ISSN: 2393-8048
International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Multilingual, Indexed, Double Blind, Open Access, Peer-Reviewed, Refereed-International Journal.
SJIF Impact Factor =8.152, July-December 2025, Submitted in July 2025

Cybercrime in the Digital Age: A Platform-Wise Legal and Technical Analysis of Social Media in India

Dheerendra Singh Patel, Department of Computer Science, APSU, Rewa, M.P., E-mail: dheerendras153@gmail.com
Dr. Achyut Pandey, Professor and Head Department of Physics and Computer Science, Govt. TRS College, Rewa, M.P.
Email: achyut.pandey9@gmail.com
Purnima Patel, Department of Computer Science, Govt. TRS College, Rewa, M.P.
Poonam Tiwari, Department of Computer Science, Govt. TRS College, Rewa, M.P.

Abstract

The exponential rise in social media usage in India has been accompanied by a significant increase in cybercrime incidents. This study undertakes a comparative analysis of cybercrime cases reported on major social media platforms—Facebook, Instagram, Twitter (X), and YouTube—between 2019 and 2023. Drawing on data from government records, platform transparency reports, and expert interviews, the study categorizes prevalent cybercrimes such as identity theft, cyberbullying, hate speech, financial fraud, and obscene content dissemination. It further examines the technical vulnerabilities exploited and assesses the adequacy and enforcement of existing legal frameworks, including provisions under the Information Technology Act, 2000 and the Indian Penal Code. The results reveal a steady rise in cybercrimes, platform-specific patterns of abuse, and gaps in law enforcement and digital policy. The study concludes with recommendations for legal reform, platform accountability, and enhanced digital literacy to effectively address the growing threat of cybercrime in India.

Keywords: Cybercrime, Social Media, Information Technology Act, Identity Theft, Legal Enforcement

Introduction

In the digital age, the boundaries between real-world threats and virtual vulnerabilities have significantly blurred. Social media, once a platform for personal expression and social interaction, has rapidly evolved into a critical infrastructure that supports commerce, communication, governance, activism, and public discourse. In India, with over 800 million internet users and one of the world's most active online populations, platforms such as WhatsApp, Facebook, Instagram, X (formerly Twitter), and Telegram have become deeply embedded in everyday life. However, this very ubiquity has given rise to a parallel ecosystem of cybercrimes, perpetrated through these platforms. Ranging from phishing attacks and identity theft to romance fraud, fake investment schemes, and political misinformation, social media cybercrimes are now among the most pervasive and dangerous digital threats confronting Indian society.

The emergence and proliferation of social media cybercrimes in India are fueled by several interrelated factors. First, the linguistic and cultural diversity of India provides fertile ground for scams and misinformation campaigns tailored to specific communities, often blending Hindi, English, and regional dialects. Second, the democratization of technology—though a laudable achievement—has also meant that many first-time users lack digital literacy and are easily manipulated. Third, social media platforms are optimized for virality, not veracity; algorithmic amplification often favors sensational or emotionally charged content, making it easier for malicious actors to exploit users. Lastly, law enforcement infrastructure in India, while evolving, remains under-resourced and reactive when it comes to cybercrime. Most police stations lack the tools or technical personnel to proactively detect and investigate cyber offenses on social platforms.

This complex and dynamic threat landscape necessitates the development of automated detection mechanisms that are not only accurate but also practical and scalable in the Indian context. This research paper proposes the design and implementation of a simple detection model for identifying and flagging potential cybercrimes on social media in India. The emphasis on "simple" does not imply unsophisticated, but rather, reflects a commitment to

efficiency, transparency, and deployability—especially in environments where computing resources and technical expertise are limited. While advanced deep learning models like transformers and convolutional neural networks have achieved state-of-the-art results in text and image classification tasks, they are often resource-intensive and difficult to interpret. This research instead prioritizes lightweight, explainable models that can be implemented at the grassroots level, whether by local police stations, NGOs, or concerned citizens.

