

Reeling You In

INTRODUCTION

Since 2010, Instagram has changed the landscape of human interaction by creating a digital face for our lives. We now have over 2 billion monthly active users, 500 million of whom engage with the service each day [1]. Operating at this scale it is unsurprising that we have faced scrutiny for ethical concerns in the past, notably a €405 million fine for the mishandling of user data [2]. Following the introduction of our Reels service there is a risk that we allow ourselves to make the same mistakes again. This report looks at three key areas of concern and explores ways to mitigate or solve them.

ADDICTION

As engineers at Instagram our fundamental objective is to capture and keep attention and we are good at our jobs [3]. However, there are concerns that this success is, in reality, a preying on the psychologically vulnerable, for example, we see greater use from people categorised as narcissists and escapists seeking the validation and detachment our platform can provide [4], and studies show a strong correlation between struggling with mental health and excessive use of our service [5][6]. Inspiring addiction and playing on our users' vulnerabilities clearly contravenes the Royal Academy of Engineering's Ethical Guidelines (RAEEG) which demand we "hold paramount the health and safety of others" [7], in our creation of a service that targets the mentally vulnerable we trample on what we are supposed to hold sacrosanct.

There is often a mitigatory argument made that any perversion is unintentional and our responsibility is with the creation of the tool, not its use. This is doubly wrong, the RAEEG makes clear our responsibility for "potential adverse effects" of our work, and, in many cases, we have actively made decisions contributing to the perversion. For example, silently raising the minimum time limit on the app from 10 to 30 minutes [8] but, most subtly yet significantly, by beginning to show users who'd exhausted all the posts from accounts they followed, posts from accounts they didn't [9].

That paradigmatic shift, from Instagram as an accumulator of explicitly preferred content (following) to a curator of implicitly preferred content (for you) inverted the roles of sculptor and sculpture between users and the algorithm. It also marked a massive empowerment of the algorithm at the expense of users. Prior to the shift users were capable of opting out of content they didn't want to see by simply removing the accounts responsible, now their shown content is determined by inferring preference from thousands of interactions fed to the user. There is now no discrete opt out operation, we instead expect the user to opt out of themselves.

Acknowledging this, it now our duty to find solutions. An easy first step here is to introduce more enforceable caps on usage, providing users with a stronger way to restrict their usage. Another solution is enhancing opt-out options, at present the user is subjected to a relentless stream of algorithmic curation that is trying to learn their brain in order to manipulate it, limiting the content available to the algorithm greatly diminishes its attack space and capacity to target vulnerabilities. Implementing these would prevent the excessive use itself, but also the mechanism by which excessive use is developed.

GLOBAL HOMOGENEITY

As a subsidiary of Meta, we at Instagram share in the goal of “bringing the world closer together” [10], but what does this mean. What happens when a single platform’s algorithm begins to dictate the cultural pulse of the world. Reels is quietly driving a phenomenon known as global homogeneity, where diverse cultural identities are overshadowed by a dominant, often Western narrative. In designing an algorithm that optimises for quick consumption and maximum interaction we amplify this trend, ending up favouring content with existing traction, often sidelining unique cultural voices in the process [11].

This unreciprocated flow from Western to non-Western is particularly intense in regions like Africa, where rapid digital adoption amplifies exposure. Creators abandon indigenous languages for English to gain reach, eroding linguistic diversity, while Western values like individualism and capitalism displace collectivist traditions [12]. It is therefore unsurprising that this trend is known by another name, cultural imperialism [13].

Acknowledging Instagram’s contribution to global homogeneity, it becomes imperative to explore remedies that restore cultural balance. One approach involves enhancing the visibility of autochthonous content by recalibrating the algorithm to prioritize Reels from local creators within their respective geographic regions. For example, users in India should encounter a greater proportion of Bollywood-inspired posts rather than an influx of American trends, thereby reinforcing local cultural expression. Another strategy entails restricting the reach of English-language content in non-English-speaking areas, employing geotargeting to ensure cultural relevance unless users explicitly opt for broader exposure. Additionally, Instagram could begin elevating underrepresented perspectives, drawing from strategies like TikTok’s promotion of hyper-local content [14].

HARMFUL CONTENT

At Instagram we are ultimately distributors, not creators, we offer a platform where users make what they want. However, in creating a digital currency of status, of likes and followers, we incentivise creators to make not what they want, but what they think will confer the most status on themselves. In this regard they become slaves to the consumer, or more accurately, slaves to the consumption. This, combined with the passive nature of Reels and absence of opportunity for explicit preference signalling, has led to a significant appetite for outrageous and extreme content.

Humans have evolved to attend to extremeness, the susceptibility of this mechanism to abuse is well documented [15] [16]. Reels has become a medium that readily supplies extremeness for vicarious, passive consumption. This hijacks the mechanism and users become avid consumers of extreme content, in unnatural quantities, that they would decry if actively prompted.

