Social Media by Design

A Literature Review on Engagement-Optimised Platforms and the Case for a Privacy-First Architecture

Anwar Syed Ali
Independent Researcher & Founder, Flair (https://useflair.app)
September 2026

Keywords: social media, engagement optimisation, dark patterns, algorithmic recommenders, digital wellbeing, attention economy, platform design, privacy-first architecture

ABSTRACT

Major social media platforms are architected to maximise time-on-screen through variable-reward recommendation algorithms, suppressed content limits, and a documented repertoire of manipulative interface patterns. Longitudinal and experimental evidence associates heavy use of these platforms with elevated rates of anxiety, depression, body dissatisfaction, and disrupted sleep, particularly among adolescents. Regulatory bodies, including the European Data Protection Board, have catalogued six categories of dark pattern employed by major platforms to override informed consent, and enforcement actions totalling several billion dollars have been brought against platform operators for related violations. This paper reviews the causal mechanisms proposed in the literature, evaluates the evidence for each, and presents Flair as a case study in privacy-first, anti-engagement platform architecture. The paper argues that meaningful reform requires action at the platform-design layer, not solely at the level of user behaviour or post-hoc regulation, and that the hypothesis of a bounded, satisfaction-signal-driven feed warrants rigorous empirical testing.

Contents

1. **Introduction**
2. **Problem Statement: The Architecture of Addiction**
   2.1. Variable-Reward Recommendation Systems
   2.2. Infinite Scroll and the Absence of Terminal Cues
   2.3. Dark Patterns in Interface Design
3. **Causes and Contributing Mechanisms**
   3.1. Dopaminergic Feedback Loops
   3.2. Social Comparison and Public Metrics
   3.3. Algorithmic Amplification of Affective Content
4. **Consequences: Evidence of Harm**
   4.1. Mental Health Outcomes
   4.2. Adolescent Vulnerability and Regulatory Response
   4.3. Sleep and Cognitive Effects
5. **Recommended Design Interventions**
   5.1. Satisfaction-Signal Ranking
   5.2. Usage Limits and Temporal Anchoring
   5.3. Removal of Public Engagement Metrics
6. **Flair: A Case Study in Anti-Engagement Architecture**
   6.1. Design Decisions and Their Theoretical Basis
   6.2. Unresolved Tensions and Hypotheses for Testing
7. **Conclusion**

References

1. Introduction

In January 2025 the global monthly active user count across the five largest social media platforms exceeded 5.4 billion (Statista, 2025). Platforms at this scale do not merely reflect culture; they actively shape the cognitive and emotional lives of their users. The question of how they are designed, and for whose benefit, has moved from the margins of academic debate to the centre of legislative agendas in the United States, European Union, and United Kingdom.

This paper proceeds from a premise supported by a growing body of empirical research: that the dominant design paradigm in social media, that is, engagement optimisation, produces measurable harms disproportionate to the benefits it claims to provide. The paper is structured as a conventional literature review, moving from the identification of the problem through its mechanisms and consequences to a set of design-level recommendations. It closes with a critical examination of Flair (https://useflair.app), a short-form video platform built on explicitly anti-engagement principles, as a practical instantiation of those recommendations.

A note on scope: this review does not argue that social media is categorically harmful, nor that all engagement is pathological. The evidence base is strongest for adolescent populations and for heavy-use patterns. The paper confines its claims to what the cited literature directly supports and flags where evidence is contested or where assertions have been made without independently verified supporting data.

2. Problem Statement: The Architecture of Addiction

The architecture of contemporary social media platforms is not a neutral substrate for human expression. It is a designed environment engineered to maximise a single metric, time-on-screen, through a combination of algorithmic recommendation, infinite scroll, and interface patterns documented by regulators as manipulative.

