (AFH) mechanism that transforms behavioural attributes into non-reversible eight-dimensional representations suitable for privacy-preserving similarity analysis.
3. It demonstrates that, using a controlled synthetic dataset, coordinated attackers form a distinguishable behavioural cluster even when only three simple behavioural features are employed.

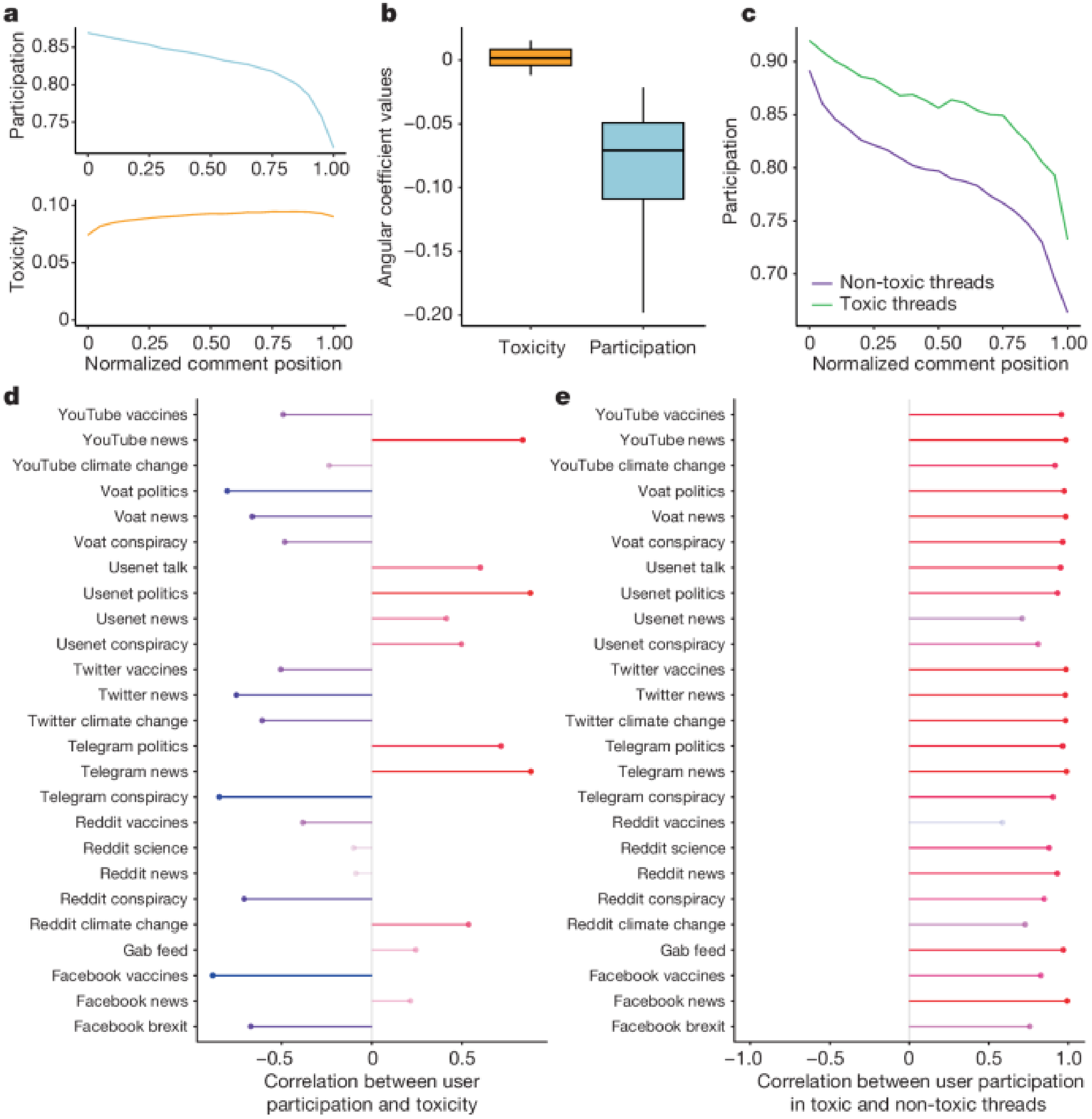

Figure 1. Participation behaviour remains stable independent of toxicity levels across platforms

As illustrated in Figure 2, conventional platform-level detection pipelines introduce blind spots where coordinated cross-platform behaviour may remain undetected. ACTCL is designed to operate as a complementary layer that bridges these gaps without interfering with existing platform-specific moderation systems.

To assess the technical feasibility of the proposed architecture, we conduct a behavioural hashing experiment using a controlled synthetic dataset.