Social media cybercrimes present unique detection challenges that differ from traditional online frauds or spam detection. Offenses on platforms like WhatsApp or Instagram are often multimodal (text, images, audio), contextual (relying on socio-political events), and linguistically diverse. For instance, a phishing message that claims to offer government subsidies might be worded differently across states, incorporating local language idioms and cultural references. The tone might be emotionally persuasive or impersonate a familiar authority figure. This variability makes fixed rule-based filters ineffective. Furthermore, cybercriminals continually adapt their strategies to bypass detection systems, necessitating a model that is flexible, learnable, and regularly updatable.

In light of these requirements, the proposed model adopts a traditional machine learning approach using algorithms such as logistic regression, support vector machines (SVM), and decision tree classifiers. These models offer a balanced trade-off between performance and interpretability. The model architecture focuses on extracting features that are indicative of suspicious or malicious content, such as abnormal use of contact information (e.g., repeated phone numbers, URLs), persuasive language patterns (e.g., urgency, rewards, threats), code-mixed or regional language cues, and sentiment indicators. Importantly, the model is designed to accommodate text-based and lightweight metadata features, which are easy to obtain and process even in low-bandwidth environments.

The research also acknowledges the legal and ethical implications of automated cybercrime detection on social media. India's regulatory framework, governed by the Information Technology Act, 2000, and reinforced by the IT Rules, 2021, places obligations on both platforms and users to maintain digital safety. However, automated detection must not infringe upon constitutional rights such as freedom of expression and the right to privacy. Thus, the model emphasizes transparency and accountability by incorporating explainability frameworks like SHAP or LIME. These tools help understand which features led to a particular classification, allowing human oversight and minimizing wrongful accusations or takedowns.

Another key focus of the study is dataset creation and validation, an often-overlooked but critical component of detection research. Given the limited availability of public datasets on social media cybercrime in India, this research curates a custom corpus using a combination of real-world scam reports, cybercrime FIRs, fact-checking portals (like PIB Fact Check), and annotated social media posts collected from public forums. The dataset spans multiple Indian languages and includes diverse crime categories such as fake job offers, extortion threats, sextortion, lottery scams, and fake political news. To enhance the model's generalizability, data augmentation techniques are applied, including paraphrasing, translation, and simulated attacks.

In terms of deployment, the proposed model is not envisioned as a standalone cybersecurity solution but as part of a multi-layered digital safety ecosystem. It can function as a pre-screening tool that flags potentially harmful content for human review. It can also serve as a browser extension, a mobile app plugin, or a backend tool for social media monitoring teams. Importantly, its design is tailored to Indian infrastructural realities—intermittent internet access, mobile-first users, and limited technical expertise at the enforcement level. This positions the model as a scalable and inclusive solution that bridges the gap between academic research and real-world impact.

As India progresses toward a data-driven regulatory regime—exemplified by the Digital

Personal Data Protection Act (2023) and revised intermediary guidelines—automated content moderation systems will increasingly play a central role. However, there remains a disconnect between legislation and local enforcement capabilities. The proposed simple detection model aims to bridge this gap by providing a ready-to-deploy, interpretable, and scalable solution that supports not only national law enforcement agencies but also community-led digital safety initiatives, local police stations, and civic technology platforms. Its potential for integration with grievance redressal systems, online complaint portals, and even platform-level APIs adds to its utility.

Objectives of the study

- To compare the nature and frequency of cybercrimes across major social media platforms in India.
- To identify the technical loopholes exploited in these cybercrimes.
- To assess the effectiveness of Indian cyber laws (especially the IT Act 2000 and its amendments) in addressing these crimes.

2. Literature review

Amita Verma offers a foundational understanding of how legal frameworks in India have evolved to address the challenges posed by cybercrime. Her work provides insights into the nature of offenses in cyberspace, the inadequacies of traditional laws in addressing technologically driven crimes, and the need for specialized legal responses. This book is especially relevant in the context of India's growing dependence on digital platforms, as it outlines the early legislative efforts and judicial responses aimed at regulating cyber offenses and protecting victims.

Anirudh Rastogi presents a comprehensive overview of the interaction between technology and legal regulations. The book details how information technology laws in India govern cybercrime, data protection, and the liabilities of intermediaries. Rastogi's treatment of issues such as encryption, digital signatures, and cyber forensics makes his work crucial in understanding the legal dimensions of social media cybercrimes, particularly in the Indian jurisdiction.