2.1 Variable-Reward Recommendation Systems

The theoretical foundation for modern recommender systems is drawn, in part, from B.F. Skinner's work on operant conditioning. Variable-ratio reinforcement schedules, where the reward is delivered after an unpredictable number of responses, produce the highest and most persistent response rates of any schedule tested (Ferster and Skinner, 1957). Haynes (2018) documents how platforms including Instagram have used variable-ratio reward delivery directly, for example withholding likes and then delivering them in bursts, drawing on the same behavioural principle.

The practitioner literature is unusually candid on this point. Documents disclosed during the United States Senate Commerce Committee hearings in 2021 reveal that Meta's internal research identified persistent negative effects on a subset of teenage users but that these findings did not precipitate design changes (United States Senate Commerce Committee, 2021). The primary optimisation target remained engagement.

It is important to note that recommendation algorithms are not inherently harmful. The harm identified in the literature is specific to algorithms optimised for engagement signals, particularly watch time, likes, and shares, rather than satisfaction signals such as rated enjoyment or stated relevance. Harris (2017), a former design ethicist at Google, frames this as the difference between optimising for time well spent rather than time on screen: platforms can be highly effective at delivering value without being effective at extracting time.

2.2 Infinite Scroll and the Absence of Terminal Cues

Aza Raskin, credited with the invention of infinite scroll, has publicly acknowledged that despite good intentions he did not foresee how the feature would be exploited by profit-driven platforms, stating in a 2026 interview that 'incentives eat intentions' and that he personally had to write software to control his own use of it (NPR, 2026c). The feature eliminates the natural stopping points, the bottom of a page or the end of a publication, that ordinarily cue cessation of a behaviour. In the absence of a terminal cue, consumption continues until an external interruption or cognitive exhaustion intervenes.

Some empirical support for this mechanism comes from Rixen et al. (2023), whose week-long field study of social media use found that sessions composed mainly of continuous scrolling were associated with significantly lower emotional valence than sessions involving other activities such as posting or viewing stories. The study did not directly compare infinite-scroll and paginated formats, so it does not by itself establish that pagination would improve outcomes, but it is consistent with the broader hypothesis that undifferentiated, terminus-free scrolling is a distinct and lower-satisfaction mode of use.

2.3 Dark Patterns in Interface Design

In 2022 the European Data Protection Board (EDPB) published guidelines cataloguing dark patterns on social media platforms, defined as interface designs that trick or manipulate users into making choices that are not in their interest (EDPB, 2022). The EDPB identified six categories, each of which appears in current major platform designs:

EDPB (2022): Six Dark Pattern Categories
Overloading: Confronting users with a large quantity of requests to encourage data-sharing or discourage opting out.
Skipping: Designing the interface so users forget or overlook privacy-relevant considerations during consent flows.
Stirring: Exploiting emotional responses, such as appeals to fear, urgency, or social norms, to steer user decisions.
Hindering: Making privacy-protective actions difficult or impossible to find and complete.
Fickle: Implementing an inconsistent interface that makes it difficult for users to understand and control their privacy settings.
Left in the Dark: Using vague or ambiguous language, or hiding information, so that users cannot make informed choices about their data.

The guidelines were adopted under the General Data Protection Regulation (GDPR) framework and constitute authoritative regulatory guidance on the application of GDPR principles to deceptive design

practices on social media platforms. Their significance extends beyond legal compliance: the taxonomy provides a language for identifying manipulative design at a granular level.

3. Causes and Contributing Mechanisms

3.1 Dopaminergic Feedback Loops

The neuroscientific literature on reward processing provides a plausible mechanism for compulsive social media use. Dopaminergic neurons in the mesolimbic pathway respond not to reward per se, but to reward prediction error, the gap between expected and received reward (Schultz, 1998). Variable social feedback, a like, a comment, a share, arriving at unpredictable intervals constitutes a variable-ratio reinforcement schedule in precisely the sense identified as maximally habit-forming in the behavioural literature.

