



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 3, March 2026

Algorithmic Amplification and Generation Z Mental Health: A Review of AI-Driven Social Media, Psychological Vulnerability, and Governance Frameworks

Apurva Shaileshbhai Patel, Dr. Sneha Patel

Student, Bhagwan Mahavir College of computer application, Bhagwan Mahavir University, Surat

Assistant Professor, Bhagwan Mahavir College of Computer Application, Bhagwan Mahavir University, Surat, India

ABSTRACT: Generation Z (Gen-Z), born between 1997 and 2012, is the first cohort to mature entirely within AI-curated digital ecosystems. Artificial intelligence-driven recommender systems deployed by platforms such as TikTok, Instagram, YouTube, and Facebook continuously optimize content feeds to maximize user engagement. Emerging evidence from 2024–2025 indicates that these algorithmic amplification mechanisms intensify emotional dysregulation, compulsive engagement, attentional fragmentation, body dissatisfaction, and exposure to online hostility among Gen-Z users. This paper synthesizes interdisciplinary findings from psychology, neurodevelopmental science, AI ethics, and public health to examine how algorithmic personalization interacts with Gen-Z’s developmental sensitivity. A five-layer vulnerability model is proposed and governance implications are evaluated within the evolving global AI regulatory landscape.

KEYWORDS: Algorithmic Amplification, Generation Z, Mental Health, Emotional Dysregulation, Adolescent Neurodevelopment, Governance, Recommender Systems

I. INTRODUCTION

In recent years, artificial intelligence (AI) has significantly transformed social media, shaping the experiences of adolescents and young adults. Generation Z (1997–2012) is the first generation to grow up fully within algorithmically curated digital environments. Unlike earlier generations, their online experiences are continuously shaped by AI-driven recommender systems.

Platforms such as TikTok, Instagram, YouTube, and Facebook use behavioral data—likes, watch time, and interactions—to personalize content and maximize engagement. While this improves user experience, growing research highlights unintended psychological effects, especially among adolescents whose brains are still developing.

A key mechanism behind this impact is algorithmic amplification, where emotionally intense and socially comparative content is prioritized. Repeated exposure to such content can lead to social comparison, body dissatisfaction, reduced attention span, and compulsive usage patterns. Features like infinite scrolling and reward-based engagement further reinforce these behaviors, making them similar to patterns observed in behavioral addiction.

II. LITERATURE REVIEW

1. Chang et al. (2025) examined how AI bias embedded within social media algorithms contributes to skewed content exposure among Gen-Z users. Their study demonstrated that biased recommender outputs disproportionately amplify negative self-comparisons and idealized lifestyle content, with measurable effects on anxiety scores and self-reported mental health in adolescent cohorts aged 14–22. [1]

2. Kelly and Sharot (2025) provided neurological and behavioral evidence demonstrating that online browsing patterns both reflect and actively shape mood states. Using longitudinal tracking of web-browsing sessions, they showed that

Volume 15, Issue 3, March 2026**[DOI: 10.15680/IJRSET.2026.1503207]**

individuals experiencing negative mood are algorithmically served more negative content, creating a recursive mood amplification cycle that is particularly dangerous for developing adolescent emotional regulation systems. [2]

3. Lakatos et al. (2025) applied the wanting-liking distinction from neurobehavioral addiction models to social media use, demonstrating that problematic Facebook use correlates more strongly with dopamine-mediated anticipatory craving than with actual satisfaction derived from use. This aligns social media reward architecture with substance-use disorder mechanisms, suggesting that AI-optimized platforms may be pharmacologically analogous in their neurological effects on adolescents. [3]

4. Thorell et al. (2024) published longitudinal findings linking intensive digital media exposure — particularly algorithmically curated short-form video — with increased ADHD-related symptoms in adolescents over a three-year follow-up period. The study controlled for prior ADHD diagnosis and concluded that algorithmically curated environments may not merely reflect pre-existing cognitive vulnerabilities but actively contribute to their reinforcement and exacerbation over time. [5]

5. Wu et al. (2024) investigated short-form video addiction specifically, demonstrating that lower psychological resilience significantly predicts stronger addictive patterns in TikTok consumption. Their findings confirm within-Gen-Z heterogeneity: risk is highest when personal resilience is low and algorithmic intensity is high, suggesting a multiplicative rather than additive vulnerability model. [4]

6. Sanzari et al. (2024) provided critical evidence reframing the body image debate: their study demonstrated that algorithmically curated content exposure predicts body dissatisfaction independently of total usage time. This finding indicates that it is the type and curation of content — driven by AI systems — rather than raw screen time, that determines psychological harm. [6]

