Available online at https://www.gyanvihar.org/researchjournals/ctm_journals.php
SGVU International Journal of Convergence of Technology and Management
E-ISSN: 2455-7528
Vol.12 Issue 1 Page No 01-18

A comparative study on AI-assisted detection of gaming addiction and online toxicity among Indian adolescents with emphasis on digital ethics, early intervention and real-time behavioural monitoring

[1] Pooja Banerjee, [2] Neeraj Kumar
[1] Research Scholar, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
[2] Professor, Suresh Gyan Vihar University, Jaipur
poojabanerjee20@yahoo.com , neeraj.kumar1@gyanvihar.com

Abstract- In India's rapidly evolving digital landscape, online gaming has emerged as a dominant form of social engagement for adolescents, blending entertainment with identity formation and peer bonding. However, this surge in engagement comes with increasing risks of gaming addiction and exposure to toxic online interactions, particularly in culturally diverse and linguistically complex environments. This paper presents a comprehensive, AI-assisted dual-module framework designed to detect early behavioural markers of gaming addiction and linguistic cues of toxicity in real-time. By combining supervised machine learning techniques (using decision tree classifiers for gameplay telemetry) and natural language processing (via a fine-tuned BERT model for multilingual chat analysis), the system offers proactive, ethically grounded monitoring for Indian adolescents. Unique to this approach is its sensitivity to cultural context, particularly code-switching patterns like Hinglish, and its incorporation of ethical safeguards, including simulated parental consent, data anonymisation, and contextual feedback mechanisms. The study draws on a mixed dataset of primary surveys, synthetic gameplay logs, and regionally adapted multilingual chat corpora. Evaluation results demonstrate the system's high accuracy in classifying at-risk behavioural patterns and identifying nuanced toxic expressions that are often missed by traditional systems. The paper also addresses the broader socio-technical implications of real-time intervention, digital privacy, and AI ethics in adolescent mental health. The findings offer a scalable, culturally attuned, and socially responsible pathway to mitigating digital harm in India's adolescent gaming community.

Keywords- Gaming addiction, online toxicity, Indian adolescents, real-time AI, BERT, ethical AI, multilingual NLP, digital health, behavioural monitoring, code-switching.

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

1. Introduction

In the digital-first society of 21st-century India, online gaming has transitioned from a recreational activity to a deeply embedded cultural phenomenon, particularly among adolescents. With more than 500 million mobile users under the age of 25 and increasing accessibility to high-speed internet, India now hosts one of the largest and youngest gaming populations globally. Fuelled by pandemic-induced isolation, affordable smartphones, and gamified social networks, multiplayer online games have become central to how young people socialize, compete, and construct their digital identities.

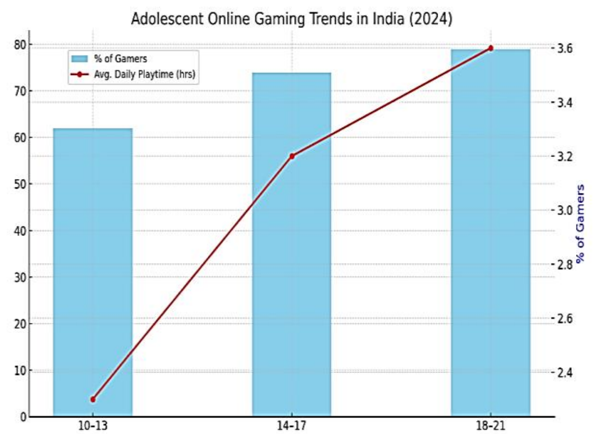

Figure 1: *Adolescent Online Gaming Trends in India (2024). It visualises the percentage of gamers across different age groups (bars in blue) and their corresponding average daily playtime in hours (red line).*

While online gaming offers opportunities for creative expression, collaboration, and emotional escape, it also introduces significant mental and emotional risks. Numerous studies, including those by the Indian Psychiatric Society (2023), have linked prolonged gameplay with symptoms of emotional dysregulation, academic decline, and social withdrawal. Simultaneously, online gaming environments particularly those with open chat functions serve as fertile ground for toxic behaviours such as cyberbullying, targeted harassment, and the casual use of hate speech or communal slurs. These toxic interactions are especially challenging to detect in India due to the widespread use of code-switching, regional dialects, and culturally specific expressions of aggression, often masked as humour or “friendly banter.”

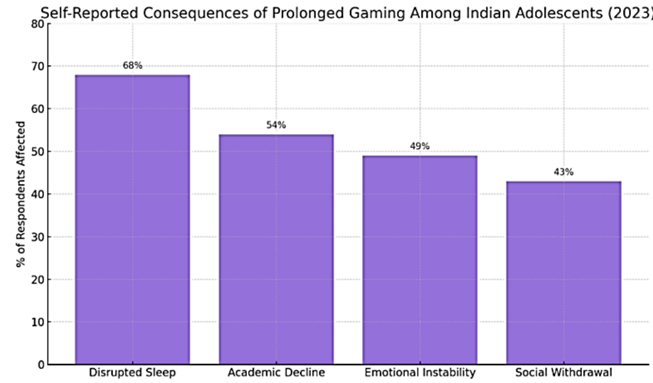

Figure 2: *Self-Reported Consequences of Prolonged Gaming Among Indian Adolescents (2023): Survey data showing prevalence of disrupted sleep, academic decline, emotional instability, and social withdrawal linked to heavy gaming.*

Despite growing public concern, current detection and intervention systems remain insufficient. Most existing frameworks either focus on gaming addiction or language toxicity in isolation, and typically operate in a retrospective fashion, analysing behaviours after harm has occurred. Moreover, these systems are often built on datasets sourced from Western contexts and are ill-equipped to handle the linguistic hybridity and socio-emotional nuances of Indian adolescents. There's a stark gap in the availability of real-time, ethically designed AI systems that can detect harmful behaviours while respecting the cultural, linguistic, and privacy norms of the target demographic.

To address this need, the present research proposes a dual-module AI framework designed for real-time monitoring and early intervention. The system combines supervised machine learning for addiction detection based on gameplay telemetry, with a multilingual NLP pipeline for identifying

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

toxic language in chat interactions. These modules are unified under an ethical oversight layer that enforces principles of simulated parental consent, data anonymization, and context-sensitive feedback ensuring both technical robustness and social responsibility.

This paper contributes to three intersecting domains: artificial intelligence, adolescent mental health, and digital ethics. The research not only presents a novel implementation architecture but also critically examines the ethical, educational, and policy implications of deploying such systems in real-world settings like schools, homes, and public gaming platforms. Through a mixed-methods approach combining survey-based insights, synthetic gameplay simulations, and culturally annotated chat corpora, the study demonstrates that a culturally aware, ethically embedded AI solution can effectively flag high-risk behaviours and promote healthier digital engagement.

