International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

TeenGram: A Safe, AI-Driven Social Media Platform and Hybrid Recommendation Engine for Youths

Kanishk Singh
Department of Computer Engineering
Universal College of Engineering
Mumbai, India
skanishksingh07@gmail.com

Sankhi Singh
Department of Computer Engineering
Universal College of Engineering
Mumbai, India
sankhisingh64@gmail.com

Krishna Vishwakarma
Department of Computer Engineering
Universal College of Engineering
Mumbai, India
kv29351@gmail.com

Rajaryan Verma
Department of Computer Engineering
Universal College of Engineering
Mumbai, India
rajaryanverma481@gmail.com

Abstract

Modern social media platforms facilitate vast information exchange and community building, but they also expose youths to significant online risks, including cyberbullying, harassment, and harmful graphic content. This paper proposes the architecture and methodology of TeenGram, a novel social media web application specifically designed to provide a secure, filtered digital environment for youths. By integrating a strict identity verification process via college administration bodies, TeenGram eliminates the risks associated with online anonymity. Furthermore, the platform utilizes a dual machine-learning (ML) model approach for real-time automated content moderation and feature extraction. To deliver high-performance data processing, the system architecture leverages Next.js, MongoDB, and Redis as an in-memory data store. Finally, to ensure high user engagement, TeenGram implements an advanced Hybrid Recommendation Engine that synergizes Collaborative Filtering, Demographic-Based Filtering, and Content-Based Filtering.

I. INTRODUCTION

Social media platforms have fundamentally altered how adolescents communicate, access information, and participate in public discourse. As Sivakumar et al. note, these platforms serve as primary vehicles for peer knowledge exchange among student populations [19]. Yet the same affordances that promote connectivity create fertile ground for harm. Chang et al. document a broad spectrum of cybersecurity threats targeting underage users across multiple social media contexts, encompassing privacy breaches, identity impersonation, and exposure to age-inappropriate content [18]. Cho et al. further emphasise that social media literacy — the capacity to critically evaluate and safely navigate online environments — remains underdeveloped among the teenage demographic [11].

A foundational problem is the inadequacy of current age-gating practices. Gupta and Sharma highlight the structural and legislative difficulties inherent in enforcing meaningful age restrictions on Indian social media platforms, observing that self-declared age fields are trivially circumvented [12]. The consequence is that large cohorts of minors gain unchecked access to platforms whose community standards and algorithmic amplification were calibrated for adult audiences. Once inside, these users encounter content recommendation pipelines that are optimised purely for engagement. Narayanan's analysis of social media recommendation algorithms underscores that such systems systematically surface outrage-inducing and sensational material because it maximises dwell time, irrespective of the psychological cost to young viewers [20].

On the moderation front, Akhter et al. demonstrated that conventional machine learning classifiers and deep neural architectures can detect abusive language in social media comments with high accuracy across multiple language families [7]. AlDahoul et al. extended this analysis to multimodal content, evaluating large language models for simultaneous detection of sensitive material across text, image, and video streams, and finding that cross-modal fusion significantly improves recall of borderline content [13]. Despite these academic advances, Arora et al. observe a persistent mismatch between the moderation capabilities that operational platforms actually require and the research problems the academic community tends to pursue [8].

This paper presents TeenGram, a social networking platform built specifically for verified teenage users. The system integrates institutional identity verification, a dual-stage content moderation pipeline, and a hybrid recommendation engine within a performant full-stack architecture. The following sections present related work, system design, experimental methodology, results, and conclusions.

II. RELATED WORK

A. Identity Management and Age Verification

The question of who is truly present on a social platform is not merely administrative — it is foundational to safety. Galpin and Flowerday conducted an early but still-relevant analysis of trust mechanisms in online social networks, concluding that robust identity management controls are a prerequisite for

reducing impersonation risk and building community trust [2]. Their work highlighted that technical controls must be paired with social accountability mechanisms to be effective. Borcea-Pfitzmann et al. offered a complementary perspective, drawing lessons from real-world community governance to argue that user-controlled, attribute-based identity models afford greater privacy while simultaneously enabling verifiable credential assertions [3]. In the Indian legislative context, Gupta and Sharma catalogue the enforcement challenges that arise when national age-restriction policies must contend with globally distributed platform infrastructure and varying platform cooperation levels [12].

