Social Media Cybersecurity in India: Threats and Regulation

Ms. Brunda Anand
Student, Bms College of Commerce and Management Bengaluru

***Abstract*—social media is India's de facto digital public square. With hundreds of millions of active identities and deeply mobile-first usage, platforms such as YouTube, Instagram, Facebook, and X underpin communication, commerce, and civic debate. This scale has brought an expanding attack surface: account takeovers, large-scale scraping, deepfake-enabled fraud, coordinated influence operations, and data protection failures. Meanwhile, India's regulatory posture has evolved quickly via the Information Technology Act, 2000; the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (as amended); the Indian Computer Emergency Response Team (CERT-In) Directions of 2022; and the Digital Personal Data Protection Act, 2023 (DPDP Act), with implementing rules under consultation in 2025. This paper surveys the threat landscape and legal regime, analyzes notable India-relevant incidents, and proposes a control framework and policy roadmap tailored to Indian platforms, enterprises, and users. It concludes with a 12-month implementation plan and a research agenda for provenance, transparency, and safety metrics.**

***Index Terms*—social media, cybersecurity, India, deepfakes, misinformation, CERT-In, DPDP Act, IT Rules 2021, incident reporting, platform governance.**

I. INTRODUCTION

India now hosts one of the world's largest online populations, with Internet and social media adoption continuing to accelerate. The security of social platforms directly affects consumers, markets, media, and public institutions. Distinct from enterprise security, social-platform risk blends identity compromise, social engineering, and content integrity threats. For example, a single compromised high-reach account (journalist, brand, politician, "finfluencer") can seed large-scale financial fraud or misinformation.

This study addresses four research questions: 1. What are the dominant cyber threats and abuse patterns on social platforms in India? 2. How do India's laws and regulations (IT Rules, CERT-In Directions, DPDP Act) shape platform obligations and enterprise practices? 3. Where are the most material gaps in prevention, detection, response, and accountability across stakeholders (platforms, brands, creators, regulators, and users)? 4. Which technical and governance controls measurably reduce harm in 2025, and how should organizations prioritize them over 12 months?

II. LITERATURE AND POLICY CONTEXT

Academic and industry research on platform safety highlight's identity-centric risks (password reuse, SIM-swap, OAuth abuse), behavioral manipulation (coordinated inauthentic behavior), and media integrity (synthetic/deepfake content). Indian policy developments—most notably the 2021 Intermediary Rules (as amended in 2022 and 2023), CERT-In's 2022 incident-reporting Directions, and the DPDP Act (2023)—establish due diligence, real-time visibility, and personal data-protection principles. Sectoral guidance (e.g., stock exchanges' public warnings on deepfakes) demonstrates the crossover between market integrity and social-media abuse.

III. METHODOLOGY

This paper synthesizes primary legal and policy documents (official gazette/MeitY texts of the DPDP Act and IT Rules; CERT-In Directions and FAQs), regulator and exchange advisories, reputable media coverage of major incidents, and cross-checked industry analyses. Where precise figures vary by source, ranges are given or conservative values are

used. Case studies focus on India-relevant incidents from 2022–2025.

Threat Landscape in India's Social Media Ecosystem

Account Takeover (ATO) and Credential Abuse

Attackers leverage password reuse, credential stuffing, OAuth token theft, SIM-swap fraud, and session hijacking to compromise high-reach accounts. Post-compromise, adversaries push crypto/forex scams, investment fraud, or malvertising; they also pivot into brand-impersonation storefronts. Given India's large creator economy, takeover of verified handles can trigger rapid victimization.

Controls: phishing-resistant MFA (FIDO2/WebAuthn passkeys) for admins and creators; hardware keys for brand accounts; risk-based session management; automated lock-down of suddenly high-reach posts; least-privilege access to ad accounts and social-management tools.

Data Scraping, Enumeration, and Privacy Harms

Automated scraping of public profiles and misconfigured APIs have led to mass-scale compilations of user metadata globally, increasing doxxing and phishing risk for Indian users. Even when datasets contain "public" fields, aggregation changes harm profiles.

Controls: abuse-resistant APIs; rotating fingerprints and bot-mitigation; privacy-by-default profile settings; monitoring for bulk enumeration; legal takedowns.

AI-Generated Content and Deepfakes

From late-2023 onward, deepfakes matured into a material security risk. Indian exchanges warned investors in 2024 after CEO-impersonation videos promoted stocks on social media. Synthetic celebrity/political content also spiked around elections.

Controls: proactive media forensics; C2PA/content provenance for official announcements; rapid impersonation takedown channels; "official sources" verification hubs; user-facing authenticity labels with clear limitations.