With such a demand, creating the outrageous content becomes lucrative in the newfangled currency of status. Yet humans are remarkably adaptable; what once shocked us dulls with repetition, desensitizing users and creators alike. To keep eyes glued, creators must escalate. This spiral brings commensurate risk to the creators themselves, who endanger their safety to

feed the machine. There is no ceiling to this ascent, no natural endpoint, only an unrelenting pressure to outdo what came before.

Our role here is murky. Are we responsible for every user? The RAEEG demands we “minimise and justify any adverse effect” of ours, and this escalation of extremeness undoubtably qualifies. UK law sheds some light: under the Online Safety Act 2023, we bear a duty of care to mitigate harm, liable not for every post but for systemic failures that enable it [17], as seen in our February 2025 apology for amplifying fatal content due to algorithmic error [18].

Mitigation, then, becomes our ethical and legal imperative. Naïve human content moderation falters before the sheer volume, 1113 posts are made each second [19]. Instead, the problem’s source, the algorithm, must also be its remedy. One approach is to recalibrate it, not to chase watch-time alone but to cap extremeness, perhaps by scoring content for intensity and throttling its spread beyond a threshold. Another is empowering users with explicit filters, letting them block categories like “shock” or “risk” upfront. These steps won’t erase the human appetite for the outrageous but will dampen its amplification on our platform.

OBSTACLES

We at Instagram are in the attention business, the algorithm we’ve constructed so complex it can exfiltrate user preferences without any conscious user preferring, but opaque enough to frustrate mechanical interpretation or understanding, and thus change. There is also the risk our solutions (cynically read *sandbags*) may push users to less scrupulous competitors, unilateral non-proliferation being ill advised in arms races. This issue is downstream of another, the RAEEG says we must be ethical, so too the Law, so too any explicit poll of users, but in the implicit, revealed preferences of users we see ambivalence. Forcing fixes therefore may put us in contention with our users, in mitigation we may consider it contention with their worst traits; but is it Instagram’s place to decide which these are, or that a person can only be their flattering side.

PROGNOSIS

“Science and technology would be used as though ... they had been made for man, not as though man were to be adapted and enslaved to them.” – *Aldous Huxley* [20]

Instagram stands at a crossroads, will we continue in passivity, allowing the bad seeds in our site to proliferate, or act despite the challenges. Instagram already does good in the world; it is our obligation and our privilege to untether it from the bad and allow it to continue connecting communities and sharing joy. While the obstacles are challenging, they are not insurmountable, Instagram has a strong social conscience, with the requisite will to change, it can be freed from these ethical blocks holding it back.

Bibliography

- [1] <https://backlinko.com/instagram-users>
- [2] <https://www.euronews.com/next/2022/09/06/ireland-fines-instagram-405-million-over-protection-of-childrens-data>
- [3] <https://www.mixbloom.com/resources/average-time-spent-on-social-media-2022>
- [4] <https://www.sciencedirect.com/science/article/pii/S2772503022000056>
- [5] <https://penntoday.upenn.edu/news/social-media-use-increases-depression-and-loneliness>
- [6] <https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2749480>
- [7] <https://www.engc.org.uk/standards-guidance/guidance/statement-of-ethical-principles/>
- [8] <https://techcrunch.com/2022/02/21/instagram-limits-daily-time-limits/>
- [9] <https://techcrunch.com/2017/12/27/instagram-will-now-add-recommended-posts-to-your-feed/>
- [10] <https://www.facebook.com/notes/393134628500376/>
- [11] <https://nsuworks.nova.edu/tqr/vol28/iss8/16/>
- [12] https://www.researchgate.net/publication/382994264_CULTURAL_HOMOGENIZATION_VS_CULTURAL_DIVERSITY_SOCIAL_MEDIA'S_DOUBLE-EDGED_SWORD_IN_THE_AGE_OF_GLOBALIZATION
- [13] <https://www.britannica.com/topic/cultural-imperialism>
- [14] <https://biggerfishmarketing.com/7-secret-strategies-to-local-tiktok-marketing-both-free-paid/>
- [15] https://en.wikipedia.org/wiki/Fire_eating
- [16] https://en.wikipedia.org/wiki/Sword_swallowing
- [17] <https://www.gov.uk/government/publications/online-safety-act-explainer/online-safety-act-explainer>
- [18] <https://www.cnn.com/2025/02/27/meta-apologizes-after-instagram-users-see-graphic-and-violent-content.html>
- [19] <https://www.marketingscoop.com/blog/how-many-posts-are-posted-on-instagram-per-day/>
- [20] https://www.academia.edu/86233849/Brave_New_World_Aldous_Huxley page x