Sherman et al. (2016) tested this mechanism directly using fMRI in a sample of 32 adolescents aged 13 to 18 viewing a simulated Instagram feed, and found significantly greater activation in the nucleus accumbens, a core reward-circuitry region, when participants viewed photographs, including their own, displaying many likes compared with few. The effect held even for neutral images, indicating that the like count itself, independent of the image's content, drives the reward response.

3.2 Social Comparison and Public Metrics

Festinger's (1954) social comparison theory predicts that individuals evaluate their own opinions and abilities by comparison with others, with a directional bias toward upward comparison. Platforms that display public engagement metrics, including follower counts, like totals, and view counts, provide continuous, quantified, and asymmetric comparison information. A creator with 200 followers sees their number in a context populated by creators with millions.

Vogel et al. (2014) demonstrated experimentally that exposure to idealised social media profiles produced significant reductions in self-evaluations of attractiveness and life satisfaction compared to a control condition. The effect was mediated by social comparison processes and moderated by the extremity of the discrepancy between the target and comparison profile. Importantly, the effect was present even for participants who rated their social media use as unproblematic.

Experimental evidence on hiding like counts specifically, rather than on public metrics in general, remains limited but is instructive where it exists. Wallace and Buil (2021) conducted a controlled experiment in which Instagram users were shown a post receiving either more or fewer likes than they had indicated would satisfy them, with the like count either visible or hidden from other viewers. Counter-intuitively, receiving more likes than desired increased negative affect rather than reducing it, and this effect was strongest when the count was publicly visible; when likes were low, visibility made little additional difference to negative affect. Higher like counts did reduce loneliness, suggesting a more complex mechanism in which users pursue visible validation to offset loneliness, at the cost of heightened comparison-driven negative affect once that validation is made public. Separately, internal Meta research disclosed through litigation and whistleblower materials and examined in the United States Senate Commerce Committee hearings found consistent harms linked to Instagram use on measures including

body image and anxiety among teenage girls (United States Senate Commerce Committee, 2021). Taken together, the evidence points toward the visibility of engagement metrics to third parties, rather than the underlying interaction itself, as the more harmful component.

3.3 Algorithmic Amplification of Affective Content

A consistent finding across multiple platforms is that algorithmically ranked feeds disproportionately surface content that elicits strong negative affect, including outrage, disgust, and anxiety, because such content reliably generates engagement signals (Brady et al., 2017). This is not a deliberate editorial choice by platform engineers; it is an emergent property of optimising a proxy metric, engagement, for a target variable, value or wellbeing, with which it is imperfectly correlated.

Brady et al. (2017) analysed 563,312 social media messages concerning three polarising public policy issues, gun control, same-sex marriage, and climate change, and found that each additional moral-emotional word in a message increased its diffusion by approximately 20 per cent. The effect was consistent across political affiliations and was most pronounced for in-group-directed moral language. The study provides empirical support for the hypothesis that engagement-optimised algorithms systematically amplify the emotional temperature of content, with downstream effects on political polarisation and affective wellbeing.

4. Consequences: Evidence of Harm

4.1 Mental Health Outcomes

The relationship between social media use and mental health has been the subject of considerable methodological debate. Early cross-sectional studies established associations between heavy use and elevated rates of anxiety and depression, but could not establish directionality. More recent longitudinal and experimental designs have strengthened the causal case.

Twenge, Martin and Campbell (2018) analysed data from over one million adolescents and found that screen time was negatively associated with psychological wellbeing in a dose-response pattern, with the largest drop occurring between moderate and heavy use. More time spent on non-screen activities, including in-person socialising and exercise, was associated with higher wellbeing. The authors discuss displacement of in-person social interaction and sleep as plausible explanations, but the study's correlational design did not formally test these as mediators. A longitudinal study by Coyne et al. (2020), following approximately 500 adolescents from age 13 to 20 over eight years, found no significant within-person association between social media use and later anxiety or depression, even though a cross-sectional association existed between individuals at a single point in time.