7. Madriaza et al. (2025) conducted a systematic review and meta-analysis confirming significant associations between online hate exposure and depression, anxiety, reduced life satisfaction, and social withdrawal across adolescent populations. Engagement-driven ranking systems elevate polarizing and hostile content because outrage generates high interaction rates, indirectly normalizing online hostility within Gen-Z digital culture. [7]

8. Nesi et al. (2025) contributed longitudinal evidence directly linking Gen-Z social media exposure to internalizing symptoms — including depression and anxiety — over extended time periods. Their findings support a causal rather than merely correlational relationship between AI-curated feed exposure and deteriorating adolescent mental health, strengthening the case for regulatory intervention. [13]

III. METHODOLOGY

This study adopts a systematic literature review of interdisciplinary research published between 2024 and 2025. Data were collected from databases such as PubMed, Scopus, Google Scholar, IEEE Xplore, and Nature using keywords related to Gen-Z mental health, AI algorithms, and social media.

Studies were included if they were peer-reviewed, focused on AI-driven platforms, targeted adolescent or Gen-Z populations, and provided measurable outcomes. Studies not related to AI-based media or focused only on older adults were excluded [14].

The five-layer vulnerability model was developed through thematic analysis, linking algorithmic mechanisms to psychological and behavioral outcomes. Governance frameworks were analyzed using policy reports from the United Nations, OECD, European Commission, WHO, and UNICEF [8,9,10,18].

3.1 Neurodevelopmental Context

Adolescence and early adulthood are characterized by ongoing maturation of the prefrontal cortex — the brain region responsible for impulse control, attention regulation, long-term planning, and emotional modulation. This maturational process is not complete until approximately age 25, meaning that the entire Gen-Z cohort encounters AI-curated social media during a period of heightened neurological vulnerability. [5]

Volume 15, Issue 3, March 2026

[DOI: 10.15680/IJIRSET.2026.1503207]

Simultaneously, the adolescent limbic system — responsible for reward processing, emotional reactivity, and social sensitivity — is fully developed and highly active during this period. This creates a well-documented imbalance between a mature, reactive reward system and an immature, developing regulatory system. AI algorithms are architecturally optimized to exploit precisely this reward system through dopamine-triggering content delivery, variable reward schedules, and social validation mechanisms. The developmental mismatch between adolescent neurology and AI-optimized platform design is therefore structurally exploitative in its effect. [3,5]

IV. COMPARATIVE ANALYSIS OF KEY STUDIES

Table 1 presents a comparative summary of key studies reviewed in this paper.

Table 1: Comparative Summary of Key Studies

Author (Year)	Focus Area	Sample	Methods	Key Finding
Chang et al. (2025)	AI bias & Gen-Z anxiety	Ages 14–22	Survey, regression	AI bias increases anxiety & negative social comparison
Kelly & Sharot (2025)	Browsing & mood feedback loops	Multi-session logs	Longitudinal tracking	Negative mood → more negative content (recursive cycle)
Lakatos et al. (2025)	Wanting vs. liking; Facebook use	Young adults	Survey, neuropsych model	Craving (not enjoyment) drives problematic social media use
Thorell et al. (2024)	ADHD symptoms & media exposure	Adolescents, 3-yr	Longitudinal study	AI media exposure reinforces ADHD symptoms over time
Wu et al. (2024)	TikTok addiction & resilience	Gen-Z users	Structural equation model	Low resilience × high algorithm = strongest addiction risk
Sanzari et al. (2024)	Body image & curated content	Adolescent females	Content analysis	Content type — not screen time — predicts body dissatisfaction
Valkenburg et al. (2024)	Differential susceptibility	Adolescents	Meta-analysis	Susceptible teens most harmed by upward social comparison
Madriaza et al. (2025)	Online hate & mental health	Global adolescent	Systematic review & meta-analysis	Hate exposure → depression, anxiety, reduced life satisfaction
Nesi et al. (2025)	Social media & internalizing	Gen-Z longitudinal	Longitudinal cohort	Causal: AI feed exposure → depression & anxiety over time
Przybylski & Weinstein (2024)	Digital engagement & wellbeing	Large-scale adolescent	Preregistered study	Passive algorithmic consumption is most psychologically harmful

V. THE FIVE-LAYER GEN-Z VULNERABILITY MODEL

Table 2: The Five-Layer Gen-Z Vulnerability Model

Layer	Layer Name	Mechanism	Impact / Outcome
Layer 1	Algorithmic Layer	AI recommender systems maximize engagement by amplifying emotionally arousing and socially comparative content	Foundation of all downstream psychological effects
Layer 2	Behavioral Layer	Infinite scrolling, variable rewards, and notifications exploit operant conditioning	Compulsive usage habits and prolonged screen time