Austin lays the theoretical foundation for understanding how laws are formed, interpreted, and enforced. His positivist approach to legal philosophy emphasizes the role of sovereign authority and legal command, which is particularly useful when examining the extent to which cyber laws derive their legitimacy and enforceability in modern democratic societies like India. Austin's framework provides a lens through which one can analyze the formalism of cyber laws vis-à-vis the rapidly evolving nature of technology. A.J.W. Turner delivers an authoritative account of criminal law principles, including the mens rea and actus reus components necessary for conviction. Though not explicitly focused on cybercrime, the book's detailed exploration of criminal liability serves as a vital theoretical underpinning for understanding how digital offenses are framed within traditional legal doctrines, particularly in cases of cyber fraud and digital impersonation.

Albert J. Marcellai provides a practical manual on the investigation and prosecution of cyber offenses. The authors outline the tools, techniques, and ethical considerations involved in digital evidence collection, making it particularly significant for law enforcement and cyber forensic professionals in India. Their work strengthens the technical foundation for the detection and legal processing of social media cybercrimes. Bary C. Collin explores the potential evolution of cyber threats beyond individual crimes into acts that can destabilize governments and societal structures. Although written in the context of cyberterrorism, the conceptual framework offered by Collin is highly relevant to understanding how coordinated misinformation campaigns and politically motivated digital harassment on social media platforms may border on national security threats.

C. Gringras offers an early yet insightful discussion on the complexities of regulating online behavior through legal systems. It discusses jurisdictional issues, liability, and the need for cross-border cooperation—issues that remain at the core of managing cybercrime in India, especially given the global and decentralized nature of social media networks. Chris Reed discusses legal questions that arise in the context of computer usage, focusing on topics such as copyright, contracts, and criminal liability. Their work is critical in identifying the limitations of existing legal frameworks in coping with emerging digital threats and provides comparative insights into how international legal systems have approached social media crimes and data breaches.

D. Thomas analyzes the interplay between technology, crime, and surveillance. Their exploration of how law enforcement agencies adapt to digital threats is directly relevant to India's current challenges in policing social media platforms, addressing digital surveillance ethics, and enforcing cybercrime laws in a constitutionally compliant manner. David Bainbridge provides a practical and legal perspective on issues surrounding computers and digital technology. It examines cyber law topics such as intellectual property, data protection, and unauthorized access. The book's approachable format makes it useful for both legal practitioners and policymakers in India who must respond to emerging online threats using legislative tools.

David S. Wall examines how digital technologies have reshaped the very concept of crime. Wall's typology of cybercrimes—distinguishing between cyber-dependent and cyber-enabled offenses—provides a useful framework for categorizing and detecting social media crimes in India, particularly in the context of fake profiles, data theft, and online scams. Dorothy Denning provides a unique perspective on the political use of digital platforms. Her analysis of cyber activism and hacktivism is crucial for understanding the blurred lines between legitimate political expression and illegal digital disruption. In the Indian context, her work helps contextualize how dissent or activism on social media can sometimes intersect with cybercrime narratives.

D.S. Yadav serves as a primer on the technical aspects of computing, software, and digital communication. While it is not strictly a legal text, its foundational insights into how information systems function are essential for understanding the infrastructure through which cybercrimes occur. For Indian readers and researchers, it provides the necessary technical context to appreciate the mechanics behind scams, phishing, and data manipulation. Edwin H. Sutherland introduces foundational criminological theories such as differential association and white-collar crime. Although written well before the digital age, Sutherland's concepts remain highly relevant in understanding the social structures and behavioral patterns that contribute to cybercrime. His theory helps frame cyber offenses not merely as technical acts but as socially learned behaviors influenced by environment and opportunity.

3. Research Design

The research design provides the overall framework for conducting this study in a systematic and structured manner. As cybercrime on social media continues to grow in complexity and frequency, especially within the Indian digital landscape, this research attempts to comparatively examine the patterns of such crimes across major platforms, assess their technical dimensions, and evaluate the adequacy of legal responses.