The most methodologically rigorous evidence comes from randomised controlled experiments. Hunt et al. (2018) randomly assigned university students to limit social media use to ten minutes per platform per day for three weeks. The experimental group showed significant reductions in loneliness and depression compared to controls, with effect sizes in the moderate range. The study is limited by its reliance on self-report and a non-clinical sample, but its experimental design provides stronger causal evidence than most observational work.

4.2 Adolescent Vulnerability and Regulatory Response

The evidence for differential effects in adolescent populations is robust. Haidt and Allen (2020), writing in a paired Nature commentary, present contrasting assessments: Haidt argues that the correlational evidence for negative effects is more consistent for girls than for boys, attributing this to girls' social lives revolving more heavily around inclusion, exclusion, and status comparison during adolescence, while Allen's companion piece in the same forum urges caution, noting that the observational evidence base remains mixed and that reported effect sizes are generally small.

The Royal Society for Public Health's (2017) survey of 1,479 young people in the United Kingdom found that Instagram was ranked most harmful for young people's mental health, primarily due to its facilitation of negative body image and social comparison.

Regulatory enforcement has followed the research. Meta, the parent company of Instagram and Facebook, settled a federal trial in August 2026 for $17 billion in response to claims that its platforms were engineered to addict children (NPR, 2026). A separate court order in the same month required Meta to pay $567 million to address children's mental health harm online (NPR, 2026). TikTok, operated by ByteDance, reached a $400 million settlement in 2026 over children's privacy violations under the Children's Online Privacy Protection Act, following an investigation and lawsuit filed by the Federal Trade Commission and the Department of Justice in 2024 (FTC, 2024; Leading Justice, 2026). The Irish Data Protection Commission had previously fined Meta's Instagram division €405 million under GDPR for mishandling children's user data (TechCrunch, 2022). Cumulatively, these enforcement actions represent the largest coordinated regulatory response to platform harm in the history of the industry.

4.3 Sleep and Cognitive Effects

Exelmans and Van den Bulck (2017) surveyed 844 Flemish adults and found that general pre-sleep mobile phone use, not social media specifically, was associated with later sleep onset, shorter sleep duration, and poorer subjective sleep quality. The authors discuss cognitive and emotional arousal from phone content as one of several plausible explanatory mechanisms, alongside light exposure and time displacement, though their survey design did not formally test mediation.

Twenge, Krizan and Hisler (2017) extended this line of evidence to adolescent sleep specifically, analysing nationally representative U.S. survey data from 1991 to 2015 (n = 369,595) and finding a clear exposure-response relationship between new media screen time and short sleep duration once daily use exceeded two hours; the prevalence of adolescents sleeping fewer than seven hours rose from 35% to 41% in one dataset and from 37% to 43% in another, coinciding with the rise of smartphone adoption. Given the well-established cognitive and physical health consequences of sleep deprivation in adolescents, this represents a meaningful public health effect independent of the direct mental health associations.

5. Recommended Design Interventions

The literature suggests a set of design-level interventions with empirical or theoretical support. These are distinguished from behavioural interventions, that is, digital detox recommendations directed at users,

on the grounds that design interventions apply systemically, without requiring individual behavioural change from each user.

5.1 Satisfaction-Signal Ranking

The core argument in the literature is that recommender systems should be retrained on satisfaction signals rather than engagement signals. Satisfaction signals, including explicit ratings of enjoyed content, stated relevance preferences, or measured indicators of positive affect, are closer proxies for genuine user benefit than watch time or like counts. The practical challenge is that satisfaction data requires active user input and is therefore harder to collect at scale than the passive engagement signals platforms already instrument.

Harris (2017) frames this intervention as a reorientation rather than a reduction in platform capability: platforms can be highly effective at delivering value without being effective at extracting time. This is treated in the present paper as the primary conceptual basis for satisfaction-signal ranking.

5.2 Usage Limits and Temporal Anchoring

The experimental literature supports the efficacy of usage limits. The Hunt et al. (2018) study cited above used a hard limit of ten minutes per platform per day and found significant mental health benefits over a three-week period. Platform-native usage dashboards, introduced by Apple and Google at the operating system level in 2018 and by Instagram in 2019, have shown more modest effects, plausibly because countdowns are easily dismissed and do not interrupt the consumption experience.