Fraud, Malvertising, and Gray-Market Affiliates

Fraudsters use paid ads and organic posts to route users to fake exchanges, "pump-and-dump" groups, or phishing pages. Weak vetting of affiliates and new domains exacerbates risk.

Controls: ad-supply-path security; allowlisted domains; affiliate KYC and clawbacks; creative hashing and reuse detection; multi-approver ad-spend workflows.

Mobile Malware and Social Engineering

Attack chains commonly begin with social DMs or posts, then pivot to encrypted messengers. Info-stealers and RATs target one-time passwords, cookies, and keystore items.

Controls: link-safeguards in UIs; OS-level warnings; app-store vetting; public advisories in Indian languages; enterprise device-health baselines for social-team laptops/phones.

Harassment, Harmful Content, and Terror/CSAM

While safety operations are broader than "cybersecurity," counter-abuse pipelines (hash-matching, classifier feedback, adversarial testing) intersect with integrity engineering. The Intermediary Rules require reasonable efforts to prevent defined harms, user grievance redressal, and timely action on lawful orders.

Indian Legal and Regulatory Framework

Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (as amended)

The IT Rules establish due-diligence obligations for intermediaries, including social media platforms: appointing grievance officers, enabling user grievance mechanisms, publishing transparency reports, acting upon lawful orders within prescribed timelines, and (for significant intermediaries) appointing compliance roles and providing an appeals path via the Grievance Appellate Committee (GAC). Amendments in 2022 and 2023 strengthened obligations around misinformation and user safety.

Digital Personal Data Protection Act, 2023 (DPDP Act)

Notified on August 11, 2023, the DPDP Act introduces consent-centric processing, children's data protections, obligations for (Significant) Data Fiduciaries, data principal rights, and penalties for non-compliance. Draft implementing rules circulated in early 2025 are expected to specify breach-notification mechanics, timelines, and thresholds, and operationalize enforcement via the Data Protection Board.

Sectoral and Market Integrity Measures

Financial-market actors and exchanges have warned about deepfake stock-tip scams and unregulated "finfluencer" activity, urging verification of official handles and cautioning that employees are not authorized to recommend stocks.

Case Studies (2022–2025)

Case 1: Deepfake CEO Stock-Tips Targeting Indian Investors (2024)

In April 2024, deepfake videos impersonating CEOs of India's major exchanges circulated on social media with stock recommendations. Exchanges publicly cautioned investors and flagged law-enforcement escalation. **Lessons:** official-source verification hubs reduce confusion; provenance watermarks on executive communications aid rapid debunking; platforms should fast-track impersonation removals for critical public-interest roles.

Case 2: Large-Scale Scraping and Public Data Aggregation (2022–2025)

Multiple global incidents produced mega-dumps of profile data (emails, phone numbers, handles). Even if sourced from "public" fields, compilation enables targeted phishing of Indian users at scale. **Lessons:** platform-side rate limits and anomaly analytics are necessary but insufficient; transparency about scope and user protection steps (e.g., HIBP-style notifications) improves trust.

Case 3: Brand-Handle Account Takeovers (Ongoing)

Indian brands, creators, and public figures have reported waves of phishing leading to ad-account hijacks and crypto scams. **Lessons:** enforce passkeys/hardware keys for social admins; segregate ad spend with approvals; rehearse rapid communications from backup channels to contain harm.

Control Framework for Indian Organizations

Identity, Access, and Administration

- Mandate phishing-resistant MFA (passkeys / FIDO2) for all brand/creator handles and admins; disable SMS-only where feasible.
- Provisioning hygiene: SCIM/JIT for social-tool access; revoke stale tokens; enforce least privilege on ad accounts, pages, and APIs.
- Session defense: device binding, step-up auth on risky actions, and rapid session revoke.

Content Authenticity and Official Communications

- Provenance: adopt C2PA signing/watermarking for executive videos and critical advisories; publish verification pages listing official channels.
- Media forensics: deploy synthetic-media detection on inbound reports; coordinate with platforms for takedowns; maintain media hashing for repeat detection.
- Public playbooks: pre-draft "deepfake/impersonation" incident statements and FAQ.

Threat Monitoring and Takedowns

- Brand-protection services for impersonation, typo-domains, and marketplace abuse; MTTA/MTTR SLAs.
- Language coverage: monitor Indic-language variants and transliterations.

Incident Response and CERT-In Alignment

- Six-hour reporting playbooks: triage → provisional report to CERT-In → preserve evidence/logs → iterative updates; maintain 180-day logs in India with synchronized time sources.
- Cross-functional response: security, legal, PR, and customer-support alignment; media monitoring to measure reach and remediation.