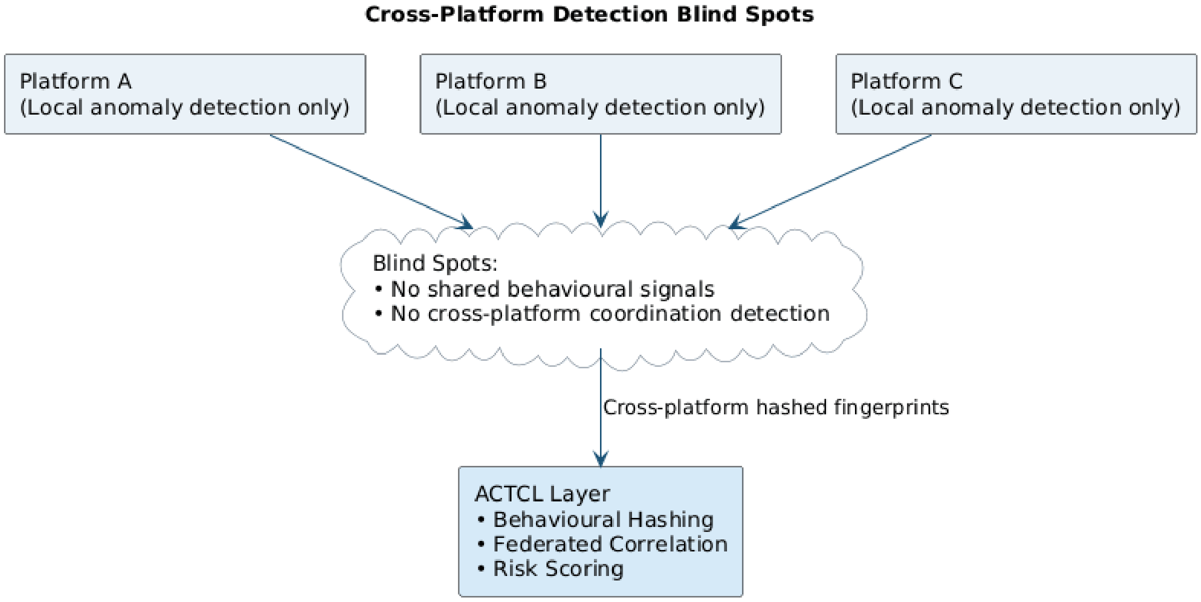
Cross-Platform Detection Blind Spots
Platform A
(Local anomaly detection only)
Platform B
(Local anomaly detection only)
Platform C
(Local anomaly detection only)
Blind Spots:
• No shared behavioural signals
• No cross-platform coordination detection
Cross-platform hashed fingerprints
ACTCL Layer
• Behavioural Hashing
• Federated Correlation
• Risk Scoring

Figure 2. Fragmented platform-level detection pipelines and cross-platform behavioural blind spots.

As shown in Figure 3, user activity and engagement exhibit high variance yet retain structurally consistent patterns across platforms. This observation motivates the use of synthetic behavioural data to evaluate ACTCL in a privacy-safe and ethically compliant manner while preserving key characteristics of real-world interaction dynamics.

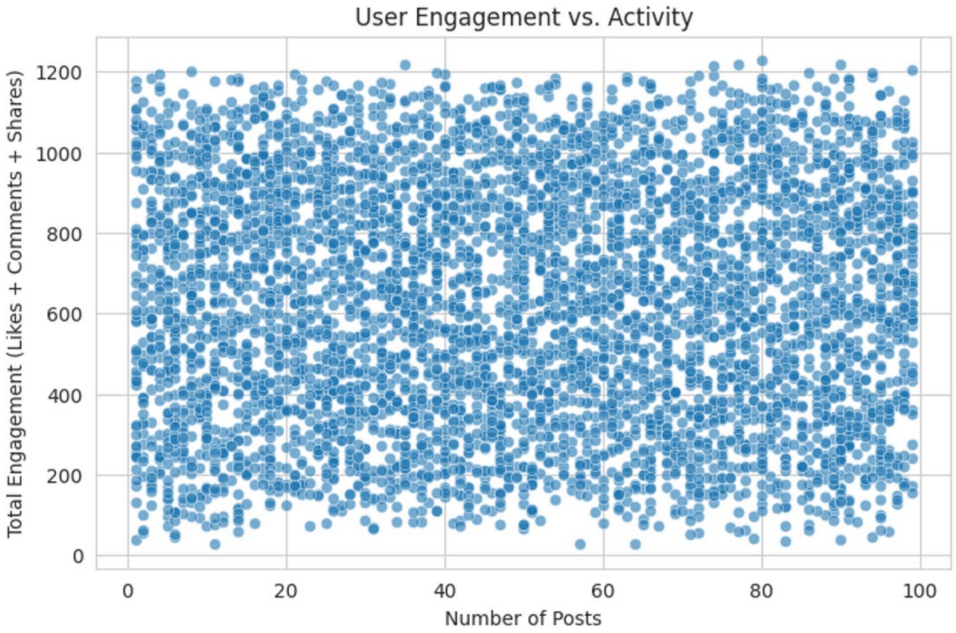

Figure 3. Empirical relationship between user activity and engagement on social media platforms.

2. Experimental Methodology

2.1 Platform-Specific Detection of Digital Violence

Most of the studies on digital violence detection have concentrated on platform-specific analysis, in which abusive or malicious behaviour is determined based on data obtained in a single social media ecosystem. Previous research on this area involves automated identification of online harassment, cyberbullying, coordinated abuse, and aggressive behaviour with content based, metadata based, and behavioural identification approaches. As an example, research conducted regarding cyberbullying and online aggression usually uses textual characteristics, sentiment analysis, or linguistic cues derived based on the platform-specific user-generated material [4, 15]. These methods have proved to be effective in detecting abusive behaviour in confined platforms like Twitter or Facebook.

In addition to content-based approaches, behavioural modelling has been considered to provide patterns of posting frequency, rhythms of temporal activities, and intensity of engagement. Behavioural features are frequently perceived to be stronger than textual indicators since they are less inclined to language and platform-related norms or moderation regulations [9]. However, in most of the current systems, behavioural analysis is restricted to data infrastructure of a single platform, making it incapable of capturing coordinated actions that cut across platforms [9].

The essential drawback of platform-specific detection pipelines is that they are architecturally isolated. Every platform has its monitoring and moderation system, which leads to the dissemination of threat intelligence and the absence of visibility regarding cross-platform coordination [5]. This fragmentation is increasingly used by malicious actors who disseminate malicious activities across websites, thereby using strategies of coordinating postings, cycling of accounts, and platform hopping. Consequently, a behaviour that seems harmless or mildly suspicious in one platform would not show its evil intent until it is evaluated in multifaceted settings [5].

Therefore, platform-level detection systems have reached great accuracy and scale, but they still fail to handle distributed aspects of digital violence that take advantage of the interoperability absence between social media systems [6]. The limitation impels the necessity of methods that will not confine one platform analysis but allow cross-platform behavioural correlation without being tied to direct identity identification or aggregate data exchange.