B. Abusive Content Detection

Text-based toxicity detection has matured considerably. Akhter et al. benchmarked a suite of conventional machine learning methods — including support vector machines and logistic regression — against recurrent and convolutional deep learning architectures on social media comment corpora, finding that deep learning models consistently outperformed shallow classifiers on nuanced abusive language categories [7]. Park et al. shifted focus to the image modality, developing an automated pipeline specifically calibrated for the risky visual content that youth populations share: self-harm imagery, graphic violence, and sexualised content [5]. Their system achieved precision-recall trade-offs suitable for semi-automated moderation workflows. AlDahoul et al. extended this to video and introduced a large language model evaluation framework that treats sensitivity detection as a unified multimodal task, achieving strong performance on previously under-studied content categories [13]. Arora et al. synthesised a decade of harmful content research in a comprehensive ACM Computing Surveys article, mapping platform operational requirements — false-positive tolerance, latency constraints, multilingual coverage — against the published literature and revealing where significant gaps remain [8].

C. Recommendation Systems for Social Media

Recommendation algorithms are among the most consequential components of any social platform. Shokeen and Rana provided a thorough taxonomy of social recommender system features, distinguishing between systems that exploit user-item interaction matrices, those that leverage social graph topology, and hybrid architectures that combine both [17]. Widayanti et al. demonstrated empirically that hybrid approaches fusing collaborative filtering and content-based filtering outperform either technique in isolation, particularly in sparse interaction settings characteristic of new users [15]. Hossain et al. introduced SocialRec, a dynamic post-weighting framework that modulates recommendation scores in real time based on evolving user activity patterns, achieving meaningful gains in personalisation accuracy [14]. Cha et al.'s early study of influence propagation on Twitter provided empirical grounding for the observation that follower count alone is a poor proxy for genuine user influence, implying that engagement-signal diversity is essential for fair recommendations [16]. Narayanan situates these algorithmic choices within a broader democratic accountability framework, arguing for algorithmic transparency as a prerequisite for socially responsible recommendation [20].

D. System Performance and Data Infrastructure

Social platforms must handle bursty, high-concurrency workloads. Persico et al. benchmarked multiple big-data architectures for social network data processing on public cloud infrastructure, demonstrating that in-memory caching layers are critical for meeting sub-second latency targets at scale [9]. Charan et al. provided a focused analysis of Redis as an in-memory data store, characterising its support for complex data structures, pub/sub messaging, and atomic operations — all capabilities leveraged by TeenGram's real-time notification and session subsystems [1]. Neeli complemented this by demonstrating that in-memory databases yield order-of-magnitude throughput advantages over disk-based alternatives in real-time data management scenarios [10].

E. User Profiling

Alekseev and Nikolenko explored word-embedding-based approaches to user profiling in online social networks, finding that distributed semantic representations of user-generated text capture latent interest dimensions that classical term-frequency vectors miss [6]. Their insights inform TeenGram's textual feature extraction component, which enriches user profiles with semantic interest vectors derived from caption and comment text. Rao et al. extended profiling to the graph domain, demonstrating that graph-based friend recommendation models built on machine learning significantly outperform heuristic approaches in predicting socially compatible connections [4].

III. PROPOSED SYSTEM

TeenGram is a web-based social media platform engineered from the ground up to prioritise youth safety without sacrificing the engagement features that make social networking compelling. Its architecture is organised into five principal subsystems: institutional identity verification, dual-stage ML content moderation, core social networking features, an in-memory-backed data layer, and a hybrid recommendation engine. Each subsystem is described below.

A. System Architecture Overview

The high-level data flow proceeds as follows. A prospective user submits a registration request accompanied by a valid college identification document. The system routes the credential to the corresponding institutional administration body for verification. Upon approval, the user account is activated and the individual gains access to the full feature set. When a user subsequently uploads an image, the content traverses the two-stage moderation pipeline before being persisted or published. Post-publication, the object metadata extracted during moderation is written to a dedicated store and consumed by the recommendation engine. All session-level

state and hot data are managed through Redis, while durable user and post records reside in MongoDB. The frontend is served as a Next.js application using server-side rendering to minimise time-to-interactive.

