List of Social Media Harms

Social media platforms, apps, and freemium games are designed to exploit attention and extract individualized data for the sake of advertising -- it is central to their business models. From January 1 to Sept. 30, 2021, Facebook made $82.3 billion from advertising.[1] During the same period, YouTube made $20.2 billion from advertising.[2] Roblox made $1.35 billion, including advertising and in-game purchases, and users, the majority of whom are under 16, spent 30.6 billion hours playing.[3] Due to lack of regulation, profit derived from attention is the foremost motivator of design. How does this impact our kids and teens?

Manipulative Design Features

The impact of these features is not universally negative, but that does not negate their inherently manipulative purpose, which is the promotion of user engagement.

Design Feature	Outcomes	Platforms
Likes, emoji reactions, comments, views	Addiction; social comparison; body image issues	Facebook; Instagram; TikTok; YouTube; Discord; Twitter; Twitch
Autoplay and infinite scroll	Addiction; fear of missing out (FOMO)	Facebook; Instagram; TikTok; YouTube; Twitch
Push notifications	Addiction; social pressure; distraction; FOMO	Facebook; Instagram; TikTok; Twitter; Roblox; Snapchat; Discord; Twitch; messaging apps
"Beautifying" filters	Social comparison; body image issues	Snapchat; Instagram; TikTok
Gamification of social interactions	Addiction; social comparison; social pressure	Snapchat; Reddit; Twitch; TikTok
Ephemeral content	Addiction; FOMO	Snapchat; Instagram; Facebook; TikTok (pilot of stories)
Marking messages as read or unread	Social pressure; distraction	Snapchat; Facebook; Instagram; messaging apps
Virtual currencies and in-app purchases	Financial exploitation; social comparison; addiction	TikTok; Roblox; Twitch; Instagram

[1] Quarterly earnings. Facebook - Financials. (n.d.). https://investor.fb.com/financials/?section=quarterlyearnings.
[2] Alphabet Investor Relations. Alphabet. (n.d.). https://abc.xyz/investor/.
[3] Quarterly results. Roblox. (n.d.). https://ir.roblox.com/financials/quarterly-results/default.aspx.

Opaque recommendation algorithms	Addiction; social comparison; body image issues; self-harm; extremism	Facebook; Instagram; TikTok; Twitter; YouTube

Opaque Recommendation Algorithms

Recommendation algorithms are difficult to study externally due to lack of transparency. Below, we discuss a selection of studies and cases that illustrate the harms caused by algorithms. Leaked internal Facebook research features prominently due to the unique research insights provided.[4]

Prioritization of divisive, hateful, sensationalist content

- Facebook's newsfeed ranking algorithm, which was altered in 2018 to ostensibly boost "meaningful interactions," actually heavily prioritized emoji reactions, long comments, and reshares of posts, resulting in the promotion of controversial and emotion-inducing content.[5] News organizations and political parties noticed this change and felt pressured to "rely more on negativity and sensationalism." Facebook researchers found increased promotion of "civic misinfo, civic toxicity, health misinfo, and health antivax content."[6] Internal researchers proposed solutions that were rejected by Mark Zuckerberg due to concerns that user engagement would decrease.[7]

Promotion of negative social comparison

- According to a 2020 Facebook study of 50,590 people in 10 countries, 33% of people compared their appearance to others on Instagram and 26% always or often saw content that made them feel worse about their appearance.[8] It's worse for teen girls: 48% compared appearance, 37% felt worse about themselves, and 34% felt “a lot” or “extreme” pressure to look perfect. In another 2020 Facebook study, 40% of teen boys experienced negative social comparison due to Instagram, but body image contributed less than other factors like economic status.[9]

[4] Pierce, D., & Kramer, A. (Oct. 28, 2021). *Here are all the Facebook Papers stories. Protocol.* https://www.protocol.com/facebook-papers.

[5] Hagey, K., & Horwitz, J. (Sept. 15, 2021). *Facebook tried to make its platform a healthier place. It got angrier instead. Wall Street Journal*. https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215.

[6] Merrill, J. B., & Oremus, W. (Oct. 26, 2021). *Five points for anger, one for a 'like': How Facebook's formula fostered rage and misinformation. Washington Post*. Retrieved Dec. 6, 2021, from https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/.

[7] Hagey, K., & Horwitz, J. (Sept. 15, 2021). *Facebook tried to make its platform a healthier place. It got angrier instead. Wall Street Journal*. https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215.

[8] *Appearance-based social comparison on Instagram*. (Sept. 29, 2021). *Wall Street Journal*. https://s.wsj.net/public/resources/documents/appearance-based-social-comparison-on-instagram.pdf.

[9] Wells, G., Horwitz, J., & Seetharaman, D. (Sept. 14, 2021). *Facebook knows Instagram is toxic for teen girls, company documents show. Wall Street Journal*. https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739.