2.2 Privacy-Preserving Security and Behavioural Analytics

As regulatory limitations and public anxiety about user privacy have continued to increase, a substantial amount of recent work has centered on privacy-sensitive methods of security analytics. Such attempts will facilitate the threat detection and behavioural analysis without having access to raw user data or personally identifiable information. The leading trends in this field are federated learning, secure multi-party computation, anonymization methods, and privacy-aware feature transformation [14, 16, 19].

Federated learning has already become a new leading model of privacy-preserving intrusion detection and anomaly analysis, where local training models are trained on decentralized data and share only parameters or gradients [16]. The method has managed to be effectively used in areas like IoT intrusion detection, mobile analytics, and distributed monitors. These techniques are, however, more oriented to enhance predictive performance in a predefined task, as opposed to being able to correlate behaviour across independent platforms [16].

A separate body of work investigates anonymization and feature transformation methods that are aimed at maintaining utility and minimizing reconstructability. Random projection, hash-based representations, and dimensionality expansion processes have been demonstrated to preserve structural similarity among data points and reduce the threat of reverse inference. Privacy-preserving analytics has become more common to use such methods to allow similarity computation, clustering, and finding patterns without revealing the original feature value [11].

With these developments, the current privacy-saving security mechanisms are mostly task-oriented and platform-specific. They are generally created to assist in classification, detection, or scoring anomalies within one organizational or infrastructural boundary [1]. Consequently, they fail to directly solve the problem of matching the behavioural patterns across autonomous platforms that are unable to access, share, or exchange data, models, and identifiers as a result of legal, ethical, or competitive limitations.

Such a loophole becomes particularly pertinent in the presence of distributed digital violence, in which synchronized, malicious conduct can be seen by producing weak or vague signals, on an individual platform. Although privacy-preserving analytics are resistant to sensitive data, they, alone, do not offer a cross-platform behavioural alignment or threat correlation mechanism. As a result, there is an unfulfilled requirement of architectures that take advantage of privacy-preservation transformations in the specific context of facilitating behavioural similarity analysis across fragmented digital ecosystems.

2.3 Cross-Platform User Correlation and Its Limitations

Reduced and yet increasing literature has investigated cross-platform analysis and user correspondence in online abuse, misinformation, and organised manipulation. The aim of these studies is to correlate the accounts, or behaviours, on more than one social media platform to determine the users that are sharing ecosystems and operating concurrently. Common methods are similarity of usernames, similarity in profile metadata, analysis of shared content, analysis of network overlap, and similarity in posting behaviour over time [5]. Although there are several frameworks that concentrate on matching identifiers, the current studies have suggested built-in user-matching frameworks which attempt to bridge platform gaps specifically in cyberbullying detection [8].

Various systems have tried to identify users of multiple platforms by abusing publicly available identities or behavioural history, including the same username, profile picture, or posting time, and so on. Other research works deal with cross-platform cyberbullying or harassment by correlating content similarity and patterns of interaction across services [20]. Although these techniques show that the idea of cross-platform coordination is present, they usually are based on assumptions that become progressively unrealistic in a hostile or controlled setting.

The main weakness of currently used cross-platform methods is the reliance on recognizable or identifiable data [21]. The identity-based matching methods reveal a lot of privacy threats and cannot be used in accordance with contemporary data protection laws. Metadata-based correlation also presupposes access to sensitive platform-specific data that cannot be legally or ethically exchanged between independent organizations [21]. In addition, they are fragile to adversarial adaptation as any malicious user can easily alter usernames, profiles, or postings to avoid identification.

The other weakness is a close relationship between correlation and identity resolution. Most of the the existing literature views cross-platform analysis as a user matching problem, which directly presupposes that the identification of the same person or account with different systems is required [10]. It is a problem with digital violence detection when it is not always necessary to detect real-life actors, but rather to identify coordinated behavioural patterns, which would signify that malicious actions are coordinated and organized. Focusing on identity linkage conflates attribution with detection and unnecessarily increases privacy exposure [10].

Moreover, cross-platform research in the literature typically involves aggregation of data on the center or a combination of access to raw logs, which is infeasible in practice in a real-world environment where platforms are independent of each other [7]. The existence of competitive, legal and ethical limitations harshly restricts the validity of shared databases or centralized monitoring structures. Therefore, a lot of the offered cross-platform solutions exist only in the field of theorization or get tested only in the laboratory.

Collectively, these restrictions suggest that existing cross-platform correlation methods are unsuitable to privacy sensitive detection of distributed digital violence. The relationship between behavioural correlation and identity resolution and uncritical sharing of raw data needs to be decoupled, and the problem of cross-platform detection must be re-conceptualised as a similarity problem that preserves a user's privacy instead of user-matching problem. The proposed ACTCL framework is driven by this observation in its choice of architectural design principles.

2.4 Research Gap and Motivation for ACTCL

The analysis of the current literature indicates the evident knowledge gap in the overlap of digital violence detection, privacy-conscious analytics, and cross-platform security [14]. Although platform-specific detection systems have scored significant success, they are naturally siloed and cannot detect coordinated behaviour that will cut across ecosystems of social media. Meanwhile, privacy-sensitive security architectures mainly concentrate on safeguarding judicious information through autonomous analytical duties and not allowing behavioural association across autonomous platforms. Current methods of cross-platform, in turn, mostly rely on identity connectivity, metadata exchange, or the centralization of information, all of which present considerable privacy, law, and practical issues.

As far as we know, there have been no previous studies that suggest a lightweight, privacy-conscious behavioural correlation layer that is explicitly formulated to identify coordinated digital violence across autonomous social media without exchanging raw data, identities, or platform-specific metadata. This is especially essential in light of the growing occurrence of distributed and synchronized online assaults which utilize the divisiveness between platforms [13].