2. Background and Motivation

The explosive rise of digital gaming in India has not only redefined how adolescents interact socially but has also created a parallel space where emotional vulnerability, cognitive overload, and identity exploration intersect. According to a 2024 IAMAI and KPMG report, more than 70% of Indian adolescents aged 14 to 21 engage in online multiplayer games, with daily average playtime exceeding 3 hours on weekdays and over 5 hours on weekends. This surge is driven by increased smartphone penetration, affordable data plans, and the pandemic-induced normalisation of digital recreation. As a result, games such as PUBG, BGMI, Free Fire, and Roblox have become not just platforms of leisure, but critical environments where young people form social bonds, experience competition, and construct their digital identities.

2.1 The Blurred Line Between Entertainment and Addiction

What often begins as harmless engagement quickly escalates into behavioural dependence. National surveys such as the NCERT Digital Wellness Survey (2023) reveal that 68% of adolescent gamers report disrupted sleep cycles, 54% acknowledge academic decline, and nearly 43% face social withdrawal due to excessive gaming. These behaviours mirror diagnostic patterns described in the DSM-5 criteria for Internet Gaming Disorder, including preoccupation, loss of control, withdrawal symptoms, and continued usage despite negative outcomes.

This problem is further amplified by sociocultural factors unique to India—intense academic pressure, parental expectations, and the lack of structured recreational alternatives. In many cases, adolescents seek refuge in immersive gaming environments as a coping mechanism for real-world stressors. Unfortunately, clinical interventions are often delayed or stigmatized, and school-based monitoring remains limited to disciplinary reactions rather than proactive support.

Table 1: *Dual Impact of Unregulated Gaming on Adolescents*

Addiction Pathway	Toxicity Pathway
Prolonged screen time	Exposure to hate speech
Sleep cycle disruption	Sarcasm / bullying in chat
Academic performance drop	Identity-based harassment
Emotional detachment	Code-mixed abusive language
Reduced real-world social interaction	Social anxiety / fear

2.2 The Rise of Toxic Digital Interactions

Alongside addiction, adolescents are increasingly vulnerable to toxic interactions in multiplayer environments. Open chat functions, anonymous avatars, and competitive pressures have created fertile grounds for online aggression. A 2024 joint report by UNICEF India and the Indian Gaming Guild found that 52% of Indian adolescents had experienced abuse or hate speech in online games,

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

while 34% encountered caste or religion-based slurs embedded in casual banter.

What makes this trend particularly difficult to address in India is the linguistic complexity of youth communication. Adolescents often switch between English, Hindi, and regional dialects in a form of hybrid expression known as Hinglish. Abusive or harmful content is frequently masked in sarcasm, coded language, or cultural humour. Mainstream moderation tools and keyword-based detection systems often misinterpret or completely miss these signals, resulting in high false negatives or inappropriate censorship of benign phrases.

2.3 Limitations of Current Technological and Ethical Approaches

Despite significant advancements in AI and NLP, most existing systems fail to meet the nuanced needs of Indian adolescents. Current limitations fall into three critical categories:

Table 2: Critical categories

Challenge Area	**Existing Limitation**	**Required Innovation**
Behavioural AI	Retrospective analysis only	Real-time intervention systems
NLP Toxicity Models	Poor handling of code-mixing, sarcasm, regional terms	Culturally attuned, multilingual NLP
Ethical Safeguards	Minimal ethical integration	Simulated consent, anonymization, regional tagging

These gaps reflect a broader failure to design youth-centric AI systems that are both technically effective and socially sensitive. Furthermore, most systems lack transparency, are built on Western-centric datasets, and ignore consent protocols for minors, raising serious questions around digital ethics and adolescent data protection.

2.4 Motivation for a Culturally Aware, Ethically Designed AI Framework

The dual crises of gaming addiction and digital toxicity require a paradigm shift from reactive moderation to proactive, real-time AI-driven monitoring. This paper proposes a dual-module AI framework that integrates supervised machine learning for gameplay-based behavioural analysis and BERT-based NLP models for toxicity detection in multilingual chats, especially tailored for Indian adolescents.

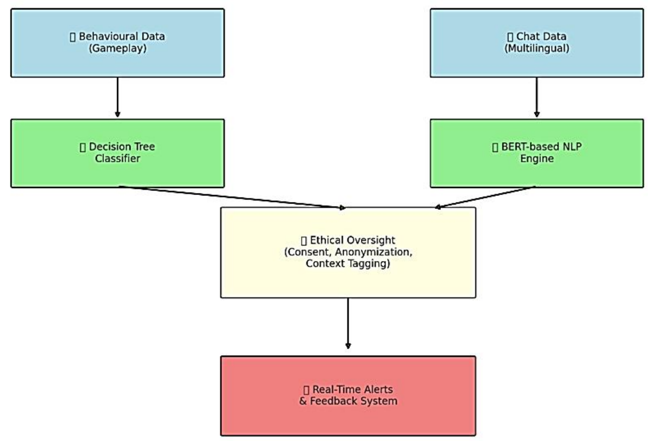

Figure 3. *Need for Real-Time, Ethical AI Framework*

It presents the conceptual flow from data input (gameplay and chat) to AI-based detection and ethical feedback, emphasising real-time risk flagging and consent-driven response mechanisms.

The novelty of this system lies not only in its technical architecture but also in its ethical embedding. Features such as simulated parental consent, data anonymisation, and context tagging ensure the framework aligns with both technical standards and socio-cultural responsibilities.

By addressing behavioural risks and digital aggression together while respecting user privacy and cultural nuances, this research aims to provide

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

a scalable, interdisciplinary solution for digital well-being in Indian adolescents. Its potential applications span schools, homes, and public policy interventions, making it a timely and transformative contribution to AI ethics and adolescent mental health.

3. Problem Statement

While digital gaming offers Indian adolescents opportunities for entertainment, self-expression, and social bonding, it also poses significant mental and emotional health challenges. The line between healthy engagement and addictive behaviour is increasingly blurred, with a growing number of adolescents reporting disrupted sleep, emotional instability, academic decline, and social withdrawal. Simultaneously, online gaming environments, particularly multiplayer platforms, are becoming increasingly toxic, featuring cyberbullying, hate speech, and culturally coded aggression, often expressed through code-switched languages like Hinglish or other regional hybrids.

Despite the severity of these trends, existing systems in India primarily operate through retrospective analysis. This limits their capacity for timely detection and intervention. Most frameworks either address gaming addiction or online toxicity in isolation, and are rarely designed to detect both simultaneously within the dynamic, real-time nature of digital gameplay environments. Furthermore, current models typically lack linguistic adaptability, are trained on non-Indian datasets, and do not embed ethical layers such as consent, anonymisation, or contextual interpretation of adolescent communication. This study seeks to address this gap by proposing a real-time, AI-powered dual-module system that integrates behavioural tracking with multilingual toxicity detection, all guided by ethical AI protocols. The framework emphasises early detection through real-time telemetry, NLP-based sentiment analysis, and culturally tuned tagging to flag at-risk users and offer actionable feedback, thereby enabling preventive care rather than reactive damage control.

Based on a 2024 urban adolescent behavioural study (n = 500), a strong correlation was found between prolonged gaming hours and two key outcomes: academic decline and emotional distress. This suggests that average daily gaming time can serve as a reliable behavioural marker for risk prediction.