Layer 3	Cognitive Layer	Rapid content switching and appearance-focused content	Attention fragmentation, reduced concentration, impaired executive function
Layer 4	Emotional Layer	Mood-responsive algorithms amplify negative emotional states	Emotional feedback loops leading to sustained negative mood
Layer 5	Clinical Outcome Layer	Combined effect of all previous layers	Anxiety, depression, ADHD symptoms, body image issues, social withdrawal

5.1 Social Comparison and Body Image

The algorithmic amplification of appearance-focused content represents one of the most clinically significant harm pathways for Gen-Z users. Sanzari et al.'s (2024) critical finding — that it is the type of algorithmically curated content, not total duration of exposure, that predicts body dissatisfaction — has profound implications for policy. It means that time-limit interventions, currently the dominant regulatory and parental response, may be largely ineffective when users receive a concentrated dose of idealized content in shorter sessions. [6]

Hyper-personalization amplifies this harm by tailoring idealization to each user's specific insecurities and aspirational reference points, identified through extensive behavioral profiling. Where earlier generations encountered generalized beauty standards, Gen-Z users are served algorithmically personalized idealization — content calibrated to individual social comparison tendencies, making upward comparison both more frequent and more psychologically potent. [16]

5.2 Online Hostility and Hate Exposure

The algorithmic amplification of hostile and polarizing content represents a distinct but equally significant harm pathway. Engagement-driven ranking systems elevate controversial and hateful content because it generates disproportionately high interaction rates — comments, shares, and emotional reactions — all interpreted by AI systems as signals of valuable content. Gen-Z users, the heaviest social media users and most vulnerable cohort, are systematically over-exposed to online hostility. [7]

Madriaza et al.'s (2025) comprehensive meta-analysis found that online hate exposure follows a dose-response pattern — the greater the frequency and intensity of exposure, the worse the mental health outcomes — suggesting that algorithmic systems amplifying hostile content function as a population-scale mental health risk factor. The WHO's (2024) formal recognition of digital hostility as a global public health concern reflects the growing institutional consensus that structural algorithmic intervention is urgently needed. [7,10]

VI. GOVERNANCE AND PROTECTION FRAMEWORKS

Protecting Gen-Z mental health at scale requires systemic regulatory accountability rather than relying solely on parental monitoring or individual digital literacy. Several international governance frameworks have emerged in direct response to growing evidence of algorithmic harm to adolescent populations.

United Nations AI Advisory Body (2024): Emphasizes transparency, accountability, and youth-centered safeguards in AI systems. Recommends that platforms disclose algorithmic ranking criteria and provide opt-out mechanisms for engagement-optimization features specifically when serving minors. [8]

OECD AI Governance Framework (2024): Recommends explainability standards and risk-based classification systems for AI platforms. Under proposed classifications, social media recommendation engines serving minors would be designated high-risk, triggering mandatory impact assessments and ongoing monitoring obligations. [9]

European Commission Youth AI Framework (2025): Introduces mandatory youth impact assessments for high-risk AI systems, requiring platforms to demonstrate that their recommendation architectures do not systematically expose minors to harmful content patterns before deployment. [15]

Volume 15, Issue 3, March 2026**[DOI: 10.15680/IJRSET.2026.1503207]**

World Health Organization (2024): Recognizes digital hostility and algorithmic mental health risks as rising global public health issues disproportionately affecting adolescent populations, calling for cross-sector collaboration between technology companies, health systems, and regulators to design protective standards. [10]

UNICEF AI and Children Policy (2025): Promotes child-rights-based AI governance principles globally, including the right of children to algorithmic transparency, the right to non-harmful curation, and the right to meaningful remediation when AI systems cause documented psychological harm. [18]

These convergent policy efforts signal a growing international consensus that Gen-Z protection from algorithmic harm requires coordinated systemic regulation addressing the structural architecture of AI platforms — not merely the behaviors of individual users.

VII. ISSUES IN RESEARCH

Despite recent progress, several limitations remain. Most studies are observational, making it difficult to establish clear causal relationships between algorithmic features and mental health outcomes [14].

There is also limited representation of non-Western populations, reducing the global applicability of findings. Additionally, research on well-being-focused algorithm design is still lacking.

Further gaps include limited neuroimaging evidence and the absence of standardized measures for “algorithmic intensity,” making cross-study comparisons difficult.

VIII. FUTURE SCOPE

The trajectory of this field points toward several promising directions. Large-scale, longitudinal cohort studies tracking both algorithmic exposure metrics and mental health outcomes across the full Gen-Z developmental window will be essential to establishing robust causal models. Such studies should incorporate passive sensor data, neuropsychological assessments, and platform behavioral logs to achieve unprecedented depth of algorithmic exposure measurement.

Researchers will increasingly focus on development and evaluation of well-being-optimized recommendation architectures. Rather than optimizing for engagement duration, next-generation algorithms could be designed to optimize for self-reported well-being, emotional balance, and educational enrichment. Pilot implementations in collaboration with platform partners could provide experimental evidence for feasibility and effectiveness of alternative design paradigms.