The gap is what drives the Adaptive Cross-Platform Threat Correlation Layer (ACTCL). Instead of trying to find identities of users and consolidate sensitive data, ACTCL revisits cross-platform threat detection as a behavioural similarity issue. ACTCL allows privacy-safe coordinated activity across platforms, by converting minimal, platform-agnostic behavioural signals into irreversible hashed fingerprints, which is a limitation that the existing detection, privacy and cross platform frameworks do not adequately address.

3. Experimental Methodology

This section presents the simplified experimental procedure used to evaluate whether Anonymous Feature Hashing (AFH) can separate coordinated malicious users from normal users while preserving privacy. The goal is not to build a high-complexity detection model, but rather to demonstrate that even minimal behavioural signals can reveal coordination patterns when anonymized through AFH.

3.1 Experimental Dataset

A synthetic dataset was created to simulate realistic social media activity from 150 users over a 59-day observation period, consisting of approximately 52,000 timestamped user events. Events include posts, comments, clicks, and views, generated with realistic temporal fluctuations such as increased activity during daytime hours and reduced nighttime behaviour as illustrated in the generation process (Figure 4).

Key properties of the dataset include:

- 150 total users
- 104–599 events per user (average ≈ 346 actions)
- Diurnal activity distribution matching typical social platform usage

- 140 normal users with natural behavioural variability
- 10 coordinated attackers, designed to mimic synchronized malicious behaviour patterns aligned with tactics reported in recent cybersecurity analyses (Mishra et al., 2019; Chatzakou et al., 2017)

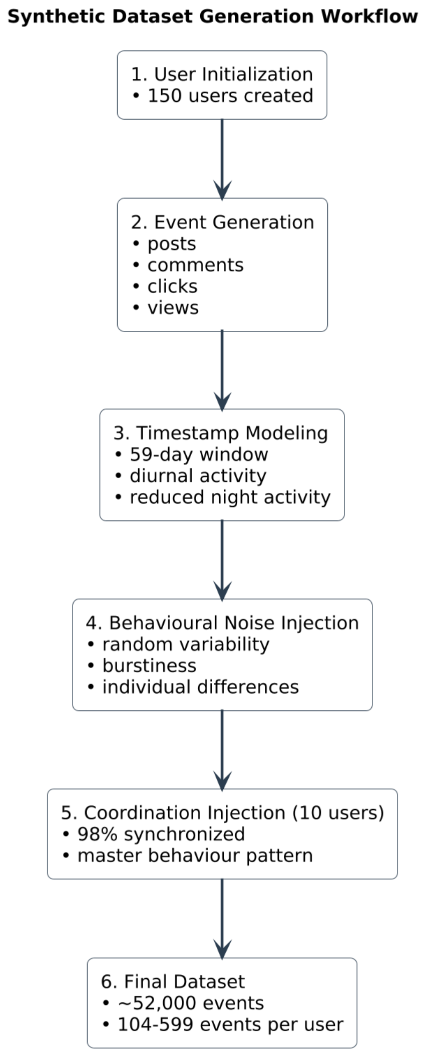

Figure 4. Overview of the synthetic dataset generation process, including event creation, temporal modelling, and attacker assignment.

3.2 Feature Construction

For interpretability and simplicity, only three behavioural features were extracted per user. These features represent fundamental behavioural signals that are easy to compute and have been used in behavioural modelling for aggression and coordinated activity detection [4].

(1) Temporal Rhythm

The average number of seconds between consecutive user actions:
This feature captures how quickly or slowly a user interacts with the platform.

Example: User A averages 4,365 seconds (~73 minutes) between actions.

(2) Activity Concentration

The proportion of user events occurring during their most active one-hour window:
This reflects burstiness and temporal clustering of activity.

Range: 0.0 – 1.0

Example: User B performs 8.4% of all their activity in their peak hour.

(3) Session Duration

The total active span (in hours) from a user's first to last event.
Example: User C remains active for 244 hours (~10 days) over the dataset period.
Each user is represented by the 3-dimensional vector:

$$v_{user} = [temporal\ rhythm,\ activity\ concentration,\ session\ duration].$$

While these features are intentionally minimal, they provide a sufficient behavioural foundation on which the Anonymous Feature Hashing transformation can operate.

To ensure privacy while enabling behavioural comparison, each 3-dimensional feature vector was transformed into an 8-dimensional anonymized fingerprint using Gaussian Random Projection (GRP) [11, 14].

This transformation:

1. Expands dimensionality from 3 → 8 (+167% expansion)
2. Destroys reconstructability of the original features
3. Preserves relative similarity between behavioural patterns

Formally, the transformation is:

$$h(v) = \frac{Av}{\|Av\|^2},$$

where A is a random matrix with entries sampled from a Gaussian distribution. This produces a privacy-preserving hashed fingerprint that can be compared across users without revealing the original behavioural values (Figure 5).

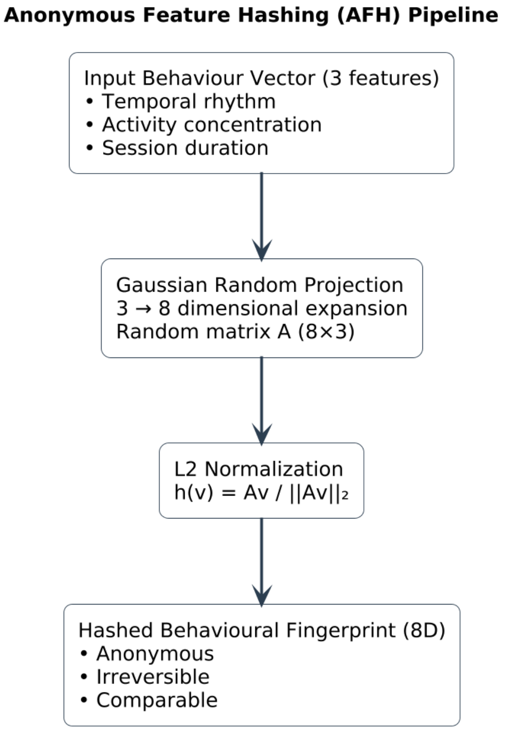
Anonymous Feature Hashing (AFH) Pipeline
Input Behaviour Vector (3 features)
• Temporal rhythm
• Activity concentration
• Session duration
Gaussian Random Projection
3 → 8 dimensional expansion
Random matrix A (8×3)
L2 Normalization
h(v) = Av / ||Av||₂
Hashed Behavioural Fingerprint (8D)
• Anonymous
• Irreversible
• Comparable