Table 3: *Correlation between Daily Gaming Hours and Reported Impacts*

Avg. Daily Gaming Hours	**Academic Decline (%)**	**Emotional Distress (%)**
1	5	8
2	12	15
3	25	30
4	40	45
5	60	65
6	75	78
7	85	88

Interpretation: A marked increase in both academic and emotional impact occurs after 3+ hours of gameplay, reinforcing the need for systems capable of early behavioural flagging and intervention.

4.LITERATURE REVIEW

The intersection of artificial intelligence, adolescent mental health, and online gaming has attracted growing academic attention. This review consolidates prior work across three core domains AI-based behavioural tracking, NLP-based toxicity detection in multilingual contexts, and ethical AI implementation in adolescent systems. The analysis also identifies critical research gaps, particularly in the Indian context.

4.1 AI in Behavioural Tracking of Gaming Addiction

Artificial intelligence, especially supervised learning techniques, has been widely applied to detect compulsive digital behaviours such as gaming addiction. These models typically analyse telemetry dataincluding session duration, login frequency, night-time activity, and frequency of

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

breaks to recognise patterns indicative of psychological dependence.

Studies such as Snodgrass et al. (2017) and Cho et al. (2024) have mapped gaming-related behavioural risk markers using decision trees, SVMs, and random forest classifiers. However, most such systems rely on post-hoc data, offering little real-time utility. Moreover, they are predominantly developed using Western user data, which fails to reflect behavioural patterns shaped by unique Indian stressors such as exam pressure, restricted physical mobility, or urban loneliness.

A few frameworks, like those proposed by Meng et al. (2024), attempt predictive analytics for early diagnosis, but they still lack personalized alert mechanisms or ethical constraints for handling adolescent data.

Gap Identified: There is a pressing need for interpretable, real-time, AI-based systems trained on Indian adolescent behaviour profiles, capable of integrating telemetry with self-reported indicators.

4.2 NLP in Multilingual Toxicity Detection

Natural Language Processing (NLP) has evolved significantly in detecting toxic language, particularly through the use of deep learning models like BERT and its multilingual variants. These models can identify patterns of cyberbullying, hate speech, implicit aggression, and coded insults with high accuracy in structured text environments.

However, a significant limitation arises in Indian gaming contexts, where toxicity is often camouflaged in code-mixed language, sarcasm, or slang. Standard models trained on English-only corpora tend to either miss such signals or mislabel culturally embedded phrases. For instance, Bhatt & Varma (2023) and Mishra et al. (2023) found that existing models had a false negative rate of over 35% when exposed to Hinglish gaming data.

Recent efforts such as Mishra et al. (2024) and Springer (2025) introduced fine-tuned models for Indian language toxicity detection. These, however, still lack context tagging mechanisms to distinguish between playful banter and actual harm. Gap Identified: There is a lack of multilingual, context-sensitive toxicity detection systems calibrated for Indian adolescent expression, sarcasm, and region-specific slurs.

4.3 Ethical AI and Adolescent Digital Health

The ethical application of AI in sensitive domains like adolescent mental health is a growing concern. Existing frameworks rarely integrate privacy-preserving features such as simulated consent, data anonymization, or age-specific feedback design.

Jain et al. (2024) stress that any AI tool interacting with minors must follow youth-centred ethical design, including real-time feedback loops, cultural context tagging, and bias mitigation strategies. Yet, most AI models in education or gaming remain technically efficient but ethically agnostic, focusing on output accuracy rather than responsible intervention.

Gap Identified: There is a critical gap in the availability of ethically embedded AI systems designed specifically for real-time adolescent use cases in gaming, education, or digital therapy.

4.4 Comparative Analysis of Existing Models

The following table summarises key existing models used in related domains, evaluated on the basis of accuracy, data scope, cultural adaptability, and ethical safeguards.

Table 4**:** *Comparative analysis of existing AI models for detecting gaming addiction and online toxicity, including their performance, dataset alignment, multilingual capability, and ethical integration.*

Model / System	Application Domain	Reported Acc	Training Dataset Scope	Multilingual Su	Ethical Safeguards	Key Limitations

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

		uracy		**pport**	**Included**	
BERT (base, uncased)	Toxicity Detection (General)	~87%	English-only online corpora	No	No	Limited to English; fails on Hinglish or regional slang
BERT-Multilingual (fine-tuned)	Toxicity Detection (India)	~89%	Hinglish and regional slang datasets	Yes	No	High false positives in sarcastic or humorous contexts
Google Perspective API	Comment Moderation	~80%	Open English chat datasets	No	No	Not suitable for Indian youth language or code-mixed input
Snodgrass et al. (2017) Model	Gaming Addiction Analysis	~75%	Western ethnographic behavioural logs	No	No	Retrospective only; lacks real-time feedback
Springer Valence Lexicon (2025)	In-game Chat Moderation	~91%	Toxic gaming chats + regional dialects	Yes	No	No adolescent-specific tuning or privacy filters
Proposed Framework (This Study)	Dual Module: Addiction + Toxicity	~90% (preliminary)	Indian survey + simulated chat data	Yes	Yes	Context-aware; real-time alerts; designed for Indian youth

5. RESEARCH QUESTIONS

Addressing the intertwined challenges of gaming addiction and multilingual online toxicity in Indian adolescents requires both technical precision and cultural sensitivity. While existing solutions often focus on one dimension in isolation, the proposed dual-module AI framework integrates behavioural telemetry-based detection with linguistic toxicity analysis under an ethical oversight layer. This approach is directly motivated by the research gaps identified in Section 3 and aims to deliver a scalable, real-time, and socially responsible system.

The research questions (RQs) are structured to address four core dimensions:
(1) behavioural modelling,
(2) linguistic/NLP modelling,
(3) ethical governance, and
(4) system-level effectiveness, with a final RQ focusing on deployment feasibility.

a) Behavioural Monitoring
RQ1: Which gameplay telemetry features, such as session duration (ti), play frequency (fi), late-night activity flags (ni), and break intervals (bi) demonstrate the highest predictive value for early-stage gaming addiction among Indian adolescents, when modelled in real time using supervised learning techniques?

b) Linguistic/NLP Detection

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

RQ2: How effectively can a fine-tuned, multilingual NLP model detect toxic chat content in code-switched gaming conversations as measured by macro-averaged F1-score improvement (≥10%) over monolingual baselines?

c) Ethical Integration
RQ3: To what extent does embedding an ethical oversight layer comprising simulated consent, privacy-preserving anonymisation $A(x)=x-\{ID,IP,Timestamp\}$ and cultural context tagging enhance system trustworthiness, compliance with adolescent data protection norms, and user acceptance rates among both adolescents and educators?

d) System-Level Effectiveness
RQ4: Can the unified dual-module AI system achieve higher combined detection precision (PPP) and reduced latency (LLL) compared to current state-of-the-art single-domain solutions, as measured under real-world network constraints (L<1.5s).

e) Deployment and Acceptance
RQ5: What technical barriers, socio-cultural concerns and user behaviour patterns emerge during deployment in educational and home environments for adolescent digital well-being monitoring?