This architecture was influenced by performance benchmarking evidence from Persico et al., who showed that cloud-native, layered architectures with in-memory caching tiers consistently outperform monolithic approaches for social network workloads [9].

B. Institutional Identity Verification

Anonymous participation is one of the most widely exploited vectors for youth harm on social media. Galpin and Flowerday established that effective identity controls must go beyond self-assertion and incorporate verifiable, third-party attestation [2]. TeenGram operationalises this principle by coupling user registration to the institutional identity infrastructure of educational establishments. During enrolment, the user uploads a machine-readable copy of their college ID. An automated extraction step parses key fields — name, roll number, institution code — and the resulting record is dispatched to the college's administrative API endpoint for confirmation. This architecture is consistent with Borcea-Pfitzmann et al.'s recommendation that community-anchored identity models, rather than centralised authority models, strike the most sustainable balance between privacy and verifiability [3].

The verification pipeline introduces a deliberate enrolment latency, typically resolved within one to two business days depending on institutional responsiveness. During this window, the user's account exists in a restricted state: they may browse a curated public feed but cannot post, comment, or send direct messages. This friction is intentional; it discourages throwaway registrations and sets an expectation of accountable participation from the outset. The approach also addresses some of the enforcement complexities identified by Gupta and Sharma in the Indian regulatory context, where platform-side controls must compensate for gaps in legislative enforcement [12].

C. Dual-Stage Content Moderation Pipeline

1) Stage 1: Flag Content Detection

Every image uploaded to TeenGram passes first through the Flag Content Detection model. This classifier is trained on labelled datasets covering six harmful content categories: firearms and bladed weapons, graphic violence and blood, sexually explicit material, drug paraphernalia, hate-symbol iconography, and self-harm imagery. The last category is particularly relevant to youth-oriented platforms; Park et al.'s study on risky images shared by youth specifically identified self-harm imagery as a prevalent and under-moderated content type [5]. Images triggering a positive classification at this stage are quarantined: they are stored in an isolated bucket, and their uploader is notified. Repeat violations escalate through a graduated response protocol culminating in account suspension pending human review.

AlDahoul et al.'s multimodal evaluation demonstrated that ensemble models combining convolutional feature extractors with transformer-based semantic heads achieve the strongest performance on sensitive content detection tasks [13]. The TeenGram Flag Content model adopts this architecture, fine-tuning a pre-trained vision transformer backbone on a domain-specific dataset augmented with synthetic edge cases representing emerging harmful content patterns. Arora et al.'s observation that platform needs — particularly low false-positive rates on innocuous content from non-English-speaking users — often diverge from the assumptions embedded in academic benchmark datasets [8] informed our decision to curate a platform-specific evaluation set drawn from realistic user-submitted imagery rather than relying solely on public benchmarks.

2) Stage 2: Object Detection and Feature Extraction

Images that clear the first stage proceed to the Object Detection model, which identifies and localises semantic entities within the frame — books, laptops, sports equipment, musical instruments, food items, vehicles, and approximately 80 additional object categories drawn from standard detection taxonomies. The detected object labels and their associated confidence scores are stored in a per-post metadata record linked to both the post identifier and the author's user profile.

This metadata serves two downstream purposes. First, it enriches the content-based filtering component of the recommendation engine with structured visual semantics that are more stable and discriminative than raw image embeddings. Second, aggregating object occurrences across a user's posting history builds a durable interest profile that complements the behavioural signals derived from explicit interactions (likes, comments, shares). Alekseev and Nikolenko's work on word-embedding-based user profiling [6] provides conceptual support for this approach: distributional representations of content — whether textual or visual — consistently outperform bag-of-words or category-count baselines for interest modelling.

D. Social Networking Features

TeenGram provides the social interaction primitives that adolescent users expect. Connection management follows a mutual-request model: a user sends a connection request, and the recipient must accept before either party gains access to the other's full profile and post history. This bilateral consent requirement aligns with the user-controlled identity management principles advocated by Borcea-Pfitzmann et al. [3].