- A 2019 study, focusing on young women, found frequency of Instagram use to be correlated with depressive symptoms, decreased self-esteem, general and physical appearance anxiety, and body dissatisfaction.[10] Additionally, exposure to beauty and fitness images, in particular, significantly decreased self-rated attractiveness.
- A 2020 study found that just seven minutes on Instagram, compared to Facebook or playing a game, led to decreased body satisfaction and increased negative emotions among college-age women.[11]
- Content moderation guidelines for TikTok, which were leaked in 2020, instructed moderators to suppress content featuring fatness, wrinkles, "abnormal body shape," "ugly facial looks," and poverty, among other categories.[12]

Promotion of disordered body image content

- Communities that encourage disordered eating use coded language to evade platform filters and to share memes, "thinspiration" photos, goal weights, and tips on Twitter, TikTok, Instagram, Facebook, YouTube, Discord, Tumblr, and Snapchat.[13] Instagram, in particular, was shown to recommend search terms like "appetite suppressant" to people with eating disorders.[14]
- Several investigations, including by Sen. Blumenthal's office and CNN, have created fake youth accounts on Instagram that like a few dieting posts and then, within a day or two, are shown majority pro-anorexia content.[15]
- A 2019 study of 12- to 13-year-olds in Australia found that girls with Snapchat, Instagram, and Tumblr accounts and boys with Snapchat, Facebook, and Instagram were significantly more likely to report disordered eating than youths not using social media. Greater daily time spent using Instagram and Snapchat was associated with significantly higher disordered eating behaviors among girls.[16]

[10] Sherlock, M., & Wagstaff, D. L. (2019). *Exploring the relationship between frequency of Instagram use, exposure to idealized images, and psychological well-being in women. Psychology of Popular Media Culture*, 8(4), 482–490.

[11] Engeln, R., Loach, R., Imundo, M. N., & Zola, A. (2020). *Compared to Facebook, Instagram use causes more appearance comparison and lower body satisfaction in college women. Body Image*, 34, 38–45.

[12] Biddle, S., Ribeiro, P. V., & Dias, T. (March 16, 2020). *TikTok told moderators to suppress posts by 'ugly' people and the poor to attract new users. The Intercept*. https://theintercept.com/2020/03/16/tiktok-app-moderators-users-discrimination/.

[13] Conger, K., Browning, K., & Woo, E. (Oct. 27, 2021). *Eating disorders and social media prove difficult to untangle. New York Times*. https://www.nytimes.com/2021/10/22/technology/social-media-eating-disorders.html; Reese, A. (May 16, 2019). *The evolution of online eating disorder communities from Tumblr to Twitter meme. Jezebel*. https://jezebel.com/the-evolution-of-online-eating-disorder-communities-fro-1833079022.

[14] Hern, A. (April 15, 2021). *Instagram apologises for promoting weight-loss content to users with eating disorders. The Guardian*. https://www.theguardian.com/technology/2021/apr/15/instagram-apologises-for-promoting-weight-loss-content-to-users-with-eating-disorders.

[15] O'Sullivan, D., Duffy, C., & Jorgensen, S. (Oct. 4, 2021). *Instagram promoted pages glorifying eating disorders to teen accounts. CNN*. Retrieved Dec. 6, 2021, from https://www.cnn.com/2021/10/04/tech/instagram-facebook-eating-disorders/index.html.

[16] Wilksch, S. M., O'Shea, A., Ho, P., Byrne, S., & Wade, T. D. (2020). *The relationship between social media use and disordered eating in young adolescents. International Journal of Eating Disorders*, 53(1), 96–106.

- YouTube, Instagram, and Facebook are central to the encouragement and sale of steroids, particularly to boys, pushed through algorithms recommending content, ads, groups to join, and accounts to follow.[17]

Promotion of self-harm and suicide content

- Internal Facebook research revealed that among teens who reported suicidal thoughts, 13% of British users and 6% of American users traced the desire to kill themselves to Instagram.[18]
- A 2018 study found that Instagram posts that mentioned suicide ideation elicited higher engagement than posts that did not.[19]
- A 2019 study found that 43% of young adults surveyed have seen self-harm content on Instagram and, of those, 80% did not seek it out (recommended or accidental).[20] The content led 39% to think about how it would feel to harm themselves. Surprisingly, 32.5% indicated that they have performed the same self-harming behavior due to seeing self-harm content on Instagram. This correlated significantly with suicide ideation and suicide risk.

Promotion of extremist content

- A 2020 study of YouTube found that 9% of users watched extremist or white supremacist videos and 22% viewed "alternative" fringe content.[21] During the two months studied, the mean numbers of videos watched were 11.5 (extremist) and 64.2 (alternative). Despite YouTube's recent algorithmic changes, 37.6% of recommendations on alternative videos and 29.3% of recommendations on videos from extremist channels were to other videos of the same type.
- A 2021 study of TikTok found that 30% of videos promoted white supremacy, 24% supported terrorists, 14.9% were antisemitic, 13.5% anti-Black, 8.7% anti-LGBTQ, and 7.9% anti-Muslim.[22] This follows 2020 findings that 14 QAnon-related hashtags had over 488 million combined views on TikTok.[23]

[17] *Digital platforms on steroids*. Digital Citizens Alliance. (September 2019). https://www.digitalcitizensalliance.org/clientuploads/directory/Reports/DCA_Platforms_on_Steroids_Report-Final.pdf.