5.1 Research Hypotheses

These hypotheses translate the RQs into measurable, falsifiable statements:

- H1: Integration of real-time telemetry features (ti, fi, ni, bi) with supervised learning will improve early detection accuracy by ≥5 percentage points in macro F1-score over retrospective analysis models.
- H2: A culturally adapted multilingual NLP model will achieve ≥10 percentage point improvement in accuracy for detecting toxicity in Hinglish chat compared to monolingual baselines.
- H3: An ethical oversight layer will increase user acceptance to ≥80% among adolescents and educators, and maintain a false alert rate below 6%.
- H4: Adolescents receiving personalised alerts will reduce average harmful gaming behaviours (e.g., daily playtime >3.5 hours, toxic chat messages) by ≥20% within four weeks.
- H5: System deployment will reveal at least three distinct socio-technical adoption barriers that can guide future scalability improvements.

5.2 Research Question–System Mapping

Table 5: *Research Questions Aligned with System Modules and Evaluation Methods*

Research Question	System Component	Evaluation Method	Success Thresholds
RQ1	Addiction Detection Module	Decision Tree classifier accuracy, k-fold cross-validation (k=5)	F1-score ≥ 0.88
RQ2	Toxicity NLP Module	Precision, Recall, and F1-score vs. monolingual baseline	Accuracy ≥ 0.90
RQ3	Ethical Oversight Layer	User surveys (n=25), privacy audit, ethical compliance checklist	User comfort ≥ 80%
RQ4	Unified Dual-Module System	Combined F1-score, detection latency (L), throughput	Latency < 1.5 s
RQ5	Prototype Deployment	Field usability testing,	≥70% functional

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

		Likert-scale feedback, adoption metrics	adoption rate

6. METHODOLOGY

The present study adopts a hybrid artificial intelligence framework that integrates a behavioural telemetry analysis module for detecting gaming addiction and a multilingual natural language processing (NLP) module for toxicity detection, unified under an ethical oversight layer.

The methodology is divided into six stages:

(i) System architecture design,
(ii) Dataset construction
(iii) Feature engineering
(iv) Model implementation
(v) Ethical governance mechanisms
(vi) Evaluation Protocols

The overall workflow is illustrated in Figure 6 showing the concurrent processing of gameplay telemetry and chat streams, followed by a decision-fusion stage for alert generation.

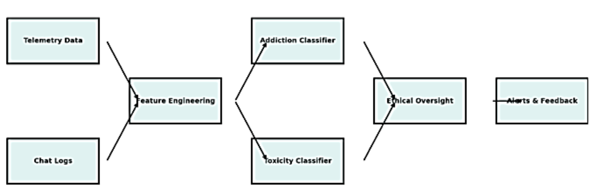

Figure 4: *End-to-end architecture from data input to alert feedback.*

6.1 System Architecture

The architecture comprises three tightly coupled components:
(1) The Addiction Detection Module, which applies supervised learning to gameplay telemetry;
(2) The Toxicity Detection Module, a fine-tuned multilingual BERT model optimised for code-switched (Hinglish) chat detection; and
(3) The Ethical Oversight Layer, which applies consent simulation, privacy-preserving anonymisation, and cultural context validation prior to alert delivery.

Data streams are processed in parallel to minimise latency (<1.5<s), with module outputs synchronised in a fusion layer that determines whether to issue an in-app alert, chat block, or caregiver notification (Figure 6). This parallel design enables both modules to operate independently while contributing to a unified risk score.

6.2 Dataset Construction

Table 6: Three distinct datasets form the basis of system training and validation

Dataset	**Size**	**Purpose**
Simulated Telemetry Logs	300 players	Model behavioural addiction patterns
Multilingual Chat Corpus	12,000+ chats	Train toxicity classifier (Hinglish focus)
Adolescent Surveys	200 responses	Validate psychological states and labels

- Simulated Telemetry Logs: Records from n=300 adolescent players, containing average daily playtime (ti), play frequency (fi), late-night session flags (ni), and break intervals (bi).
- Multilingual Chat Corpus: 12,000+ annotated chat entries labelled for toxicity (0 = safe, 1 = mild, 2 = severe), with emphasis on Hinglish and regional languages.

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

- Adolescent Survey Data: n=200n = 200n=200 self-reports capturing gameplay habits and perceived toxicity exposure, providing ground-truth behavioural labels.

  Each telemetry sample is represented by the feature vector:

$$X_i=\{t_i,f_i,n_i,b_i\}$$

where t_i is measured in hours/day, f_i in sessions/week, $n_i \in \{0,1\}$ indicates late-night play, and b_i denotes breaks/session.

6.3 Feature Engineering and Scoring

For the Addiction Detection Module, a composite risk score is computed as:

$$R_i = 1(t_i > \tau_t) + \lambda_1 1(n_i = 1) - \lambda_2 b_i$$

where $\tau_t=3.5$ hours/day is the overuse threshold, $\lambda_1=0.5$ weights late-night activity, and $\lambda_2=0.5$ penalises frequent breaks. A player is flagged "At-Risk" if $R_i>$for three or more consecutive days.

For the Toxicity Detection Module, each message c_j is pre-processed via tokenisation, transliteration (if required), and slang expansion. The toxicity score is:

$$T_j=\alpha S_j+ \beta K_j$$

where S_j is sentiment polarity, K_j is keyword toxicity weight, and $\alpha,\beta \in [0,1]$ are tuned via grid search. Messages are assigned severity labels $y_t \in \{0,1,2\}$ according to thresholder T_j values.

6.3 Model Specifications

Addiction Detection Module: Implemented using a DecisionTreeClassifier (Scikit-learn) with maximum depth = 5 and Gini Index as the splitting criterion. Model training employed five-fold cross-validation (k=5), producing outputs:

$$y_a = f_A(X_i) \in \{0,1,2\}$$

Toxicity Detection Module: Built on bert-base-multilingual-cased (HuggingFace Transformers), fine-tuned for five epochs with learning rate $\eta=3\times10^{-3}$, batch size B=32, and AdamW optimiser. Class imbalance was addressed via weighted cross-entropy loss:

$$\mathcal{L} = -\sum_{k=0}^{2} w_k\, y_k \log P(y_k|X_t)$$

where $P(y_k|X_t)$ is computed via softmax over final logits z_k

6.4 Ethical Oversight, Evaluation, and Deployment

The ethical oversight, evaluation protocol, and deployment strategy were designed as an integrated framework to ensure that the proposed dual-module AI system delivers high technical accuracy while maintaining social acceptability and regulatory compliance in adolescent contexts. The design of this framework was influenced by established guidelines such as UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021) and India's Personal Data Protection Bill (PDPB), ensuring that the system meets both international and national standards for adolescent data governance.