A theoretically stronger intervention is temporal anchoring, replacing a countdown timer with a specific clock time. For example, a platform might present, after 60 minutes of use: “You’ve had a good session. Take a break. Come back at 9:30 PM.” This approach draws on Kahneman et al.'s (1993) peak-end rule, which suggests that the final moments of an experience disproportionately shape its retrospective evaluation. The inference that a concrete future reference point interrupts habitual behaviour more effectively than an abstract countdown is this paper's own extrapolation from that principle, and has not itself been tested experimentally in this context.

5.3 Removal of Public Engagement Metrics

The case for removing public like and follower counts rests on the social comparison literature reviewed in section 3.2. Removing these metrics does not prevent users from knowing that their content is performing well; it prevents the continuous, quantified, and socially visible ranking of individuals against one another.

The experimental evidence reviewed in section 3.2 supports this design choice. Wallace and Buil (2021) found that the negative-affect cost of an unexpectedly high like count, and the social-comparison pressure it creates, were concentrated in conditions where the count was visible to others rather than hidden. Removing public counts does not remove the underlying interaction; on this evidence, it removes the third-party visibility component that appears to carry most of the harm. Independent academic investigation of platform-level, default-off implementations remains limited, since most real-world deployments of hidden likes have been opt-in toggles rather than controlled, default-level experiments.

6. Flair: A Case Study in Anti-Engagement Architecture

Flair (https://useflair.app) is a short-form social media platform, currently in development, built on the explicit design principle that technology should be designed to be enjoyed rather than to maximise time-on-screen. The platform is developed by Anwar Syed Ali as a sole-trader product. This section reviews Flair's core design decisions against the evidence base established in the preceding sections.

6.1 Design Decisions and Their Theoretical Basis

Ten-Second Content Format

Flair limits all video content to ten seconds per note (Flair's term for a short-form video post). This constraint is intended to preserve a natural decision point after every piece of content, rather than relying on the viewer to interrupt an ongoing video voluntarily. Schaffner et al. (2025) found, in a controlled study of Netflix viewing behaviour, that removing autoplay, which otherwise removes the natural stopping cue between episodes, reduced average daily viewing by 21 minutes and average session length by 17 minutes, as users took longer to decide whether to continue watching. The specific hypothesis under test is whether this format, combined with the bounded feed and satisfaction-signal ranking described below, reduces compulsive use relative to platforms offering equivalent content without a duration cap. This hypothesis is discussed further, including a countervailing consideration, in section 6.2.

Bounded Feed Architecture

Flair implements a feed that has a genuine end state. Users receive a finite set of notes per session, with a real end-of-feed state rather than an infinite loop. This directly addresses the terminal-cue deficit central to Raskin's critique of infinite scroll, and is consistent with field evidence that scrolling-dominated sessions are associated with lower emotional valence than sessions with a natural endpoint or varied activity (Rixen et al., 2023).

Satisfaction-Signal Ranking

Flair's ranking algorithm is designed to incorporate per-item satisfaction signals, including completion rate, saves, and shares, alongside options for users to signal disinterest in specific content categories. The ranking system is under active development and not yet finalised: the current implementation draws primarily on engagement signals as a baseline while the satisfaction-signal layer is being built out. This trajectory is directionally consistent with the recommendations in section 5.1.

Honest Creator Analytics

Flair does not display public like counts, follower counts, or view totals in the social feed. This decision is grounded in the Vogel et al. (2014) and Festinger (1954) evidence base on social comparison. Creator-facing analytics are available in a private dashboard showing quality signals, including audience retention, returning viewers, and saves. These signals are visible only to the creator rather than published for public comparison, which distinguishes them from the continuously visible public ranking the literature in section 3.2 identifies as the harmful mechanism.