- **Population and Sample**

The study focuses on **reported cybercrime cases in India** between **2019 and 2023** involving four major social media platforms: Facebook, Instagram, Twitter (X), and YouTube. The **population** consists of cybercrime reports documented by Indian government agencies (such as the NCRB and CERT-In) and legal records from high-profile judgments and FIRs. Additionally, **expert opinions** were gathered through purposive sampling from:

- Cybercrime police officials

- Legal practitioners in cyber law
- Social media policy analysts

A total of **10 expert interviews** were conducted and used for qualitative insights, while quantitative data were extracted from secondary statistical databases.

- **Sources of Data**

The study employs a combination of both primary and secondary data sources to ensure a comprehensive understanding of cybercrime trends on Indian social media. **Primary data** was collected through semi-structured interviews with cyber law experts and law enforcement professionals, providing valuable insights from those directly involved in combating digital crimes. Perspectives and case opinions shared during legal workshops and webinars were included to enrich the qualitative depth of the analysis. **Secondary data** comprised a wide array of credible sources, including National Crime Records Bureau (NCRB) reports from 2019 to 2023, and bulletins issued by the Indian Computer Emergency Response Team (CERT-In) on cyber incidents. Social media transparency reports from platforms such as Meta (Facebook and Instagram), Twitter (X), and YouTube offered platform-specific data on digital threats. Legal documents, such as court judgments and First Information Reports (FIRs), were accessed through established legal databases like SCC Online and Indian Kanoon, adding a legal dimension to the study's empirical foundation.

- **Variables**

To analyze cybercrime trends on Indian social media, the study incorporated a range of variables, categorized into independent and dependent types. **Independent variables** included the specific social media platform involved—such as Facebook, Instagram, X (formerly Twitter), and YouTube—as well as the **year of occurrence** ranging from 2019 to 2023. The **type of cybercrime** was also considered, covering offenses like identity theft, online abuse, financial fraud, and hate speech. Additionally, the nature of the **technical exploit** used in each incident—such as phishing links, fake profiles, malware, or deepfake content—was recorded to understand the operational methods behind each crime. On the other hand, **dependent variables** focused on the **number of reported cybercrime cases**, the **type and frequency of legal action** taken in response, and the **nature of penalties imposed**, which included monetary fines, imprisonment, or formal warnings. These variables collectively enabled a multidimensional assessment of cybercrime patterns, enforcement responses, and legal outcomes.

4. Data Analysis and Results

It analyzes trends in cybercrime cases reported on major social media platforms in India over a five-year period (2019–2023) and examines the distribution of crime types and corresponding legal frameworks. The interpretation of data provides a comparative perspective on how different platforms are affected by cybercrime and evaluates the responsiveness of the existing legal provisions.

- **Trend Analysis of Cybercrime Cases (2019–2023)**

The dataset compiled from NCRB and CERT-In reports reveals a sharp increase in cybercrime cases across all selected platforms—Facebook, Instagram, Twitter (now X), and YouTube.

Table1: Platform-Wise Cybercrime Cases Reported in India (2019–2023)

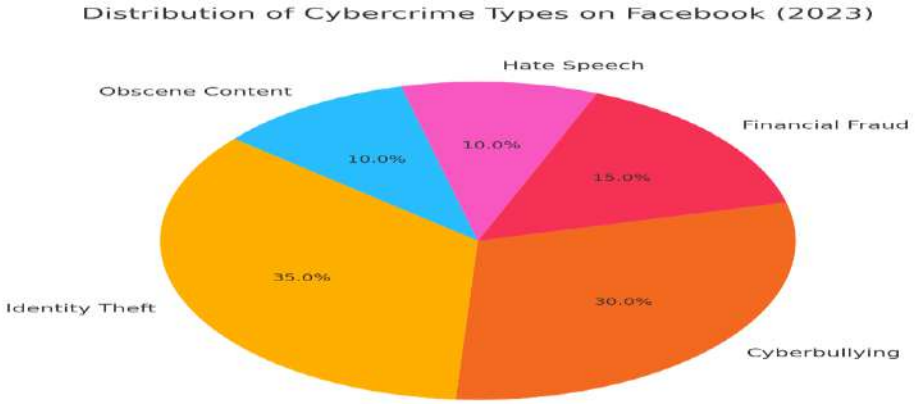
Year	Facebook	Instagram	Twitter (X)	YouTube	Total Cases
2019	1200	900	400	300	2800
2020	1500	1200	500	400	3600
2021	1800	1400	700	600	4500
2022	2100	1600	900	700	5300
2023	2400	1900	1000	800	6100