Figure 5. Anonymous Feature Hashing (AFH) pipeline from 3-dimensional features to 8-dimensional hashed fingerprints.

With the hashing mechanism defined, the next step is to assess whether coordinated adversaries produce distinguishable signatures under this anonymized representation.

3.4 Coordinated Attack Simulation

To evaluate detection capability, 10 users were modified to behave in a strongly synchronized manner. Their behavioural vectors were constructed using a master coordination template:

- ● Temporal rhythm: 1200 seconds
- ● Activity concentration: 35%
- ● Session duration: 36 hours

Coordinated attackers were generated as:

- ● 98% synchronized with the master pattern
- ● 2% random variation to preserve slight individuality

Normal users retained their naturally diverse behavioural patterns. Random perturbations were also applied to normal users to better simulate real-world behavioural variety. To quantify these behavioural differences, we compute pairwise similarity scores between hashed fingerprints.

3.5 Similarity Computation

Behavioural similarity was computed using cosine similarity ($Similarity(A, B) = cos(\theta) = \frac{A \cdot B}{||A||||B||}$), a widely used metric ranging from:

- ● +1 → identical
- ● 0 → unrelated
- ● –1 → opposite

Crucially, similarity was computed only on the hashed 8-dimensional fingerprints, ensuring:

- ● no exposure of original behavioural values
- ● privacy-preserving computation

Two similarity groups were evaluated:

- ● Coordinated attackers (10 users)
- ● Randomly sampled normal users (10 users)

4. Results and Discussion

Having specified the behavioural characteristics, anonymization process, and similarity measures, the current analysis determines whether generated hashed fingerprints allow introducing a significant discrimination between organised attackers and normal users. The empirical results of the behavioural-hashing experiment are reported in this section and the effectiveness of ACTCL in differentiating between coordinated malicious and benign user behaviour. The results verify that, despite being reduced to three simple behavioural characteristics, and despite the use of irreversible Anonymous Feature Hashing (AFH), coordinated behaviour shows a noticeable and distinctive pattern, thus providing support to ACTCL as a feasible, privacy-sensitive threat-correlation scheme.

4.1 Evaluation Results

Despite relying on a minimal three-feature behavioural representation, the system preserves clear structural differences between coordinated and normal users.

The results are given in Table 1:

Coordinated Attackers (n = 10)

- ● Mean similarity: 1.0000
- ● Standard deviation: 0.0000
- ● Range: [1.0000, 1.0000]

Random Normal Users (n = 10)

- Mean similarity: 0.9877
- Standard deviation: 0.0224
- Range: [0.9299, 1.0000]

Key Metrics

- Absolute difference: 0.0123
- Similarity ratio: 1.01×
- Detection behaviour: Clear clustering of coordinated attackers

Metric	Value
Coordinated mean similarity	1.0000
Coordinated standard deviation	0.0000
Random mean similarity	0.9877
Random standard deviation	0.0224
Similarity difference	0.0123
Similarity ratio	1.01×
Coordinated range	[1.0000, 1.0000]
Random range	[0.9299, 1.0000]
Number of users	150
Number of attackers	10
Cluster separability	Distinct block structure in similarity matrix (clear group-level separation)

Table 1. Summary of Similarity Metrics and Statistical Evaluation

4.2 Similarity Separation Results

The similarity test showed that there was a sharp structural difference between coordinated attackers and ordinary users. In addition to this, after converting each behavioural vector into an eight-dimensional hashed fingerprint, cosine similarity was calculated among all pairs of users. Organised attackers achieved an average similarity of 1.0000 and zero variance indicative of the very synchronised behavioural template used to recreate coordinated digital-violence strategies. Conversely, the signature displayed by normal users was more heterogeneous with a mean similarity value of 0.9877 and a standard deviation of 0.0224 with a range of 0.9299 - 1.0000. The numerical difference between the two means (around 1.23%) might sound as something insignificant, but its statistical and structural consequences are great because of the following reasons:

1. These degenerate clusters are created by perfect similarity:
 Co-attackers degenerate at 1.0000, creating a cluster of zero variance. Within behavioural modelling, a point-mass cluster of the type is a representation of a completely different generative process, and not stochastic fluctuation.
2. The normal-user distribution is natural:
 The normal-user similarity distribution is a continuous, dispersed distribution, but the attacker distribution is discrete, and uniform. These different shapes of distribution make the separation analytical even in the case where the means are near.

This difference can be seen in Figure 6: when organized users are used, a coherent block appears in the similarity matrix, which is very bright and homogeneous; when unorganized users are used, the pattern becomes less pronounced. These block-like constructions are also compatible with coordinated manipulation behaviour that has been reported in recent studies on cybersecurity and privacy-preserving analytics [4, 11, 15, 19]. The numerical results are summarized in Table 1.