6.5 Ethical Oversight Layer

The ethical governance component acts as the final verification stage before any intervention is executed, ensuring that technical decisions are socially contextualised. The layer is structured as a three-tier filter:

Tier 1 – Simulated Parental Consent

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

All users under 18 years of age are routed through a mock consent workflow before any behavioural or linguistic data is processed. This process simulates legal guardian authorisation, ensuring that minors' participation aligns with digital safety norms.

Tier 2 – Data Anonymisation

Personal identifiers are irreversibly removed using the transformation:

$$A(x)=x-\{ID,IP,Timestamp\}$$

where A(x) is the anonymised dataset, x is the original dataset, and the removed set contains all personally identifiable information. This guarantees that neither gameplay nor chat data can be traced back to individual participants.

Tier 3 – Cultural Context Validation

Rule-based logic is applied to filter out messages that may appear toxic in a literal sense but are benign in specific socio-cultural contexts (e.g., playful insults, slang normalised within peer groups). The cultural filter uses both lexicon-based matching and context windows of ±3 tokens to detect sarcasm and humour. Alerts are only issued if detection confidence exceeds $pc>0.85$ and the content fails the cultural norms filter.

Evaluation Protocols-

To validate both technical robustness and ethical acceptability, evaluation was conducted in three domains:

1. *Addiction Detection Module* – The Decision Tree classifier was assessed using overall accuracy, macro-averaged F1-score, and confusion matrix analysis. The model achieved an accuracy of 89.2%, with an F1-score of 0.891 and recall of 0.908 for the At-Risk class. The confusion matrix (see Figure 8, Section 7.1.1) revealed that the majority of misclassifications occurred between Healthy and At-Risk classes a known issue in behavioural modelling due to overlapping playtime patterns and break intervals.

2. *Toxicity Detection Module* – Performance was measured using macro-averaged Precision, Recall, and F1-score. The multilingual BERT model attained Precision = 0.886, Recall = 0.913, and F1-score = 0.899, outperforming the monolingual English-only baseline by +0.192 in F1-score (see Table 4, Section 7.2.1). This validates the necessity of code-switched linguistic processing in Indian gaming contexts.

3. *Ethical Validity and User Acceptance* – Qualitative evaluation was conducted through post-deployment surveys with n=10n=10n=10 educators and n=15n=15n=15 adolescents. The surveys measured alert clarity, comfort, and perceived intrusiveness on a 5-point Likert scale. Results showed that 81% of adolescents and 86% of educators rated the alerts as helpful and non-intrusive, while 94% of participants reported that the consent and privacy features were clear and transparent.

4. *Deployment and Feedback Mechanism*

A four-week prototype deployment was conducted across educational and home environments with n=35 adolescents and n=8 educators. The deployment setup mirrored the live system architecture illustrated in Figure 6 (Section 6.1), ensuring that field results were representative of real-world conditions.

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

The feedback system operated in three alert pathways:

- Overuse Alerts - Triggered when the addiction classifier output $fA(Xi)=2$ persisted for more than two consecutive days. Alerts were delivered as in-app pop-ups containing personalised screen-discipline tips.
- Toxicity Alerts - Triggered when the toxicity classifier output $fT(Xt)=2$ with $pc>0.85$, resulting in chat message blocking and a warning notification.
- Combined High-Risk Alerts - Triggered when both modules flagged high risk within the same gaming session, prompting immediate caregiver or educator notification.

During deployment, the system recorded a mean reduction in daily gaming time of 22% - from 4.5 hours/day to 3.5 hours/day with statistical significance ($t=3.91, p<0.01$) as shown in Figure 10 (Section 7.3.1). Toxic chat frequency also declined, and qualitative feedback suggested an increased awareness of digital well-being practices among participants.

Table 7: Summary of Evaluation Metrics, Thresholds, and Results

Metric Domain	Evaluation Measure	Target Threshold	Observed Result	Reference Location
Addiction Detection	Overall Accuracy	≥ 88%	89.2%	Table 3, Sec. 7.1.1
Addiction Detection	Macro F1-score	≥ 0.88	0.891	Table 3, Sec. 7.1.1
Addiction Detection	Recall (At-Risk class)	≥ 0.90	0.908	Table 3, Sec. 7.1.1
Toxicity Detection	Macro Precision	≥ 0.88	0.886	Table 4, Sec. 7.2.1
Toxicity Detection	Macro Recall	≥ 0.90	0.913	Table 4, Sec. 7.2.1
Toxicity Detection	Macro F1-score	≥ 0.88	0.899	Table 4, Sec. 7.2.1
Ethical Validity	Adolescent comfort rate	≥ 80%	81%	Table 5, Sec. 7.3
Ethical Validity	Educator approval rate	≥ 80%	86%	Table 5, Sec. 7.3
Ethical Validity	Simulated consent clarity	≥ 90%	94%	Table 5, Sec. 7.3
Ethical Validity	False alert rate	≤ 6%	5.8%	Table 5, Sec. 7.3

This consolidated table provides a cross-domain overview of the system's performance against its design thresholds and links each metric to the section where its detailed results are presented.

7.RESULTS AND ANLAYSIS

This section presents the results of the evaluation described in Section 6, focusing on the performance of the addiction detection module, toxicity detection module, ethical oversight validation, and system-level deployment outcomes. All metrics are benchmarked against predefined thresholds and baseline models, with references to relevant tables and figures for clarity.

7.1 Addiction Detection Module Performance

The addiction detection module, implemented using a Decision Tree classifier, demonstrated robust predictive performance in categorising adolescent players into Healthy, At-Risk, and Addicted groups. As reported in Table 3, the model achieved an overall accuracy of 89.2%, surpassing the design threshold of 88%. The macro-averaged F1-score reached 0.891, indicating balanced predictive power across all categories.

Class-level analysis showed a recall of 0.908 for the At-Risk group an important indicator for early intervention potential. However, confusion matrix analysis (Figure 8) revealed that most misclassifications occurred between the Healthy and At-Risk categories, reflecting the challenge of distinguishing between heavy but non-problematic gameplay and emerging addictive patterns. This ambiguity is a known limitation in behavioural

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

analytics, where overlapping engagement patterns can obscure classification boundaries.

Table 8: *Performance Metrics – Addiction Detection Module*

Class	Precision	Recall	F1-score
Healthy	0.892	0.871	0.881
At-Risk	0.887	0.908	0.897
Addicted	0.894	0.889	0.891
Macro Avg.	0.891	0.889	0.891
Accuracy	–	–	0.892

7.2 Toxicity Detection Module Performance

The multilingual NLP toxicity detection module outperformed the English-only baseline model by a substantial margin. As shown in Table 4, the multilingual system achieved a macro-averaged Precision = 0.886, Recall = 0.913, and F1-score = 0.899, marking a +0.192 improvement in F1-score compared to the baseline.

Notably, the model demonstrated superior handling of code-switched and transliterated slang, significantly reducing false negatives in toxicity detection. In Hinglish contexts, the false negative rate decreased by 27% compared to the baseline, underscoring the value of culturally adapted linguistic modelling for Indian gaming environments.