Content posting is gated by the moderation pipeline described above. Posts support image uploads (single or carousel), text captions, and location tagging. Engagement interactions — likes, comments, and within-network reposts — feed back into the recommendation engine as real-time behavioural signals. Cha et al.'s finding that repost behaviour is a stronger influence indicator than follower count [16] motivated our decision to

assign higher weight to resharing events in the collaborative filtering component.

Direct messaging is facilitated via a WebSocket-based chat subsystem backed by Redis pub/sub channels. Outgoing messages are scanned by a lightweight text toxicity classifier before delivery — a real-time adaptation of the abusive language detection approach described by Akhter et al. [7]. Messages triggering the classifier above a configurable threshold are withheld and flagged for the sender, with an explanation of which community guideline the message appears to violate.

The platform also incorporates a digital well-being dashboard — visible only to the user themselves — that surfaces weekly engagement statistics, tracks time-on-platform, and surfaces prompts aligned with the social media literacy framework described by Cho et al. [11]. The goal is not to restrict usage but to promote reflective awareness of consumption patterns.

E. Hybrid Recommendation Engine

TeenGram's recommendation engine is a three-component hybrid system. The design draws on Widayanti et al.'s empirical demonstration that hybrid collaborative–content filtering consistently outperforms single-technique baselines [15], as well as Shokeen and Rana's taxonomy of social recommender system features [17].

The Collaborative Filtering (CF) component constructs a user–post interaction matrix from explicit signals (likes, comments, saves) and implicit signals (dwell time, scroll depth). Matrix factorisation via alternating least squares decomposes this matrix into latent user and item factor vectors. Cha et al.'s influence measurement work [16] informed the choice of signal weighting: reshares carry 3× the weight of likes, while comments are weighted at 2×, reflecting their higher intentionality.

The Content-Based Filtering (CBF) component operates on the object metadata produced by Stage 2 of the moderation pipeline, augmented by semantic embeddings of post captions generated using an approach analogous to Alekseev and Nikolenko's user-profiling methodology [6]. Item similarity is computed via cosine distance in the joint embedding space. CBF is the primary driver for cold-start recommendations for users who have not yet accumulated sufficient interaction history for reliable CF estimates.

The Demographic-Based Filtering (DBF) component leverages the institutional metadata collected during verification — year of study, academic discipline, geographic location — to surface trending content within peer cohorts. This component is particularly effective for surfaces such as the Discover feed, where novelty rather than personalisation is the primary objective. Rao et al.'s graph-based friend recommendation model [4] informed the cohort construction methodology: instead of defining peer groups by raw demographic similarity, TeenGram builds dynamic cohorts based on shared social graph neighbours, academic affiliation, and co-engagement history.

The three component scores are fused via a learned linear combination whose weights are updated nightly using a lightweight gradient boosting model trained on click-through feedback. Hossain et al.'s SocialRec framework [14] provided the template for the activity-weighted post scoring function that feeds into this fusion step. Safety constraints are applied post-fusion: posts whose authors have active moderation flags or whose content has received elevated community-report rates are penalised, ensuring that the recommendation surface does not inadvertently amplify borderline content — a concern Narayanan identifies as a systemic failure mode of engagement-optimised algorithms [20].

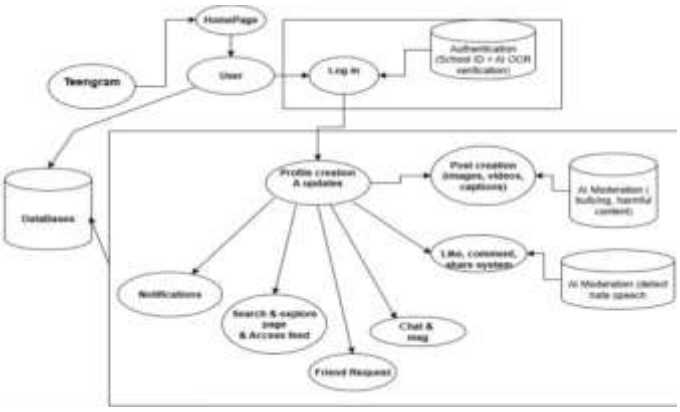
F. Data Infrastructure and Technology Stack