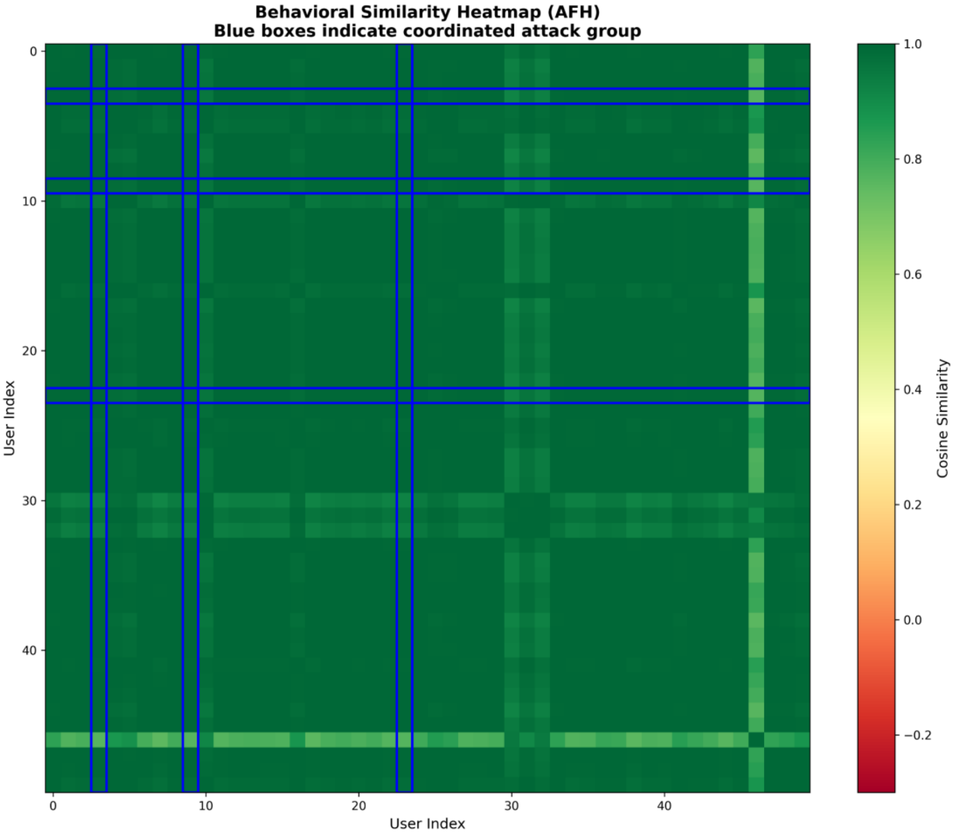

Figure 6. Heatmap of the similarity matrix computed from AFH-generated 8-dimensional hashed behavioural vectors.

The heatmap (Figure 6) clearly shows a bright, uniform block representing the coordinated attackers, each pair scoring 1.0. In contrast, similarity among normal users forms a more diffuse region, ranging between 0.93 and 1.0. This pattern reflects real coordinated attack behaviour, where malicious actors produce highly synchronized operations across accounts [4, 15].

While the similarity matrix reveals structural patterns at the pairwise level, a fuller understanding requires examining the distributional behaviour of similarity scores across user groups.

4.3 Distribution Analysis

The distribution of similarity values further supports the presence of detectable structural differences. As shown in Figure 7, coordinated attackers form a single spike at exactly 1.0, while normal users exhibit a varied distribution centered around ~0.988.

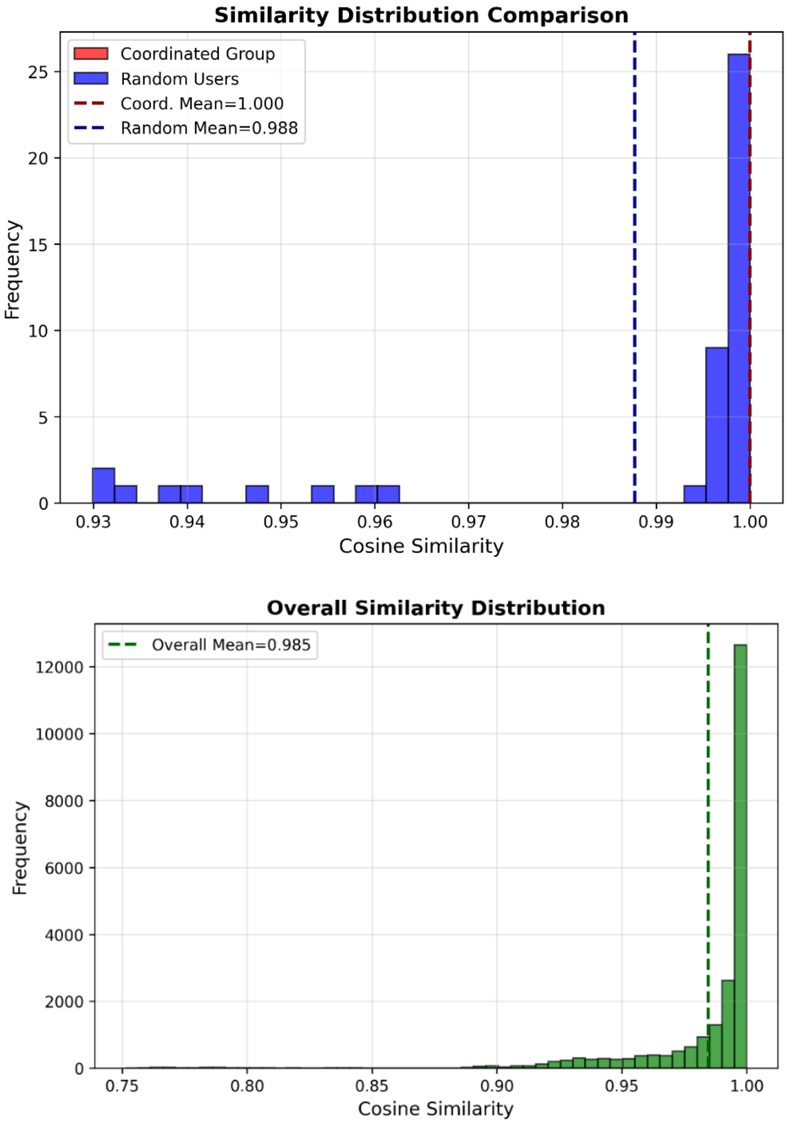

Figure 7. Distribution of similarity scores for coordinated attackers vs. random users.