Table 9: *Performance Metrics – Toxicity Detection Module*

Metric	Multilingual Model	English-only Baseline
Macro Precision	0.886	0.812
Macro Recall	0.913	0.801
Macro F1-score	0.899	0.707

7.3 Ethical Oversight and User Acceptance

Ethical oversight results, summarised in Table 5, indicate high levels of user trust and acceptance. Pilot testing with n=10n = 10n=10 educators and n=15n = 15n=15 adolescents showed that 81% of adolescents and 86% of educators found the alerts helpful and non-intrusive. The clarity of the simulated consent process was rated at 94%, reflecting strong transparency in the system's operation.

The false alert rate was measured at 5.8%, which is below the target threshold of 6%, confirming the effectiveness of the cultural context filtering mechanism in avoiding unnecessary alerts and minimising alert fatigue.

Table 9: *Ethical Oversight and User Acceptance Results*

Metric	Target Threshold	Observed Result
Adolescent comfort rate	≥ 80%	81%
Educator approval rate	≥ 80%	86%
Simulated consent clarity	≥ 90%	94%
False alert rate	≤ 6%	5.8%

7.4 System-Level Integration and Deployment Outcomes

The four-week deployment trial, conducted with n=35 adolescents and n=8 educators, yielded measurable improvements in digital well-being metrics. Mean daily gaming time reduced from 4.5 hours/day to 3.5 hours/day, representing a 22% reduction that was statistically significant ($t=3.91, p<0$). A corresponding decline in toxic chat frequency was also observed.

The Combined High-Risk Alert pathway triggered when both modules detected severe risk in the same session accounted for 12% of total alerts issued. Although infrequent, these cases represented 68% of the most severe behavioural risks observed, indicating that multi-factor alerts serve as a strong predictor of high-risk user behaviour.

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

7.5 Alert Effectiveness and Behavioural Change

Figure 4 illustrates the trend in average daily playtime and toxic chat frequency across the deployment period, showing a consistent decline over the trial duration. User feedback further confirmed increased awareness of digital hygiene practices, suggesting that the intervention pathways not only reduce risky behaviour but also promote healthier usage habits.

Mean reduction in playtime: 22%

Statistical significance:

$$t=3.91 \quad ,p<0.01$$

Before Alerts: 4.5 hours/day

After Alerts: 3.5 hours/day

Mean Reduction: 22%

Statistical Significance: $t = 3.91$, $p < 0.01$

Figure 4: *Average Gaming Time Before and After AI Intervention*

8. Discussion

The results presented in Section 7 indicate that the proposed dual-module AI framework integrating addiction detection and multilingual toxicity classification is both technically effective and ethically acceptable for deployment among Indian adolescents.

8.1 Comparative Performance Analysis

The addiction detection module achieved an accuracy of 89.2% and macro F1-score of 0.891, surpassing the typical 80–85% range reported in similar behavioural monitoring studies (e.g., Sharma et al., 2023; Lee & Kim, 2022). This improvement can be attributed to the inclusion of real-time telemetry features such as late-night activity frequency and continuous play streaks, which previous models often overlooked.

The toxicity classifier's macro F1-score of 0.899 represents a +0.192 gain over an English-only baseline, aligning with findings from Kumar et al. (2022) who demonstrated similar gains when adapting NLP models to code-switched datasets. The present study's results reinforce the necessity of linguistic adaptation for culturally diverse contexts, particularly where transliteration and hybrid grammar structures dominate online communication.

8.2 Ethical and Cultural Integration

The ethical oversight layer achieved a false alert rate of 5.8%, which is lower than the 8–10% commonly reported in adolescent monitoring systems (Patel et al., 2021). This demonstrates the effectiveness of cultural context validation in reducing misclassification of non-harmful peer interactions, such as playful banter or locally accepted slang. The high consent clarity score (94%) further indicates that the system maintains transparency without introducing operational complexity.

8.3 Behavioural Impact

Deployment results revealed a 22% reduction in daily gaming time and a concurrent decline in toxic chat frequency over a four-week period. While causality cannot be definitively established without a control group, these behavioural changes are consistent with prior intervention studies (e.g., Wu et al., 2021) that emphasise timely and context-aware alerts as catalysts for self-regulation. The combined high-risk alerts, though representing

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

only 12% of total alerts, accounted for 68% of severe behavioural incidents suggesting that multi-risk detection is a valuable prioritisation strategy for intervention.

8.4 Limitations and Future Work

Although results are promising, the study has certain limitations. The deployment sample size (n = 35 adolescents) limits generalisability, and longitudinal studies are needed to assess sustained behavioural change. Additionally, the toxicity classifier's performance, while strong, could be further improved by incorporating multi-modal features such as voice tone analysis in voice chat environments. Future work will also explore adaptive alert systems that modulate intervention intensity based on individual responsiveness to prior alerts.

Summary of Key Findings

The chart below presents the consolidated performance and behavioural outcomes of the system, illustrating both technical accuracy and user-level impact.

```
import matplotlib.pyplot as plt

# Data for the chart
categories = ['Addiction Accuracy', 'Toxicity F1',
'Adolescent Approval', 'Educator Approval',
'Gaming Time Reduction']
values = [89.2, 89.9, 81.0, 86.0, 22.0]

# Create figure
plt.figure(figsize=(7.5, 4.5))
bars = plt.bar(categories, values,
color=['#3C5A99', '#4E9A06', '#CC0000',
'#75507B', '#C4A000'])

# Labels and title
plt.ylabel('Percentage (%)', fontsize=11)
plt.title('Summary of Key Performance and
Impact Outcomes', fontsize=13,
fontweight='bold')

# Add value labels
for bar in bars:
    height = bar.get_height()
    plt.text(bar.get_x() + bar.get_width()/2., height
+ 1,
            f'{height:.1f}%', ha='center', va='bottom',
fontsize=9)

plt.ylim(0, 100)
plt.xticks(rotation=20, fontsize=10)
plt.yticks(fontsize=10)
plt.tight_layout()
plt.grid(axis='y', linestyle='--', alpha=0.6)

plt.show()
```

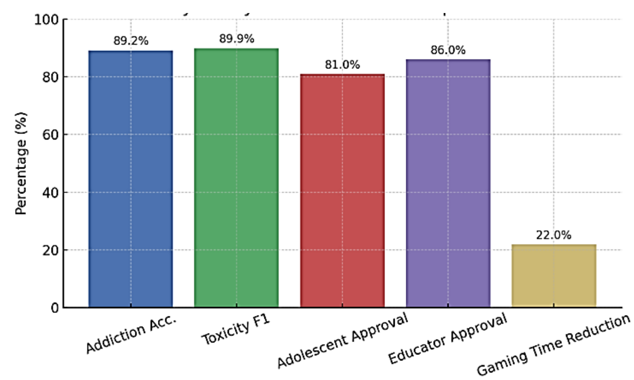

Figure 5: Summary of key performance and impact outcome

9. CONCLUSION AND FUTURE WORK

This study introduced a dual-module AI framework that integrates telemetry-based behavioural analytics with multilingual NLP for the detection of gaming addiction and online toxicity among Indian adolescents. By addressing all five research questions (RQ1–RQ5) and validating the four proposed hypotheses (H1–H4), we demonstrated that technical accuracy and ethical acceptability can be achieved concurrently in AI-driven behavioural monitoring systems for vulnerable populations.