[18] Teen mental health deep dive. Published by the *Wall Street Journal*. (Sept. 29, 2021). https://s.wsj.net/public/resources/documents/teen-mental-health-deep-dive.pdf.

[19] Carlyle, K. E., Guidry, J. P., Williams, K., Tabaac, A., & Perrin, P. B. (2018). *Suicide conversations on Instagram™: contagion or caring? Journal of Communication in Healthcare*, 11(1), 12–18.

[20] Arendt, F., Scherr, S., & Romer, D. (2019). *Effects of exposure to self-harm on social media: Evidence from a two-wave panel study among young adults. New Media & Society*, 21(11–12), 2422–2442.

[21] Chen, A. Y., Nyhan, B., Reifler, J., Robertson, R. E., & Wilson, C. (Feb. 12, 2021). *Exposure to alternative & extremist content on YouTube*. Anti-Defamation League. https://www.adl.org/resources/reports/exposure-to-alternative-extremist-content-on-youtube.

[22] O'Connor, C. (August 2021). *Hatescape: An in-depth analysis of extremism and hate speech on TikTok. Institute for Strategic Dialogue*. Retrieved from https://www.isdglobal.org/wp-content/uploads/2021/08/HateScape_v5.pdf.

[23] Little, O. (Oct. 6, 2020). *Spread of a conspiracy theory about Trump's COVID-19 diagnosis shows why TikTok must be proactive about QAnon misinformation*. Media Matters for America. https://www.mediamatters.org/qanon-conspiracy-theory/spread-conspiracy-theory-about-trumps-covid-19-diagnosis-shows-why-tiktok.

Inadequate Content Moderation

Hate speech

- By internal measures, Facebook's hate speech algorithm prevents only 3–5% of hate speech views and 0.6% of all content that violates policies against violence and incitement.[24] Publicly, Mark Zuckerberg had claimed 94% removal.[25]
- Facebook knew that its hate speech algorithm disproportionately protected White people and men, and inadequately protected those who are Black, Jewish, Muslim, and LGBTQ, but didn't prioritize changing the algorithm, resulting in greater harms and increased prevalence of racist language.[26]
- Facebook spends 84% of its misinformation budget on the U.S.[27] and does not have translators for the majority of languages,[28] so hate content is worse abroad.

Sexual abuse online

- In a 2020 study, 25% of U.S. 9- to 17-year-olds reported having a sexual encounter online with someone who they believed to be an adult.[29] Of youths 9–12 years old, 19% had interacted with an adult sexually online. 55% of youths were recontacted by individuals they had blocked and/or reported (45% on the same platform under a new account, 43% on a different platform). 83% of kids will not tell a trusted adult about abuse online. 94% will not tell a trusted adult about unsolicited nudes they have received from adults online.
- There was a 77% increase in "self-generated" child sexual content from 2019 to 2020.[30] In 80% of the cases, victims were 11- to 13-year-old girls.
- Most companies use tools to detect child sexual abuse material (87% use image "hash-matching"); only 37% currently use tools to detect the online grooming of children.

[24] Seetharaman, D., Horwitz, J., & Scheck, J. (Oct. 17, 2021). *Facebook says AI will clean up the platform. Its own engineers have doubts*. *Wall Street Journal*. https://www.wsj.com/articles/facebook-ai-enforce-rules-engineers-doubtful-artificial-intelligence-11634338184.

[25] Dwoskin, E., Newmyer, T., & Mahtani, S. (Oct. 26, 2021). *The case against Mark Zuckerberg: Insiders say Facebook's CEO chose growth over safety*. *Washington Post*. https://www.washingtonpost.com/technology/2021/10/25/mark-zuckerberg-facebook-whistleblower/.

[26] Dwoskin, E., Tiku, N., & Timberg, C. (Nov. 22, 2021). *Facebook's race-blind practices around hate speech came at the expense of Black users, new documents show*. Washington Post. https://www.washingtonpost.com/technology/2021/11/21/facebook-algorithm-biased-race/.

[27] Zakrzewski, C., Vynck, G., Masih, N., & Mahtani, S. (Oct. 24, 2021). *How Facebook neglected the rest of the world, fueling hate speech and violence in India*. *Washington Post*. https://www.washingtonpost.com/technology/2021/10/24/india-facebook-misinformation-hate-speech/.

[28] Facebook audit: Reporting system. Next Billion Network. (Oct. 24, 2021). https://nextbillion.network/?p=529863313343.

[29] Thorn & Benenson Strategy Group. (May 2021). *Responding to online threats: Minors' perspectives on disclosing, reporting, and blocking*. Thorn. https://info.thorn.org/hubfs/Research/Responding%20to%20Online%20Threats_2021-Full-Report.pdf.

[30] *Face the facts: Internet Watch Foundation annual report 2020*. Internet Watch Foundation. (April 2021). https://www.iwf.org.uk/about-us/who-we-are/annual-report/.