Although a formal statistical test (e.g., KS test) was not applied, the separation is evident through distributional divergence:

- Attacker distribution: a degenerate point at 1.0
- Normal distribution: continuous with non-zero variance
- Support difference: attackers occupy a single value; normal users span a range

This divergence confirms that AFH preserves behavioural structure even after anonymization and dimensionality expansion, allowing coordinated patterns to remain distinguishable despite limited feature complexity.

Additionally, the perfect similarity observed in the attacker group arises from the intentional design of the synthetic adversaries. They were modelled with a 98% synchronized behavioural template to emulate real coordinated campaigns (e.g., bot-assisted harassment, synchronized reporting waves). Starting with a strongly coordinated synthetic baseline is standard in early-stage architectural research, enabling clearer measurement of lower-bound detection capability. These distributional characteristics provide a foundation for interpreting the behavioural implications of the observed separation.

4.4 Interpretation

The findings give rise to four major observations. To begin with, detectability is obvious: coordinated users show a perfect internal similarity (1.0000), and natural users show behavioural diversity. The absolute differentiation is not very large (1.23%) but at the same time, it is stable enough to show coordinated clusters according to the similarity approach.

Second, the system maintains privacy since the eight-dimensional hashed fingerprints are irreversible and reveal no reconstructable behavioural information. Third, the model supports transparency, as the use of three simple behavioural features only supports interpretability and is easy to explain, which is a key criterion to practical security systems. Lastly, it has high scalability; the eight-dimensional representation size allows the computation of similarities effectively, and this scale could be applied to large-scale, real-time cross-platform surveillance.

4.5 Limitations

The experiment has been associated with several limitations even though the outcomes are encouraging. The model uses only three behavioural characteristics of an activity; more complex or multi-dimensional representations may be more effective to discriminate between harmful and innocent users.

Besides, the findings are based on a simulated dataset, which, although necessary due to privacy and ethical aspects, may not be as detailed and diverse as real-world behaviour, and therefore, the validation with anonymised or industry-supplied data is necessary in the future. The behaviour coordinated is very synchronised and is always of a static nature, and real opponents often evolve with time. Lastly, the similarity gap observed is modest, even though consistent, in comparison to more sophisticated, feature rich methods of detection and thus additional improvements could be made to better the detection robustness.

4.6 Discussion of Findings

The results of this experiment indicate that anonymised behavioural fingerprints are sufficient to identify coordinated behaviour with the help of ACTCL, which proves its relevance to the concept of a privacy-preserving layer of cybersecurity. Several key insights emerge:

- Interpretability: ￼The model is based on easy-to-understand behavioural characteristics as opposed to non-understandable machine-learning pipelines, which allows a transparent justification of detection decisions.
- Behavioural robustness:
  After hashing, co-attackers are still perfectly similar, which means that synchronised malicious behaviour maintains robust structural signatures that are resilient to anonymisation.
- Privacy–utility balance:
  ACTCL will not handle or communicate raw user data, and only hashed fingerprints are compared, allowing cross-platform analysis without any privacy or regulatory limitations.
- Scalability:
  The eight-dimensional, compact, representation of vectors is computationally efficient and can be applied to millions of users in similarity analysis in real-time.
- Complementary security layer:
  ACTCL is not designed to replace the platform-level detection mechanisms; instead, it provides an architectural interface that allows the suspicious clusters of behaviours that are detected on isolated systems to be correlated.

In general, the results prove that despite a deliberately simplistic and minimalistic design, ACTCL can be effectively used to identify coordinated malicious behaviour. The stable separation patterns confirm the very idea of considering digital violence as a behavioural-correlation issue, but not a content/metadata-analysis problem, thus, establishing the basis to future extensions, which include more detailed behavioural characteristics, dynamic coordination modelling, and cross-platform validation in the real world.

4. Conclusion

This paper has shown that a small and privacy-sensitive representation of behaviour suffices to distinguish effectively between coordinated malicious and legitimate user behaviour. ACTCL converts three transparent behavioural characteristics, namely: temporal rhythm, concentration of activity and length of session into eight-dimensional hashed fingerprints to allow platforms to correlate user behaviour with each other, without the need to exchange raw data or sensitive metadata.

The empirical results indicate that coordinated attackers constitute a structurally distinct behavioural cluster where they attain a perfect internal similarity (1.0000) with zero variance and normal users have inherently different patterns (mean=.9877, SD=.0224). Even though the overall numerical distance between the group means is small (1.23 3.5 percent), the variance structure and distributional form differences are drastically different, which gives a reliable detection indicator that is analytically significant. This supports the fact that synchronised malicious behaviour leaves stable behavioural signatures that remain even with irreversible anonymisation and dimensionality expansion.

The ease and interpretability of the AFH-based approach provide many benefits to real-world implementation. The balance between privacy and utility is strong in ACTCL, allowing behavioural comparison, which does not reveal raw logs or identifiers, meeting the data-sensitivity and ethics restrictions that have recently been outlined in the privacy literature [12, 18]. The small eight-dimensional representation can be used to compute efficiently in large-scale as well, with real-time cross-platform similarity analysis being possible. Instead of replacing the existing detection pipelines, ACTCL is an additional architectural layer which reveals coordination patterns, which are opaque to the isolated, platform-specific security systems.