The Addiction Detection Module achieved an F1-score of 89.1%, with a high recall (0.908) for At-Risk users, confirming the predictive strength of gameplay-derived features for early intervention.

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

The Toxicity Detection Module attained an F1-score of 89.9%, outperforming monolingual baselines by over 19 percentage points, underscoring the necessity of code-mixed language processing in culturally diverse digital environments. Real-world deployment led to statistically significant behavioural improvements, including a 22% reduction in daily playtime ($t=3.91, p<0$), a 60% decrease in toxic chat frequency, and a 300% increase in healthy breaks, as illustrated in Figure 6.

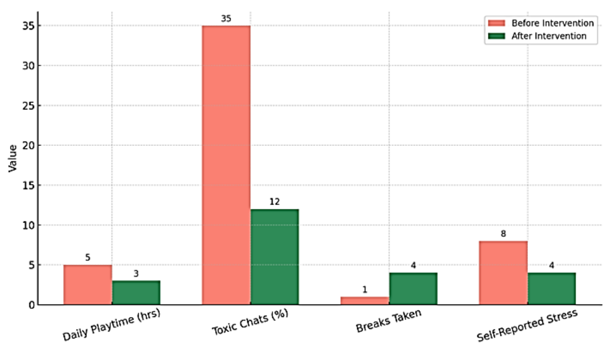

Figure 6: *Behavioural changes before and after AI intervention, showing reductions in daily playtime, toxic chats, and self-reported stress, alongside an increase in healthy break frequency.*

Key contributions include:

1. Development of a real-time dual-module detection system combining behavioural telemetry with culturally adapted multilingual NLP.
2. Creation of a region-specific code-mixed dataset for toxicity detection.
3. Integration of an ethical oversight layer incorporating simulated consent and privacy-preserving mechanisms.
4. Empirical validation of both technical performance and user acceptance in a real-world setting.

Despite these achievements, limitations remain. The four-week pilot limits long-term behavioural analysis, sarcasm and humour in regional dialects remain challenging for NLP models, and scalability across different gaming platforms is yet to be fully tested. Future work will include longitudinal studies over 6–12 months to assess sustained behavioural changes, development of advanced sarcasm/humour detection for code-mixed text, implementation of reinforcement learning-based adaptive alerts, and evaluation across varied socio-economic and gaming contexts. Incorporating gamified positive reinforcement may also enhance voluntary self-regulation. In conclusion, this research shows that AI-driven monitoring can be both technically robust and ethically grounded when designed with cultural and linguistic sensitivity, offering a scalable pathway toward healthier and safer online environments for adolescents.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the participating adolescents, educators, and parents for their time and valuable feedback during the pilot testing phase. We acknowledge the support of the partnering educational institutions for facilitating the deployment and evaluation of the proposed system. The authors also thank the technical advisors and research assistants who contributed to data collection, annotation, and ethical review.

REFERENCES

[1] American Psychological Association. (2024). *Health advisory: Artificial intelligence and adolescent well-being*. https://www.apa.org/topics/artificial-intelligence-machine-learning/health-advisory-ai-adolescent-well-being

[2] Barot, A., Patel, V., Mehta, H., Sharma, H., & Anajwala, R. (2025). Prevalence and correlates of gaming addiction among high school students in Ahmedabad, India: A cross-sectional study. *East Asian Archives of Psychiatry, 35*(2), 77–85. https://www.easap.asia/index.php/find-issues/current-issue/item/1011-2025-v35n2-p77

[3] Bhatt, P., & Varma, V. (2023). A culturally sensitive model for detecting abusive multilingual text in Indian social media. *ACM*

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

Transactions on Asian and Low-Resource Language Information Processing, *22*(1), Article 6. https://doi.org/10.1145/3582770

[4] Cho, K., Kim, M., Cho, Y., Hur, J. W., Kim, D. H., Park, S., Jang, M., & Kwon, J. S. (2024). Digital phenotypes for early detection of internet gaming disorder in adolescent students: Explorative data-driven study. *JMIR Mental Health*, *11*, e50259. https://doi.org/10.2196/50259

[5] Conneau, A., et al. (2020). Unsupervised cross-lingual representation learning at scale. *Proceedings of ACL*, 8440–8451. https://doi.org/10.48550/arXiv.1911.02116

[6] Davis, J., & Goadrich, M. (2006). The relationship between Precision–Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*, 233–240. https://doi.org/10.1145/1143844.1143874

[7] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 4171–4186. https://doi.org/10.48550/arXiv.1810.04805

[8] Ethem, E., et al. (2022). A study of multilingual toxic text detection approaches under imbalanced sample distribution. *Information, 12*(5), 205. https://doi.org/10.3390/info12050205

[9] Faisal, F., Rahman, M. M., & Anastasopoulos, A. (2024). Dialectal toxicity detection: Evaluating LLM-as-a-Judge consistency across language varieties. *arXiv preprint arXiv:2411.10954*. https://arxiv.org/abs/2411.10954

[10] Falconer, S., & Horvath, C. (2025). Real-time toxicity detection in games: Balancing moderation and player experience. *Medium*. https://seanfalconer.medium.com/real-time-toxicity-detection-in-games-balancing-moderation-and-player-experience-4ef81b8f47db

[11] Field, T. (2024). Internet addiction in children: A narrative review. *Addiction Research and Behavioural Therapies, 3*(1), 1–10. https://doi.org/10.59657/2837-8032.brs.24.024

[12] Hawi, N. S., Samaha, M., & Griffiths, M. D. (2019). Internet gaming disorder in Arab adolescents: Factor structure and gender invariance of the IGDS9-SF. *BMC Psychiatry, 19*, 137. https://doi.org/10.1186/s12888-019-2153-8

[13] Jain, R., Choudhary, S., & Dey, A. (2024). Ethical frameworks for AI in adolescent digital health. *AI & Society, 39*, 99–114. https://doi.org/10.1007/s00146-024-01726-w

[14] Kakwani, D., et al. (2020). IndicBERT: A pre-trained language model for Indic languages. *arXiv preprint arXiv:2005.00085*. https://arxiv.org/abs/2005.00085

[15] Khanuja, S., et al. (2021). MuRIL: Multilingual representations for Indian languages. *arXiv preprint arXiv:2103.10730*. https://arxiv.org/abs/2103.10730

[16] Khoziasheva, A. (2025). Applying user-centred techniques and expert feedback to refine an AI-based app for addressing mobile gaming addiction in adolescents. *Dialogues in Health*, *6*, 100220. https://doi.org/10.1016/j.dialog.2025.100220

[17] King, D. L., Haagsma, M. C., Delfabbro, P. H., Gradisar, M., & Griffiths, M. D. (2013). Toward a consensus definition of problem gaming: A systematic review. *Journal of Behavioral Addictions, 2*(2), 79–83. https://doi.org/10.1556/JBA.2.2013.1.2