From a governance perspective, future work should develop standardized youth algorithmic impact assessment methodologies — analogous to environmental impact assessments — that platforms must complete before deploying or significantly modifying recommendation systems. International harmonization of these standards, coordinated through the UN, OECD, and WHO, will be essential to prevent regulatory arbitrage by platforms operating across multiple jurisdictions. Interdisciplinary research partnerships between computer scientists, psychologists, neuroscientists, ethicists, and policy makers will be indispensable, as no single discipline possesses the combined expertise needed to simultaneously redesign algorithmic systems, establish clinical frameworks, and build adequate regulatory instruments. [8,9,10]

IX. CONCLUSION

Generation Z is the first generation to have grown up entirely surrounded by AI-driven digital platforms, and research from recent years makes it increasingly clear that the algorithms powering these platforms do not simply reflect human behavior — they actively shape it. The five-layer model presented in this paper traces how platform architecture leads to real psychological harm: from the way recommendation engines are built, to the habits they form, the attention they fragment, the emotions they amplify, and the mental health conditions they ultimately contribute to.

The dangers Gen-Z faces online are not random or accidental. Compulsive usage patterns, shortened attention spans, relentless social comparison, and exposure to hostility and hate are all foreseeable outcomes of systems designed to

Volume 15, Issue 3, March 2026**| DOI: 10.15680/IJRSET.2026.1503207 |**

keep users engaged at any cost. When the goal of a platform is to maximize time spent rather than user well-being, harm becomes a built-in feature, not a side effect.

Telling young people to simply “use their phones less” is not a real solution. The problem is structural, and so must be the response. What is needed is a three-pronged approach: first, platforms must be redesigned so their algorithms prioritize psychological health over engagement numbers; second, young people need meaningful education that helps them understand and navigate the algorithmic environments they live in; and third, governments and international bodies must establish and enforce clear legal standards that hold technology companies accountable for the measurable harm their systems cause. A generation that has grown up inside the algorithm deserves a digital world built around human values — not profit-driven metrics.

REFERENCES

- [1] J. P.-C. Chang, S.-W. Cheng, S. M.-J. Chang, and K.-P. Su, “Navigating the digital maze: AI bias, social media, and mental health in Generation Z,” *AI*, vol. 6, 2025.
- [2] C. A. Kelly and T. Sharot, “Web-browsing patterns reflect and shape mood and mental health,” *Nature Human Behaviour*, vol. 9, 2025.
- [3] D. Lakatos et al., “Wanting and liking of Facebook functions and their correlation to problematic use,” *Scientific Reports*, vol. 15, 2025.
- [4] S. Wu, Z. Shafait, and K. Bao, “Proactive personality and short-form video addiction,” *PLoS ONE*, vol. 19, 2024.
- [5] L. B. Thorell et al., “Longitudinal associations between digital media use and ADHD symptoms,” *European Child & Adolescent Psychiatry*, vol. 33, 2024.
- [6] C. M. Sanzari et al., “Social media content exposure and body image outcomes,” *Eating Behaviors*, 2024.
- [7] P. Madriaza et al., “Exposure to hate in online and traditional media: A systematic review and meta-analysis,” *Campbell Systematic Reviews*, vol. 21, 2025.
- [8] United Nations AI Advisory Body, *Governing AI for Humanity: Final Report*, United Nations, 2024.
- [9] OECD, *AI Governance Framework Update*, Paris, 2024.
- [10] World Health Organization, *Adolescent Mental Health in the Digital Era: Global Report*, Geneva, 2024.
- [11] American Psychological Association, “Health advisory on social media use in adolescence: 2024 update,” 2024.
- [12] C. Montag, H. Yang, and J. D. Elhai, “Psychological mechanisms underlying TikTok use,” *Frontiers in Public Health*, 2024.
- [13] R. B. Nesi et al., “Adolescent social media use and internalizing symptoms,” *Journal of Adolescent Health*, 2025.
- [14] M. Odgers and C. Jensen, “Annual research review: Adolescent mental health in the age of algorithmic media,” *Journal of Child Psychology and Psychiatry*, 2024.
- [15] European Commission, *AI Risk Classification and Youth Protection Framework*, Brussels, 2025.
- [16] T. Valkenburg, P. Beyens, and I. Pouwels, “Differential susceptibility to media effects,” *Current Opinion in Psychology*, 2024.
- [17] A. Przybylski and N. Weinstein, “Digital engagement and adolescent well-being,” *Nature Human Behaviour*, 2024.
- [18] UNICEF, *AI and Children: Policy Guidance 2025 Edition*, New York, 2025.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details