Data Protection and DPDP Readiness

- Consent & notices: align collection/use notices for social sign-ins; data minimization; children's data safeguards.
- Breach preparedness: templates for user notifications; evidence packs for authorities; clear role mapping for Data Fiduciary vs. Processor.
- Vendor governance: assess social-tool processors; localize logs and telemetry as required.

Ads and Affiliate Integrity

- Allowlisted domains and brand-verified landing pages; block newly registered domains until vetted.
- Affiliate KYC and contractual clawbacks; automated detection of reused scam creatives.

Metrics and Assurance

Track: % handles on passkeys; impersonation-takedown median hours; number of scam clusters neutralized; % incidents reported within 6 hours; % logs retrievable; transparency report cadence; user complaint outcomes (including GAC escalations).

User-Focused Guidance (India)

- Enable passkeys or app-based MFA on all social accounts; avoid SMS-only.
- Treat sensational "CEO tips" and celebrity endorsements with skepticism; verify via official websites before acting.
- Do not click shortened links from unknown DMs; use in-platform reporting tools in Indian languages when available.
- Use a password manager and unique passwords; regularly review connected apps.

Policy Discussion and Recommendations

India's layered approach—intermediary due diligence (IT Rules), real-time visibility (CERT-In Directions), and data-protection principles (DPDP)—provides an evolving foundation. To balance safety, privacy, and innovation:

For Government and Regulators

1. Finalize DPDP Rules with phased timelines and clear breach-notification triggers and formats.
2. Standardize CERT-In reporting templates (IODEF/STIX) and provide safe-harbor incentives for rapid disclosure.
3. Pilot cross-platform provenance/watermarking for designated public-interest communications (disaster alerts, market integrity), with robust transparency about limitations.
4. Publish anonymized incident trend reports and promote researcher access that respects privacy.

For Platforms

1. Default to passkeys for high-reach accounts; enforce device attestation for admin consoles.
2. Strengthen anti-scraping posture (rate limits, behavioral ML, "privacy by default").
3. Launch India-specific impersonation fast lanes tied to verified public-interest roles.
4. Expand Indic-language safety operations and disclosures.

For Enterprises and Creators

1. Secure social toolchains (SSO, SCIM, least privilege).
2. Pre-register official channels with customers/investors; sign executive media.
3. Rehearse six-hour CERT-In reporting; maintain evidence packs.
4. Measure outcomes (harm reduction), not only takedown counts.

12-Month Implementation Roadmap (Indian Enterprises)

Quarter 1: Inventory all handles and admins; mandate passkeys/hardware keys; designate CERT-In liaison; publish an "official sources" page; contract a takedown vendor with SLAs.
Quarter 2: Complete DPDP gap analysis; implement breach-notification playbooks; roll out provenance signing for executive media; segregate ad-account spend with approvals.
Quarter 3: Conduct a red-team on social-channel takeover; simulate deepfake/impersonation response and CERT-In six-hour reporting; launch investor/customer verification campaigns.
Quarter 4: Audit log retention and time-sync; review cross-border data flows; release the first safety & integrity transparency note.

Research Agenda (India)

1. Robust deepfake provenance & detection across compressions, translations, and edits; evaluate watermark durability in the wild.
2. Impact metrics for interventions (fraud losses averted, impersonation dwell-time reduction) rather than content-removal counts.
3. Privacy-preserving open datasets for abuse research (graph snapshots, synthetic corpora).
4. Localized threat modeling for Indic languages and cultural contexts.
5. Efficacy of six-hour reporting on containment and public transparency.

IV. CONCLUSION

Securing India's social-media ecosystem requires simultaneous advances in platform security engineering, regulatory clarity, and user literacy.

Near-term priorities include phishing-resistant authentication, anti-scraping defenses, deepfake response workflows with provenance for official communications, and transparent, privacy-respecting incident handling aligned with Indian requirements. With coordinated action among government, platforms, enterprises, and users, India can preserve the benefits of social media while measurably reducing cyber harm.

REFERENCES

[1] Ministry of Electronics & IT (MeitY). Digital Personal Data Protection Act, 2023 (No. 22 of 2023). Official Gazette/MeitY texts.
[2] MeitY. Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (updated through 2023).
[3] Indian Computer Emergency Response Team (CERT-In). Directions under Section 70B (April 28, 2022); related FAQs and guidance.
[4] DataReportal. Digital 2025: India (Feb. 2025) and Global Overview Report (2025).
[5] Reuters coverage of Indian stock exchanges' 2024 warnings about deepfake CEO videos and related market-integrity issues.
[6] Industry analyses and legal commentaries on IT Rules, CERT-In Directions, and DPDP operationalization (2022–2025).