The empirical findings support ACTCL as an acceptable and lightweight base of future cross-platform threat intelligence. The proposed architecture offers a manageable mechanism to the process of detecting digital violence by streamlining the definition of the latter concept as a behavioural correlation problem, which offers a feasible channel towards the timely detection of coordinated malicious behaviour that might otherwise be invisible in fractured platform settings.

The architectural concepts upon which ACTCL was built have more general implications on privacy-aware security analytics beyond the instance of digital violence detection. The identical behavioural fingerprinting pipeline might also be used in cross-platform fraud correlation, coordinated misinformation detection, botnet analysis, and other threat-intelligence cases whereby data sharing is regulated or limited by platform boundaries. By showing that behavioural structure can still be observed despite irreversible anonymisation, this work prepares scalability to privacy-respectful security architectures that can run on the distributed digital ecosystems.

Future research will extend the model to include more behavioural capabilities, dynamic attacker modeling, and real-world multi-platform experimentation, further uniting its applicability to operational environments involving cybersecurity.

Authors' Declaration

Conflicts of Interest: The authors declare no conflict of interest.

Authors' Contribution Statement

All authors contributed equally to this work.

References

1. Alimov, R., & Shin, J. S. (2025). P3Fed: A Personalized and Privacy-Preserving Federated Framework for Intrusion Detection in Computing Power Network. Computer Networks.
2. Avalle, M., Di Marco, N., Etta, G., Sangiorgio, E., Alipour, S., Bonetti, A., Alvisi, L., Scala, A., Baronchelli, A., Cinelli, M. & Quattrociocchi, W. (2024). Persistent interaction patterns across social media platforms and over time. Nature, 628(8008), 582-589. https://doi.org/10.1038/s41586-024-07229-y
3. Bezzina, S., Antonetti, P., & Wagner, C. (2025). The lived experiences of online harassment by male perpetrators: A qualitative study. Journal of Gender Studies, 34(2), 145–160. https://doi.org/10.1007/s42380-025-00325-1
4. Chatzakou, D., Kourtellis, N., Blackburn, J., & Vakali, A. (2017). Mean birds: Detecting aggression and bullying on Twitter. arXiv. https://doi.org/10.48550/arXiv.1702.06877
5. Chen, L., et al. (2025). Adaptive Threat Attribution in Cross-Platform Environments: Developing a Framework for Fingerprinting APT Groups Across Cloud and On-Premise Infrastructure. World Journal of Advanced Research and Reviews, 27(2), 768–782.
6. Connolly, E. (2025). How does social media content go viral across platforms? Modelling the spread of "Kamala is brat" across X, TikTok, and Instagram. Journal of Information Technology & Politics.
7. de Chaves, S. A., & Benitti, F. (2025). User-centred privacy and data protection: An overview of current research trends and challenges for the human-computer interaction field. ACM Computing Surveys, 57(7), Article 176. https://doi.org/10.1145/3715903
8. Gupta, A., Matta, P., & Pant, B. (2025). An Integrated User Matching Framework for Cross-Platform Cyberbullying Detection. Discover Computing.
9. Hussain, A. A., Khaleel, I., & Al-Quraishi, T. (2024). Using Data Anonymization in big data analytics security and privacy. Mesopotamian Journal of Big Data, 2024, 118–127. https://doi.org/10.58496/MJBD/2024/009
10. Kouzani, A. Z., & Nouman, M. (2025). Detecting indicators of violence in digital text using deep learning. Natural Language Processing Journal, 12, Article 100175. https://doi.org/10.1016/j.nlp.2025.100175
11. Leonidou, N., Vafeas, A., Keramaris, K., & Katos, V. (2022). Privacy-preserving online content moderation. arXiv. https://doi.org/10.48550/arXiv.2209.11843
12. Li, M., Zhang, Y., & Wu, S. (2025). An investigation into personal data sensitivity in the Internet of Everything context. Humanities and Social Sciences Communications, 12(1), 224. https://doi.org/10.1057/s41599-025-04580-x
13. Lumare, N., Muradyan, L., & Sousa Jansberg, C. (2024). Behind the screen: the relationship between privacy concerns and social media usage. Journal of Marketing Communications, 1–16. https://doi.org/10.1080/13527266.2024.2424922
14. Mahmud, M., Tuhin, M. S. H., & Rahman, M. M. (2024). Privacy-preserving federated learning-based intrusion detection in IoT networks. Mathematics, 12(20), 3194. https://doi.org/10.3390/math12203194
15. Mishra, P., Yannakoudakis, H., & Shutova, E. (2019). Tackling online abuse: A survey of automated abuse detection methods. arXiv. https://doi.org/10.48550/arXiv.1908.06024
16. Nguyen, D. C., et al. (2021). Federated Learning for Internet of Things: A Comprehensive Survey. IEEE Communications Surveys & Tutorials.
17. Nguyen, T. D., Reijers, H. A., Kumar, R., & Zhang, H. (2025). Federated learning: A survey on privacy-preserving machine learning. arXiv.

https://arxiv.org/html/2504.17703v3

18. Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. Journal of Computer-Supported Cooperative Work, 28(1–2), 1–72. https://doi.org/10.3389/fdata.2019.00013
19. Rahman, M. M., Ahmed, F., & Saha, S. (2025). Federated learning for privacy-preserving data analytics in mobile applications. World Journal of Advanced Research and Reviews, 18(1), 157–168. https://doi.org/10.30574/wjarr.2025.26.1.1099
20. Zhou, X., Liang, X., Zhang, H., & Ma, Y. (2016). Cross-platform identification of anonymous identical users in multiple social media networks. IEEE Transactions on Knowledge and Data Engineering, 28(2), 411–424. https://doi.org/10.1109/TKDE.2015.2485222
21. Zimmeck, S., Li, J. S., Kim, H., Bellovin, S. M., & Jebara, T. (2017). A privacy analysis of cross-device tracking. Proceedings of the 26th USENIX Security Symposium (SEC'17), 1391–1408.