Timed Break with Temporal Anchoring

After sixty minutes of use, Flair presents a full-screen break prompt: “You've had a good session. Take a break. Come back at [specific time]”. The specific return time operationalises the temporal anchoring principle drawn from Kahneman et al.'s (1993) peak-end rule. The use of a concrete clock time rather than a countdown is a deliberate design choice to make the pre-commitment device more effective.

Privacy-First Data Architecture

Flair's data architecture follows a minimal-collection principle consistent with Article 5(1)(c) of the GDPR (data minimisation) and aligned with the EDPB's (2022) anti-dark-pattern guidelines. The platform does not sell user data to third parties. Advertising, where implemented, takes the form of native ten-second sponsored content, subject to the same completion-signal ranking as organic content, with explicit prior user consent.

6.2 Unresolved Tensions and Hypotheses for Testing

The paper would be incomplete without acknowledging the tensions within Flair's design that are not yet resolved by the existing evidence base.

The ten-second content format is hypothesised to reduce compulsive use by ensuring a natural stopping point occurs after every piece of content, addressing the terminal-cue removal mechanism that Schaffner et al. (2025) found extends viewing time on autoplay-driven platforms. However, algorithmically curated short-form feeds carry their own addictive potential independent of individual content length: Qin, Omar and Musetti (2022) found that TikTok's recommendation system drives states of high concentration and time distortion among adolescent users (n = 659), and that these flow states were the strongest predictor of addictive use in their sample. The bounded feed and satisfaction-signal ranking are intended as layered mitigations, but the interaction of the ten-second format with these mitigations has not been empirically tested. The hypothesis that a ten-second cap, when combined with a bounded feed and satisfaction-signal ranking, reduces compulsive use relative to a standard algorithmic short-form feed is falsifiable and warrants empirical investigation.

Flair’s Design Hypothesis

Flair is designed around the proposition that a short-form video platform combining: (1) a bounded per-session feed, (2) satisfaction-signal rather than engagement-signal ranking, and (3) a temporal-anchoring break at 60 minutes, will produce lower rates of unintended overuse and higher post-session satisfaction than a platform optimised for engagement. This is the design hypothesis Flair is built to test, and the platform commits to evaluating it empirically as it reaches sufficient scale.

7. Conclusion

The reviewed literature establishes that the dominant design paradigm in social media, that is, engagement optimisation through variable-reward algorithms, infinite scroll, and documented dark patterns, is associated with measurable harms, particularly to adolescent populations and heavy users. The mechanisms proposed in the literature, including dopaminergic feedback loops, social comparison amplification, and algorithmic affective escalation, are supported by a combination of neuroimaging, behavioural, and large-scale computational evidence.

The regulatory response has intensified in scale. Enforcement actions in the United States, Ireland, and across the European Union have resulted in settlements and fines exceeding $18 billion against Meta and TikTok alone between 2022 and 2026. These actions confirm that platform conduct has caused legally actionable harm, not merely academic concern.

Nonetheless, the existing regulatory response operates primarily at the level of content moderation and data protection law. It has not yet addressed the design-level mechanisms through which engagement optimisation produces harm. The EDPB's (2022) dark pattern guidelines are a significant step, but their scope is limited to GDPR-relevant consent manipulation rather than the broader architecture of addictive design.

Design-level interventions, including satisfaction-signal ranking, bounded feeds, removal of public metrics, and temporal anchoring, offer a path to platforms that are commercially viable without being architecturally exploitative. The hypothesis that such interventions are compatible with sustainable platform economics is plausible but empirically undertested.

Flair represents one attempt to operationalise these principles at the product level. Its success or failure as a commercial product will provide real-world evidence on whether privacy-first, anti-engagement architecture can attract and retain users at scale. That evidence, when available, will be a meaningful contribution to the empirical debate, regardless of the commercial outcome.

