myresearchgo

myresearchgo Volume 1, September Issue 6, 2025, ISSN: 3107-3816 (Online)

Data Security on Social Media Under Threat for Training AI Modules

[1] Atharva R. Khot,[2] Mansi D. Agrawal
[1]Student, [2]Student
Department of Information Technology
S. D. S. M. College, Palghar, Maharashtra, India

Abstract

Intelligence (AI) systems increasingly rely on large-scale data drawn from social media platforms. While these datasets enable powerful predictive models, they also introduce substantial risks to privacy, security, and user consent. This research paper explores the vulnerabilities associated with social media data when used to train AI modules. Drawing on existing scholarship, we highlight the ethical, legal, and technical challenges posed by unauthorized scraping, re-identification, and governance gaps. Through the works of Robertson et al., Frichot, and Gerbrandt & Howard, we emphasize how current practices undermine user trust and call for stronger data governance and privacy-preserving mechanisms. The paper concludes that without meaningful safeguards, AI development risks exacerbating digital exploitation and eroding privacy rights.

Keywords: Data Security, Social Media, AI Training, Privacy, Governance, Ethics

myresearchgo Volume 1, September Issue 6, 2025, ISSN: 3107-3816 (Online)

Introduction

The explosive growth of Artificial Intelligence (AI) has been fueled by the availability of massive datasets, much of which originates from social media platforms. Social media provides a continuous flow of diverse, real-time content, including text, images, videos, and behavioral data. For AI developers, such material offers invaluable insights for improving natural language processing (NLP), image recognition, recommendation systems, and generative models. The richness and volume of this content enable the training of powerful algorithms capable of mimicking human communication and predicting behavior patterns with unprecedented accuracy.

However, this dependence on social media introduces critical risks. Social media content is often shared by users for personal communication or entertainment, not with the intention of being harvested for AI training. Once extracted, data can be stored indefinitely, stripped of its original context, and repurposed without consent. This leads to a host of issues, including the violation of intellectual property rights, misuse of personally identifiable information (PII), and the possibility of re-identifying users from supposedly anonymized datasets. As Robertson et al. argue, the reframing of social media posts into training modules often strips nuance and exposes users to unanticipated harms, such as profiling or surveillance.

The ethical implications of such practices are equally pressing. Frichot emphasizes the power asymmetry between technology corporations and ordinary users. Platforms act not only as gatekeepers of massive amounts of user data but also as arbiters of how that data is monetized, filtered, and shared with AI developers. This dynamic leaves users with little agency, as terms of service agreements rarely provide genuine choice or transparency. The commodification of user activity raises significant questions about consent, ownership, and the erosion of digital autonomy in the age of AI.

Furthermore, governance frameworks and legal safeguards are still struggling to keep pace. While regulations like the European Union's GDPR and India's DPDP Act (2023) attempt to enforce transparency and accountability, enforcement remains inconsistent, especially across global platforms. Gerbrandt & Howard's work on moderation exemptions highlights how even content regulation loopholes, such as those created under the "newsworthiness" clause, can allow harmful or privacy-infringing content to circulate unchecked. This complicates the use of such content in training datasets, where bias, misinformation, or harmful materials can be unintentionally encoded into AI systems.

This paper argues that the intersection of AI training and social media data represents a new frontier of cybersecurity and ethical risk. By examining the vulnerabilities and governance gaps identified in existing scholarship, we aim to build a comprehensive threat model for understanding the dangers of repurposing social media data. The ultimate objective is to highlight the urgent need for privacy-preserving mechanisms, transparent governance structures, and stronger user protections in order to balance AI innovation with ethical responsibility.

Literature Review

Robertson et al. argue that social media serves as a rich, yet problematic, reservoir for AI training. The authors show how social media data, when transformed into training modules, risks stripping context, distorting meaning, and creating vulnerabilities around identity exposure. Their study emphasizes the need to

re-evaluate how AI researchers collect, clean, and reframe such data to mitigate unintended harm.
Frichot's contribution addresses the ethical and governance implications of social media surveillance. The paper critiques the concentration of power in platform owners who not only collect vast quantities of data but also decide how it is monetized and repurposed. This creates ethical dilemmas regarding consent, user autonomy, and the commodification of online behaviors.
Gerbrandt & Howard analyze content moderation policies, specifically focusing on exemptions granted under the 'newsworthiness' criterion. They demonstrate how such policies enable harmful or privacy-violating content to persist, complicating accountability. This is directly relevant to AI training, as datasets often incorporate unmoderated or biased content, perpetuating systemic risks.

Research Methodology

This research adopts a qualitative approach by analyzing existing literature and case studies to construct a threat model for social media data security in AI training. The methodology involves:
- Reviewing scholarly works (Robertson et al., Frichot, Gerbrandt & Howard) to identify recurring themes of risk.
- Mapping governance gaps where legal and ethical safeguards fall short.
- Synthesizing insights to propose a framework for balancing AI innovation with user rights.

By combining ethical analysis with governance critiques, this approach ensures a holistic understanding of how social media data becomes vulnerable when used for AI development.

Findings

The literature reveals several core threats:
- Unauthorized Scraping and Data Harvesting: Platforms struggle to control third-party collection of user data.
- Loss of Context and Meaning: When social media content is reframed as training data, nuances are stripped, increasing risks of misrepresentation.
- Power Imbalances: Users lack meaningful consent mechanisms, while corporations retain disproportionate control.
- Moderation Gaps: Loopholes in content regulation (e.g., 'newsworthiness' exemptions) undermine accountability.
- Governance Challenges: Current legal frameworks, such as GDPR and DPDP 2023, remain insufficiently enforced.

Discussion

The findings highlight a pressing dilemma: while AI benefits from large-scale social media datasets, the absence of robust governance structures puts user data at risk. Robertson et al. reveal the technical risks of decontextualizing content, while Frichot exposes the ethical power imbalances that leave users vulnerable to exploitation. Gerbrandt & Howard add a legal dimension, showing how moderation gaps and selective enforcement enable privacy violations. Together, these insights underscore the need for interdisciplinary

solutions. Future work must integrate differential privacy, provenance frameworks, and user-centric governance models to align AI development with ethical standards.

Conclusion

This paper concludes that social media data security is critically endangered when repurposed for AI training modules. The review of Robertson et al., Frichot, and Gerbrandt & Howard demonstrates that technical vulnerabilities, ethical dilemmas, and governance gaps converge to amplify risks. Without stronger legal safeguards, transparency in AI pipelines, and privacy-preserving training methods, users remain exposed to exploitation. The future of AI must balance innovation with trust by embedding data security and ethical responsibility into every stage of model development.

References

- Robertson, et al. (2025). Social Media, Social Intelligence and Training Modules.
- Frichot, Emma. (2025).
- Gerbrandt, J., & Howard, P. (2025). Should Social Media Platforms Permit Violating Content That Is ‘Newsworthy’?