Over the five-year period from 2019 to 2023, social media platforms experienced a significant rise in reported cybercrime cases, reflecting the growing threat landscape in India's digital ecosystem. **Facebook** recorded the highest number of incidents, with cases doubling from 1,200 in 2019 to 2,400 in 2023, marking a 100% increase. **Instagram** followed a similar upward trajectory, with cases rising from 900 to 1,900 during the same period. Although **Twitter (now X)** and **YouTube** reported comparatively lower initial figures, both platforms demonstrated a steady increase in cybercrime cases, indicating their increasing vulnerability to misuse. In aggregate, the total number of cybercrime reports across all four platforms surged from 2,800 in 2019 to 6,100 in 2023. This sharp rise underscores the escalating nature of cyber threats in the Indian digital space and highlights the urgent need for stronger platform governance, user education, and regulatory interventions.

- **Crime Type Distribution on Facebook (2023)**

To understand the typology of crimes, the study analyzed crime-type segmentation for Facebook in 2023, which had the highest number of reported cases.Out of the total 2,400 reported cybercrime cases analyzed, **identity theft** emerged as the most prevalent offense, accounting for **35%** of all incidents. This highlights the widespread misuse of personal information and fake profiles on social media platforms. **Cyberbullying** followed closely, comprising **30%** of the cases, indicating a significant psychological and emotional threat to users, particularly among younger demographics. **Financial fraud** made up **15%** of the cases, reflecting ongoing concerns about online scams, phishing, and deceptive schemes. Both **hate speech** and **obscene or deepfake content** accounted for **10%** each, revealing the persistent challenges of harmful and manipulated content circulating in digital spaces. This distribution emphasizes the multifaceted nature of cybercrime and the need for tailored prevention strategies addressing each category.

Figure 1: Distribution of Cybercrime Types on Facebook (2023)



Identity-related offenses such as fake profiles, impersonation, and phishing scams dominate the landscape, reflecting users' vulnerability to data theft. Cyberbullying, particularly among younger users, also accounts for a substantial share. Financial fraud cases—often linked to phishing links and scam ads—form a notable portion. Hate speech and the circulation of explicit content, though less frequent, are critical issues due to their social impact.

- **Legal Provisions and Platform-Wise Applicability**

The study also compared the legal remedies available under Indian law for common types of cybercrime occurring on social media. These are summarized below:

Table 2: Applicable Legal Provisions for Social Media Cybercrimes in India

Type of Crime	Applicable Laws	Relevant Sections	Penalty
Identity Theft	IT Act 2000	Section 66C, 66D	Up to 3 years imprisonment + fine
Cyberbullying	IPC + Juvenile Justice Act	IPC Section 509, Sec 354D	Up to 3 years imprisonment + fine

Financial Fraud	IT Act + IPC	Section 66D, IPC 420	Up to 7 years imprisonment + fine
Hate Speech	IPC	Sections 153A, 295A	Up to 3 years imprisonment
Obscene/Deepfake Content	IT Act	Sections 67, 67A	5–7 years imprisonment + fine

While the legal framework exists, enforcement is uneven across platforms. A major concern is the continued mention of **Section 66A of the IT Act**, which was declared unconstitutional in Shreya Singhal vs. Union of India (2015), but is still wrongly cited in FIRs and complaints. Moreover, newer challenges like **AI-generated deepfakes** require legislative attention, as the existing IT Act may not comprehensively address them.