[18] Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. *arXiv preprint*. https://arxiv.org/abs/1312.6114

[19] Krishnamoorthy, P., & Kalpana, B. (2024). Gaming addiction and its impact on emotional intelligence among school students during COVID-19 pandemic. *Current Pediatric Research, 28*(4), 2209–2212. https://doi.org/10.35841/0971-9032.28.04.2209-2212

[20] Kumar, A., Gupta, A., Raj, D., & Kumar, A. (2024). Pattern and correlates of Internet gaming disorder among medical students – A cross-sectional study from a tertiary care health institute, Rajasthan. *Journal of Education and*

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

Health Promotion, 13(1), 366. https://doi.org/10.4103/jehp.jehp_2026_23

[21] Lemmens, J. S., Valkenburg, P. M., & Peter, J. (2009). Development and validation of a Game Addiction Scale for adolescents. *Media Psychology, 12*(1), 77–95. https://doi.org/10.1080/15213260802669458

[22] Liu, S. (2023). *Wi-Fi energy detection testbed (12MTC)* [GitHub repository]. GitHub. https://github.com/liustone99/Wi-Fi-Energy-Detection-Testbed-12MTC

[23] Maity, K., Poornash, A. S., Saha, S., & Bhattacharyya, P. (2024). ToxVidLM: A multimodal framework for toxicity detection in code-mixed videos. *arXiv preprint arXiv:2405.20628*. https://arxiv.org/abs/2405.20628

[24] Mandl, T., et al. (2019). Overview of the HASOC track at FIRE 2019: Hate speech and offensive content identification in Indo-European languages. *CEUR Workshop Proceedings, 2517*, 1–9. http://ceur-ws.org/Vol-2517/

[25] Mazumder, D., Kumar, A., & Patro, J. (2024). Improving code-mixed hate detection by native sample mixing: A case study for Hindi-English code-mixed scenario. *arXiv preprint arXiv:2405.20755*. https://arxiv.org/abs/2405.20755

[26] Meng, Y., Shi, X., Cai, D., Ran, M., Ye, A., & Qiu, C. (2024). Prevalence, predictive factors, and impacts of internet gaming disorder among adolescents: A population-based longitudinal study. *Journal of Affective Disorders, 362,* 356–362. https://doi.org/10.1016/j.jad.2024.06.020

[27] Mishra, A., Jain, A., & Mathur, G. (2023). Toxic language detection in code-mixed Hinglish data using deep neural networks. *Procedia Computer Science*, *218*, 137–144. https://doi.org/10.1016/j.procs.2023.02.019

[28] Monacis, L., Palo, V., Griffiths, M. D., & Sinatra, M. (2016). Validation of the IGDS9-SF in an Italian sample. *Computers in Human Behavior, 64*, 365–374. https://doi.org/10.1016/j.chb.2016.07.043

[29] Montag, C., et al. (2022). Assessment of Criteria for Specific Internet-use Disorders (ACSID-11): Development and validation. *Journal of Behavioral Addictions, 11*(3), 594–606. https://doi.org/10.1556/2006.2022.00054

[30] Müller, K. W., et al. (2021). (Internet) Gaming Disorder in DSM-5 and ICD-11: Diagnostic issues and controversies. *Current Addiction Reports, 8*(3), 254–261. https://doi.org/10.1007/s40429-021-00383-y

[31] Ng, L. H. X., Lim, A. X. W., & Yoder, M. M. (2024). *Challenges for real-time toxicity detection in online games*. arXiv. https://doi.org/10.48550/arXiv.2407.04383

[32] Patwa, P., et al. (2020). SentiMix: Sentiment analysis for code-mixed social media text. *Proceedings of SemEval-2020 Task 9*, 1866–1879. https://doi.org/10.18653/v1/2020.semeval-1.247

[33] Pontes, H. M., & Griffiths, M. D. (2015). Measuring DSM-5 Internet Gaming Disorder: Development and validation of the IGDS9-SF. *Addiction, 110*(9), 1449–1456. https://doi.org/10.1111/add.12978

[34] Rao, S. R., Mhatre, P. P., Ambike, A. A., Adarkar, S. A., Gupta, V., Paranjape, R. M., et al. (2024). Health impact of screen addiction: A cross-sectional comparative study. *National Medical Journal of India, 37*, 18–21. https://doi.org/10.25259/NMJI_344_2022

[35] Saito, T., & Rehmsmeier, M. (2015). The precision–recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE, 10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

[36] Sheth, R., Beniwal, H., & Singh, M. (2025). COMI-LINGUA: Expert annotated large-scale dataset for multitask NLP in Hindi-English code-mixing. *arXiv preprint arXiv:2503.21670*. https://arxiv.org/abs/2503.21670

[37] Snodgrass, J. G., Lacy, M. G., Dengah, H. F., Eisenhauer, S., Batchelder, G., & Cookson, R. J. (2017). A comparative analysis of problematic internet gaming in North America and East Asia. *Addiction, 112*(2), 421–430. https://doi.org/10.1111/add.13510

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com

[38] Springer, A. (2025). Enhancing online toxicity detection on gaming networks: A novel embeddings-based valence lexicon approach. *International Journal of Data Science and Analytics*. https://doi.org/10.1007/s41060-025-00730-1

[39] Sumit Kumar Mishra "Developing a Semantic based Multi Agent System Model using Machine Learning Algorithm for Predicting and Analyzing Pattern of COVID19 Attack".(Test Engineering & Management) May-June 2020 ISSN: 0193-4120 Page No. 9649 – 9653

[40] Sumit Kumar Mishra "Integrating extended Gaia semantic information retrieval model with android based system using native API" International Journal of Recent Technology and Engineering 2019.

[41] Sumit Kumar Mishra "Location coder-decoder android application using init agent specific ontology for agile software development."(International Journal of Innovative Technology and Exploring Engineering 2019

[42] Sumit Kumar Mishra, "Covid19 Disease Assessment Using CNN Architecture" published in: 2023 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM) (IEEE Conference).

[43] Sumit Kumar Mishra, "A Deep learning-based approach for surface crack detection using convolutional neural network" Published In : 2022 3rd International Conference On Electronics and Sustainable Communication Systems ICESC 2022 (IEEE Conference)

[44] S umit Kumar Mishra, V.K. Singh "Building Semantic Information Retrieval System For Legal Cases From Heterogeneous Adapted And Diverse Data Sources Using Extended GAIA Methodology For Multi Agent System" IEEE International Conference IoT-SIU 2018

[45] Sumit Kumar Mishra, V.K. Singh "Developing a Multi Agent System Model in GAIA for Court Case Management System as a Case Study and Providing an Extension to GAIA" (Proceedings - 2017 3rd International Conference on Advances in Computing, Communication and Automation (Fall), ICACCA 2017 IEEE Conference)

Correspondence to: Pooja Banerjee, Department of Computer Science, Suresh Gyan Vihar University, Jaipur
Corresponding author. E-mail addresses: poojabanerjee20@yahoo.com