The central claim of this paper is modest and, the author believes, well-supported: that the question of how social media platforms are designed is not separable from the question of what effects they have. Architecture is not neutral. Design is a choice. And the choice to optimise for engagement rather than for genuine user benefit is one that platforms can, and should, revisit.

References

Brady, W.J., Wills, J.A., Jost, J.T., Tucker, J.A. and Van Bavel, J.J. (2017) 'Emotion shapes the diffusion of moralized content in social networks', Proceedings of the National Academy of Sciences, 114(28), pp. 7313-7318. Available at: https://doi.org/10.1073/pnas.1618923114

Coyne, S.M., Rogers, A.A., Zurcher, J.D., Stockdale, L. and Booth, M. (2020) 'Does time spent using social media impact mental health?: an eight year longitudinal study', Computers in Human Behavior, 104, p. 106160. Available at: https://doi.org/10.1016/j.chb.2019.106160

European Data Protection Board (EDPB) (2022) Guidelines 3/2022 on Dark Patterns in Social Media Platform Interfaces. Version 2.0. Brussels: EDPB. Available at: https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-32022-dark-patterns-social-media-platform_en

Exelmans, L. and Van den Bulck, J. (2017) 'Bedtime mobile phone use and sleep in adults', Social Science & Medicine, 148, pp. 93-101. Available at: https://doi.org/10.1016/j.socscimed.2015.11.037

Federal Trade Commission (FTC) (2024) FTC Investigation Leads to Lawsuit Against TikTok and ByteDance for Flagrantly Violating Children's Privacy Law. Press release, 2 August. Available at: https://www.ftc.gov/news-events/news/press-releases/2024/08/ftc-investigation-leads-lawsuit-against-tiktok-bytedance-flagrantly-violating-childrens-privacy-law

Ferster, C.B. and Skinner, B.F. (1957) Schedules of Reinforcement. New York: Appleton-Century-Crofts.

Festinger, L. (1954) 'A theory of social comparison processes', Human Relations, 7(2), pp. 117-140. Available at: https://doi.org/10.1177/001872675400700202

Haidt, J. and Allen, N.B. (2020) 'Scrutinizing the effects of digital technology on mental health', Nature, 578, pp. 226-227. Available at: https://doi.org/10.1038/d41586-020-00296-x

Harris, T. (2017) How a handful of tech companies control billions of minds every day. TED Talk, April 2017. Available at: https://www.ted.com/talks/tristan_harris_how_a_handful_of_tech_companies_control_billions_of_minds_every_day

Haynes, T. (2018) 'Dopamine, smartphones & you: a battle for your time', Harvard University: Science in the News, 1 May. Available at: https://www.scribd.com/document/460458389/Dopamine-Smartphones-You-A-battle-for-your-time-Science-in-the-News-pdf

Hunt, M.G., Marx, R., Lipson, C. and Young, J. (2018) 'No more FOMO: limiting social media decreases loneliness and depression', Journal of Social and Clinical Psychology, 37(10), pp. 751-768. Available at: https://doi.org/10.1521/jscp.2018.37.10.751

Kahneman, D., Fredrickson, B.L., Schreiber, C.A. and Redelmeier, D.A. (1993) 'When more pain is preferred to less: adding a better end', Psychological Science, 4(6), pp. 401-405. Available at: https://doi.org/10.1111/j.1467-9280.1993.tb00589.x

Leading Justice (2026) TikTok agrees to $400 million settlement over alleged children's privacy violations. 10 September. Available at: https://www.leadingjustice.net/news/09/10/2026/tiktok-agrees-to-400-million-settlement-over-alleged-childrens-privacy-violations/

NPR (2026a) Court orders Instagram and Facebook's Meta to pay $567M to address kids' mental health online. 7 August. Available at: https://www.npr.org/2026/08/07/nx-s1-5924581/court-orders-instagram-and-facebooks-meta-to-pay-567m-to-address-kids-mental-health-online

NPR (2026b) Meta, states agree to $17 billion settlement in child safety trial. 26 August. Available at: https://www.npr.org/2026/08/26/nx-s1-5944781/meta-settlement-child-safety-lawsuit