- **Comparative Platform Analysis**

Based on both technical trends and legal provisions, the platforms exhibit unique cybercrime patterns:

Table 3: Comparative Platform Analysis

Platform	Predominant Crimes	Technical Vulnerability	Legal Action Frequency
Facebook	Identity Theft, Cyberbullying	Account cloning, phishing links	Moderate
Instagram	Cyberbullying, Obscene Content	DMs, story-based trolling	Moderate
Twitter	Hate Speech, Misinformation	Virality of content, fake accounts	Low
YouTube	Obscene Content, Financial Fraud	Video descriptions, ads	Low

The table highlights that **Facebook and Instagram** are more prone to crimes targeting individuals directly (like bullying and identity theft), while **Twitter and YouTube** tend to be exploited for spreading hate speech and misleading content. The legal action frequency is relatively lower for Twitter and YouTube, possibly due to the **difficulty of tracking anonymous content** and **lack of clear content moderation standards**.

Conclusion

The findings clearly demonstrate a sharp increase in the number of reported cases, with Facebook and Instagram emerging as the most affected platforms. Identity theft, cyberbullying, and the spread of obscene content were identified as the dominant forms of cybercrime, particularly on platforms with a high user base and interactive features such as personal messaging and content sharing. The analysis also reveals critical technical vulnerabilities that are frequently exploited, including phishing links, fake profiles, and the misuse of private messaging features. Twitter and YouTube, while showing lower overall volumes, presented significant issues related to hate speech, misinformation, and deepfake content. These crimes pose not only legal but also social and psychological challenges for users and enforcement agencies alike. From a legal standpoint, the study highlights the presence of relevant laws under the IT Act, 2000 and the Indian Penal Code, yet underscores serious concerns regarding their implementation. Despite judicial interventions, outdated or repealed provisions like Section 66A of the IT Act continue to be misapplied. Moreover, the rapidly evolving nature of technology—such as AI-generated deepfakes—exposes gaps in current legislation, calling for urgent policy updates and specialized digital law enforcement training.

While India has established foundational frameworks for dealing with cybercrime, the current mechanisms are insufficient to address the complex and dynamic challenges presented by social media. There is a pressing need for coordinated efforts involving legal reform, technological safeguards, platform accountability, and public awareness to foster a safer digital

ecosystem.

References

1. Albert J. Marcellai and Robert S. Greenfield, *Cyber Forensics: A Field Manual for Collecting, Examining and Processing Evidence of Computer Crime* (Aurebuch Publications, London, 2002).
2. Amita Verma, *Cyber Crimes & Law* (Central Law House Publications, Allahabad, 1st edn., 2009).
3. Anirudh Rastogi, *Cyber Law – Law of Information Technology and Internet* (LexisNexis, Reed Elsevier India Pvt. Ltd., Gurgaon, 1st edn., 2014).
4. A.J.W. Turnere, *Kenny's Outlines of Criminal Law* (Cambridge University Press, U.K., 1st edn., 1962).
5. Austin, *Lectures on Jurisprudence: The Philosophy of Positive Law* (J. Murray, London, 1st edn., 1920).
6. Bary C. Collin, *The Future of Cyber Terrorism* (University of Illinois, Chicago, 1996).
7. C. Gringras, *The Laws of Internet* (Butterworth, U.K., 1997).
8. Chris Reed and John Angel, *Computer Law* (Oxford University Press, U.K., 3rd edn., 2003).
9. D. S. Yadav, *Foundation of Information Technology* (New Age International Publication Ltd., 3rd edn., 2007).
10. David Bainbridge, *Introduction to Computer Law* (Financial Times Management, 4th edn., 2000).
11. David S. Wall, *Cybercrime: The Transformation of Crime in the Information Age* (Polity Press, Cambridge, U.K., 2007).
12. Dorothy Denning, *Activism, Hacktivism and Cyberterrorism: The Internet as a Tool for Influencing Foreign Policy* (AMC Press, New York, 2001).
13. D. Thomas and B. D. Loader, *Cybercrime: Law Enforcement, Security and Surveillance in the Information Age* (Routledge, 1995).
14. Edwin H. Sutherland, *Principles of Criminology* (Chicago Lippincott, 1947, 1965).