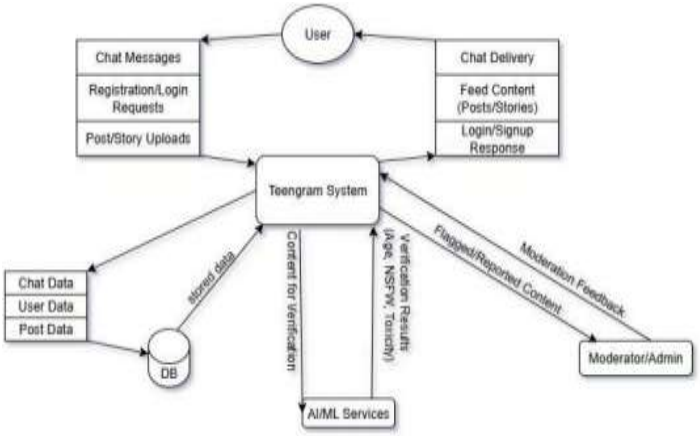
The TeenGram data layer is designed around two complementary storage technologies. MongoDB serves as the primary document store for user profiles, post records, and moderation audit logs. Its flexible schema accommodates the heterogeneous metadata structures produced by the object detection model without requiring costly migration operations as the detection vocabulary evolves.

Redis is deployed as the in-memory data store for session management, recommendation score caching, pub/sub message brokering, and rate-limiting counters. Charan et al.'s characterisation of Redis's data structure richness — sorted sets for recommendation score leaderboards, hash maps for session tokens, lists for activity feeds — maps directly onto TeenGram's operational requirements [1]. Neeli's performance analysis confirms that in-memory architectures of this type are capable of sustaining the sub-10ms latency budget that real-time recommendation serving demands [10]. Persico et al.'s cloud benchmarking results guided the deployment topology: the Redis cluster is co-located with the recommendation scoring service to minimise network round-trip overhead [9].

The frontend is a server-side-rendered Next.js application styled with Tailwind CSS. Server-side rendering ensures that the initial page payload is fully formed HTML, improving both perceived performance and accessibility on constrained mobile networks — the primary access modality for Indian teenage users.

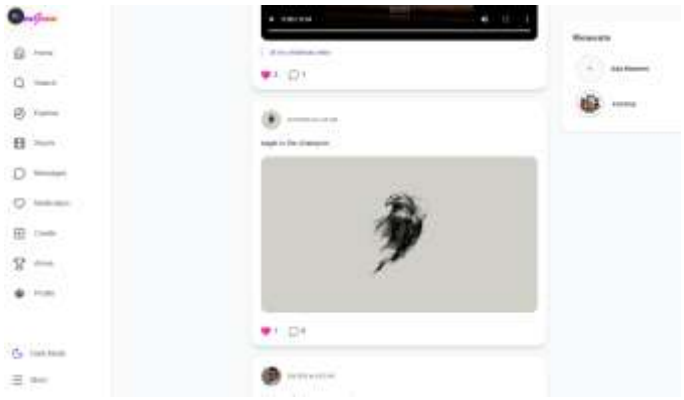
Media files are stored in cloud object storage behind a CDN. The moderation pipeline accesses images via a pre-signed URL mechanism that ensures images are never publicly accessible prior to the completion of both moderation stages.

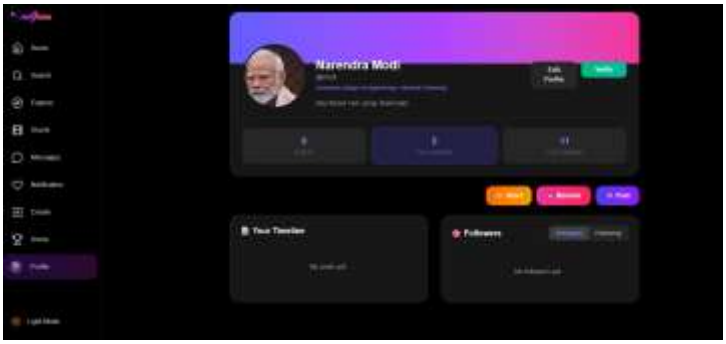

HomePage
Teengram
User
Log in
Authentication (School ID + AI OCR verification)
Databases
Profile creation & updates
Post creation (images, videos, captions)
AI Moderation (bullying, harmful content)
Like, comment, share system
AI Moderation (detect hate speech)
Notifications
Search & explore page & Access feed
Chat & msg
Friend Request