NPR (2026c) ‘Why infinite scroll’s inventor wants to kill his creation’, The Indicator from Planet Money, 7 April. Available at: https://www.npr.org/2026/04/07/nx-s1-5775917/why-infinite-scrolls-inventor-wants-to-kill-his-creation

Qin, Y., Omar, B. and Musetti, A. (2022) 'The addiction behavior of short-form video app TikTok: the information quality and system quality perspective', Frontiers in Psychology, 13, 932805. Available at: https://doi.org/10.3389/fpsyg.2022.932805

Rixen, J.O., Meinhardt, L.-M., Glöckler, M., Ziegenbein, M.-L., Schlothauer, A., Colley, M., Rukzio, E. and Gugenheimer, J. (2023) 'The loop and reasons to break it: investigating infinite scrolling behaviour in social media applications and reasons to stop', Proceedings of the ACM on Human-Computer Interaction, 7(MHCI), pp. 1-22. Available at: https://doi.org/10.1145/3604275

Royal Society for Public Health (RSPH) (2017) #StatusOfMind: Social Media and Young People's Mental Health and Wellbeing. London: RSPH. Available at: https://www.rsph.org.uk/media/filer_public/e5/c1/e5c1bd64-2ab9-42ed-921b-f06566970da7/status_of_mind.pdf

Schaffner, B., Ulloa, Y., Sahni, R., Li, J., Cohen, A.K., Messier, N., Gao, L. and Chetty, M. (2025) 'An experimental study of Netflix use and the effects of autoplay on watching behaviors', Proceedings of the ACM on Human-Computer Interaction, 9(2), Article CSCW030. Available at: https://doi.org/10.1145/3710928

Schultz, W. (1998) 'Predictive reward signal of dopamine neurons', Journal of Neurophysiology, 80(1), pp. 1-27. Available at: https://doi.org/10.1152/jn.1998.80.1.1

Sherman, L.E., Payton, A.A., Hernandez, L.M., Greenfield, P.M. and Dapretto, M. (2016) 'The power of the like in adolescence: effects of peer influence on neural and behavioral responses to social media', Psychological Science, 27(7), pp. 1027-1035. Available at: https://doi.org/10.1177/0956797616645673

Statista (2025) Global social media statistics. Available at: https://www.statista.com/statistics/617136/digital-population-worldwide/

TechCrunch (2022) Instagram fined €405M in EU over children's privacy. 5 September. Available at: https://techcrunch.com/2022/09/05/instagram-gdpr-fine-childrens-privacy/

Twenge, J.M., Krizan, Z. and Hisler, G. (2017) 'Decreases in self-reported sleep duration among U.S. adolescents 2009-2015 and links to new media screen time', Sleep Medicine, 39, pp. 47-53. Available at: https://doi.org/10.1016/j.sleep.2017.08.013

Twenge, J.M., Martin, G.N. and Campbell, W.K. (2018) 'Decreases in psychological well-being among American adolescents after 2012 and links to screen time during the rise of smartphone technology', Emotion, 18(6), pp. 765-780. Available at: https://doi.org/10.1037/emo0000403

United States Senate Commerce Committee (2021) Protecting Kids Online: Testimony from a Facebook Whistleblower. Hearing, 5 October. Washington DC: United States Senate. Available at: https://www.commerce.senate.gov/2021/10/protecting-kids-online-testimony-from-a-facebook-whistleblower

Vogel, E.A., Rose, J.P., Roberts, L.R. and Eckles, K. (2014) 'Social comparison, social media, and self-evaluation', Psychology of Popular Media Culture, 3(4), pp. 206-222. Available at: https://doi.org/10.1037/ppm0000047

Wallace, E. and Buil, I. (2021) 'Hiding Instagram Likes: Effects on negative affect and loneliness', Personality and Individual Differences, 170, 110509. Available at: https://doi.org/10.1016/j.paid.2020.110509