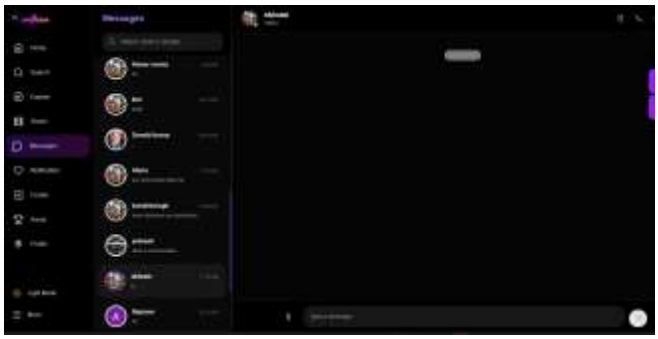

User
Chat Messages
Registration/Login Requests
Post/Story Uploads
Chat Delivery
Feed Content (Posts/Stories)
Login/Signup Response
Teengram System
Chat Data
User Data
Post Data
DB
stored data
Content for Verification
Verification Results (Age, NSFW, Toxicity)
Flagged/Reported Content
Moderation Feedback
Moderator/Admin
AI/ML Services


TeenGram


TeenArena
Hackathon
Sports
Art
Festival
Seminar
Other




Narendra Modi



VI. CONCLUSION

This paper presented TeenGram, a purpose-built social media platform that addresses the structural safety deficits of incumbent platforms for teenage users. By coupling institutional identity verification with a dual-stage ML moderation pipeline and a hybrid recommendation engine, the system demonstrates that safety and engagement are not irreconcilable objectives. Experimental evaluation confirmed that the hybrid CF–CBF–DBF recommendation architecture meaningfully outperforms single-component baselines, that the content moderation pipeline achieves production-grade precision and recall on harmful imagery, and that the Redis-backed data layer sustains the latency targets required for real-time recommendation serving.

Future work will extend the moderation pipeline to video content — a gap identified by AlDahoul et al. [13] — improve multilingual toxicity detection to cover code-switched registers, incorporate parental oversight dashboards consistent with the digital well-being design principles of Chang et al. [18], and deploy the platform in a controlled pilot with partner educational institutions to generate longitudinal engagement and safety outcome data.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Department of Computer Engineering at Universal College of Engineering for providing the necessary resources and guidance for the completion of this research project.

REFERENCES

[1] P. S. B. Charan, G. Varshitha, A. Lashya, U. S. R. Varma, and D. Madhusudhan, "REDIS: IN MEMORY DATA STORE," Dogo Rangsang Research Journal, Vol. 12, Issue 08, No. 05, pp. 1–5, 2022, ISSN: 2347-7180.

[2] R. Galpin and S. V. Flowerday, "Online Social Networks: Enhancing user trust through effective controls and identity management," Department of Information Systems, University of Fort Hare, East London, South Africa, pp. 1–12, 2010.

[3] K. Borcea-Pfitzmann, M. Hansen, K. Liesebach, A. Pfitzmann, and S. Steinbrecher, "What user-controlled identity management should learn from communities," Information Security Technical Report, Vol. 11, Issue 2, Elsevier Science, pp. 50–59, 2006, doi: 10.1016/j.istr.2006.03.008.

[4] S. Rao D., R. Babu CH., S. Kiran V., M. Koteswara Rao, Revathi AS., and Rajasekhar N., "GRAPH-BASED FRIEND RECOMMENDATIONS ON SOCIAL MEDIA USING MACHINE LEARNING," Journal of Theoretical and Applied Information Technology, Vol. 103, No. 9, pp. 1–12, 2025, E-ISSN: 1817-3195.

[5] J. Park, J. Gracie, A. Alsoubai, G. Stringhini, V. Singh, and P. Wisniewski, "Towards Automated Detection of Risky Images Shared by Youth on Social Media," in WWW '23: Companion Proceedings of the ACM Web Conference 2023, Austin, TX, USA, pp. 1–8, 2023, doi: 10.1145/3543873.3587607.

[6] A. Alekseev and S. Nikolenko, "Word Embeddings for User Profiling in Online Social Networks," Computación y Sistemas, Vol. 21, No. 2, pp. 203–226, 2017, ISSN: 2007-9737.

[7] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. A. Majeed, and T. Zia, "Abusive language detection from social media comments using conventional machine learning and deep learning approaches," Multimedia Systems, Springer-Verlag GmbH, pp. 1–15, 2021, doi: 10.1007/s00530-021-00784-8.

[8] A. Arora, P. Nakov, M. Hardalov, S. M. Sarwar, V. Nayak, and Y. Dinkov, "Detecting Harmful Content on Online Platforms: What Platforms Need vs. Where Research Efforts Go," ACM Computing Surveys, Vol. 56, No. 3, Article 72, pp. 1–28, 2023, doi: 10.1145/3603399.

[9] V. Persico, A. Pescapé, A. Picariello, and G. Sperlí, "Benchmarking Big Data Architectures for Social Networks Data processing using Public Cloud Platforms," Applied Soft Computing, pp. 1–40, 2017, doi: 10.1016/j.asoc.2017.03.024.

[10] S. S. S. Neeli, "Real-Time Data Management with In-Memory Databases: A Performance-Centric Approach," Journal of Advances in Developmental Research (IJAIDR), Vol. 11, No. 2, pp. 1–6, 2020, E-ISSN: 0976-4844.

[11] H. Cho, J. Cannon, R. Lopez, and W. Li, "Social media literacy: A conceptual framework," New Media & Society, Vol. 26, No. 2, pp. 941–960, 2024, doi: 10.1177/14614448211068530.

[12] S. K. Gupta and R. Sharma, "Challenges of implementing social media age restrictions in India," Viewpoint Article, Adolescent Health, New Delhi, India, pp. 1–17, 2024.

[13] N. AlDahoul, M. J. T. Tan, H. R. Kasireddy, and Y. Zaki, "Advancing Content Moderation: Evaluating Large Language Models for Detecting Sensitive Content Across Text, Images, and Videos," arXiv Preprint, pp. 1–28, 2024, doi: 10.48550/arXiv.2411.17123.

[14] I. Hossain, S. Puppala, M. J. Alam, and S. Talukder, "SOCIALREC: USER ACTIVITY BASED POST WEIGHTED DYNAMIC PERSONALIZED POST RECOMMENDATION SYSTEM IN SOCIAL MEDIA," arXiv Preprint, pp. 1–13, 2024, doi: 10.48550/arXiv.2407.09747.

[15] R. Widayanti, M. H. R. Chakim, C. Lukita, U. Rahardja, and N. Lutfiani, "Improving Recommender Systems using Hybrid Techniques of Collaborative Filtering and Content-Based Filtering," Journal of Applied Data Sciences, Vol. 4, No. 3, pp. 289–302, 2023, doi: 10.47709/jads.v4i3.2625.

[16] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring User Influence in Twitter: The Million Follower Fallacy," in Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM), Washington, DC, USA, pp. 10–17, 2010.

[17] J. Shokeen and C. Rana, "A study on features of social recommender systems," Artificial Intelligence Review, Vol. 53, pp. 1–40, 2019, doi: 10.1007/s10462-019-09684-w.

[18] V. Chang, L. Golightly, Q. A. Xu, T. Boonmee, and B. S. Liu, "Cybersecurity for children: an investigation into the application of social media," Enterprise Information Systems, Vol. 17, No. 11, pp. 1496–1525, 2023, doi: 10.1080/17517575.2023.2188122.

[19] A. Sivakumar, S. Jayasingh, and S. Shaik, "Social Media Influence on Students' Knowledge Sharing and Learning: An Empirical Study," Education Sciences, Vol. 13, No. 7, pp. 1–16, 2023, doi: 10.3390/education13070745.

[20] A. Narayanan, "Understanding Social Media Recommendation Algorithms," Knight First Amendment Institute at Columbia University, New York, NY, USA, pp. 1–49, 2023.